The Annals of Applied Probability
2015, Vol. 25, No. 5, 2809–2866
DOI: 10.1214/14-AAP1061

CAN LOCAL PARTICLE FILTERS BEAT THE CURSE OF DIMENSIONALITY?[1]

BY PATRICK REBESCHINI AND RAMON VAN HANDEL

Princeton University

The discovery of particle filtering methods has enabled the use of nonlinear filtering in a wide array of applications. Unfortunately, the approximation error of particle filters typically grows exponentially in the dimension of the underlying model. This phenomenon has rendered particle filters of limited use in complex data assimilation problems. In this paper, we argue that it is often possible, at least in principle, to develop local particle filtering algorithms whose approximation error is dimension-free. The key to such developments is the decay of correlations property, which is a spatial counterpart of the much better understood stability property of nonlinear filters. For the simplest possible algorithm of this type, our results provide under suitable assumptions an approximation error bound that is uniform both in time and in the model dimension. More broadly, our results provide a framework for the investigation of filtering problems and algorithms in high dimension.

.

CONTENTS

Received March 2013; revised May 2014.

[1]Supported in part by NSF Grants DMS-10-05575 and CAREER-DMS-1148711, and by the ARO through PECASE award W911NF-14-1-0094.

MSC2010 subject classifications. 60G35, 60K35, 62M20, 65C05, 68Q87.

Key words and phrases. Filtering in high dimension, local particle filters, curse of dimensionality, interacting Markov chains, decay of correlations, filter stability, data assimilation.

1. Introduction and background. A fundamental problem in a broad range of applications is the combination of observed data and dynamical models. Particularly in highly complex systems with partial observations, the effective extraction and utilization of the information contained in observed data can only be accomplished by exploiting the availability of accurate predictive models of the underlying dynamical phenomena of interest. Such problems arise in applications that range from classical tracking problems in navigation and robotics to extremely large-scale problems such as weather forecasting. In the latter setting, and in other complex applications in the geophysical, atmospheric and ocean sciences, incorporating observed data into dynamical models is called *data assimilation*.

From a probabilistic perspective, it is in principle simple to formulate the optimal solution to the data assimilation problem. We model the dynamics and observations jointly as a bivariate Markov chain $(X_n, Y_n)_{n\geq 0}$ taking values in a possibly high-dimensional state space $\mathbb{X} \times \mathbb{Y}$ (throughout this paper we will consider discrete time models for simplicity; continuous time models may also be considered). The process $(X_n)_{n\geq 0}$ describes the underlying dynamics of interest, while the process $(Y_n)_{n\geq 0}$ denotes the observed data. To estimate the hidden state X_n based on the observation history $Y_1, \ldots, Y_n$ to date, we introduce the *nonlinear filter*

$$\pi_n = \mathbf{P}[X_n \in \cdot | Y_1, \ldots, Y_n].$$

If the conditional distribution π_n can be computed, it yields an optimal (least mean square) estimate of X_n as well as a complete representation of the uncertainty in this estimate. Moreover, an important property of the filter is that it is recursive: π_n depends only on π_{n-1} and the new observation Y_n. This is crucial in practice, as it allows the filter to be implemented on-line over a long time horizon.

In practice, however, the optimal filter is almost never directly computable: it requires the propagation of an entire conditional distribution, which generally does not admit any efficiently computable sufficient statistics. The practical implementation of nonlinear filtering was therefore long considered to be intractable until the discovery of a class of surprisingly efficient sequential Monte Carlo algorithms, known as *particle filters*, for approximating the filter. The simplest such algorithm simply inserts a random sampling step in the filtering recursion and approximates the filter π_n by the resulting empirical measure $\hat{\pi}_n$ (cf. Section 1.1 below). It is not difficult to show that this gives rise to a standard Monte Carlo error

$$\sup_{|f|\leq 1} \mathbf{E}|\pi_n(f) - \hat{\pi}_n(f)| \leq \frac{C}{\sqrt{N}},$$

where N denotes the number of particles. Moreover, a crucial insight is that the constant C typically does not depend on time n due to the stability property of nonlinear filters [5, 6], so that the particle filter can indeed function in an online fashion. Particle filters have proved to perform extraordinarily well in many classical applications and are widely used in practice. We refer to [5] for a detailed overview of particle filtering algorithms and their analysis.

Unfortunately, despite their widespread success, particle filters have nonetheless proved to be essentially useless in truly complex data assimilation problems. The reason for this, long known to practitioners, has only recently been subjected to mathematical analysis in the work of Bickel et al. [3, 16]. Roughly speaking, the constant C in the above bound, while independent of time n, must typically be exponential in the dimension of the state space of the underlying model. This *curse of dimensionality* does not affect most classical tracking problems, whose dimension is typically of order unity, but becomes absolutely prohibitive in large-scale data assimilation problems such as weather forecasting where model dimensions of order 10^7 are routinely encountered [1]. While the curse of dimensionality problem in particle filters is now fairly well understood, there exists no rigorous approach to date for alleviating this problem [2, 15, 19]. Practical data assimilation in high-dimensional models is therefore generally performed by means of ad-hoc algorithms, frequently based on (questionable) Gaussian approximations, that possess limited theoretical justification [1, 9, 11]. The development of ideas that could enable the principled use of particle filters in high-dimensional settings remains a fundamental open problem in data assimilation and in numerous other complex filtering problems (e.g., multitarget tracking, tracking the spread of epidemics, traffic flow prediction in freeway networks, etc.).

At the same time, the mathematical theory of nonlinear filtering in high dimension has remained essentially in its infancy. Despite that the study of large-scale interacting systems is an important topic in contemporary probability theory (frequently motivated by problems in statistical mechanics, e.g., [8, 12]), almost nothing is known about the emergence of high-dimensional phenomena in the setting of conditional distributions. It is not even entirely clear how filtering problems in high dimension can be fruitfully formulated, and what type of models should be investigated in this setting. Moreover, most mathematical tools used in nonlinear filtering theory (cf. [5]) are ill-suited to the investigation of the much more delicate problems that arise in high dimension. We have recently begun to explore high-dimensional probabilistic phenomena in nonlinear filtering [13, 18]. The present paper arose from the realization that such phenomena are not only of interest in their own right, but that they can provide mechanisms that enable the development and analysis of particle filtering algorithms in high dimension.

The central idea of this paper is that the *decay of correlations* property of high-dimensional filtering models, which is in essence a spatial counterpart of the much better understood stability property of nonlinear filters, can be exploited to develop *local* particle filters that avoid the curse of dimensionality. For the simplest

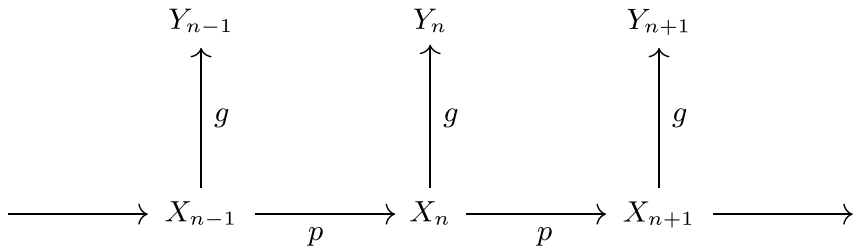

FIG. 1. *Dependency graph of a hidden Markov model.*

possible algorithm of this type, we will prove under suitable assumptions an approximation error bound that is uniform both in time and in the model dimension. While it is far from clear whether this simple algorithm is of immediate practical utility in the most complex real-world applications (a question far beyond the scope of this paper; cf. Section 2.3), our results provide the first rigorous proof of concept that it is in fact possible, at least in principle, to develop particle filtering algorithms whose approximation error is dimension-free. A broader goal of this paper is to introduce a natural foundation for the investigation of filtering problems and algorithms in high dimension, as well as some basic mathematical tools for this purpose.

In the remainder of this section, we provide some essential background on nonlinear filtering, particle filtering algorithms and the curse of dimensionality, as well as a brief overview of the general ideas and contributions of this paper.

1.1. *Classical filtering models and particle filters.* A *hidden Markov model* is a Markov chain $(X_n, Y_n)_{n\geq 0}$ whose transition probability P can be factored as

$$P\big((x,y),A\big) = \int \mathbf{1}_A(x',y')\,p(x,x')\,g(x',y')\,\psi(dx')\,\varphi(dy').$$

Thus, $(X_n)_{n\geq 0}$ is itself a Markov chain in a Polish state space $\mathbb{X}$ with transition density $p:\mathbb{X}\times\mathbb{X}\to\mathbb{R}_+$ with respect to a given reference measure ψ, while $(Y_n)_{n\geq 0}$ are conditionally independent given $(X_n)_{n\geq 0}$ in a Polish state space $\mathbb{Y}$ with transition density $g:\mathbb{X}\times\mathbb{Y}\to\mathbb{R}_+$ with respect to a reference measure φ. This dependence structure is illustrated in Figure 1. We interpret $(X_n)_{n\geq 0}$ as an underlying dynamical process that is not directly observable, while the observable process $(Y_n)_{n\geq 0}$ consists of partial and noisy observations of $(X_n)_{n\geq 0}$.

In the following, we will assume that the process $(X_n, Y_n)_{n\geq 0}$ is realized on its canonical probability space, and denote for any probability measure μ on $\mathbb{X}$ by $\mathbf{P}^\mu$ the probability measure under which $(X_n, Y_n)_{n\geq 0}$ is a hidden Markov model with transition probability P as above and with initial condition $X_0\sim\mu$. For $x\in\mathbb{X}$, we write for simplicity $\mathbf{P}^x := \mathbf{P}^{\delta_x}$. As the process $(X_n)_{n\geq 0}$ is unobservable, a central

problem in this setting is to track the unobserved state X_n given the observation history $Y_1, \ldots, Y_n$, that is, we aim to compute the *nonlinear filter*

$$\pi_n^\mu := \mathbf{P}^\mu[X_n \in \cdot | Y_1, \ldots, Y_n].$$

It is well known, and easily verified using the Bayes formula, that the filter π_n^μ can be computed recursively, that is, we have the recursion (see, e.g., [5])

$$\pi_0^\mu = \mu, \qquad \pi_n^\mu = \mathsf{F}_n \pi_{n-1}^\mu \qquad (n \geq 1),$$

where

$$(\mathsf{F}_n \rho)(f) := \frac{\int f(x') g(x', Y_n) p(x, x') \psi(dx') \rho(dx)}{\int g(x', Y_n) p(x, x') \psi(dx') \rho(dx)}.$$

It is instructive to write the recursion $\mathsf{F}_n := \mathsf{C}_n \mathsf{P}$ in two steps

$$\pi_{n-1}^\mu \overset{\text{prediction}}{\longrightarrow} \pi_{n-}^\mu = \mathsf{P} \pi_{n-1}^\mu \overset{\text{correction}}{\longrightarrow} \pi_n^\mu = \mathsf{C}_n \pi_{n-}^\mu,$$

where

$$(\mathsf{P}\rho)(f) := \int f(x') p(x, x') \psi(dx') \rho(dx),$$

$$(\mathsf{C}_n \rho)(f) := \frac{\int f(x) g(x, Y_n) \rho(dx)}{\int g(x, Y_n) \rho(dx)}.$$

In the prediction step, the filter π_{n-1}^μ is propagated forward using the dynamics of the underlying unobserved process $(X_n)_{n \geq 0}$ to compute the predictive distribution $\pi_{n-}^\mu := \mathbf{P}^\mu[X_n \in \cdot | Y_1, \ldots, Y_{n-1}]$. Then, in the correction step, the predictive distribution is conditioned on the new observation Y_n to obtain the filter π_n^μ.

The recursive structure of the nonlinear filter is of central importance, as it allows the filter to be computed on-line over a long time horizon. Nonetheless, the recursion is still at the level of probability measures, and in general no finite-dimensional sufficient statistics exist. Therefore, the practical implementation of nonlinear filters typically proceeds by Monte Carlo approximation. The most common algorithm of this type simply inserts a sampling step in the filtering recursion: π_n^μ is approximated by the empirical distribution $\hat{\pi}_n^\mu$ computed by the recursion

$$\hat{\pi}_0^\mu = \mu, \qquad \hat{\pi}_n^\mu = \hat{\mathsf{F}}_n \hat{\pi}_{n-1}^\mu \qquad (n \geq 1),$$

where $\hat{\mathsf{F}}_n := \mathsf{C}_n \mathsf{S}^N \mathsf{P}$ consists of three steps

$$\hat{\pi}_{n-1}^\mu \overset{\text{prediction}}{\longrightarrow} \mathsf{P}\hat{\pi}_{n-1}^\mu \overset{\text{sampling}}{\longrightarrow} \hat{\pi}_{n-}^\mu = \mathsf{S}^N \mathsf{P} \hat{\pi}_{n-1}^\mu \overset{\text{correction}}{\longrightarrow} \hat{\pi}_n^\mu = \mathsf{C}_n \hat{\pi}_{n-}^\mu.$$

Here, $N \geq 1$ is the number of particles used in the algorithm, and S^N is the sampling operator that defines for a probability measure ρ the random measure

$$\mathsf{S}^N \rho := \frac{1}{N} \sum_{i=1}^N \delta_{x(i)}, \qquad (x(i))_{i=1,\ldots,N} \text{ are i.i.d. samples} \sim \rho$$

Algorithm 1: Bootstrap particle filter

Let $\hat{\pi}_0^\mu = \mu$;
for $k = 1, \ldots, n$ **do**
  Sample i.i.d. $\hat{x}_{k-1}(i)$, $i = 1, \ldots, N$ from the distribution $\hat{\pi}_{k-1}^\mu$;
  Sample $x_k(i) \sim p(\hat{x}_{k-1}(i), \cdot)\, d\psi$, $i = 1, \ldots, N$;
  Compute $w_k(i) = g(x_k(i), Y_k) / \sum_{\ell=1}^N g(x_k(\ell), Y_k)$, $i = 1, \ldots, N$;
  Let $\hat{\pi}_k^\mu = \sum_{i=1}^N w_k(i) \delta_{x_k(i)}$;
end

FIG. 2. *The classical bootstrap particle filtering algorithm.*

[if ρ is a random measure, then $(x(i))_{i=1,\ldots,N}$ are drawn conditionally given ρ]. This yields the bootstrap particle filtering algorithm described in Figure 2. This algorithm is exceedingly simple to implement, and it is easily shown that the particle filter $\hat{\pi}_n^\mu$ converges to the exact filter π_n^μ as $N \to \infty$. We refer to [5] for a detailed overview of particle filtering algorithms and their analysis.

To gain some insight into the approximation properties of the particle filter, let us perform the simplest possible error analysis. We define the distance

$$|||\rho - \rho'||| := \sup_{|f| \le 1} \mathbf{E}\big[|\rho(f) - \rho'(f)|^2\big]^{1/2}$$

between two random measures ρ, ρ' on $\mathbb{X}$. It is an easy exercise to show that

$$|||\mathsf{P}\rho - \mathsf{P}\rho'||| \le |||\rho - \rho'|||, \qquad |||\rho - \mathsf{S}^N \rho||| \le \frac{1}{\sqrt{N}}.$$

Let us assume for simplicity that the observation density g is bounded away from zero and infinity, that is, $\kappa \le g(x, y) \le \kappa^{-1}$ for some $0 < \kappa < 1$. As

$$(\mathsf{C}_n\rho)(f) - \big(\mathsf{C}_n\rho'\big)(f) = \frac{\kappa^{-1}}{\rho(g_n)}\big\{\rho(\kappa f g_n) - \rho'(\kappa f g_n)\big\} + \frac{\rho'(f g_n)}{\rho'(g_n)} \frac{\kappa^{-1}}{\rho(g_n)}\big\{\rho'(\kappa g_n) - \rho(\kappa g_n)\big\}$$

with $g_n(x) := g(x, Y_n)$, and as $|\kappa g_n| \le 1$ and $\rho(g_n) \ge \kappa$, we obtain

$$|||\mathsf{C}_n\rho - \mathsf{C}_n\rho'||| \le 2\kappa^{-2}|||\rho - \rho'|||.$$

Putting these bounds together, we find that

$$|||\pi_n^\mu - \hat{\pi}_n^\mu||| \le 2\kappa^{-2}\left\{\frac{1}{\sqrt{N}} + |||\pi_{n-1}^\mu - \hat{\pi}_{n-1}^\mu|||\right\} \le \frac{\sum_{i=1}^n (2\kappa^{-2})^i}{\sqrt{N}},$$

where the second inequality is obtained by iterating the first inequality n times. We therefore find that the bootstrap particle filter does indeed approximate the exact nonlinear filter with the typical Monte Carlo $1/\sqrt{N}$-rate.

It should be noted that our crude error bound grows exponentially in time n. If the error were in fact to grow exponentially in time, this would make the particle

filter largely useless in practice as it could not be run reliably for more than a few time steps (in particular, it could not be run on-line over a long time horizon). Fortunately, however, the exponential growth of the error is an artifact of our crude bound and typically does not occur in practice. We have omitted to take into account an essential phenomenon: ergodicity of the underlying model will cause the filter to be *stable*, that is, π_n^μ forgets its initial condition μ as $n \to \infty$. The stability property provides a dissipation mechanism that mitigates the accumulation of approximation errors over time. A more sophisticated analysis that exploits this idea yields a time-uniform error bound; see Section 3.1 below.

1.2. *The curse of dimensionality.* We have stated that particle filters suffer from the curse of dimensionality. It is, however, far from obvious at this point why this should be the case: no explicit notion of dimension appears in the above error bound. To understand why the above bound is typically exponential in the model dimension, we must consider a suitable class of models in which the dependence on dimension can be explicitly investigated. In Section 2, we will introduce a general class of high-dimensional filtering models that is prototypical of many data assimilation problems. In the present section, however, we consider a much simpler class of *trivial* models that is useless in any application, but is nonetheless helpful for developing intuition for dimensionality issues in particle filters.

In a d-dimensional model, X_n, Y_n are each described by d coordinates X_n^i, Y_n^i, $i = 1, \ldots, d$. To construct a trivial d-dimensional model, we simply start with a given one-dimensional model and duplicate it d times. That is, let $(\tilde{X}_n, \tilde{Y}_n)_{n\geq 0}$ be a hidden Markov model on $\tilde{\mathbb{X}} \times \tilde{\mathbb{Y}}$ with transition density $\tilde{p}$ and observation density $\tilde{g}$ with respect to reference measures $\tilde{\psi}$ and $\tilde{\varphi}$, respectively. Then we set

$$\mathbb{X} = \tilde{\mathbb{X}}^d, \qquad \mathbb{Y} = \tilde{\mathbb{Y}}^d, \qquad \psi = \tilde{\psi}^{\otimes d}, \qquad \varphi = \tilde{\varphi}^{\otimes d}$$

and

$$p(x, z) = \prod_{i=1}^{d} \tilde{p}(x^i, z^i), \qquad g(x, y) = \prod_{i=1}^{d} \tilde{g}(x^i, y^i),$$

so that each coordinate $(X_n^i, Y_n^i)_{n\geq 0}$ is an independent copy of $(\tilde{X}_n, \tilde{Y}_n)_{n\geq 0}$. Note that we have used the term d-dimensional in the sense that our model has d independent degrees of freedom: each degree of freedom can itself in principle take values in a high- or even infinite-dimensional state space $\tilde{\mathbb{X}} \times \tilde{\mathbb{Y}}$. This is, however, precisely the notion of dimension that is relevant to the curse of dimensionality (in [3, 16] this idea is sharpened by a notion of "effective dimension").

In this trivial setting, it is now easily seen how the curse of dimensionality arises in our error bound. Indeed, let us assume again for simplicity that $\kappa \leq \tilde{g}(\tilde{x}, \tilde{y}) \leq \kappa^{-1}$ for some $0 < \kappa < 1$. Then $\kappa^d \leq g(x, y) \leq \kappa^{-d}$, so we obtain a bound that is exponential in the dimension d even after only one time step:

$$\|\!|\pi_1^\mu - \hat{\pi}_1^\mu|\!\| \leq \frac{2\kappa^{-2d}}{\sqrt{N}}.$$

An inspection of our bound clarifies the source of this exponential growth: even though the Monte Carlo sampling itself is dimension-free ($|||\rho - \mathsf{S}^N\rho||| \le N^{-1/2}$ independent of dimension), the correction operator C_n blows up the sampling error exponentially in high dimension (this is a manifestation of the fact that the prior ρ and posterior $\mathsf{C}_n\rho$ measures are nearly singular in high dimension, so that random samples drawn from ρ have exponentially small likelihood under $\mathsf{C}_n\rho$). In particular, it is evidently the dimension of the observations, rather than that of the underlying model, that controls the exponential growth in our error bound.

Of course, the above analysis is far from convincing. First, we have only proved a rather crude upper bound on the approximation error: could a more sophisticated bound eliminate the exponential dependence on dimension as was done using the filter stability property to eliminate the exponential dependence on time? Second, one could argue that a good approximation of $\pi_n(f)$ for *any* function f (as is implicit in the definition of the $|||\cdot|||$-norm) is too much to ask for in high dimension: could a *local* notion of approximation avoid the exponential dependence on dimension? Unfortunately, neither of these ideas can help us avoid the curse of dimensionality of the bootstrap particle filter, which is a genuine phenomenon and not a mathematical deficiency of our analysis. As a simple illustration of this phenomenon, we note that even if $f(x)$ is a function that depends on a single dimension x^i only [any reasonable approximation of $\pi_n(f)$ should work at least for such local functions] and if $\mu = \delta_x$, the asymptotic variance σ_f^2 in the central limit theorem

$$\sqrt{N}\{\pi_1^\mu(f) - \hat{\pi}_1^\mu(f)\} \Longrightarrow N(0, \sigma_f^2) \qquad \text{as } N \to \infty$$

grows exponentially in the dimension d (the computation of σ_f is a simple exercise that is left to the interested reader), which suggests that our crude upper bound is qualitatively correct. The more delicate analysis of Bickel et al. [3, 16], which allows d to grow with N, demonstrates conclusively that the bootstrap particle filter cannot approximate the filter unless the number of particles N grows exponentially in the dimension d. Nonetheless, both the ideas raised above to eliminate the exponential dependence on dimension will play an important role in the remainder of this paper, as will be explained in the next section.

REMARK 1.1. The problem of sampling from a weighted measure of the form $(\mathsf{C}\rho)(dx) := g(x)\rho(dx)/\int g(z)\rho(dz)$ appears in numerous applications in statistics, computer science and physics. The naive approximation $\mathsf{C}\rho \approx \mathsf{C}\mathsf{S}^N\rho$ is well known to be useless in large-scale problems: instead, Markov Chain Monte Carlo (MCMC) methods are almost universally used for this purpose. However, even if we were able to sample *exactly* from the weighted measure $\mathsf{C}\rho$, this would still not resolve our problems in the filtering context. Indeed, if we implement the "optimal proposal" (cf. [15]) particle filtering recursion $\hat{\mathsf{F}}_n = \mathsf{S}^N\mathsf{C}_n\mathsf{P}$ rather than the bootstrap filter $\hat{\mathsf{F}}_n = \mathsf{C}_n\mathsf{S}^N\mathsf{P}$, then the error between $\hat{\pi}_1^\mu = \hat{\mathsf{F}}_1\mu$ and $\pi_1^\mu = \mathsf{F}_1\mu$ would

be dimension-free, but the error between $\hat{\pi}_2^\mu = \hat{\mathsf{F}}_2 \hat{\pi}_1^\mu$ and $\pi_2^\mu = \mathsf{F}_2 \pi_1^\mu$ would again exhibit exponential dependence on the dimension due to the sampling performed in the first time step. The curse of dimensionality would therefore still arise essentially as above due to the recursive nature of the filtering problem.

If, instead of computing the filter $\mathbf{P}[X_n \in \cdot | Y_1, \ldots, Y_n]$, we wish to compute the full conditional path distribution $\mathbf{P}[X_0, \ldots, X_n \in \cdot | Y_1, \ldots, Y_n]$ (known as the smoothing problem), MCMC methods can be successfully employed in high dimension. However, this procedure requires the entire history of observations and is not recursive, so that it cannot be implemented on-line and is impractical over a long time horizon (cf. [2]). The crucial question to be addressed is therefore whether it is possible to develop filtering algorithms that are both recursive and that admit error bounds that are uniform in time and in the model dimension.

1.3. *Contributions of this paper.* While the curse of dimensionality in particle filters is now fairly well understood, it is far from clear how one could go about addressing this problem. Several fundamental questions arise directly:

1. What sort of filtering models are natural to investigate in high dimension?
2. What sort of mechanism might allow to surmount the curse of dimensionality? How can such a mechanism be exploited algorithmically?
3. What sort of mathematical tools are needed to address such problems?

We aim to address each of these questions in the sequel. We will presently provide an informal discussion of some basic ideas in this paper; much of the remainder of the paper will be devoted to making these ideas precise.

Some basic insight can be obtained by considering again the trivial model of the previous section. Despite that the bootstrap particle filter suffers from the curse of dimensionality when applied to the full model, it is obvious in this case that one can surmount this problem in a trivial fashion: as each of the coordinates is independent, one can simply run an independent bootstrap filter in each coordinate. It is evident that the local error of this algorithm (i.e., the error of the marginal of the filter in each coordinate) is, by construction, independent of the model dimension (i.e., the number of coordinates). Even though this approach exploits a very special property of the trivial model—the independence of the coordinates—we will see that the same basic idea can be implemented in a much more general setting.

In most data assimilation problems, the high-dimensional nature of the model is essentially due to its spatial structure: the aim of the problem is to track the dynamics of a random field (e.g., the atmospheric pressure and temperature fields in the case of weather forecasting). In this paper, we take as a starting point the notion that the coordinates X_n^v, Y_n^v $(v \in V)$ of our hidden Markov model are indexed by a large graph $G = (V, E)$ that represents the spatial degrees of freedom of the model, and that its interactions are local: the dynamics and observations at a spatial location depends only on the states at locations in a neighborhood, as is illustrated in Figure 3 below. While the law of the model at each spatial location

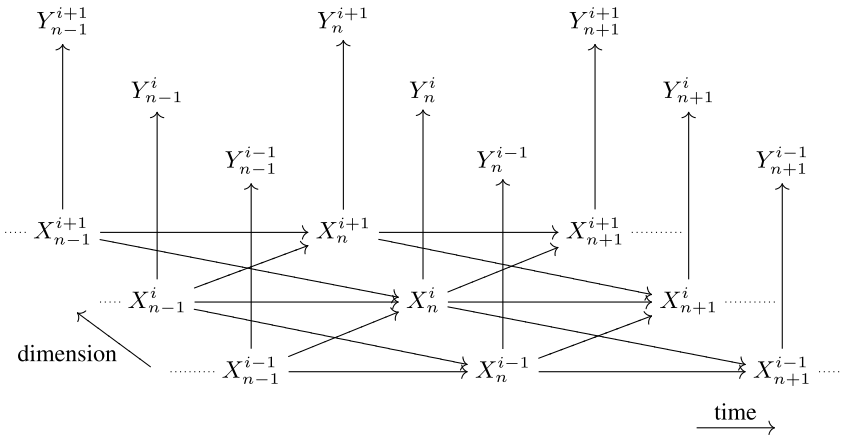

FIG. 3. *Dependency graph of a high-dimensional filtering model of the type considered in this paper.*

is no longer independent as in the trivial model of the previous section, large-scale interacting systems can nonetheless exhibit an approximate version of this property: this is the *decay of correlations* phenomenon that has been particularly well studied in statistical mechanics [8]. Informally speaking, while the states X_n^v and X_n^w at two sites $v, w \in V$ are probably quite strongly correlated when v and w are close together, one might expect that X_n^v and X_n^w are nearly independent when v and w are far apart with respect to the natural distance in the graph G.[2]

The core idea of this paper is that the decay of correlations property can provide a mechanism to surmount the curse of dimensionality. A speculative back-of-the-envelope computation explains how this might work. Due to the decay of correlations, the conditional distribution of the site X_n^v given the new observation Y_n should not depend significantly on observations Y_n^w at sites w distant from v. Suppose we can develop a *local* particle filtering algorithm that at each site v only uses observations in a local neighborhood K of v to update the filtering distribution. As we have seen in the previous section, the sampling error is controlled by the dimension of the observations: as we have now restricted to observations in K, the sampling error at each site will be exponential only in $\operatorname{card} K$ rather than in the full dimension $\operatorname{card} V$. On the other hand, the truncation to observations in K is only approximate: the decay of correlations property suggests that the bias introduced by this truncation should decay exponentially in $\operatorname{diam} K$. Therefore,

$$\text{error} = \text{bias} + \text{variance} \approx e^{-\operatorname{diam} K} + \frac{e^{\operatorname{card} K}}{\sqrt{N}}.$$

[2]The precise formulation of the decay of correlations property that will be used in our analysis is determined by the mathematical machinery that will be used in the proofs; cf. Sections 3.2 and 4.2.

If the size of the neighborhoods K is chosen so as to optimize the error, then the resulting algorithm is evidently consistent (with a slower convergence rate than the standard $1/\sqrt{N}$ Monte Carlo rate: this is likely unavoidable in high dimension) with an error bound that is independent of the model dimension card V.

The main result of this paper is that these speculative ideas can be made precise at least for one particularly simple local filtering algorithm: the block particle filter (Section 2.2). While the above back-of-envelope computation provides a basic template for our approach, the rigorous implementation of these ideas requires the introduction of mathematical machinery that has not previously been applied in the study of nonlinear filtering. Just as in the case of the filter stability property (see [18] and the references therein), it is far from clear that any decay of correlations properties of the underlying model are inherited by the filter as we have taken for granted above: in fact, striking counterexamples show that such inheritance can fail in surprising ways [13]. More generally, the development of machinery for the *local* analysis of high-dimensional filtering problems forms an essential part of our proofs. An outline of the main steps and ideas in the proof of our main result will be given in Section 3; detailed proofs are given in Section 4.

It should be emphasized that our result, while providing a first rigorous analysis of a local particle filtering algorithm in high dimension, is essentially a proof of concept. The general idea to exploit decay of correlations provides a promising approach to the curse of dimensionality problem (such a possibility has also been occasionally mentioned in the applied literature, e.g., [16, 19]); however, the block particle filter that we analyze is the simplest possible algorithm of its type, and possesses some inherent limitations that can potentially be addressed by the development of more sophisticated local particle filters. In Section 2.3, we will discuss some limitations of our results and potential directions for further investigation.

2. Main result and discussion.

2.1. *Filtering models in high dimension.* This paper is concerned with filtering problems in high dimension. In order to investigate such problems systematically, we presently introduce a class of high-dimensional filtering models that will provide the basic framework to be investigated throughout this paper. In these models, the state (X_n, Y_n) at each time n is a random field $(X_n^v, Y_n^v)_{v\in V}$ indexed by a (finite) undirected graph $G = (V, E)$. The graph G describes the spatial degrees of freedom of the model, and the underlying dynamics and observations are local with respect to the graph structure in a sense to be made precise below. The dimension of the model should be interpreted as the cardinality of the vertex set V, which is typically assumed to be large. Our aim is to develop quantitative results that are, under appropriate assumptions, independent of the dimension card V.

We now define the hidden Markov model $(X_n, Y_n)_{n\geq 0}$ to be considered in the sequel (we will adopt throughout the basic setting and notation introduced in Section 1.1). The state spaces $\mathbb{X}$ and $\mathbb{Y}$ of X_n and Y_n, and the reference measures ψ

and φ of the transition densities p and g, respectively, are of product form

$$\mathbb{X} = \prod_{v \in V} \mathbb{X}^v, \qquad \mathbb{Y} = \prod_{v \in V} \mathbb{Y}^v, \qquad \psi = \bigotimes_{v \in V} \psi^v, \qquad \varphi = \bigotimes_{v \in V} \varphi^v,$$

where ψ^v and φ^v are reference measures on the Polish spaces $\mathbb{X}^v$ and $\mathbb{Y}^v$, respectively. The transition densities p and g are given by

$$p(x, z) = \prod_{v \in V} p^v(x, z^v), \qquad g(x, y) = \prod_{v \in V} g^v(x^v, y^v),$$

where $p^v : \mathbb{X} \times \mathbb{X}^v \to \mathbb{R}_+$ and $g^v : \mathbb{X}^v \times \mathbb{Y}^v \to \mathbb{R}_+$ are transition densities with respect to the reference measures ψ^v and φ^v, respectively.

The spatial graph G is endowed with its natural distance d [i.e., $d(v, v')$ is the length of the shortest path in G between $v, v' \in V$]. Let us fix throughout a neighborhood size $r \in \mathbb{N}$, and define for each vertex $v \in V$ the r-neighborhood

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

We will assume that the dynamics of the underlying process $(X_n)_{n \ge 0}$ is *local* in the sense that $p^v(x, z^v)$ depends on $x^{N(v)}$ only [we write $x^J = (x^j)_{j \in J}$ for $J \subseteq V$]:

$$p^v(x, z^v) = p^v(\tilde{x}, z^v) \qquad \text{whenever } x^{N(v)} = \tilde{x}^{N(v)}.$$

That is, the conditional distribution of X_n^v given $X_0, \dots, X_{n-1}$ depends on $X_{n-1}^{N(v)}$ only. Similarly, by construction, the observations are local in that the conditional distribution of Y_n^v given X_n depends on X_n^v only. This dependence structure is illustrated in Figure 3 (in the simplest case of a linear graph G with $r = 1$).

Markov models of the form introduced above appear in the literature under various names, such as locally interacting Markov chains or probabilistic cellular automata [7, 10]. Such models arise naturally in numerous complex and large-scale applications, including percolation models of disease spread or forest fires, freeway traffic flow models, probabilistic models on networks and large-scale queueing systems, and various biological, ecological and neural models. Moreover, local Markov processes of this type arise naturally from finite-difference approximation of stochastic partial differential equations, and are therefore in principle applicable to a diverse set of data assimilation problems that arise in areas such as weather forecasting, oceanography and geophysics (cf. Section 2.3.3). While more general models are certainly of substantial interest, the model defined above is prototypical of a broad range of high-dimensional data assimilation problems and provides a basic setting for the investigation of filtering problems in high dimension.

2.2. *Block particle filter*: *Dimension-free bounds*. As was explained in Section 1.2, the bootstrap particle filter is not well suited to high-dimensional models: the approximation error generally grows exponentially in the model dimension

card V. To surmount this problem, we aim to develop *local* particle filtering algorithms that can exploit decay of correlations properties of the underlying filtering model. In this paper, we will investigate in detail the simplest possible algorithm of this type, the *block particle filter*, that will be introduced presently. While this algorithm possesses some inherent limitations (see below), it is the simplest local algorithm both mathematically and computationally and, therefore, provides an ideal starting point for the investigation of particle filters in high dimension.

To define the block particle filtering algorithm, we begin by introducing a partition $\mathcal{K}$ of the vertex set V into nonoverlapping blocks, that is, we have

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

We now define the blocking operator

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

where for any measure ρ on $\mathbb{X} = \prod_{v\in V} \mathbb{X}^v$ and $J \subseteq V$ we denote by $\mathsf{B}^J\rho$ the marginal of ρ on $\prod_{v\in J} \mathbb{X}^v$. The random field described by the measure $\mathsf{B}\rho$ on $\mathbb{X}$ is independent across different blocks defined by the partition $\mathcal{K}$, while the marginal on each block agrees with the original measure ρ. The block particle filter inserts an additional blocking step into the bootstrap particle filter recursion, that is,

$$\hat\pi_0^\mu = \mu, \qquad \hat\pi_n^\mu = \hat{\mathsf{F}}_n \hat\pi_{n-1}^\mu \qquad (n \geq 1),$$

where $\hat{\mathsf{F}}_n := \mathsf{C}_n \mathsf{B} \mathsf{S}^N \mathsf{P}$ consists of four steps

$$\hat\pi_{n-1}^\mu \overset{\text{prediction/sampling}}{\longrightarrow} \hat\pi_{n-}^\mu = \mathsf{S}^N \mathsf{P} \hat\pi_{n-1}^\mu \overset{\text{blocking/correction}}{\longrightarrow} \hat\pi_n^\mu = \mathsf{C}_n \mathsf{B} \hat\pi_{n-}^\mu.$$

The resulting algorithm is given in Figure 4. In the special case $\mathcal{K} = \{V\}$, the block particle filter reduces to the bootstrap particle filter, so that the former is a strict

Algorithm 2: Block particle filter

Let $\hat\pi_0^\mu = \mu$;
for $k = 1, \ldots, n$ **do**
  Sample i.i.d. $\hat x_{k-1}(i)$, $i = 1, \ldots, N$ from the distribution $\hat\pi_{k-1}^\mu$;
  Sample $x_k^v(i) \sim p^v(\hat x_{k-1}(i), \cdot)\, d\psi^v$, $i = 1, \ldots, N$, $v \in V$;
  Compute $w_k^K(i) = \frac{\prod_{v\in K} g^v(x_k^v(i), Y_k^v)}{\sum_{\ell=1}^N \prod_{v\in K} g^v(x_k^v(\ell), Y_k^v)}$, $i = 1, \ldots, N$, $K \in \mathcal{K}$;
  Let $\hat\pi_k^\mu = \bigotimes_{K\in\mathcal{K}} \sum_{i=1}^N w_k^K(i) \delta_{x_k^K(i)}$;
end

FIG. 4. *The block particle filtering algorithm considered in this paper. Note that sampling $\hat x$ from a product distribution $\bigotimes_{K\in\mathcal{K}} \rho^K$ is implemented by sampling independently $\hat x^K \sim \rho^K$, $K \in \mathcal{K}$.*

generalization of the latter (we have therefore not introduced a separate notation for the bootstrap particle filter: in the sequel, the notation $\hat{\pi}_n^\mu$ *always refers to the block particle filter*). The introduction of independent blocks allows to localize the algorithm, however, which will be crucial in the high-dimensional setting.

It is immediately evident from inspection of the block particle filtering algorithm that only observations in block K are used by the algorithm to update the filtering distribution in block K. Therefore, following the heuristic ideas of Section 1.3, we expect that the sampling error of the algorithm is exponential in $\operatorname{card} K$ rather than in the model dimension $\operatorname{card} V$. To control the bias introduced by the blocking step, note that the blocking operator $\mathsf{B}\rho$ decouples the distribution ρ at the boundaries of the blocks. The decay of correlations property (if it can be established) should cause the influence of such a perturbation on the marginal distribution at a vertex $v \in K$ to decay exponentially in the distance from v to the boundary of the block K. Thus, the back-of-the-envelope computation in Section 1.3 applies to the local error at "most" vertices, as the boundaries of the blocks only constitute a small fraction of the total number of vertices. On the other hand, the error will necessarily be larger for vertices closer to the block boundaries. This spatial inhomogeneity of the local error is an inherent limitation of the block particle filter that one might hope to alleviate by the development of more sophisticated local particle filters. We postpone further discussion of this point to Section 2.3.2.

Having introduced the block particle filtering algorithm, we now proceed to formulate the main result of this paper (Theorem 2.1 below).

Recall that we have introduced the neighborhoods

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

above, where the neighborhood size r is fixed throughout this paper [in our model, the state of vertex v depends only on the states of vertices in $N(v)$ in the previous time step]. Given a set $J \subseteq V$, we denote the r-inner boundary of J as

$$\partial J := \{v \in J : N(v) \not\subseteq J\}$$

(i.e., ∂J is the subset of vertices in J that can interact with vertices outside J in one step of the dynamics). We also define the following quantities:

$$|\mathcal{K}|_\infty := \max_{K \in \mathcal{K}} \operatorname{card} K,$$

$$\Delta := \max_{v \in V} \operatorname{card}\{v' \in V : d(v, v') \le r\},$$

$$\Delta_{\mathcal{K}} := \max_{K \in \mathcal{K}} \operatorname{card}\{K' \in \mathcal{K} : d(K, K') \le r\},$$

where we define as usual $d(J, J') := \min_{v \in J} \min_{v' \in J'} d(v, v')$ for $J, J' \subseteq V$. Thus, $|\mathcal{K}|_\infty$ is the maximal size of a block in $\mathcal{K}$, while Δ ($\Delta_{\mathcal{K}}$) is the maximal number of vertices (blocks) that interact with a single vertex (block) in one step of the dynamics. It should be emphasized that r, Δ and $\Delta_{\mathcal{K}}$ are *local* quantities that depend on the geometry but not on the size of the spatial graph G.

Finally, we introduce for $J \subseteq V$ the local distance

$$|||\rho - \rho'|||_J := \sup_{f \in \mathcal{X}^J : |f| \le 1} \mathbf{E}\big[|\rho(f) - \rho'(f)|^2\big]^{1/2}$$

between random measures ρ, ρ' on $\mathbb{X}$, where $\mathcal{X}^J$ denotes the class of measurable functions $f : \mathbb{X} \to \mathbb{R}$ such that $f(x) = f(\tilde{x})$ whenever $x^J = \tilde{x}^J$.

THEOREM 2.1. *There exists a constant $0 < \varepsilon_0 < 1$, depending only on the local quantities Δ and $\Delta_{\mathcal{K}}$, such that the following holds.*

Suppose there exist $\varepsilon_0 < \varepsilon < 1$ and $0 < \kappa < 1$ such that

$$\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1}, \qquad \kappa \le g^v(x^v, y^v) \le \kappa^{-1} \qquad \forall v \in V, x, z \in \mathbb{X}, y \in \mathbb{Y}.$$

Then for every $n \ge 0$, $x \in \mathbb{X}$, $K \in \mathcal{K}$ and $J \subseteq K$ we have

$$|||\pi_n^x - \hat{\pi}_n^x|||_J \le \alpha \operatorname{card} J \left[e^{-\beta_1 d(J, \partial K)} + \frac{e^{\beta_2 |\mathcal{K}|_\infty}}{\sqrt{N}} \right],$$

where the constants $0 < \alpha, \beta_1, \beta_2 < \infty$ depend only on ε, κ, r, Δ and $\Delta_{\mathcal{K}}$.

The key point of this result is that both the assumptions and the resulting error bound depend only on *local* quantities. In particular, the assumptions and error bound depend neither on time n nor on the model dimension $\operatorname{card} V$.

REMARK 2.2. A threshold requirement of the form $\varepsilon > \varepsilon_0$ is essential in order to obtain the decay of correlations property, which can fail if $\varepsilon > 0$ is too small (a phenomenon known as *phase transition* in statistical mechanics). Otherwise, the assumptions of Theorem 2.1 are comparable to assumptions commonly imposed in the literature to obtain error bounds for the bootstrap particle filter [5, 6] and possess similar limitations. We postpone discussion of these issues to Section 2.3.1.

REMARK 2.3. In Theorem 2.1, we have considered $\pi_n^x := \pi_n^{\delta_x}$ and $\hat{\pi}_n^x := \hat{\pi}_n^{\delta_x}$ with a nonrandom initial condition $x \in \mathbb{X}$. This is a choice of convenience: the proof of Theorem 2.1 yields the same conclusion for more general initial conditions that satisfy a suitable decay of correlations property. On the other hand, the stability property of the filter (Corollary 4.7 below) ensures that π_n^μ forgets its initial condition μ exponentially fast uniformly in the dimension, so there is little loss of generality in choosing a computationally convenient initial condition.

REMARK 2.4. The particle filter $\hat{\pi}_n^\mu$ depends both on the random samples that are drawn in the algorithm and on the random sequence of the observations. However, the randomness of the observations plays no role in our proofs. One can therefore interpret the expectation in the definition of $|||\cdot|||_J$ as being taken only with respect to the random sampling mechanism in the block particle filter, and the bound of Theorem 2.1 as holding uniformly with respect to the observation sequence.

To provide a concrete illustration of Theorem 2.1, we consider in the remainder of this section the example where the spatial graph G is a square lattice, that is,

$$V = \{-d, \dots, d\}^q \qquad (d, q \in \mathbb{N})$$

endowed with its natural edge structure. Note that in this case, the graph distance $d(v, v')$ is simply the ℓ_1-distance between the corresponding vectors of integers. To define the partition $\mathcal{K}$, we cover V by blocks of radius $b \in \mathbb{N}$, that is,

$$\mathcal{K} = \big\{(x + \{-b, \dots, b\}^q) \cap V : x \in (2b+1)\mathbb{Z}^q\big\}.$$

We assume for simplicity in the sequel that $b \geq r$, and that $(2d+1)/(2b+1) \in \mathbb{N}$ is integer so that all $K \in \mathcal{K}$ are translates of $\{-b, \dots, b\}^q$ (this slightly simplifies our arguments below but is not essential to our results). We can easily compute

$$|\mathcal{K}|_\infty = (2b+1)^q, \qquad \Delta \leq (2r+1)^q, \qquad \Delta_{\mathcal{K}} \leq 3^q.$$

Note that these local quantities do not depend on the size d of our lattice. In a data assimilation application one might have, for example, $q = 2$, $r = 1$, $d \sim 10^3$.

Consider the block $K = \{-b, \dots, b\}^q$. Note that for $u = 0, \dots, b - r$

$$\big\{v \in K : d(v, \partial K) > u\big\} = \big\{-(b - r - u), \dots, b - r - u\big\}^q.$$

Fix $0 < \delta < 1$ and choose $u = \lfloor \delta(2b+1)/2q - r \rfloor$. Then

$$\frac{\operatorname{card}\{v \in K : d(v, \partial K) > u\}}{\operatorname{card} K} = \left(\frac{2(b - r - u) + 1}{2b+1}\right)^q \geq 1 - \delta,$$

where we have used $1 - (1-\delta)^{1/q} \geq \delta/q$. The same conclusion evidently holds for every block $K \in \mathcal{K}$. Thus, Theorem 2.1 gives the following corollary.

COROLLARY 2.5. *In the square lattice setting $V = \{-d, \dots, d\}^q$, there exists a constant $0 < \varepsilon_0 < 1$, depending only on r and q, such that the following holds.*

Suppose there exist $\varepsilon_0 < \varepsilon < 1$ and $0 < \kappa < 1$ such that

$$\varepsilon \leq p^v(x, z^v) \leq \varepsilon^{-1}, \qquad \kappa \leq g^v(x^v, y^v) \leq \kappa^{-1} \qquad \forall v \in V, x, z \in \mathbb{X}, y \in \mathbb{Y}.$$

Then for every $x \in \mathbb{X}$, $n \geq 0$ and $0 < \delta < 1$ we have

$$\operatorname{card}\left\{v \in V : \big|\!\big|\!\big|\pi_n^x - \hat{\pi}_n^x\big|\!\big|\!\big|_v \leq \alpha' e^{-\beta_1' \delta (2b+1)} + \alpha' \frac{e^{\beta_2'(2b+1)^q}}{\sqrt{N}}\right\} \geq (1-\delta) \operatorname{card} V,$$

where the constants $0 < \alpha', \beta_1', \beta_2' < \infty$ depend only on ε, κ, r and q.

In particular, if we choose the block size $b = \lfloor \frac{1}{2}(4\beta_2')^{-1/q} \log^{1/q} N - \frac{1}{2} \rfloor$, then

$$\operatorname{card}\big\{v \in V : \big|\!\big|\!\big|\pi_n^x - \hat{\pi}_n^x\big|\!\big|\!\big|_v \leq c_1 e^{-c_2 \delta \log^{1/q} N}\big\} \geq (1-\delta) \operatorname{card} V$$

and (using $\mathbf{E}|Z| = \int_0^\infty \mathbf{P}[|Z| \geq t]\, dt$)

$$\frac{1}{\operatorname{card} V} \sum_{v \in V} \big|\!\big|\!\big|\pi_n^x - \hat{\pi}_n^x\big|\!\big|\!\big|_v \leq \frac{c_3}{\log^{1/q} N},$$

where the constants $0 < c_1, c_2, c_3 < \infty$ depend only on ε, κ, r and q.

Corollary 2.5 makes precise the notion that a properly tuned block particle filter can avoid the curse of dimensionality: choosing the block size $b \sim \log^{1/q} N$, we obtain a local error that can be made arbitrarily small, uniformly both in time n and in the lattice size d, by choosing a sufficiently large sample size N. More precisely, we see that the local error at *most* locations is of order $e^{-c\log^{1/q} N}$, which is polynomial for $q = 1$ and subpolynomial otherwise, while the average local error is similarly uniform in n and d albeit with a very slow convergence rate. It appears that these results are chiefly limited by the spatial inhomogeneity that is inherent in the block particle filtering algorithm, as will be discussed in Section 2.3.2 below.

REMARK 2.6. Theorem 2.1 and Corollary 2.5 should be viewed as a theoretical proof of concept that it is possible, in principle, to design particle filters that avoid the curse of dimensionality. In practice, the slow rate $b \sim \log^{1/q} N$ suggests that the block size must typically be quite small (of order unity) for realistic values of the sample size N, which yields a large bias term in our bounds. We have nonetheless observed in simple simulations that the algorithm can work quite well even with the choice $b = 0$, so that the practical utility of the algorithm may not be fully captured by our mathematical results. Moreover, specific features of certain data assimilation applications, such as sparsity of observations, could make it possible to choose substantially larger blocks. A systematic investigation of the empirical performance of local particle filtering algorithms in applications is beyond the scope of this paper, however. The practical implementation of local particle filters for data assimilation will likely require further advances in all mathematical, methodological and applied aspects of high-dimensional filtering.

2.3. *Discussion.*

2.3.1. *Mixing assumptions and the ergodicity threshold.* The basic assumption of Theorem 2.1 is that the local transition densities are bounded above and below:

$$\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1}, \qquad \kappa \le g^v(x^v, y^v) \le \kappa^{-1}.$$

This is a local counterpart of the mixing assumptions that are routinely employed in the analysis of particle filters [5, 6]. The global mixing assumption $\varepsilon \le p(x, z) \le \varepsilon^{-1}$ would imply that the underlying Markov chain is strongly ergodic (in the sense that its transition kernel is a strict contraction with respect to the total variation distance) and is often used to establish the stability property of the filter; this is essential to obtain a time-uniform bound on the particle filter error. See Section 3.1 below. The local mixing assumption $\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1}$ employed here should similarly be viewed as a local ergodicity assumption on the model.

It is well known that strong mixing assumptions impose some constraints on the underlying model. In particular, they typically hold only in a compact state space:

in a noncompact state space the likelihood ratio $p(x,z)/p(x',z)$ is typically unbounded as $|z|\to\infty$, while $\varepsilon \le p(x,z) \le \varepsilon^{-1}$ would imply that $p(x,z)/p(x',z)$ is uniformly bounded. While qualitative results in this area have been obtained in much more general settings (cf. [18] and the references therein), it has proved to be more difficult to obtain quantitative results under assumptions weaker than strong mixing conditions. These technical issues are however unrelated to the problems that arise in high dimension, and we do not address them here.

On the other hand, there is a crucial assumption in Theorem 2.1 that does not arise in finite dimension. In classical results on particle filters, it is assumed that $\varepsilon \le p(x,z) \le \varepsilon^{-1}$ with $\varepsilon > 0$. For the local assumption $\varepsilon \le p^v(x,z^v) \le \varepsilon^{-1}$, however, it is not sufficient to assume that $\varepsilon > 0$; we must assume that $\varepsilon > \varepsilon_0$ for some strictly positive threshold $\varepsilon_0 > 0$. Some assumption of this form is absolutely essential in the high-dimensional setting. Unlike the global mixing assumption, the local mixing assumption is not in itself sufficient to ensure that the underlying model will remain ergodic as the dimension $\operatorname{card} V \to \infty$: the cumulative effect of the interactions can create long-range correlations that break both ergodicity and any decay of correlations properties. Typically, the model is ergodic when the mixing constant ε is sufficiently large, but ergodicity breaks abruptly as ε drops below a threshold value ε_0. Such phenomena, called *phase transitions* in statistical mechanics, are very common in large-scale interacting systems; see [7, 10] for a number of examples. When the underlying model fails to exhibit ergodicity and decay of correlations, we lack the mechanism that we aim to exploit by developing local particle filters. Therefore, some assumption of the form $\varepsilon > \varepsilon_0$ is essential in Theorem 2.1 in order to ensure the presence of decay of correlations.

Unfortunately, the actual constant ε_0 that arises in the proof of Theorem 2.1 is almost certainly far from optimal. The Dobrushin machinery [8], Chapter 8, that forms the basis of our proof already does not yield sharp estimates of the phase transition point even in the simplest classical models of statistical mechanics. It is also far from clear whether the block particle filter should necessarily possess the same phase transition point as the underlying model: it may be that the algorithm only works in a strict subset of the regime in which the underlying model possesses the decay of correlations property. The mathematical tools used in this paper are not sufficiently powerful to address questions of this type. The practical relevance of Theorem 2.1 is therefore of a qualitative nature—we show that the block particle filter can beat the curse of dimensionality above a certain phase transition point—but should not be relied upon to provide quantitative guidance in specific situations. The development of sharper quantitative results will require new probabilistic tools for the investigation of filtering problems in high dimension.

One drawback of the assumptions of Theorem 2.1 is that mixing in space and time are treated on the same footing: as $\varepsilon \to 1$, both the spatial and temporal interactions disappear. To ensure that ergodicity and decay of correlations hold, it should suffice to assume only that the spatial interactions are weak. Such an improvement can be obtained using more refined mathematical tools that make it possible to separate the temporal and spatial ergodicity assumptions [14].

2.3.2. *Local algorithms and spatial homogeneity.* The major drawback of the block particle filtering algorithm is the spatial inhomogeneity of the bias. The consequences of this inhomogeneity are manifested quantitatively in Corollary 2.5. Near the block boundaries, Theorem 2.1 gives a bound of order unity. By excluding a small fraction of spatial locations, however, we eliminate the block boundaries to retain an error of order $e^{-c\log^{1/q} N}$ at "most" spatial locations:

$$\operatorname{card}\bigl\{v \in V : \bigl|\!\bigl|\!\bigl|\pi_n^x - \hat{\pi}_n^x\bigr|\!\bigr|\!\bigr|_v \lesssim e^{-c\delta\log^{1/q} N}\bigr\} \geq (1-\delta)\operatorname{card} V.$$

If, on the other hand, we compute the spatial average of the error, we obtain an exceedingly slow convergence rate that is much worse than the "typical" rate:

$$\frac{1}{\operatorname{card} V}\sum_{v\in V}\bigl|\!\bigl|\!\bigl|\pi_n^x - \hat{\pi}_n^x\bigr|\!\bigr|\!\bigr|_v \lesssim \frac{1}{\log^{1/q} N}.$$

Note that the block boundaries constitute a fraction $\sim 1/b$ of spatial locations, where b is the block size; therefore, as $b \sim \log^{1/q} N$ in Corollary 2.5, we see that the error at the block boundaries dominates our bound on the average error.

The behavior of the errors described above seems to be an inherent limitation of the block particle filtering algorithm. It is therefore of significant interest to explore the possibility that one could develop alternative local particle filtering algorithms that are spatially homogeneous. Conceptually, as explained in Section 1.3, such an algorithm should update the filtering distribution at each site v using sites in a centered neighborhood $N_b(v) := \{v' \in V : d(v,v') \leq b\}$; the decay of correlations should then yield a bias that decays exponentially in b. In this case, we would expect to obtain a spatially uniform error bound of the form

$$\sup_{v\in V}\bigl|\!\bigl|\!\bigl|\pi_n^x - \hat{\pi}_n^x\bigr|\!\bigr|\!\bigr|_v \lesssim e^{-c\log^{1/q} N}$$

for the optimized neighborhood size $b \sim \log^{1/q} N$. Whether it is in fact possible to design a local particle filtering algorithm that attains such a uniform error bound is perhaps the most immediate open question that arises from our results.

It is, of course, not at all obvious how one might go about developing a spatially homogeneous algorithm. We will presently discuss one possible idea that could be of interest in this setting. It should be emphasized the following discussion is intended to be heuristic, as we do not know how to analyze algorithms of the type that we will discuss. However, our aim is to illustrate that the general idea of local particle filters could be much broader than is suggested by the block particle filtering algorithm—and that the mathematical analysis developed in this paper could in itself provide inspiration for further methodological developments.

At the heart of our results lies the decay of correlations. In our proofs, we will use an intuitive notion of decay of correlations of essentially the following form: a probability measure ρ on $\mathbb{X}$ possesses the decay of correlations property if the effect on the conditional distribution $\rho(X^v \in \cdot\,|X^{V\setminus\{v\}} = x^{V\setminus\{v\}})$ of a perturbation

to $x^{v'}$ decays exponentially in the distance $d(v,v')$ (cf. Sections 3.2 and 4.2). The blocking operation evidently replaces these conditional distributions by

$$(\mathsf{B}\rho)\big(X^v \in A | X^{V\setminus\{v\}} = x^{V\setminus\{v\}}\big) = \rho\big(X^v \in A | X^{K\setminus\{v\}} = x^{K\setminus\{v\}}\big)$$

for every $K \in \mathcal{K}$ and $v \in K$. Therefore, if ρ possesses the decay of correlations property, then the bias at site $v \in K$ incurred by the blocking operation decays exponentially in the distance between v and the boundary of K. From this perspective, an approach to spatially homogeneous algorithms readily suggests itself: we should aim to replace B with another operator M that satisfies

$$(\mathsf{M}\rho)\big(X^v \in A | X^{V\setminus\{v\}} = x^{V\setminus\{v\}}\big) = \rho\big(X^v \in A | X^{N_b(v)\setminus\{v\}} = x^{N_b(v)\setminus\{v\}}\big)$$

for every $v \in V$. The bias incurred by this operation decays exponentially in b uniformly for all v (it is therefore spatially homogeneous). On the other hand, as

$$\begin{aligned}&(\mathsf{C}_n\mathsf{M}\mathsf{P}\rho)\big(X^v \in A | X^{V\setminus\{v\}} = x^{V\setminus\{v\}}\big)\\ &\quad= \frac{\int \mathbf{1}_A(x^v) g^v(x^v, Y_n^v) \prod_{w\in N_b(v)} p^w(z, x^w)\rho(dz)\psi^v(dx^v)}{\int g^v(x^v, Y_n^v) \prod_{w\in N_b(v)} p^w(z, x^w)\rho(dz)\psi^v(dx^v)},\end{aligned}$$

the sampling error incurred if we replace ρ by $\mathsf{S}^N\rho$ in this expression should only be exponential in $\operatorname{card} N_b(v)$ (which is $\sim b^q$ for the square lattice) rather than in the model dimension $\operatorname{card} V$. This suggests that the local particle filter defined by the recursion $\hat{\mathsf{F}}_n = \mathsf{S}^N\mathsf{C}_n\mathsf{M}\mathsf{P}$ should yield a spatially homogeneous algorithm in accordance with our intuition. To implement this algorithm, one needs to sample from the measure $\mathsf{C}_n\mathsf{M}\mathsf{P}\rho$, which we have defined only implicitly in terms of its conditional distributions. This is however precisely the task to which MCMC methods such as the Gibbs sampler are well suited. One would therefore ostensibly obtain a spatially homogeneous local particle filtering algorithm that is recursive in time and that uses MCMC to sample the spatial degrees of freedom (regularization using M is still key to avoiding the curse of dimensionality; cf. Remark 1.1).

Conceptually, the idea introduced here is quite natural. The general idea of local particle filters is that one should introduce a spatial regularization step into the filtering recursion that enables local sampling. In the block particle filter, this regularization is provided by the blocking operation B that projects a probability measure on the class of measures that are independent across blocks. In the above algorithm, we aim to regularize instead by the operation M that projects a probability measure on the class of Markov random fields of order b. The fatal flaw in our reasoning is that the operator M that we have defined implicitly above does not exist: the truncated conditional distributions $\rho(X^v \in \cdot | X^{N_b(v)\setminus\{v\}} = x^{N_b(v)\setminus\{v\}})$ are typically not consistent, so there exists no single probability measure that satisfies our definition of $\mathsf{M}\rho$. Nonetheless, the basic idea introduced here could be fruitful if one can develop a practical approach to approximating random fields by Markov random fields [e.g., one could attempt to substitute the above expression for $(\mathsf{C}_n\mathsf{M}\mathsf{P}\rho)(X^v \in \cdot | X^{V\setminus\{v\}})$ in a Gibbs sampler regardless of its inconsistency].

The development of such ideas evidently presents some interesting mathematical as well as methodological challenges that should be investigated further.

Let us finally observe that, by their nature, local particle filtering algorithms are well suited to distributed computation: as the particles are updated locally in the spatial graph, this opens the possibility of implementing each local neighborhood on a separate processor. While this was not the original intention of the algorithms we propose, such properties could prove to be advantageous in their own right for the practical implementation of filtering algorithms in very large-scale systems.

2.3.3. *High-dimensional models in data assimilation.* The basic model that we have introduced in Section 2.1 is prototypical of many data assimilation problems and provides a particularly convenient mathematical setting for the investigation of filtering problems in high dimension. While such models could be directly relevant to many high-dimensional applications, there remains a substantial gap between relatively simple models of this form and realistic models used in the most complex applications, particularly in the geophysical, atmospheric and ocean sciences, that frequently consist of coupled systems of partial differential equations. The investigation of such complex problems, and the associated numerical, physical and practical issues, is far beyond the scope of this paper. We therefore restrict our discussion of such problems to a few brief comments.

In principle, discrete models as defined in Section 2.1 arise naturally as finite-difference approximations of stochastic partial differential equations with space–time white noise forcing. As the resulting state spaces $\mathbb{X}^v$ are not compact, such systems cannot satisfy strong mixing assumptions (cf. Section 2.3.1), but this is likely a mathematical rather than a practical problem. More importantly, it is not clear whether the discretized models will be in the regime of decay of correlations (i.e., above the phase transition point) even if the original continuum model possesses such properties. It is possible that this requirement would place constraints on the spatial and temporal discretization steps, in the spirit of the von Neumann stability criterion in numerical analysis. The physics of such problems could also impose constraints on the design of local particle filters; for example, it is suggested in [19], page 4107, that discontinuities (such as might be introduced at the block boundaries in the block particle filtering algorithm) could generate spurious gravity waves in ocean models. Such numerical and practical issues are distinct from the fundamental problems in high dimension that we aim to address in this paper, but can ultimately play an equally important role in complex applications.

Let us also note that models considered in the data assimilation literature are often deterministic partial differential equations without stochastic forcing; the only randomness in such models comes from the initial condition (cf. [1, 9]). In deterministic chaotic dynamical systems, it is impossible to obtain time-uniform approximations using classical particle filters as there is no dissipation mechanism for approximation errors (the filter cannot be stable in this case; cf. Section 3.1). This issue is not directly related to dimensionality issues in particle filters: such

problems arise in every deterministic filtering problem. It is natural to regularize deterministic systems by adding dynamical noise to the model (there is an extensive literature on random perturbations of chaotic dynamics; see, e.g., [4]); a similar observation has been made by practitioners in the context of ad-hoc filtering algorithms; cf. [9], Section 5. To our knowledge, a rigorous analysis of such ideas in the setting of particle filters has yet to be performed.

3. Outline of the proof.

3.1. *Error decomposition.* The goal of Theorem 2.1 is to bound the error between the filter π_n^μ and the block particle filter $\hat{\pi}_n^\mu$. Recall that both the filter (Section 1.1) and block particle filter (Section 2.2) are defined recursively:

$$\pi_n^\mu = \mathsf{F}_n \cdots \mathsf{F}_1 \mu, \qquad \hat{\pi}_n^\mu = \hat{\mathsf{F}}_n \cdots \hat{\mathsf{F}}_1 \mu,$$

where $\mathsf{F}_n := \mathsf{C}_n \mathsf{P}$ and $\hat{\mathsf{F}}_n := \mathsf{C}_n \mathsf{B} \mathsf{S}^N \mathsf{P}$. We introduce also the *block filter*

$$\tilde{\pi}_n^\mu = \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \mu$$

with $\tilde{\mathsf{F}}_n := \mathsf{C}_n \mathsf{B} \mathsf{P}$. By the triangle inequality, we have

$$\|\!|\pi_n^\mu - \hat{\pi}_n^\mu|\!\|_J \le \|\!|\pi_n^\mu - \tilde{\pi}_n^\mu|\!\|_J + \|\!|\tilde{\pi}_n^\mu - \hat{\pi}_n^\mu|\!\|_J.$$

The first term on the right-hand side quantifies the bias introduced by the projection on independent blocks, while the second term quantifies the error due to the variance of the random sampling in the algorithm. Each term will be bounded separately to obtain the two terms in the error bound of Theorem 2.1.

The challenges encountered in bounding the bias term (cf. Section 3.3) and the variance term (cf. Section 3.4) are quite different in nature. Nonetheless, both bounds are based on a basic scheme of proof that was invented in order to prove time-uniform bounds for the bootstrap particle filter [5, 6]. We therefore begin by reviewing this general idea, which is based on a simple error decomposition.

Suppose for sake of illustration that we aim to bound directly the error between π_n^μ and $\hat{\pi}_n^\mu$. The basic idea is to write $\pi_n^\mu - \hat{\pi}_n^\mu$ as a telescoping sum:

$$\pi_n^\mu - \hat{\pi}_n^\mu = \sum_{s=1}^n \{\mathsf{F}_n \cdots \mathsf{F}_{s+1} \mathsf{F}_s \hat{\mathsf{F}}_{s-1} \cdots \hat{\mathsf{F}}_1 \mu - \mathsf{F}_n \cdots \mathsf{F}_{s+1} \hat{\mathsf{F}}_s \hat{\mathsf{F}}_{s-1} \cdots \hat{\mathsf{F}}_1 \mu\}.$$

By the triangle inequality,

$$\|\!|\pi_n^\mu - \hat{\pi}_n^\mu|\!\| \le \sum_{s=1}^n \|\!|\mathsf{F}_n \cdots \mathsf{F}_{s+1} \mathsf{F}_s \hat{\pi}_{s-1}^\mu - \mathsf{F}_n \cdots \mathsf{F}_{s+1} \hat{\mathsf{F}}_s \hat{\pi}_{s-1}^\mu|\!\|.$$

The term s in this sum could be interpreted as the contribution to the total error at time n due to the filter approximation made in time step s.

The key insight is now that one can employ the *filter stability* property to control this sum uniformly in time. In its simplest form, this property can be proved in the following form: if $\varepsilon \le p(x,z) \le \varepsilon^{-1}$ for all $x, z \in \mathbb{X}$, then [5, 6]

$$\|\!|\mathsf{F}_n \cdots \mathsf{F}_{s+1}\rho - \mathsf{F}_n \cdots \mathsf{F}_{s+1}\rho'|\!\| \le \varepsilon^{-2}(1-\varepsilon^2)^{n-s}\|\!|\rho-\rho'|\!\|.$$

Thus, the filter forgets its initial condition at an exponential rate. However, this also means that past approximation errors are forgotten at an exponential rate: if we substitute the stability property in the above error decomposition, we obtain

$$\|\!|\pi_n^\mu - \hat\pi_n^\mu|\!\| \le \sum_{s=1}^n \varepsilon^{-2}(1-\varepsilon^2)^{n-s}\|\!|\mathsf{F}_s\hat\pi_{s-1}^\mu - \hat{\mathsf{F}}_s\hat\pi_{s-1}^\mu|\!\| \le \varepsilon^{-4}\sup_{n,\rho}\|\!|\mathsf{F}_n\rho - \hat{\mathsf{F}}_n\rho|\!\|.$$

Thus, if we can control the error $\|\!|\mathsf{F}_n\rho - \hat{\mathsf{F}}_n\rho|\!\|$ in a single time step, we obtain a time-uniform bound of the same order. In the case of the bootstrap particle filter, if $\kappa \le g(x,y) \le \kappa^{-1}$, we proved that $\|\!|\mathsf{F}_n\rho - \hat{\mathsf{F}}_n\rho|\!\| \le 2\kappa^{-2}/\sqrt{N}$ in Section 1.1, and we obtain a time-uniform version of the crude error bound given there.

The basic error decomposition discussed above allows us to separate the problem of obtaining time-uniform bounds into two parts: the one-step approximation error and the stability property. It is important to note, however, that both parts become problematic in high dimension. We have already seen (Section 1.2) that the one-step approximation error of the bootstrap particle filter is exponential in the model dimension; we will surmount this problem by working with the block particle filtering algorithm and performing a local analysis of the one-step error using the decay of correlations property (which must itself be established). On the other hand, the filter stability bound used above also becomes exponentially worse in high dimension: a local bound of the form $\varepsilon \le p^v(x,z^v) \le \varepsilon^{-1}$ only yields $\varepsilon^{\operatorname{card} V} \le p(x,z) \le \varepsilon^{-\operatorname{card} V}$, which is exponential in the model dimension $\operatorname{card} V$. To surmount this problem, we must develop a much more precise understanding of the filter stability property in high dimension, which proves to be closely related to the decay of correlations property. The development of these ingredients constitutes the bulk of the proof of Theorem 2.1.

3.2. *Dobrushin comparison method.* How can one control the approximation error of high-dimensional distributions? The basic idea that we aim to exploit, both algorithmically and mathematically, is that the decay of correlations property leads to a form of localization: the effect on the distribution in some spatial set J of a perturbation made in another set J' decays rapidly in the distance $d(J,J')$. Therefore, as long as we measure the error locally (in $\|\!|\cdot|\!\|_J$ rather than $\|\!|\cdot|\!\|$), one would hope to control the spatial accumulation of approximation errors much as we controlled the accumulation of approximation errors in time using the filter stability property. We will presently introduce a powerful (albeit blunt) tool—the Dobrushin comparison theorem—that makes this idea precise in a very general

setting. This fundamental result, which plays an important role in statistical mechanics [8], Chapter 8, is the main workhorse that will be used repeatedly in our proofs.

Let I be a finite set, and 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 $X^i : x \mapsto x^i$ for $x \in \mathbb{S}$ an $i \in I$. For any probability ρ on $\mathbb{S}$, we fix a version $\rho^i_\cdot$ of the regular conditional probability

$$\rho^i_x(A) = \rho\big(X^i \in A | X^{I\setminus\{i\}} = x^{I\setminus\{i\}}\big).$$

We also define for $J \subseteq I$ the local total variation distance

$$\|\rho - \rho'\|_J := \sup_{f\in\mathcal{S}^J\,:\,|f|\le 1} |\rho(f) - \rho'(f)|,$$

where $\mathcal{S}^J$ is the class of measurable functions $f : \mathbb{S} \to \mathbb{R}$ such that $f(x) = f(z)$ whenever $x^J = z^J$. For $J = I$, we write $\|\rho - \rho'\|$ for simplicity.

We can now state the Dobrushin comparison theorem [8], Theorem 8.20.[3]

THEOREM 3.1 (Dobrushin). *Let $\rho, \tilde\rho$ be probability measures on $\mathbb{S}$. Define*

$$C_{ij} = \frac{1}{2} \sup_{x,z\in\mathbb{S}\,:\,x^{I\setminus\{j\}}=z^{I\setminus\{j\}}} \|\rho^i_x - \rho^i_z\|, \qquad b_j = \sup_{x\in\mathbb{S}} \|\rho^j_x - \tilde\rho^j_x\|.$$

Suppose that the Dobrushin condition holds:

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

Then the matrix sum $D := \sum_{n\ge 0} C^n$ is convergent, and we have for every $J \subseteq I$

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

This result could be informally interpreted as follows. C_{ij} measures the degree to which a perturbation of site j directly affects site i under the distribution ρ. However, perturbing site j might also indirectly affect i: it could affect another site k which in turn affects i, etc. The aggregate effect of a perturbation of site j on site i is captured by the quantity D_{ij}. In this setting, a useful manifestation of the decay of correlations property is that D_{ij} decays exponentially in the distance $d(i, j)$. If this is in fact the case, then Theorem 3.1 yields, for example, $\|\rho - \tilde\rho\|_i \lesssim \sum_j e^{-d(i,j)} b_j$, where b_j measures the local error at site j between ρ and $\tilde\rho$ (in terms of the conditional distributions $\rho^j_\cdot$ and $\tilde\rho^j_\cdot$). The decay of correlations property therefore controls the accumulation of local errors much as one might expect.

[3]Note that our definition of $\|\cdot\|_J$ differs by a factor 2 from that in [8].

Let us now explain how Theorem 3.1 will be applied in the filtering setting. For sake of illustration, consider the problem of obtaining a local filter stability bound: that is, we would like to bound $\|\pi_n^x - \pi_n^{\tilde{x}}\|_J$ for $x, \tilde{x} \in \mathbb{X}$ and $J \subseteq V$. It would seem natural to apply Theorem 3.1 directly with $I = V$, $\mathbb{S} = \mathbb{X}$, and $\rho = \pi_n^x$, $\tilde{\rho} = \pi_n^{\tilde{x}}$. This is not useful, however, as we do not know how to control the corresponding local quantities such as $\rho_z^v = \mathbf{P}^x[X_n^v \in \cdot | Y_1, \ldots, Y_n, X_n^{V\setminus\{v\}} = z^{V\setminus\{v\}}]$.

Instead, define $I = \{0, \ldots, n\} \times V$ and $\mathbb{S} = \mathbb{X}^{n+1}$, and let

$$\rho = \mathbf{P}^x\big[(X_0, \ldots, X_n) \in \cdot | Y_1, \ldots, Y_n\big],$$

$$\tilde{\rho} = \mathbf{P}^{\tilde{x}}\big[(X_0, \ldots, X_n) \in \cdot | Y_1, \ldots, Y_n\big].$$

As

$$\big\|\pi_n^x - \pi_n^{\tilde{x}}\big\|_J = \|\rho - \tilde{\rho}\|_{\{n\}\times J},$$

we can now apply Theorem 3.1 to the *smoothing* distributions $\rho, \tilde{\rho}$. Unlike the filters $\pi_n^x, \pi_n^{\tilde{x}}$, however, ρ and ρ' are Markov random fields on I (cf. Figure 3), so that the conditional distributions $\rho_z^{k,v}$ and $\tilde{\rho}_z^{k,v}$ can be easily computed and controlled in terms of the local densities $p^v(x, z^v)$ and $g^v(x^v, y^v)$. For example, as

$$\rho(A) \propto \int \mathbf{1}_A(x, x_1, \ldots, x_n) \prod_{k=1}^n \prod_{v\in V} p^v\big(x_{k-1}, x_k^v\big) g^v\big(x_k^v, Y_k^v\big) \psi^v\big(dx_k^v\big),$$

and as $p^v(x_{k-1}, x_k^v)$ depends only on x_{k-1}^w for $d(w, v) \le r$, we obtain

$$\rho_z^{k,v}(B) \propto \int \mathbf{1}_B\big(z_k^v\big) p^v\big(z_{k-1}, z_k^v\big) g^v\big(z_k^v, Y_k^v\big) \prod_{w\in N(v)} p^w\big(z_k, z_{k+1}^w\big) \psi^v\big(dz_k^v\big)$$

for $0 < k < n$ and $v \in V$ (the proportionality is up to a normalization factor). We will repeatedly exploit expressions of this type to obtain explicit bounds on the quantities C_{ij} and b_j that appear in Theorem 3.1. It should be emphasized that $\rho_z^{k,v}$ is a genuinely local quantity: the product inside the integral contains at most card $N(v) \le \Delta$ factors. We will consequently be able to use Theorem 3.1 to obtain bounds that do not depend on the model dimension card V.

3.3. *Bounding the bias*: *Decay of correlations*. To bound the bias $\|\pi_n^x - \tilde{\pi}_n^x\|_J$, we follow the basic error decomposition scheme described above, that is,

$$\big\|\pi_n^x - \tilde{\pi}_n^x\big\|_J \le \sum_{s=1}^n \big\|\mathsf{F}_n \cdots \mathsf{F}_{s+1} \mathsf{F}_s \tilde{\pi}_{s-1}^x - \mathsf{F}_n \cdots \mathsf{F}_{s+1} \tilde{\mathsf{F}}_s \tilde{\pi}_{s-1}^x\big\|_J.$$

To implement our program, we must now obtain suitable local bounds on the stability of the filter and on the one-step approximation error. Both these problems will be approached by application of the Dobrushin comparison theorem.

In its most basic form, one can prove a filter stability property of the following type: provided $\varepsilon > \varepsilon_0$, there exists $\beta > 0$ (depending only on Δ and r) such that

$$\|\mathsf{F}_n \cdots \mathsf{F}_{s+1}\mu - \mathsf{F}_n \cdots \mathsf{F}_{s+1}\nu\|_J \le 4 \operatorname{card} J e^{-\beta(n-s)}$$

for any probability measures μ, ν on $\mathbb{X}$ and $J \subseteq V$, $n \ge 0$ (cf. Corollary 4.7). This bound is evidently dimension-free, unlike the crude filter stability bound described in Section 3.1. Nonetheless, this filter stability bound would yield a trivial result when substituted in the error decomposition, as it does not provide any control in terms of the distance between μ and ν (and, therefore, in terms of the one-step error). Instead, we will prove in Section 4.2 the local stability bound

$$\|\mathsf{F}_n \cdots \mathsf{F}_{s+1}\mu - \mathsf{F}_n \cdots \mathsf{F}_{s+1}\nu\|_J \le 2e^{-\beta(n-s)} \sum_{v \in J} \max_{v' \in V} e^{-\beta d(v,v')} D_{v'}(\mu, \nu),$$

where $D_{v'}(\mu, \nu)$ is a suitable measure of the local error between μ and ν at site v' that arises naturally from the Dobrushin comparison theorem (see Proposition 4.4 for precise expressions). This filter stability bound is genuinely local: the stability on the spatial set $J \subseteq V$ depends predominantly on the local distance of the initial conditions near J (i.e., the spatial accumulation of errors is mitigated). This localization comes at a price, however; the local filter stability bound holds only if the initial condition μ satisfies a priori a decay of correlations property.

Once the local filter stability bound is substituted in the error decomposition, it remains to prove a bound on the one-step error $D_v(\mathsf{F}_s\tilde{\pi}^x_{s-1}, \tilde{\mathsf{F}}_s\tilde{\pi}^x_{s-1})$ with respect to the local distance prescribed by the filter stability bound. This will be done in Section 4.3: we will show that for a constant C that depends only on Δ, r, ε,

$$D_v(\mathsf{F}_s\mu, \tilde{\mathsf{F}}_s\mu) \le Ce^{-\beta d(v,\partial K)}$$

for every $K \in \mathcal{K}$ and $v \in K$, provided again that μ satisfies a priori a decay of correlations property. This is precisely what we expect: as B only introduces errors at the block boundaries, the decay of correlations should ensure that the error at site v decays exponentially in the distance to the nearest block boundary. The Dobrushin comparison theorem allows to make this intuition precise.

The decay of correlations property evidently plays a dual role in our setting: it controls the approximation error of the block filter, which is the basic principle behind the block particle filtering algorithm; at the same time, it mitigates the spatial accumulation of approximation errors, which is essential for proving dimension-free bounds. In order to apply the above bounds, the key step that remains is to prove that the appropriate decay of correlations property does in fact hold, uniformly in time, for the block filter $\tilde{\pi}^x_n$. The latter will be shown in Section 4.4 by iterating a one-step decay of correlations bound that is obtained once again using the Dobrushin comparison theorem. We conclude by putting together all these ingredients in Section 4.5 to obtain a bound on the bias of the form

$$\|\pi^x_n - \tilde{\pi}^x_n\|_J \le C \operatorname{card} J e^{-\beta d(J,\partial K)}$$

for $J \subseteq K$ (Theorem 4.14). This proves the first half of Theorem 2.1 (note that, as the bias does not depend on the random sampling in the block particle filtering algorithm, we can trivially replace $\|\pi_n^x - \tilde{\pi}_n^x\|_J$ by $|||\pi_n^x - \tilde{\pi}_n^x|||_J$ in this bound).

3.4. *Bounding the variance*: *The computation tree*. To bound the variance term $|||\tilde{\pi}_n^x - \hat{\pi}_n^x|||_J$, we once again start from the basic error decomposition

$$|||\tilde{\pi}_n^x - \hat{\pi}_n^x|||_J \le \sum_{s=1}^{n} |||\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1} \tilde{\mathsf{F}}_s \hat{\pi}_{s-1}^x - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1} \hat{\mathsf{F}}_s \hat{\pi}_{s-1}^x|||_J.$$

The difficulties encountered in controlling this expression are quite different in nature, however, than what was needed to control the bias term.

Dimension-free bounds on the bias exploit decay of correlations: the core difficulty is to obtain local control of the error inside the blocks. The variance term, on the other hand, will already grow exponentially in the size of the blocks due to the exponential dependence of the sampling error on the dimension of the observations. There is therefore no need bound the error on a finer scale than a single block. This makes the analysis of the variance much less delicate than controlling the bias, and it is indeed not difficult to obtain a variance bound of the right order on a finite time horizon (but growing exponentially in time n).

The chief difficulty in controlling the variance is to obtain a time-uniform bound. Note that, in the error decomposition for the variance term, it is not stability of the filter π_n^μ that enters the picture but rather stability of the block filter $\tilde{\pi}_n^\mu$. Unlike the filter, however, which has by construction an interpretation as the marginal of a smoothing distribution, the block filter is defined by a recursive algorithm and not as a conditional expectation. It is therefore not entirely obvious how one could adapt the approach outlined in Section 3.2 to this setting.

The key idea that will be used to establish stability is that the block filter can nonetheless be viewed as the marginal of a suitably defined Markov random field, just like the filter can be viewed as the marginal of a smoothing distribution. This random field, however, lives on a much larger index set than the original model. The basic idea behind the construction is illustrated in Figure 5 (disregarding the observations for simplicity of exposition). When we apply the transition operator P, each block interacts with its $\Delta_{\mathcal{K}}$ neighbors in the previous time step. However, if we subsequently apply the blocking operator B, then each block is replaced by an independent copy. This could be modeled equivalently by introducing independent duplicates of the blocks in the previous time step, and having each block interact with its own set of duplicates. This unravels the original dependency graph into a tree. By iterating this process, we can express the block filter as the marginal of a Markov random field defined on a tree that contains many independent duplicates of each block. We call this construction the *computation tree* in analogy with a similar notion that arises in the analysis of belief propagation algorithms [17].

With this construction in place, we can now obtain a stability bound for the block filter by applying the Dobrushin comparison theorem to the computation

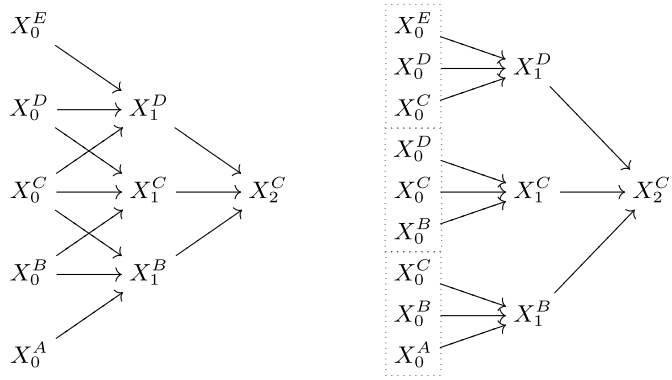

FIG. 5. *For a linear spatial graph G partitioned into blocks A–E (with $r = 1$), the dependencies between the blocks at subsequent times are illustrated here. The left dependency graph represents $\mathsf{B}^C\mathsf{P}^2\mu$, the right graph represents $\mathsf{B}^C\mathsf{PBP}\mu$. The blocking operation unravels the original graph into a tree by introducing independent duplicates (dotted boxes) of blocks in the previous time step.*

tree. This will be done in Section 4.6 to obtain a bound of the following form: provided $\varepsilon > \varepsilon_0$, there exist $\beta, \beta' > 0$ (depending only on $\Delta, \Delta_{\mathcal{K}}, r$) such that

$$\max_{K\in\mathcal{K}} \|\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1}\mu - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1}\nu\|_K \le e^{\beta'|\mathcal{K}|_\infty} e^{-\beta(n-s)} \max_{K\in\mathcal{K}} \|\mu^K - \nu^K\|$$

for any pair of initial conditions of product form $\mu = \bigotimes_{K\in\mathcal{K}} \mu^K$, $\nu = \bigotimes_{K\in\mathcal{K}} \nu^K$ (cf. Corollary 4.18). Combining this bound with the error decomposition, we obtain in Section 4.7 a time-uniform bound on the variance term of the form

$$\max_{K\in\mathcal{K}} \|\!|\tilde{\pi}_n^x - \hat{\pi}_n^x|\!\|_K \le C\frac{e^{\beta'|\mathcal{K}|_\infty}}{\sqrt{N}},$$

where we bound the one-step error in the same spirit as the computation for the bootstrap particle filter in Section 1.1 (however, a more involved argument is needed here to surmount the fact that the block filter stability bound is given in a total variation norm rather than the weaker norm $\|\!|\cdot|\!\|_K$). Thus, Theorem 2.1 is proved.

4. Proof of Theorem 2.1. Theorem 2.1 yields a bound on $\|\!|\pi_n^\mu - \hat{\pi}_n^\mu|\!\|_J$. As

$$\|\!|\pi_n^\mu - \hat{\pi}_n^\mu|\!\|_J \le \|\!|\pi_n^\mu - \tilde{\pi}_n^\mu|\!\|_J + \|\!|\tilde{\pi}_n^\mu - \hat{\pi}_n^\mu|\!\|_J,$$

it suffices to bound each term in this inequality. As was explained in Section 3.1, the first term quantifies the bias of the block particle filter, while the second term quantifies the variance of the random sampling. The bias term will be bounded in Theorem 4.14 below, while the variance will be bounded in Theorem 4.23. The combination of these two results immediately yields Theorem 2.1.

4.1. *Preliminary lemmas.* The Dobrushin comparison method introduced in Section 3.2 is the main workhorse of our proof. To use this method, we must be able to bound the quantities C_{ij}, b_j and D_{ij} that appear in Theorem 3.1. The goal of this preliminary section is to collect some elementary lemmas for this purpose.

We start with a rather trivial lemma that will be used to bound C_{ij}.

LEMMA 4.1. *Let probability measures $\nu, \nu', \gamma, \gamma'$ and $\varepsilon > 0$ be such that $\nu(A) \geq \varepsilon\gamma(A)$ and $\nu'(A) \geq \varepsilon\gamma'(A)$ for every measurable set A. Then*

$$\|\nu - \nu'\| \leq 2(1-\varepsilon) + \varepsilon\|\gamma - \gamma'\|.$$

In particular, if $\gamma = \gamma'$, then $\|\nu - \nu'\| \leq 2(1-\varepsilon)$.

PROOF. As $\mu = (1-\varepsilon)^{-1}(\nu - \varepsilon\gamma)$ and $\mu' = (1-\varepsilon)^{-1}(\nu' - \varepsilon\gamma')$ are probability measures and $\nu - \nu' = (1-\varepsilon)(\mu - \mu') + \varepsilon(\gamma - \gamma')$, the result follows readily. □

Next, we state a simple lemma on the distance between weighted measures. We have already used this result in Section 1.1 to bound $|||\mathsf{C}_n\rho - \mathsf{C}_n\rho'|||$.

LEMMA 4.2. *Let μ, ν be (possibly random) probability measures and let Λ be a bounded and strictly positive measurable function. Define*

$$\mu_\Lambda(A) := \frac{\int \mathbf{1}_A(x)\Lambda(x)\mu(dx)}{\int \Lambda(x)\mu(dx)}, \qquad \nu_\Lambda(A) := \frac{\int \mathbf{1}_A(x)\Lambda(x)\nu(dx)}{\int \Lambda(x)\nu(dx)}.$$

Then

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

The same conclusion holds if the $\|\cdot\|$-norm is replaced by the $|||\cdot|||$-norm.

PROOF. The result follows readily from the identity

$$\mu_\Lambda(f) - \nu_\Lambda(f) = \frac{1}{\mu(\Lambda)}\left[\{\mu(f\Lambda) - \nu(f\Lambda)\} + \frac{\nu(f\Lambda)}{\nu(\Lambda)}\{\nu(\Lambda) - \mu(\Lambda)\}\right]$$

using the definition of the norms $\|\cdot\|$ or $|||\cdot|||$. □

Finally, we give a lemma that will be essential for bounding D_{ij}. In essence, the lemma states that if C_{ij} decays exponentially in the distance between i and j at a sufficiently rapid rate, then D_{ij} will also decay exponentially in the distance between i and j. This is essential in order to establish the decay of correlations property using only bounds on C_{ij}, which can be obtained in explicit form. While the lemma should be interpreted in the spirit of decay of correlations, it is essentially a simple lemma about matrices and will be stated as such.

LEMMA 4.3. *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

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

In particular, this implies that

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

for every $J \subseteq I$.

PROOF. Define for any matrix A with nonnegative entries the norm

$$\|A\|_m := \max_{i\in I} \sum_{j\in I} e^{m(i,j)} A_{ij}.$$

Using $m(i,j) \le m(i,k) + m(k,j)$, we compute

$$\begin{aligned}
\|AB\|_m &= \max_{i\in I} \sum_{j\in I} e^{m(i,j)} \sum_{k\in I} A_{ik} B_{kj} \\
&\le \max_{i\in I} \sum_{k\in I} e^{m(i,k)} A_{ik} \sum_{j\in I} e^{m(k,j)} B_{kj} \\
&\le \|A\|_m \|B\|_m,
\end{aligned}$$

so $\|A\|_m$ is a matrix norm. Therefore,

$$\|D\|_m \le \sum_{n\ge 0} \|C\|_m^n \le \sum_{n\ge 0} c^n = \frac{1}{1-c}.$$

As

$$e^{m(i,J)} \sum_{j\in J} D_{ij} \le \sum_{j\in J} e^{m(i,j)} D_{ij} \le \|D\|_m,$$

the last statement of the lemma follows immediately. □

4.2. *Local stability of the filter.* The main goal of this section is to prove a local stability bound for the nonlinear filter. We begin, however, by introducing a number of objects that will appear several times in the sequel.

For any probability measure μ on $\mathbb{X}$ and $x, z \in \mathbb{X}$, $v \in V$, we define

$$\mu^v_{x,z}(A) := \mathbf{P}^\mu[X^v_0 \in A | X_0^{V\setminus\{v\}} = x^{V\setminus\{v\}}, X_1 = z] = \frac{\int \mathbf{1}_A(x^v) \prod_{w\in N(v)} p^w(x, z^w)\mu^v_x(dx^v)}{\int \prod_{w\in N(v)} p^w(x, z^w)\mu^v_x(dx^v)}$$

(recall the notation $\mu^v_x := \mathbf{P}^\mu[X^v_0 \in \cdot | X_0^{V\setminus\{v\}} = x^{V\setminus\{v\}}]$ in Section 3.2). Let

$$C^\mu_{vv'} := \frac{1}{2} \sup_{z\in\mathbb{X}} \sup_{x,\tilde{x}\in\mathbb{X}: x^{V\setminus\{v'\}}=\tilde{x}^{V\setminus\{v'\}}} \|\mu^v_{x,z} - \mu^v_{\tilde{x},z}\|$$

for $v, v' \in V$. The quantity

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

could be viewed as a measure of the degree of correlation decay of the measure μ at rate $\beta > 0$. It will turn out that this (not entirely obvious) measure of decay of correlations is precisely tuned to the needs of the proof of Theorem 2.1. This is due to the fact that the measures $\mu^v_{x,z}$ arise naturally when applying the Dobrushin comparison method to the smoothing distributions as discussed in Section 3.2.

We recall once and for all that the interaction radius r and neighborhood size Δ that will appear repeatedly in the following results are defined in Section 2.2.

PROPOSITION 4.4 (Local filter stability). *Suppose there exists $\varepsilon > 0$ such that*

$$\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1} \qquad \text{for all } v \in V, x, z \in \mathbb{X}.$$

Let μ, ν be probability measures on $\mathbb{X}$, and suppose that

$$\mathrm{Corr}(\mu, \beta) + 3(1 - \varepsilon^{2\Delta})e^{2\beta r}\Delta^2 \le \tfrac{1}{2}$$

for a sufficiently small constant $\beta > 0$. Then we have

$$\|\mathsf{F}_n \cdots \mathsf{F}_{s+1}\mu - \mathsf{F}_n \cdots \mathsf{F}_{s+1}\nu\|_J \le 2e^{-\beta(n-s)} \sum_{v\in J} \max_{v'\in V} e^{-\beta d(v,v')} \sup_{x,z\in\mathbb{X}} \|\mu^{v'}_{x,z} - \nu^{v'}_{x,z}\|$$

for every $J \subseteq V$ and $s < n$.

REMARK 4.5. There is nothing magical about the constant $1/2$ in the decay of correlations assumption; any constant $c < 1$ would work at the expense of a constant $1/(1-c)$ rather than 2 in the filter stability bound. As our methods are not expected to yield tight quantitative bounds, we have taken the liberty to fix various constants of this sort throughout the following sections for aesthetic purposes.

REMARK 4.6. Note that by Lemma 4.2

$$\|\mu^{v'}_{x,z} - \nu^{v'}_{x,z}\| \le \frac{2}{\varepsilon^{2\Delta}} \|\mu^{v'}_x - \nu^{v'}_x\|.$$

This yields a slightly cleaner bound in Proposition 4.4 with a worse constant. For our purposes, however, it will be just as easy to bound $\|\mu^{v'}_{x,z} - \nu^{v'}_{x,z}\|$ directly.

PROOF OF PROPOSITION 4.4. Define the smoothing distributions

$$\rho = \mathbf{P}^{\mu}[X_0, \dots, X_n \in \cdot | Y_1, \dots, Y_n],$$
$$\tilde{\rho} = \mathbf{P}^{\nu}[X_0, \dots, X_n \in \cdot | Y_1, \dots, Y_n].$$

We will apply Theorem 3.1 to $\rho, \tilde{\rho}$ with $I = \{0, \dots, n\} \times V$ and $\mathbb{S} = \mathbb{X}^{n+1}$ as discussed in Section 3.2. To this end, we must bound the quantities C_{ij} and b_j. We begin by bounding C_{ij} with $i = (k, v)$ and $j = (k', v')$. We distinguish three cases.

Case $k = 0$. The key observation in this case is that $\rho^i_x = \mu^v_{x_0,x_1}$ by the Markov property (or by direct computation). Note that as $\operatorname{card} N(v) \le \Delta$, we have

$$\mu^v_{x,z}(A) = \frac{\int \mathbf{1}_A(x^v) \prod_{w \in N(v)} p^w(x, z^w) \mu^v_x(dx^v)}{\int \prod_{w \in N(v)} p^w(x, z^w) \mu^v_x(dx^v)} \ge \varepsilon^{2\Delta} \mu^v_x(A),$$

so $\|\mu^v_{x,z} - \mu^v_{x,z'}\| \le 2(1 - \varepsilon^{2\Delta})$ for any $z, z' \in \mathbb{X}$ by Lemma 4.1. Therefore,

$$C_{ij} \le \begin{cases} C^{\mu}_{vv'}, & \text{if } k' = 0, \\ 1 - \varepsilon^{2\Delta}, & \text{if } k' = 1 \text{ and } v' \in N(v), \\ 0, & \text{otherwise.} \end{cases}$$

This evidently implies that

$$\sum_{(k',v') \in I} e^{\beta k'} e^{\beta d(v,v')} C_{(0,v)(k',v')} \le \operatorname{Corr}(\mu, \beta) + (1 - \varepsilon^{2\Delta}) e^{\beta(r+1)} \Delta.$$

Case $0 < k < n$. Now we have (cf. Section 3.2)

$$\rho^i_x(A) = \frac{\int \mathbf{1}_A(x^v_k) p^v(x_{k-1}, x^v_k) g^v(x^v_k, Y^v_k) \prod_{w \in N(v)} p^w(x_k, x^w_{k+1}) \psi^v(dx^v_k)}{\int p^v(x_{k-1}, x^v_k) g^v(x^v_k, Y^v_k) \prod_{w \in N(v)} p^w(x_k, x^w_{k+1}) \psi^v(dx^v_k)}.$$

By inspection, ρ^i_x does not depend on $x^{v'}_{k'}$ except in the following cases: $k' = k - 1$ and $v' \in N(v)$; $k' = k + 1$ and $v' \in N(v)$; $k' = k$ and $v' \in \bigcup_{w \in N(v)} N(w)$. As

$$\rho^i_x(A) \ge \varepsilon^{2\Delta} \frac{\int \mathbf{1}_A(x^v_k) p^v(x_{k-1}, x^v_k) g^v(x^v_k, Y^v_k) \psi^v(dx^v_k)}{\int p^v(x_{k-1}, x^v_k) g^v(x^v_k, Y^v_k) \psi^v(dx^v_k)}$$

as well as

$$\rho^i_x(A) \ge \varepsilon^2 \frac{\int \mathbf{1}_A(x^v_k) g^v(x^v_k, Y^v_k) \prod_{w \in N(v)} p^w(x_k, x^w_{k+1}) \psi^v(dx^v_k)}{\int g^v(x^v_k, Y^v_k) \prod_{w \in N(v)} p^w(x_k, x^w_{k+1}) \psi^v(dx^v_k)},$$

we can use Lemma 4.1 to estimate

$$C_{ij} \le \begin{cases} 1-\varepsilon^2, & \text{if } k' = k-1 \text{ and } v' \in N(v), \\ 1-\varepsilon^{2\Delta}, & \text{if } k' = k+1 \text{ and } v' \in N(v), \\ 1-\varepsilon^{2\Delta}, & \text{if } k' = k \text{ and } v' \in \bigcup_{w \in N(v)} N(w), \\ 0, & \text{otherwise.} \end{cases}$$

This yields

$$\begin{aligned} \sum_{(k',v') \in I} e^{\beta|k-k'|} e^{\beta d(v,v')} C_{(k,v)(k',v')} &\le \bigl(1-\varepsilon^{2\Delta}\bigr)\bigl\{e^{2\beta r}\Delta^2 + 2e^{\beta(r+1)}\Delta\bigr\} \\ &\le 3\bigl(1-\varepsilon^{2\Delta}\bigr)e^{2\beta r}\Delta^2, \end{aligned}$$

where we have used that $r \ge 1$ and $\Delta \ge 1$ in the last inequality.

Case $k = n$. Now we have

$$\begin{aligned} \rho_x^i(A) &= \frac{\int \mathbf{1}_A(x_n^v) p^v(x_{n-1}, x_n^v) g^v(x_n^v, Y_n^v) \psi^v(dx_n^v)}{\int p^v(x_{n-1}, x_n^v) g^v(x_n^v, Y_n^v) \psi^v(dx_n^v)} \\ &\ge \varepsilon^2 \frac{\int \mathbf{1}_A(x_n^v) g^v(x_n^v, Y_n^v) \psi^v(dx_n^v)}{\int g^v(x_n^v, Y_n^v) \psi^v(dx_n^v)}, \end{aligned}$$

and we obtain precisely as above

$$C_{ij} \le \begin{cases} 1-\varepsilon^2, & \text{if } k' = n-1 \text{ and } v' \in N(v), \\ 0, & \text{otherwise.} \end{cases}$$

We therefore find

$$\sum_{(k',v') \in I} e^{\beta|k-k'|} e^{\beta d(v,v')} C_{(n,v)(k',v')} \le \bigl(1-\varepsilon^2\bigr)e^{\beta(r+1)}\Delta.$$

Combining the above three cases and the assumption of the proposition yields

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

Thus, Lemma 4.3 gives

$$\max_{(k,v) \in I} \sum_{(k',v') \in I} e^{\beta\{|k-k'|+d(v,v')\}} D_{(k,v)(k',v')} \le 2.$$

Now consider the quantities b_j in Theorem 3.1. By the Markov property, it is evident that $\rho_x^i = \tilde{\rho}_x^i$ whenever $i = (k, v)$ with $k \ge 1$. On the other hand, for $k = 0$ we obtain $\rho_x^i = \mu_{x_0,x_1}^v$ and $\tilde{\rho}_x^i = \nu_{x_0,x_1}^v$. Applying Theorem 3.1 therefore yields

$$\bigl\|\pi_n^\mu - \pi_n^\nu\bigr\|_J = \|\rho - \tilde{\rho}\|_{\{n\}\times J} \le \sum_{v \in J} \sum_{v' \in V} D_{(n,v)(0,v')} \sup_{x,z \in \mathbb{X}} \bigl\|\mu_{x,z}^{v'} - \nu_{x,z}^{v'}\bigr\|.$$

However, note that

$$\begin{aligned}
&\sum_{v'\in V} D_{(n,v)(0,v')} \sup_{x,z\in\mathbb{X}} \|\mu^{v'}_{x,z} - \nu^{v'}_{x,z}\| \\
&\quad = e^{-\beta n} \sum_{v'\in V} e^{\beta\{n+d(v,v')\}} D_{(n,v)(0,v')} e^{-\beta d(v,v')} \sup_{x,z\in\mathbb{X}} \|\mu^{v'}_{x,z} - \nu^{v'}_{x,z}\| \\
&\quad \le 2e^{-\beta n} \max_{v'\in V} e^{-\beta d(v,v')} \sup_{x,z\in\mathbb{X}} \|\mu^{v'}_{x,z} - \nu^{v'}_{x,z}\|,
\end{aligned}$$

using the above estimate on the matrix D. Substituting this into the bound for $\|\pi^\mu_n - \pi^\nu_n\|_J$ yields the statement of the proposition for the special case $s = 0$.

To obtain the result for any $s < n$, note that $\mathsf{F}_n \cdots \mathsf{F}_{s+1}\mu$ and π^μ_{n-s} differ only in that a different sequence of observations ($Y_{s+1}, \ldots, Y_n$ versus $Y_1, \ldots, Y_{n-s}$) is used in the computation of these quantities. As our bound holds uniformly in the observation sequence, however, the general result follows immediately. □

As a corollary of Proposition 4.4, let us derive a simple filter stability statement that illustrates the role of decay of correlations (this will not be used elsewhere).

COROLLARY 4.7 (Filter stability). *Suppose there exists $\varepsilon > 0$ such that*

$$\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1} \qquad \text{for all } v \in V, x, z \in \mathbb{X},$$

and such that

$$\varepsilon > \varepsilon_0 = \left(1 - \frac{1}{6\Delta^2}\right)^{1/2\Delta}.$$

Then for any probability measures μ, ν on $\mathbb{X}$ and $J \subseteq V$, $n \ge 0$, we have

$$\|\pi^\mu_n - \pi^\nu_n\|_J \le 4 \operatorname{card} J \gamma^{n/2r},$$

where $\gamma = 6\Delta^2(1 - \varepsilon^{2\Delta}) < 1$.

PROOF. We first apply Proposition 4.4 with $\mu = \delta_x$. Then $\operatorname{Corr}(\mu, \beta) = 0$ for any $\beta > 0$. Choosing $\beta = -(2r)^{-1} \log \gamma > 0$, we find that

$$\operatorname{Corr}(\mu, \beta) + 3(1 - \varepsilon^{2\Delta}) e^{2\beta r} \Delta^2 = \tfrac{1}{2},$$

so that the assumption of Proposition 4.4 is satisfied. Therefore,

$$\|\pi^x_n - \pi^\nu_n\|_J \le 4 \operatorname{card} J e^{-\beta n} = 4 \operatorname{card} J \gamma^{n/2r}.$$

To obtain the result for arbitrary μ, note that

$$\begin{aligned}
\pi^\mu_n(A) &= \mathbf{P}^\mu[X_n \in A | Y_1, \ldots, Y_n] \\
&= \mathbf{E}^\mu\big[\mathbf{P}^\mu[X_n \in A | X_0, Y_1, \ldots, Y_n] | Y_1, \ldots, Y_n\big] \\
&= \mathbf{E}^\mu\big[\pi^{\delta_{X_0}}_n(A) | Y_1, \ldots, Y_n\big].
\end{aligned}$$

Therefore, by Jensen's inequality,

$$\|\pi_n^\mu - \pi_n^\nu\|_J \le \mathbf{E}^\mu\big[\|\pi_n^{\delta_{X_0}} - \pi_n^\nu\|_J | Y_1, \ldots, Y_n\big] \le \sup_{x\in\mathbb{X}} \|\pi_n^x - \pi_n^\nu\|_J,$$

which yields the result. □

While Proposition 4.4 requires a decay of correlations assumption on the initial condition [Corr(μ, β) must be sufficiently small], Corollary 4.7 works for any initial condition provided that $\varepsilon > \varepsilon_0$ is sufficiently large (which is necessary in general, see Section 2.3.1). Thus, no assumption is needed on the initial condition if we want to show only that the filter is stable in time. On the other hand, Proposition 4.4 controls not only the stability in time, but also the spatial accumulation of error between μ and ν by virtue of the damping factor $e^{-\beta d(v,v')}$: the decay of correlations property of the initial condition is essential to obtain this type of local control. The latter is of central importance if we wish to obtain local error bounds for filter approximations that are uniform in time and in the model dimension.

4.3. *The block projection error.* The proof of a time-uniform error bound between π_n^μ and $\tilde\pi_n^\mu$ requires two ingredients: we need the filter stability property of π_n^μ, developed in the previous section, in order to mitigate the accumulation of approximation errors over time; and we need to control the approximation error between π_n^μ and $\tilde\pi_n^\mu$ in one time step. The latter is the purpose of this section.

We will in fact consider two separate cases. To control the total error $\|\pi_n^\mu - \tilde\pi_n^\mu\|_J$, we need to consider the one-step error made in each time step $s = 1, \ldots, n$. For time steps $s < n$ (for which the error is dissipated by the stability of the filter), the error must be measured in terms of the quantities that appear in Proposition 4.4: that is, we must control $\|(\mathsf{F}_s\nu)^v_{x,z} - (\tilde{\mathsf{F}}_s\nu)^v_{x,z}\|$. On the other hand, in the last time step $s = n$, we must control directly $\|\mathsf{F}_n\nu - \tilde{\mathsf{F}}_n\nu\|_J$. While the proofs of these cases are quite similar, each must be considered separately in the following.

We begin by bounding the error in time steps $s < n$.

PROPOSITION 4.8 (Block error, $s < n$). *Suppose there exists $\varepsilon > 0$ such that*

$$\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1} \qquad \text{for all } v \in V, x, z \in \mathbb{X}.$$

Let ν be a probability measure on $\mathbb{X}$, and suppose that

$$\mathrm{Corr}(\nu, \beta) + (1 - \varepsilon^2)e^{\beta(r+1)}\Delta \le \tfrac{1}{2}$$

for a sufficiently small constant $\beta > 0$. Then we have

$$\sup_{x,z\in\mathbb{X}} \|(\mathsf{F}_s\nu)^v_{x,z} - (\tilde{\mathsf{F}}_s\nu)^v_{x,z}\| \le 4e^{-\beta}(1 - \varepsilon^{2\Delta})e^{-\beta d(v,\partial K)}$$

for every $s \in \mathbb{N}$, $K \in \mathcal{K}$ and $v \in K$.

This result makes precise the idea that was heuristically expressed in Section 2.2: if the measure ν possesses the decay of correlations property, then the error at site v incurred by applying the block filter rather than the true filter decays exponentially in the distance between v and the boundary of the block that it is in.

PROOF OF PROPOSITION 4.8. We begin by writing out the definitions

$$(\mathsf{F}_s\nu)(A) = \frac{\int \mathbf{1}_A(x)\prod_{w\in V} p^w(x_0,x^w)g^w(x^w,Y_s^w)\nu(dx_0)\psi(dx)}{\int \prod_{w\in V} p^w(x_0,x^w)g^w(x^w,Y_s^w)\nu(dx_0)\psi(dx)},$$

$$(\tilde{\mathsf{F}}_s\nu)(A) = \frac{\int \mathbf{1}_A(x)\prod_{K'\in\mathcal{K}}[\int \prod_{w\in K'} p^w(x_0,x^w)g^w(x^w,Y_s^w)\nu(dx_0)]\psi(dx)}{\int \prod_{K'\in\mathcal{K}}[\int \prod_{w\in K'} p^w(x_0,x^w)g^w(x^w,Y_s^w)\nu(dx_0)]\psi(dx)}.$$

Let us fix $K\in\mathcal{K}$, $v\in K$ throughout the proof. Then

$$(\mathsf{F}_s\nu)^v_x(A) = \frac{\int \mathbf{1}_A(x^v)g^v(x^v,Y_s^v)\prod_{w\in V} p^w(x_0,x^w)\nu(dx_0)\psi^v(dx^v)}{\int g^v(x^v,Y_s^v)\prod_{w\in V} p^w(x_0,x^w)\nu(dx_0)\psi^v(dx^v)},$$

$$(\tilde{\mathsf{F}}_s\nu)^v_x(A) = \frac{\int \mathbf{1}_A(x^v)g^v(x^v,Y_s^v)\prod_{w\in K} p^w(x_0,x^w)\nu(dx_0)\psi^v(dx^v)}{\int g^v(x^v,Y_s^v)\prod_{w\in K} p^w(x_0,x^w)\nu(dx_0)\psi^v(dx^v)}.$$

Define $I=(\{0\}\times V)\cup(1,v)$ and $\mathbb{S}=\mathbb{X}\times\mathbb{X}^v$, and the probability measures on $\mathbb{S}$

$$\rho(A) = \frac{\int \mathbf{1}_A(x_0,x^v)g^v(x^v,Y_s^v)\prod_{w\in V} p^w(x_0,x^w)\prod_{u\in N(v)} p^u(x,z^u)\nu(dx_0)\psi^v(dx^v)}{\int g^v(x^v,Y_s^v)\prod_{w\in V} p^w(x_0,x^w)\prod_{u\in N(v)} p^u(x,z^u)\nu(dx_0)\psi^v(dx^v)},$$

$$\tilde\rho(A) = \frac{\int \mathbf{1}_A(x_0,x^v)g^v(x^v,Y_s^v)\prod_{w\in K} p^w(x_0,x^w)\prod_{u\in N(v)} p^u(x,z^u)\nu(dx_0)\psi^v(dx^v)}{\int g^v(x^v,Y_s^v)\prod_{w\in K} p^w(x_0,x^w)\prod_{u\in N(v)} p^u(x,z^u)\nu(dx_0)\psi^v(dx^v)}.$$

Then we have by construction

$$\big\|(\mathsf{F}_s\nu)^v_{x,z} - (\tilde{\mathsf{F}}_s\nu)^v_{x,z}\big\| = \|\rho-\tilde\rho\|_{(1,v)}.$$

We will apply Theorem 3.1 to bound $\|\rho-\tilde\rho\|_{(1,v)}$. To this end, we must bound C_{ij} and b_i with $i=(k',v')$ and $j=(k'',v'')$. We distinguish two cases.

Case $k'=0$. In this case, we have

$$\rho^i_{(x_0,x^v)}(A) = \frac{\int \mathbf{1}_A(x_0^{v'})\prod_{w\in N(v')} p^w(x_0,x^w)\nu^{v'}_{x_0}(dx_0^{v'})}{\int \prod_{w\in N(v')} p^w(x_0,x^w)\nu^{v'}_{x_0}(dx_0^{v'})},$$

$$\tilde\rho^i_{(x_0,x^v)}(A) = \frac{\int \mathbf{1}_A(x_0^{v'})\prod_{w\in N(v')\cap K} p^w(x_0,x^w)\nu^{v'}_{x_0}(dx_0^{v'})}{\int \prod_{w\in N(v')\cap K} p^w(x_0,x^w)\nu^{v'}_{x_0}(dx_0^{v'})}.$$

In particular, $\rho^i_{(x_0,x^v)}=\nu^{v'}_{x_0,x}$, so $C_{ij}\le C^v_{v'v''}$ if $k''=0$. Moreover, as

$$\rho^i_{(x_0,x^v)}(A) \ge \varepsilon^2\frac{\int \mathbf{1}_A(x_0^{v'})\prod_{w\in N(v')\setminus\{v\}} p^w(x_0,x^w)\nu^{v'}_{x_0}(dx_0^{v'})}{\int \prod_{w\in N(v')\setminus\{v\}} p^w(x_0,x^w)\nu^{v'}_{x_0}(dx_0^{v'})},$$

we have $C_{ij} \le 1-\varepsilon^2$ if $k''=1$ (so $v''=v$) and $v \in N(v')$ by Lemma 4.1, and $C_{ij}=0$ otherwise. We therefore immediately obtain the estimate

$$\sum_{(k'',v'')\in I} e^{\beta k''} e^{\beta d(v',v'')} C_{(0,v')(k'',v'')} \le \mathrm{Corr}(\nu,\beta) + (1-\varepsilon^2)e^{\beta(r+1)}.$$

On the other hand, note that $\rho^i_{(x_0,x^v)} = \tilde{\rho}^i_{(x_0,x^v)}$ if $N(v') \subseteq K$, and that we have $\rho^i_{(x_0,x^v)} \ge \varepsilon^{2\Delta} \nu^{v'}_{x_0}$ and $\tilde{\rho}^i_{(x_0,x^v)} \ge \varepsilon^{2\Delta}\nu^{v'}_{x_0}$. Therefore, by Lemma 4.1

$$b_i = \sup_{(x_0,x^v)\in\mathbb{S}} \|\rho^i_{(x_0,x^v)} - \tilde{\rho}^i_{(x_0,x^v)}\| \le \begin{cases} 0, & \text{for } v' \in K \setminus \partial K, \\ 2(1-\varepsilon^{2\Delta}), & \text{otherwise.} \end{cases}$$

Case $k'=1$. In this case, we have

$$\begin{aligned}\rho^i_{(x_0,x^v)}(A) &= \tilde{\rho}^i_{(x_0,x^v)}(A) \\ &= \frac{\int \mathbf{1}_A(x^v) g^v(x^v, Y^v_s) p^v(x_0,x^v) \prod_{u\in N(v)} p^u(x,z^u)\psi^v(dx^v)}{\int g^v(x^v,Y^v_s) p^v(x_0,x^v)\prod_{u\in N(v)} p^u(x,z^u)\psi^v(dx^v)}.\end{aligned}$$

Thus, $b_i=0$, and estimating as above we obtain $C_{ij} \le 1-\varepsilon^2$ whenever $k''=0$ and $v'' \in N(v)$, and $C_{ij}=0$ otherwise. In particular, we obtain

$$\sum_{(k'',v'')\in I} e^{\beta|1-k''|} e^{\beta d(v,v'')} C_{(1,v)(k'',v'')} \le (1-\varepsilon^2)e^{\beta(r+1)}\Delta.$$

Combining the above two cases and the assumption of the proposition yields

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

Applying Theorem 3.1 and Lemma 4.3 gives

$$\begin{aligned}\|(\mathsf{F}_s\nu)^v_{x,z} - (\tilde{\mathsf{F}}_s\nu)^v_{x,z}\| &= \|\rho-\tilde{\rho}\|_{(1,v)} \\ &\le 2(1-\varepsilon^{2\Delta}) \sum_{v'\in V\setminus(K\setminus\partial K)} D_{(1,v)(0,v')} \\ &\le 4e^{-\beta}(1-\varepsilon^{2\Delta})e^{-\beta d(v,\partial K)}.\end{aligned}$$

As the choice of $x,z\in\mathbb{X}$ was arbitrary, the proof is complete. □

We now use a similar argument to bound the error in time step n.

PROPOSITION 4.9 (Block error, $s=n$). *Suppose there exists* $\varepsilon>0$ *such that*

$$\varepsilon \le p^v(x,z^v) \le \varepsilon^{-1} \qquad \text{for all } v\in V, x,z\in\mathbb{X}.$$

Let ν *be a probability measure on* $\mathbb{X}$, *and suppose that*

$$\mathrm{Corr}(\nu,\beta) + (1-\varepsilon^2)e^{\beta(r+1)}\Delta \le \tfrac{1}{2}$$

for a sufficiently small constant $\beta > 0$. Then we have

$$\|\mathsf{F}_n\nu - \tilde{\mathsf{F}}_n\nu\|_J \le 4e^{-\beta}(1-\varepsilon^{2\Delta})e^{-\beta d(J,\partial K)}\operatorname{card} J$$

for every $K \in \mathcal{K}$ and $J \subseteq K$.

PROOF. Define $I = \{0,1\} \times V$ and $\mathbb{S} = \mathbb{X}^2$. Fix $K \in \mathcal{K}$, and let

$$\rho(A) = \frac{\int \mathbf{1}_A(x_0,x_1)\prod_{v\in V} p^v(x_0,x_1^v)g^v(x_1^v,Y_n^v)\nu(dx_0)\psi(dx_1)}{\int \prod_{v\in V} p^v(x_0,x_1^v)g^v(x_1^v,Y_n^v)\nu(dx_0)\psi(dx_1)},$$

$$\tilde\rho(A) = \frac{\int \mathbf{1}_A(x_0,x_1)\prod_{v\in K} p^v(x_0,x_1^v)\prod_{w\in V} g^w(x_1^w,Y_n^w)\nu(dx_0)\psi(dx_1)}{\int \prod_{v\in K} p^v(x_0,x_1^v)\prod_{w\in V} g^w(x_1^w,Y_n^w)\nu(dx_0)\psi(dx_1)}.$$

Then for any $J \subseteq K$, we have

$$\|\mathsf{F}_n\nu - \tilde{\mathsf{F}}_n\nu\|_J = \|\rho - \tilde\rho\|_{\{1\}\times J}.$$

We will apply Theorem 3.1 to bound $\|\rho - \tilde\rho\|_{\{1\}\times J}$. To this end, we must bound C_{ij} and b_i with $i = (k,v)$ and $j = (k',v')$. We distinguish two cases.

Case $k = 0$. In this case, we have

$$\rho_x^i(A) = \frac{\int \mathbf{1}_A(x_0^v)\prod_{w\in N(v)} p^w(x_0,x_1^w)\nu_{x_0}^v(dx_0^v)}{\int \prod_{w\in N(v)} p^w(x_0,x_1^w)\nu_{x_0}^v(dx_0^v)},$$

$$\tilde\rho_x^i(A) = \frac{\int \mathbf{1}_A(x_0^v)\prod_{w\in N(v)\cap K} p^w(x_0,x_1^w)\nu_{x_0}^v(dx_0^v)}{\int \prod_{w\in N(v)\cap K} p^w(x_0,x_1^w)\nu_{x_0}^v(dx_0^v)}.$$

In particular, $\rho_x^i = \nu_{x_0,x_1}^v$, so $C_{ij} \le C_{vv'}^{\nu}$ if $k' = 0$. Moreover, as

$$\rho_x^i(A) \ge \varepsilon^2\frac{\int \mathbf{1}_A(x_0^v)\prod_{w\in N(v)\setminus\{v'\}} p^w(x_0,x_1^w)\nu_{x_0}^v(dx_0^v)}{\int \prod_{w\in N(v)\setminus\{v'\}} p^w(x_0,x_1^w)\nu_{x_0}^v(dx_0^v)},$$

we have $C_{ij} \le 1-\varepsilon^2$ if $k' = 1$ and $v' \in N(v)$ by Lemma 4.1, and $C_{ij} = 0$ otherwise. We therefore immediately obtain the estimate

$$\sum_{(k',v')\in I} e^{\beta k'}e^{\beta d(v,v')}C_{(0,v)(k',v')} \le \operatorname{Corr}(\nu,\beta) + (1-\varepsilon^2)e^{\beta(r+1)}\Delta.$$

On the other hand, note that $\rho_x^i = \tilde\rho_x^i$ if $N(v) \subseteq K$, and that we have $\rho_x^i \ge \varepsilon^{2\Delta}\nu_{x_0}^v$ and $\tilde\rho_x^i \ge \varepsilon^{2\Delta}\nu_{x_0}^v$. Therefore, we obtain by Lemma 4.1

$$b_i = \sup_{x\in\mathbb{S}}\|\rho_x^i - \tilde\rho_x^i\| \le \begin{cases} 0, & \text{for } v \in K\setminus\partial K,\\ 2(1-\varepsilon^{2\Delta}), & \text{otherwise.}\end{cases}$$

Case $k = 1$. In this case, we have

$$\rho_x^i(A) = \frac{\int \mathbf{1}_A(x_1^v)p^v(x_0,x_1^v)g^v(x_1^v,Y_n^v)\psi^v(dx_1^v)}{\int p^v(x_0,x_1^v)g^v(x_1^v,Y_n^v)\psi^v(dx_1^v)},$$

while $\tilde{\rho}_i^x = \rho_i^x$ if $v \in K$ and

$$\tilde{\rho}_x^i(A) = \frac{\int \mathbf{1}_A(x_1^v) g^v(x_1^v, Y_n^v) \psi^v(dx_1^v)}{\int g^v(x_1^v, Y_n^v) \psi^v(dx_1^v)},$$

otherwise. Thus, we obtain from Lemma 4.1

$$b_i = \sup_{x \in \mathbb{S}} \|\rho_x^i - \tilde{\rho}_x^i\| \leq \begin{cases} 0, & \text{for } v \in K, \\ 2(1-\varepsilon^2), & \text{otherwise.} \end{cases}$$

On the other hand, we can readily estimate as above

$$\sum_{(k',v') \in I} e^{\beta|1-k'|} e^{\beta d(v,v')} C_{(1,v)(k',v')} \leq (1-\varepsilon^2) e^{\beta(r+1)} \Delta.$$

Combining the above two cases and the assumption of the proposition yields

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

Applying Theorem 3.1 and Lemma 4.3 gives

$$\begin{aligned} \|\mathsf{F}_n \nu - \tilde{\mathsf{F}}_n \nu\|_J &= \|\rho - \tilde{\rho}\|_{\{1\} \times J} \\ &\leq 2(1-\varepsilon^{2\Delta}) \sum_{v \in J} \left\{ \sum_{v' \in (V \setminus K) \cup \partial K} D_{(1,v)(0,v')} + \sum_{v' \in V \setminus K} D_{(1,v)(1,v')} \right\} \\ &\leq 4e^{-\beta}(1-\varepsilon^{2\Delta}) e^{-\beta d(J, \partial K)} \operatorname{card} J \end{aligned}$$

for every $J \subseteq K$. $\square$

4.4. *Decay of correlations of the block filter.* To idea behind the block filter $\tilde{\pi}_n^\mu$ is that the error should decay exponentially in the block size by virtue of the decay of correlations property. While we have developed above the two ingredients (filter stability and one-step error bound) required to obtain a time-uniform error bound between π_n^μ and $\tilde{\pi}_n^\mu$, we have done this by imposing the decay of correlations property as an assumption. Thus, perhaps the crucial point remains to be proved: we must show that decay of correlations does indeed hold, that is, $\operatorname{Corr}(\tilde{\pi}_n^\mu, \beta)$ can be controlled uniformly in time. This is the goal of the present section.

Unfortunately, $\operatorname{Corr}(\tilde{\pi}_n^\mu, \beta)$ is not straightforward to control directly. We therefore introduce an alternative measure of correlation decay that will be easier to control. For any probability measure μ on $\mathbb{X}$ and $x, z \in \mathbb{X}$, $v \in V$, $K \in \mathcal{K}$, let

$$\begin{aligned} \mu_{x,z}^{v,K}(A) &:= \mathbf{P}^\mu[X_0^v \in A | X_0^{V \setminus \{v\}} = x^{V \setminus \{v\}}, X_1^K = z^K] \\ &= \frac{\int \mathbf{1}_A(x^v) \prod_{w \in N(v) \cap K} p^w(x, z^w) \mu_x^v(dx^v)}{\int \prod_{w \in N(v) \cap K} p^w(x, z^w) \mu_x^v(dx^v)}. \end{aligned}$$

We now define

$$\tilde{C}^{\mu}_{vv'} := \frac{1}{2} \max_{K\in\mathcal{K}} \sup_{z\in\mathbb{X}} \sup_{x,\tilde{x}\in\mathbb{X}:x^{V\setminus\{v'\}}=\tilde{x}^{V\setminus\{v'\}}} \|\mu^{v,K}_{x,z} - \mu^{v,K}_{\tilde{x},z}\|$$

for $v, v' \in V$. The quantity

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

is a measure of correlation decay that is well adapted to the block filter. In order for this quantity to be useful, we must first show that it controls $\mathrm{Corr}(\mu,\beta)$.

LEMMA 4.10. *For any probability measure μ and $\beta > 0$, we have*

$$\mathrm{Corr}(\mu,\beta) \le (1-\varepsilon^{2\Delta}) e^{2\beta r}\Delta^2 + 2\varepsilon^{-2\Delta}\,\widetilde{\mathrm{Corr}}(\mu,\beta).$$

PROOF. By definition

$$\mu^{v}_{x,z}(A) = \frac{\int \mathbf{1}_A(x^v)\prod_{w\in N(v)\setminus K} p^w(x,z^w)\mu^{v,K}_{x,z}(dx^v)}{\int \prod_{w\in N(v)\setminus K} p^w(x,z^w)\mu^{v,K}_{x,z}(dx^v)}.$$

Let $x, \tilde{x} \in \mathbb{X}$ be such that $x^{V\setminus\{v'\}} = \tilde{x}^{V\setminus\{v'\}}$. If $v' \notin \bigcup_{w\in N(v)} N(w)$, then

$$\|\mu^{v}_{x,z} - \mu^{v}_{\tilde{x},z}\| \le 2\varepsilon^{-2\Delta}\|\mu^{v,K}_{x,z} - \mu^{v,K}_{\tilde{x},z}\|$$

by Lemma 4.2. On the other hand, note that

$$\mu^{v}_{x,z}(A) \ge \varepsilon^{2\Delta}\mu^{v,K}_{x,z}(A), \qquad \mu^{v}_{\tilde{x},z}(A) \ge \varepsilon^{2\Delta}\mu^{v,K}_{\tilde{x},z}(A).$$

We can therefore estimate using Lemma 4.1 for $v' \in \bigcup_{w\in N(v)} N(w)$

$$\|\mu^{v}_{x,z} - \mu^{v}_{\tilde{x},z}\| \le 2(1-\varepsilon^{2\Delta}) + \varepsilon^{2\Delta}\|\mu^{v,K}_{x,z} - \mu^{v,K}_{\tilde{x},z}\|.$$

Thus, we obtain

$$\begin{aligned}\mathrm{Corr}(\mu,\beta) &\le (1-\varepsilon^{2\Delta}) \max_{v\in V} \sum_{v'\in\bigcup_{w\in N(v)} N(w)} e^{\beta d(v,v')} + 2\varepsilon^{-2\Delta}\,\widetilde{\mathrm{Corr}}(\mu,\beta)\\ &\le (1-\varepsilon^{2\Delta}) e^{2\beta r}\Delta^2 + 2\varepsilon^{-2\Delta}\,\widetilde{\mathrm{Corr}}(\mu,\beta).\end{aligned}$$

As μ and β were arbitrary, the proof is complete. □

We now aim to establish a time-uniform bound on $\widetilde{\mathrm{Corr}}(\tilde{\pi}^{\mu}_n,\beta)$. To this end, we first prove a one-step bound which will subsequently be iterated.

PROPOSITION 4.11. *Suppose there exists $\varepsilon > 0$ such that*

$$\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1} \qquad \text{for all } v \in V, x, z \in \mathbb{X}.$$

Let ν be a probability measure on $\mathbb{X}$, and suppose that

$$\widetilde{\mathrm{Corr}}(\nu, \beta) + (1 - \varepsilon^2) e^{\beta(r+1)} \Delta \le \tfrac{1}{2}$$

for a sufficiently small constant $\beta > 0$. Then we have

$$\widetilde{\mathrm{Corr}}(\tilde{\mathsf{F}}_s \nu, \beta) \le 2(1 - \varepsilon^{2\Delta}) e^{2\beta r} \Delta^2$$

for any $s \in \mathbb{N}$.

PROOF. Let $K, K' \in \mathcal{K}$, $v \in K$, $v' \in V$ $(v' \neq v)$, and let $z, x, \tilde{x} \in \mathbb{X}$ such that $x^{V\setminus\{v'\}} = \tilde{x}^{V\setminus\{v'\}}$. These choices will be fixed until further notice.

Define $I = (\{0\} \times V) \cup (1, v)$ and $\mathbb{S} = \mathbb{X} \times \mathbb{X}^v$, and let

$$\rho(A) = \frac{\int \mathbf{1}_A(x_0, x^v) g^v(x^v, Y_s^v) \prod_{w\in K} p^w(x_0, x^w) \prod_{u\in N(v)\cap K'} p^u(x, z^u) \nu(dx_0) \psi^v(dx^v)}{\int g^v(x^v, Y_s^v) \prod_{w\in K} p^w(x_0, x^w) \prod_{u\in N(v)\cap K'} p^u(x, z^u) \nu(dx_0) \psi^v(dx^v)},$$

$$\tilde{\rho}(A) = \frac{\int \mathbf{1}_A(x_0, \tilde{x}^v) g^v(\tilde{x}^v, Y_s^v) \prod_{w\in K} p^w(x_0, \tilde{x}^w) \prod_{u\in N(v)\cap K'} p^u(\tilde{x}, z^u) \nu(dx_0) \psi^v(d\tilde{x}^v)}{\int g^v(\tilde{x}^v, Y_s^v) \prod_{w\in K} p^w(x_0, \tilde{x}^w) \prod_{u\in N(v)\cap K'} p^u(\tilde{x}, z^u) \nu(dx_0) \psi^v(d\tilde{x}^v)}.$$

Then we have by construction

$$\|(\tilde{\mathsf{F}}_s \nu)_{x,z}^{v,K'} - (\tilde{\mathsf{F}}_s \nu)_{\tilde{x},z}^{v,K'}\| = \|\rho - \tilde{\rho}\|_{(1,v)}.$$

We will apply Theorem 3.1 to bound $\|\rho - \tilde{\rho}\|_{(1,v)}$. To this end, we must bound C_{ij} and b_i with $i = (k, t)$ and $j = (k', t')$. We distinguish two cases.

Case $k = 0$. In this case, we have

$$\rho^i_{(x_0,x^v)}(A) = \frac{\int \mathbf{1}_A(x_0^t) \prod_{w\in N(t)\cap K} p^w(x_0, x^w) \nu^t_{x_0}(dx_0^t)}{\int \prod_{w\in N(t)\cap K} p^w(x_0, x^w) \nu^t_{x_0}(dx_0^t)},$$

$$\tilde{\rho}^i_{(x_0,\tilde{x}^v)}(A) = \frac{\int \mathbf{1}_A(x_0^t) \prod_{w\in N(t)\cap K} p^w(x_0, \tilde{x}^w) \nu^t_{x_0}(dx_0^t)}{\int \prod_{w\in N(t)\cap K} p^w(x_0, \tilde{x}^w) \nu^t_{x_0}(dx_0^t)}.$$

Note that $\rho^i_{(x_0,x^v)} = \nu^{t,K}_{x_0,x}$. We therefore have $C_{ij} \le \tilde{C}^v_{tt'}$ when $k' = 0$. Moreover,

$$\rho^i_{(x_0,x^v)}(A) \ge \varepsilon^2 \frac{\int \mathbf{1}_A(x_0^t) \prod_{w\in N(t)\cap(K\setminus\{v\})} p^w(x_0, x^w) \nu^t_{x_0}(dx_0^t)}{\int \prod_{w\in N(t)\cap(K\setminus\{v\})} p^w(x_0, x^w) \nu^t_{x_0}(dx_0^t)}$$

implies $C_{ij} \le 1 - \varepsilon^2$ if $k' = 1$ and $v \in N(t)$ by Lemma 4.1, and $C_{ij} = 0$ otherwise. On the other hand, note that as $x^{V\setminus\{v'\}} = \tilde{x}^{V\setminus\{v'\}}$ we have $\rho^i_{(x_0,x^v)} = \tilde{\rho}^i_{(x_0,\tilde{x}^v)}$ if $v' \notin N(t) \cap K$, while both $\rho^i_{(x_0,x^v)}(A)$ and $\tilde{\rho}^i_{(x_0,\tilde{x}^v)}(A)$ dominate

$$\varepsilon^2 \frac{\int \mathbf{1}_A(x_0^t) \prod_{w\in N(t)\cap(K\setminus\{v'\})} p^w(x_0, x^w) \nu^t_{x_0}(dx_0^t)}{\int \prod_{w\in N(t)\cap(K\setminus\{v'\})} p^w(x_0, x^w) \nu^t_{x_0}(dx_0^t)}.$$

Therefore, by Lemma 4.1

$$b_{(0,t)} \le \begin{cases} 0, & \text{for } v' \notin N(t) \cap K, \\ 2(1-\varepsilon^2), & \text{otherwise.} \end{cases}$$

Case $k = 1$. In this case, we have

$$\rho^i_{(x_0,x^v)}(A) = \frac{\int \mathbf{1}_A(x^v) g^v(x^v, Y_s^v) p^v(x_0, x^v) \prod_{u \in N(v)\cap K'} p^u(x, z^u) \psi^v(dx^v)}{\int g^v(x^v, Y_s^v) p^v(x_0, x^v) \prod_{u \in N(v)\cap K'} p^u(x, z^u) \psi^v(dx^v)},$$

$$\tilde\rho^i_{(x_0,\tilde x^v)}(A) = \frac{\int \mathbf{1}_A(\tilde x^v) g^v(\tilde x^v, Y_s^v) p^v(x_0, \tilde x^v) \prod_{u \in N(v)\cap K'} p^u(\tilde x, z^u) \psi^v(d\tilde x^v)}{\int g^v(\tilde x^v, Y_s^v) p^v(x_0, \tilde x^v) \prod_{u \in N(v)\cap K'} p^u(\tilde x, z^u) \psi^v(d\tilde x^v)}.$$

Estimating as above, we obtain $C_{ij} \le 1 - \varepsilon^2$ whenever $k' = 0$ and $t' \in N(v)$, and $C_{ij} = 0$ otherwise. Similarly, arguing again as above, we obtain

$$b_{(1,v)} \le \begin{cases} 0, & \text{for } v' \notin \bigcup_{w \in N(v)\cap K'} N(w), \\ 2(1-\varepsilon^{2\Delta}), & \text{otherwise.} \end{cases}$$

Define the matrix $\{C_{ij}(v)\}_{i,j\in I}$ with the following entries:

$$\begin{aligned} C_{(0,t)(0,t')}(v) &= \tilde C^v_{tt'}, \\ C_{(0,t)(1,v)}(v) &= C_{(1,v)(0,t)}(v) = (1-\varepsilon^2)\mathbf{1}_{t\in N(v)}, \\ C_{(1,v)(1,v)}(v) &= 0. \end{aligned}$$

Combining the above two cases yields $C_{ij} \le C_{ij}(v)$, and we readily compute

$$\sum_{(k',t')\in I} e^{\beta\{|k-k'|+d(t,t')\}} C_{(k,t)(k',t')}(v) \le \widetilde{\mathrm{Corr}}(\nu,\beta) + (1-\varepsilon^2)e^{\beta(r+1)}\Delta \le \frac{1}{2}$$

where we have used the assumption of the proposition. By Theorem 3.1

$$\begin{aligned} \big\|(\tilde{\mathsf{F}}_s\nu)^{v,K'}_{x,z} - (\tilde{\mathsf{F}}_s\nu)^{v,K'}_{\tilde x,z}\big\| &= \|\rho - \tilde\rho\|_{(1,v)} \\ &\le 2(1-\varepsilon^2)\mathbf{1}_{v'\in K} \sum_{t'\in N(v')} D_{(1,v)(0,t')}(v) \\ &\quad + 2(1-\varepsilon^{2\Delta})\mathbf{1}_{v'\in\bigcup_{w\in N(v)\cap K'} N(w)} D_{(1,v)(1,v)}(v), \end{aligned}$$

where $D(v) := \sum_{n\ge 0} C(v)^n$. But note that the right-hand side does not depend on K' or $z, x, \tilde x$ (provided $x^{V\setminus\{v'\}} = \tilde x^{V\setminus\{v'\}}$). We therefore obtain

$$\begin{aligned} \tilde C^{\tilde{\mathsf{F}}_s\nu}_{vv'} &\le (1-\varepsilon^2)\mathbf{1}_{v'\in K} \sum_{t'\in N(v')} D_{(1,v)(0,t')}(v) \\ &\quad + (1-\varepsilon^{2\Delta})\mathbf{1}_{v'\in\bigcup_{w\in N(v)\cap K'} N(w)} D_{(1,v)(1,v)}(v) \end{aligned}$$

for every $K \in \mathcal{K}$, $v \in K$ and $v' \in V$. In particular, we have

$$\sum_{v' \in V} e^{\beta d(v,v')} \tilde{C}^{\tilde{\mathsf{F}}_s \nu}_{vv'} \le (1-\varepsilon^2) \sum_{v' \in K} e^{\beta d(v,v')} \sum_{t' \in N(v')} D_{(1,v)(0,t')}(v) + (1-\varepsilon^{2\Delta}) D_{(1,v)(1,v)}(v) \sum_{v' \in \bigcup_{w \in N(v) \cap K'} N(w)} e^{\beta d(v,v')}.$$

To proceed, we note that

$$\sum_{v' \in K} e^{\beta d(v,v')} \sum_{t' \in N(v')} D_{(1,v)(0,t')}(v) \le e^{\beta r} \Delta \sum_{v' \in V} e^{\beta d(v,v')} D_{(1,v)(0,v')}(v),$$

where we have used that $d(v,v') \le d(v,t') + r$ for $t' \in N(v')$. Similarly, we have

$$\sum_{v' \in \bigcup_{w \in N(v) \cap K'} N(w)} e^{\beta d(v,v')} \le e^{2\beta r} \Delta^2.$$

We can therefore estimate

$$\sum_{v' \in V} e^{\beta d(v,v')} \tilde{C}^{\tilde{\mathsf{F}}_s \nu}_{vv'} \le (1-\varepsilon^{2\Delta}) e^{2\beta r} \Delta^2 \sum_{(k',v') \in I} e^{\beta\{|1-k'| + d(v,v')\}} D_{(1,v)(k',v')}(v).$$

Applying Lemma 4.3 to $C(v)$ yields the result. □

We now iterate the above result.

COROLLARY 4.12. *Suppose there exists $\varepsilon > 0$ such that*

$$\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1} \qquad \text{for all } v \in V, x, z \in \mathbb{X},$$

and such that

$$\varepsilon > \varepsilon_0 = \left(1 - \frac{1}{16\Delta^2}\right)^{1/2\Delta}.$$

Let μ be a probability measure on $\mathbb{X}$ such that

$$\widetilde{\mathrm{Corr}}(\mu, \beta) \le \tfrac{1}{8},$$

where $\beta = -(2r)^{-1} \log 16\Delta^2(1-\varepsilon^{2\Delta}) > 0$. Then

$$\widetilde{\mathrm{Corr}}(\tilde{\pi}^\mu_n, \beta) \le \tfrac{1}{8} \qquad \text{for all } n \ge 0.$$

In particular, the latter holds whenever $\mu = \delta_x$ for any $x \in \mathbb{X}$.

PROOF. The assumption $\varepsilon > \varepsilon_0$ implies $\beta > 0$ and

$$(1-\varepsilon^2) e^{\beta(r+1)} \Delta \le \tfrac{1}{16}.$$

Therefore, if $\widetilde{\mathrm{Corr}}(\nu, \beta) \le 1/8$, then Proposition 4.11 yields

$$\widetilde{\mathrm{Corr}}(\tilde{\mathsf{F}}_s \nu, \beta) \le 2(1-\varepsilon^{2\Delta}) e^{2\beta r} \Delta^2 \le \tfrac{1}{8}.$$

Thus, if $\widetilde{\mathrm{Corr}}(\mu,\beta)\le 1/8$, then $\widetilde{\mathrm{Corr}}(\tilde\pi_n^\mu,\beta)\le 1/8$ for all $n\ge 0$. Moreover, as $\widetilde{\mathrm{Corr}}(\delta_x,\beta)=0$, the result holds automatically for $\mu=\delta_x$. $\square$

We finally obtain the requisite bound on $\mathrm{Corr}(\tilde\pi_n^\mu,\beta)$ using Lemma 4.10.

COROLLARY 4.13 (Decay of correlations). *Suppose there exists $\varepsilon>0$ with*

$$\varepsilon\le p^v(x,z^v)\le\varepsilon^{-1}\qquad\text{for all } v\in V,\ x,z\in\mathbb{X},$$

such that

$$\varepsilon>\varepsilon_0=\left(1-\frac{1}{16\Delta^2}\right)^{1/2\Delta}.$$

Let $\beta=-(2r)^{-1}\log 16\Delta^2(1-\varepsilon^{2\Delta})>0$. Then

$$\mathrm{Corr}(\tilde\pi_n^x,\beta)\le\tfrac{1}{3}$$

for every $n\ge 0$ and $x\in\mathbb{X}$.

PROOF. By Corollary 4.12 and Lemma 4.10, we can estimate

$$\mathrm{Corr}(\tilde\pi_n^x,\beta)\le\tfrac{1}{16}+\tfrac{1}{4}\varepsilon^{-2\Delta}\le\tfrac{1}{3},$$

where we used that $\varepsilon^{2\Delta}\ge 1-1/16$. $\square$

4.5. *Bounding the bias.* In the previous sections, we have proved a local filter stability bound (Proposition 4.4), a local one-step error bound (Propositions 4.8 and 4.9), and decay of correlations of the block filter (Corollary 4.13). We can now combine these results to obtain a time-uniform error bound between the filter and the block filter; this controls the bias of the block particle filtering algorithm.

THEOREM 4.14 (Bias term). *Suppose there exists $\varepsilon>0$ such that*

$$\varepsilon\le p^v(x,z^v)\le\varepsilon^{-1}\qquad\text{for all } v\in V,\ x,z\in\mathbb{X},$$

and such that

$$\varepsilon>\varepsilon_0=\left(1-\frac{1}{18\Delta^2}\right)^{1/2\Delta}.$$

Let $\beta=-(2r)^{-1}\log 18\Delta^2(1-\varepsilon^{2\Delta})>0$. Then

$$\|\pi_n^x-\tilde\pi_n^x\|_J\le\frac{8e^{-\beta}}{1-e^{-\beta}}(1-\varepsilon^{2\Delta})\,\mathrm{card}\,J\,e^{-\beta d(J,\partial K)}$$

for every $n\ge 0$, $x\in\mathbb{X}$, $K\in\mathcal{K}$ and $J\subseteq K$.

PROOF. We begin with the elementary error decomposition

$$\|\pi_n^x - \tilde{\pi}_n^x\|_J \le \sum_{s=1}^{n} \|\mathsf{F}_n \cdots \mathsf{F}_{s+1}\mathsf{F}_s\tilde{\pi}_{s-1}^x - \mathsf{F}_n \cdots \mathsf{F}_{s+1}\tilde{\mathsf{F}}_s\tilde{\pi}_{s-1}^x\|_J.$$

We will bound each term in the sum.

Case $s = n$. To bound this term, note that

$$\mathrm{Corr}(\tilde{\pi}_{n-1}^x, \beta) + (1-\varepsilon^2)e^{\beta(r+1)}\Delta \le \tfrac{1}{3} + \tfrac{1}{18} \le \tfrac{1}{2}$$

by Corollary 4.13. Therefore, applying Proposition 4.9 with $\nu = \tilde{\pi}_{n-1}^x$, we obtain

$$\|\mathsf{F}_n\tilde{\pi}_{n-1}^x - \tilde{\mathsf{F}}_n\tilde{\pi}_{n-1}^x\|_J \le 4e^{-\beta}(1-\varepsilon^{2\Delta})e^{-\beta d(J,\partial K)}\,\mathrm{card}\,J.$$

Case $s < n$. To bound this term, note that by Corollary 4.13

$$\mathrm{Corr}(\tilde{\pi}_s^x, \beta) + 3(1-\varepsilon^{2\Delta})e^{2\beta r}\Delta^2 \le \tfrac{1}{3} + \tfrac{1}{6} = \tfrac{1}{2}.$$

Applying Proposition 4.4 with $\mu = \tilde{\pi}_s^x$ and $\nu = \mathsf{F}_s\tilde{\pi}_{s-1}^x$ yields

$$\begin{aligned}&\|\mathsf{F}_n \cdots \mathsf{F}_{s+1}\mathsf{F}_s\tilde{\pi}_{s-1}^x - \mathsf{F}_n \cdots \mathsf{F}_{s+1}\tilde{\mathsf{F}}_s\tilde{\pi}_{s-1}^x\|_J \\ &\quad\le 2e^{-\beta(n-s)}\sum_{v\in J}\max_{v'\in V} e^{-\beta d(v,v')}\sup_{x,z\in\mathbb{X}}\|(\mathsf{F}_s\tilde{\pi}_{s-1}^x)_{x,z}^{v'} - (\tilde{\mathsf{F}}_s\tilde{\pi}_{s-1}^x)_{x,z}^{v'}\|.\end{aligned}$$

On the other hand, as by Corollary 4.13

$$\mathrm{Corr}(\tilde{\pi}_{s-1}^x, \beta) + (1-\varepsilon^2)e^{\beta(r+1)}\Delta \le \tfrac{1}{3} + \tfrac{1}{18} \le \tfrac{1}{2},$$

we have by Proposition 4.8 with $\nu = \tilde{\pi}_{s-1}^x$

$$\sup_{x,z\in\mathbb{X}}\|(\mathsf{F}_s\tilde{\pi}_{s-1}^x)_{x,z}^{v'} - (\tilde{\mathsf{F}}_s\tilde{\pi}_{s-1}^x)_{x,z}^{v'}\| \le 4e^{-\beta}(1-\varepsilon^{2\Delta})e^{-\beta d(v',\partial K)}.$$

We therefore obtain the estimate

$$\begin{aligned}&\|\mathsf{F}_n \cdots \mathsf{F}_{s+1}\mathsf{F}_s\tilde{\pi}_{s-1}^x - \mathsf{F}_n \cdots \mathsf{F}_{s+1}\tilde{\mathsf{F}}_s\tilde{\pi}_{s-1}^x\|_J \\ &\quad\le 8e^{-\beta}(1-\varepsilon^{2\Delta})e^{-\beta(n-s)}e^{-\beta d(J,\partial K)}\,\mathrm{card}\,J,\end{aligned}$$

where we have used $d(v,v') + d(v',\partial K) \ge d(v,\partial K)$.

Substituting the above two cases into the error decomposition and summing the geometric series yields the statement of the theorem. □

4.6. *Local stability of the block filter.* As was explained in Section 3.4, the chief difficulty in obtaining a time-uniform bound on the variance term is to establish stability of the block filter. This will be done in the present section.

We first establish a stability bound for nonrandom initial conditions.

PROPOSITION 4.15. *Suppose there exists $\varepsilon > 0$ such that*

$$\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1} \qquad \text{for all } v \in V, x, z \in \mathbb{X},$$

and such that

$$\varepsilon > \varepsilon_0 = \left(1 - \frac{1}{6\Delta^2}\right)^{1/2\Delta}.$$

Let $\beta = -\log 6\Delta^2(1-\varepsilon^{2\Delta}) > 0$. Then

$$\left\|\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1}\delta_z - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1}\delta_{z'}\right\|_J \le 4 \operatorname{card} J e^{-\beta(n-s)}$$

for every $s < n$, $z, z' \in \mathbb{X}$, $K \in \mathcal{K}$, and $J \subseteq K$.

PROOF. Fix throughout the proof $n > 0$, $K \in \mathcal{K}$, and $J \subseteq K$. We will also assume throughout the proof for notational simplicity that $s = 0$ (the ultimate conclusion will extend to any $s < n$ as in the proof of Proposition 4.4).

We begin by constructing the computation tree as explained in Section 3.4. For future reference, let us work first in the more general setting where the initial distributions $\mu = \bigotimes_{K' \in \mathcal{K}} \mu^{K'}$ and $\nu = \bigotimes_{K' \in \mathcal{K}} \nu^{K'}$ are independent across the blocks (rather than the special case of point masses δ_x and $\delta_{x'}$). Define for $K' \in \mathcal{K}$

$$N(K') = \{K'' \in \mathcal{K} : d(K', K'') \le r\},$$

that is, $N(K')$ is the collection of blocks that interact with block K' in one step of the dynamics [recall that $\operatorname{card} N(K') \le \Delta_{\mathcal{K}}$]. Then we can evidently write

$$\mathsf{B}^{K'}\tilde{\mathsf{F}}_s\mu = \mathsf{C}_s^{K'}\mathsf{P}^{K'} \bigotimes_{K'' \in N(K')} \mu^{K''},$$

where we have defined for any probability η on $\mathbb{X}^{K'}$

$$(\mathsf{C}_s^{K'}\eta)(A) := \frac{\int \mathbf{1}_A(x^{K'}) \prod_{v \in K'} g^v(x^v, Y_s^v)\eta(dx^{K'})}{\int \prod_{v \in K'} g^v(x^v, Y_s^v)\eta(dx^{K'})},$$

and for any probability η on $\mathbb{X}^{\bigcup_{K'' \in N(K')} K''}$

$$(\mathsf{P}^{K'}\eta)(A) := \int \mathbf{1}_A(x^{K'}) \prod_{v \in K'} p^v(z, x^v)\psi^v(dx^v)\eta(dz).$$

We therefore have

$$\begin{aligned}
&\mathsf{B}^K\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1\mu \\
&\quad = \mathsf{C}_n^K\mathsf{P}^K \bigotimes_{K_{n-1} \in N(K)} \Bigg[\mathsf{C}_{n-1}^{K_{n-1}}\mathsf{P}^{K_{n-1}} \\
&\qquad\qquad \bigotimes_{K_{n-2} \in N(K_{n-1})} \Bigg[\mathsf{C}_{n-2}^{K_{n-2}}\mathsf{P}^{K_{n-2}} \cdots \\
&\qquad\qquad\qquad \bigotimes_{K_1 \in N(K_2)} \Bigg[\mathsf{C}_1^{K_1}\mathsf{P}^{K_1} \bigotimes_{K_0 \in N(K_1)} \mu^{K_0}\Bigg] \cdots\Bigg]\Bigg].
\end{aligned}$$

The structure of the computation tree is now readily visible in this expression. To formalize the construction, we introduce the tree index set

$$T := \big\{[K_u \cdots K_{n-1}] : 0 \le u < n,\, K_s \in N(K_{s+1}) \text{ for } u \le s < n\big\} \cup \big\{[\varnothing]\big\},$$

where we write $K_n := K$ for simplicity (recall that K and n are fixed throughout). The root of the tree $[\varnothing]$ represents the block K at time n, while $[K_u \cdots K_{n-1}]$ represents the duplicate of block K_u at time u that affects block K at time n along the branch $K_u \to K_{u+1} \to \cdots \to K_{n-1} \to K$ (cf. Figure 5 for a simple illustration). The vertex set corresponding to the computation tree is defined as

$$I = \big\{[K_u \cdots K_{n-1}]v : [K_u \cdots K_{n-1}] \in T,\, v \in K_u\big\} \cup \big\{[\varnothing]v : v \in K\big\},$$

and the corresponding state space is given by

$$\mathbb{S} = \prod_{i \in I} \mathbb{X}^i, \qquad \mathbb{X}^{[t]v} = \mathbb{X}^v \qquad \text{for } [t]v \in I.$$

It will be convenient in the sequel to introduce some additional notation. First, we will specify the children $c(i)$ of an index $i \in I$ as follows:

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

and similarly for $c([\varnothing]v)$. Denote the depth $d(i)$ and location $v(i)$ of $i \in I$ as

$$d\big([K_u \cdots K_{n-1}]v\big) := u, \qquad \big([\varnothing]v\big) := n, \qquad v\big([t]v\big) := v.$$

We define the index set of nonleaf vertices in I as

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

and the set of leaves of the tree T as

$$T_0 := \big\{[K_0 \cdots K_{n-1}] : K_s \in N(K_{s+1}) \text{ for } 0 \le s < n\big\}.$$

Finally, it will be natural to identify $[t] \in T$ with the corresponding subset of I:

$$[K_u \cdots K_{n-1}] = \big\{[K_u \cdots K_{n-1}]v : v \in K_u\big\},$$

together with the analogous identification for $[\varnothing]$.

We now define the probability measures $\rho, \tilde{\rho}$ on $\mathbb{S}$ as follows:

$$\rho(A) = \frac{\int \mathbf{1}_A(x) \prod_{i \in I_+} p^{v(i)}(x^{c(i)}, x^i) g^{v(i)}(x^i, Y^{v(i)}_{d(i)}) \psi^{v(i)}(dx^i) \prod_{[t] \in T_0} \mu^{[t]}(dx^{[t]})}{\int \prod_{i \in I_+} p^{v(i)}(x^{c(i)}, x^i) g^{v(i)}(x^i, Y^{v(i)}_{d(i)}) \psi^{v(i)}(dx^i) \prod_{[t] \in T_0} \mu^{[t]}(dx^{[t]})},$$

$$\tilde{\rho}(A) = \frac{\int \mathbf{1}_A(x) \prod_{i \in I_+} p^{v(i)}(x^{c(i)}, x^i) g^{v(i)}(x^i, Y^{v(i)}_{d(i)}) \psi^{v(i)}(dx^i) \prod_{[t] \in T_0} \nu^{[t]}(dx^{[t]})}{\int \prod_{i \in I_+} p^{v(i)}(x^{c(i)}, x^i) g^{v(i)}(x^i, Y^{v(i)}_{d(i)}) \psi^{v(i)}(dx^i) \prod_{[t] \in T_0} \nu^{[t]}(dx^{[t]})},$$

where we write $\mu^{[K_0\cdots K_{n-1}]} := \mu^{K_0}$ and $\nu^{[K_0\cdots K_{n-1}]} := \nu^{K_0}$ for simplicity. Then, by construction, the measure $\mathsf{B}^K\tilde{\mathsf{F}}_n\cdots\tilde{\mathsf{F}}_1\mu$ coincides with the marginal of ρ on the root of the computation tree, while $\mathsf{B}^K\tilde{\mathsf{F}}_n\cdots\tilde{\mathsf{F}}_1\nu$ coincides with the marginal of $\tilde{\rho}$ on the root of the computation tree. In particular, we obtain

$$\|\tilde{\mathsf{F}}_n\cdots\tilde{\mathsf{F}}_1\mu - \tilde{\mathsf{F}}_n\cdots\tilde{\mathsf{F}}_1\nu\|_J = \|\rho-\tilde{\rho}\|_{[\varnothing]J}.$$

We will use Theorem 3.1 to obtain a bound on this expression.

Throughout the remainder of the proof, we specialize to the case that $\mu=\delta_z$ and $\nu=\delta_{z'}$. To apply Theorem 3.1, we must bound the quantities C_{ij} and b_i with $i=[K_u\cdots K_{n-1}]v$ and $j=[K'_{u'}\cdots K'_{n-1}]v'$. We distinguish three cases.

Case $u=0$. As $\mu=\delta_z$ is nonrandom we evidently have $\rho^i_x=\delta_{z^v}$, so that $C_{ij}=0$. On the other hand, as $\tilde{\rho}^i_x=\delta_{z'^v}$, we cannot do better than $b_i\le 2$.

Case $0<u<n$. Now we have

$$\rho^i_x(A)=\tilde{\rho}^i_x(A)$$
$$=\frac{\int \mathbf{1}_A(x^i)g^v(x^i,Y^v_u)p^v(x^{c(i)},x^i)\prod_{\ell\in I_+:\,i\in c(\ell)}p^{v(\ell)}(x^{c(\ell)},x^\ell)\psi^v(dx^i)}{\int g^v(x^i,Y^v_u)p^v(x^{c(i)},x^i)\prod_{\ell\in I_+:\,i\in c(\ell)}p^{v(\ell)}(x^{c(\ell)},x^\ell)\psi^v(dx^i)}.$$

Thus, $b_i=0$. Moreover, by inspection, ρ^i_x does not depend on x^j except in the following cases: $j\in c(i)$; $i\in c(j)$; $j\in c(\ell)$ for some $\ell\in I_+$ such that $i\in c(\ell)$. As $\operatorname{card}c(\ell)\le\Delta$ for every $\ell\in I_+$, we estimate using Lemma 4.1

$$C_{ij}\le\begin{cases}1-\varepsilon^2, & \text{if } j\in c(i),\\ 1-\varepsilon^2, & \text{if } i\in c(j),\\ 1-\varepsilon^{2\Delta}, & \text{if } j\in\bigcup_{\ell\in I_+:\,i\in c(\ell)}c(\ell),\\ 0, & \text{otherwise.}\end{cases}$$

This yields

$$\sum_{j\in I}e^{\beta|d(i)-d(j)|}C_{ij}\le 2(1-\varepsilon^2)e^\beta\Delta+(1-\varepsilon^{2\Delta})\Delta^2\le 3(1-\varepsilon^{2\Delta})e^\beta\Delta^2,$$

where we have used that $\beta>0$ and $\Delta\ge1$ in the last inequality.

Case $u=n$. Now $i=[\varnothing]v$, so we have

$$\rho^i_x(A)=\tilde{\rho}^i_x(A)=\frac{\int\mathbf{1}_A(x^i)g^v(x^i,Y^v_n)p^v(x^{c(i)},x^i)\psi^v(dx^i)}{\int g^v(x^i,Y^v_n)p^v(x^{c(i)},x^i)\psi^v(dx^i)}.$$

Arguing precisely as above, we obtain $b_i=0$ and

$$\sum_{j\in I}e^{\beta|d(i)-d(j)|}C_{ij}\le(1-\varepsilon^2)e^\beta\Delta.$$

Combining the above three cases, we obtain

$$\max_{i\in I}\sum_{j\in I}e^{\beta|d(i)-d(j)|}C_{ij}\le 3(1-\varepsilon^{2\Delta})e^\beta\Delta^2=\frac{1}{2}$$

by the assumption of the proposition. Thus, by Theorem 3.1

$$\|\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \delta_z - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \delta_{z'}\|_J = \|\rho - \tilde{\rho}\|_{[\varnothing]J} \le 4 \operatorname{card} J e^{-\beta n},$$

where we have used Lemma 4.3 with $m(i, j) = \beta|d(i) - d(j)|$. The proof is completed by extending to general $s < n$ as in the proof of Proposition 4.4. □

The proof of Proposition 4.15 was simplified by the fact that the resulting bound holds uniformly for all point mass initial conditions (this could be used to obtain a uniform bound for all initial measures along the same lines as the proof of Corollary 4.7). To obtain a bound on the variance term, however, we require a more precise stability bound for the block filter that provides explicit control in terms of the initial conditions. We will shortly deduce such a bound from Proposition 4.15. Before we can do so, however, we must prove a refinement of Lemma 4.2.

LEMMA 4.16. *Let $\mu = \mu^1 \otimes \cdots \otimes \mu^d$ and $\nu = \nu^1 \otimes \cdots \otimes \nu^d$ 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. Define the probability measures*

$$\mu_\Lambda(A) := \frac{\int \mathbf{1}_A(x)\Lambda(x)\mu(dx)}{\int \Lambda(x)\mu(dx)}, \qquad \nu_\Lambda(A) := \frac{\int \mathbf{1}_A(x)\Lambda(x)\nu(dx)}{\int \Lambda(x)\nu(dx)}.$$

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

$$\varepsilon \Lambda^i(x) \le \Lambda(x) \le \varepsilon^{-1} \Lambda^i(x) \qquad \text{for all } x \in \mathbb{S}$$

and such that $\Lambda^i(x) = \Lambda^i(\tilde{x})$ whenever $x^{\{1,\ldots,d\}\setminus\{i\}} = \tilde{x}^{\{1,\ldots,d\}\setminus\{i\}}$. Then

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

PROOF. Define for $i = 0, \ldots, d$ the measures

$$\rho_i := \nu^1 \otimes \cdots \otimes \nu^i \otimes \mu^{i+1} \otimes \cdots \otimes \mu^d, \qquad \rho_{i,\Lambda}(A) := \frac{\int \mathbf{1}_A(x)\Lambda(x)\rho_i(dx)}{\int \Lambda(x)\rho_i(x)}$$

(by convention, $\rho_0 = \mu$ and $\rho_d = \nu$). Then we can estimate

$$\|\mu_\Lambda - \nu_\Lambda\| \le \sum_{i=1}^d \|\rho_{i,\Lambda} - \rho_{i-1,\Lambda}\|.$$

Now note that we can estimate for $|f| \le 1$

$$|\rho_{i,\Lambda}(f) - \rho_{i-1,\Lambda}(f)| \le \frac{1}{\varepsilon \rho_i(\Lambda^i)} \big[|\rho_i(f\Lambda) - \rho_{i-1}(f\Lambda)| + |\rho_i(\Lambda) - \rho_{i-1}(\Lambda)|\big]$$

as in the proof of Lemma 4.2. Moreover, we can write

$$|\rho_i(f\Lambda) - \rho_{i-1}(f\Lambda)| = \frac{\rho_i(\Lambda^i)}{\varepsilon}\left|\int f^i(x)\nu^i(dx^i) - \int f^i(x)\mu^i(dx^i)\right|,$$

$$|\rho_i(\Lambda) - \rho_{i-1}(\Lambda)| = \frac{\rho_i(\Lambda^i)}{\varepsilon}\left|\int g^i(x)\nu^i(dx^i) - \int g^i(x)\mu^i(dx^i)\right|,$$

where f^i and g^i are functions on $\mathbb{S}^i$ defined by

$$f^i(x^i) := \frac{\varepsilon}{\rho_i(\Lambda^i)}\int f(x)\Lambda(x)\nu^1(dx^1)\cdots\nu^{i-1}(dx^{i-1})\mu^{i+1}(dx^{i+1})\cdots\mu^d(dx^d),$$

$$g^i(x^i) := \frac{\varepsilon}{\rho_i(\Lambda^i)}\int \Lambda(x)\nu^1(dx^1)\cdots\nu^{i-1}(dx^{i-1})\mu^{i+1}(dx^{i+1})\cdots\mu^d(dx^d).$$

Evidently $|f^i| \le 1$ and $|g^i| \le 1$, and the proof follows directly. □

We can now obtain a stability bound with control on the initial conditions.

PROPOSITION 4.17. *Suppose there exists $\varepsilon > 0$ with*

$$\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1} \qquad \text{for all } v \in V, x, z \in \mathbb{X}$$

such that

$$\varepsilon > \varepsilon_0 = \left(1 - \frac{1}{6\Delta^2}\right)^{1/2\Delta}.$$

Let $\beta = -\log 6\Delta^2(1-\varepsilon^{2\Delta}) > 0$. Then for any product probability measures

$$\mu = \bigotimes_{K\in\mathcal{K}} \mu^K, \qquad \nu = \bigotimes_{K\in\mathcal{K}} \nu^K,$$

we have

$$\|\tilde{\mathsf{F}}_n\cdots\tilde{\mathsf{F}}_{s+1}\mu - \tilde{\mathsf{F}}_n\cdots\tilde{\mathsf{F}}_{s+1}\nu\|_J \le \frac{4}{\varepsilon^{2|\mathcal{K}|_\infty}}\operatorname{card} J e^{-\beta(n-s)}\sum_{K\in\mathcal{K}}\alpha_K\|\mu^K - \nu^K\|$$

for every $s < n$, $K \in \mathcal{K}$, and $J \subseteq K$. Here, $(\alpha_K)_{K\in\mathcal{K}}$ are nonnegative integers, depending on J and $n-s$ only, such that $\sum_{K\in\mathcal{K}}\alpha_K \le \Delta_{\mathcal{K}}^{n-s}$.

PROOF. We fix $s = 0$, $n > 0$, $K \in \mathcal{K}$, $J \subseteq K$ as in the proof of Proposition 4.15, and adopt the notation used there. Define the functions

$$h_A(x^{T_0}) := \int \mathbf{1}_A(x^{[\varnothing]J})\prod_{i\in I_+} p^{v(i)}(x^{c(i)}, x^i)g^{v(i)}(x^i, Y^{v(i)}_{d(i)})\psi^{v(i)}(dx^i),$$

$$h(x^{T_0}) := \int \prod_{i\in I_+} p^{v(i)}(x^{c(i)}, x^i)g^{v(i)}(x^i, Y^{v(i)}_{d(i)})\psi^{v(i)}(dx^i)$$

on the leaves T_0 of the computation tree, for every measurable $A \subseteq \mathbb{X}^J$. Then

$$(\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \mu)(A) = \frac{\int h_A(x^{T_0}) \prod_{[t]\in T_0} \mu^{[t]}(dx^{[t]})}{\int h(x^{T_0}) \prod_{[t]\in T_0} \mu^{[t]}(dx^{[t]})} = \int \frac{h_A(x^{T_0})}{h(x^{T_0})} \tilde{\mu}(dx^{T_0}),$$

where we define the measure

$$\tilde{\mu}(A) := \frac{\int \mathbf{1}_A(x^{T_0}) h(x^{T_0}) \prod_{[t]\in T_0} \mu^{[t]}(dx^{[t]})}{\int h(x^{T_0}) \prod_{[t]\in T_0} \mu^{[t]}(dx^{[t]})}.$$

The measure $\tilde{\nu}$ is defined analogously, and we have

$$\|\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \mu - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \nu\|_J = 2 \sup_{A\subseteq\mathbb{X}^J} \left| \int \frac{h_A}{h} d\tilde{\mu} - \int \frac{h_A}{h} d\tilde{\nu} \right|,$$

where the supremum is taken only over measurable sets. But note that h_A/h is precisely the filter obtained when the initial condition is a point mass on the leaves of the computation tree (albeit not with the special duplication pattern induced by the unravelling of the original model; however, this was not used in the proof of Proposition 4.15). Therefore, the proof of Proposition 4.15 yields

$$2 \sup_{z,\tilde{z}\in\mathbb{X}^{T_0}} \sup_{A\subseteq\mathbb{X}^J} \left| \frac{h_A(z)}{h(z)} - \frac{h_A(\tilde{z})}{h(\tilde{z})} \right| \le 4 \operatorname{card} J e^{-\beta n}.$$

In particular, using the identity $|\mu(f) - \nu(f)| \le \frac{1}{2} \operatorname{osc} f \, \|\mu - \nu\|$, we obtain

$$\|\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \mu - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \nu\|_J \le 2 \operatorname{card} J e^{-\beta n} \|\tilde{\mu} - \tilde{\nu}\|.$$

We now aim to apply Lemma 4.16 to estimate $\|\tilde{\mu} - \tilde{\nu}\|$.

To this end, consider a block $[t] \in T_0$. The integrand in the definition of $h(x^{T_0})$ depends only on $x^{[t]}$ through the terms $p^{v(i)}(x^{c(i)}, x^i)$ with $c(i) \cap [t] \neq \varnothing$. If we write $[t] = [K_0 \cdots K_{n-1}]$, then $c(i) \cap [t] \neq \varnothing$ requires at least $i \in [K_1 \cdots K_{n-1}]$ and therefore $\operatorname{card}\{i \in I_+ : c(i) \cap [t] \neq \varnothing\} \le \operatorname{card} K_1 \le |\mathcal{K}|_\infty$. Thus, we have

$$\varepsilon^{|\mathcal{K}|_\infty} h^{[t]}(z) \le h(z) \le \varepsilon^{-|\mathcal{K}|_\infty} h^{[t]}(z)$$

for every $z \in \mathbb{X}^{T_0}$ and $[t] \in T_0$, where

$$h^{[t]}(x^{T_0}) := \int \prod_{i\in I_+ : c(i)\cap[t]=\varnothing} p^{v(i)}(x^{c(i)}, x^i) \prod_{i\in I_+} g^{v(i)}(x^i, Y^{v(i)}_{d(i)}) \psi^{v(i)}(dx^i)$$

does not depend on $x^{[t]}$. By Lemma 4.16, we obtain

$$\|\tilde{\mu} - \tilde{\nu}\| \le \frac{2}{\varepsilon^{2|\mathcal{K}|_\infty}} \sum_{[t]\in T_0} \|\mu^{[t]} - \nu^{[t]}\| = \frac{2}{\varepsilon^{2|\mathcal{K}|_\infty}} \sum_{K'\in\mathcal{K}} \alpha_{K'} \|\mu^{K'} - \nu^{K'}\|,$$

where we define $\alpha_{K'} = \operatorname{card}\{[K_0 \cdots K_{n-1}] \in T_0 : K_0 = K'\}$. As the computation tree has a branching factor of at most $\Delta_{\mathcal{K}}$, we evidently have $\sum_{K\in\mathcal{K}} \alpha_K =$

card $T_0 \le \Delta_{\mathcal{K}}^n$. The result therefore follows directly for the case $s = 0$, and the general case $s < n$ is immediate as in the proof of Proposition 4.4. $\square$

We finally state the block filter stability bound in its most useful form.

COROLLARY 4.18 (Block filter stability). *Suppose there exists $\varepsilon > 0$ with*

$$\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1} \qquad \text{for all } v \in V, x, z \in \mathbb{X}$$

such that

$$\varepsilon > \varepsilon_0 = \left(1 - \frac{1}{6\Delta_{\mathcal{K}}\Delta^2}\right)^{1/2\Delta}.$$

Let $\beta = -\log 6\Delta_{\mathcal{K}}\Delta^2(1 - \varepsilon^{2\Delta}) > 0$.

Then for any (possibly random) product probability measures

$$\mu = \bigotimes_{K\in\mathcal{K}} \mu^K, \qquad \nu = \bigotimes_{K\in\mathcal{K}} \nu^K,$$

we have

$$\begin{aligned}&\mathbf{E}\big[\|\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1}\mu - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1}\nu\|_J^2\big]^{1/2}\\ &\quad\le \frac{4}{\varepsilon^{2|\mathcal{K}|_\infty}} \operatorname{card} J\, e^{-\beta(n-s)} \max_{K\in\mathcal{K}} \mathbf{E}\big[\|\mu^K - \nu^K\|^2\big]^{1/2}\end{aligned}$$

for every $s < n$, $K \in \mathcal{K}$, and $J \subseteq K$.

PROOF. The result follows readily from Proposition 4.17 (note that we have now absorbed the branching factor $\Delta_{\mathcal{K}}^{n-s}$ in the definition of β). $\square$

4.7. *Bounding the variance.* To complete the proof of Theorem 2.1, it now remains to bound the variance term $|||\tilde{\pi}_n - \hat{\pi}_n|||_J$ uniformly in time. This is the goal of the present section. We will first obtain bounds on the one-step error, and then combine these with the block filter stability bound of Corollary 4.18 to obtain time-uniform control of the error. The main remaining difficulty is to properly account for the fact that Corollary 4.18 is phrased in terms of the total variation norm $\|\cdot\|_J$, which is too strong to control the sampling error (we do not know how to prove an analogous result to Corollary 4.18 in the weaker $|||\cdot|||_J$-norm). To this end, we retain one time step of the block filter dynamics in the one-step error (we control $\|\tilde{\mathsf{F}}_{s+1}\tilde{\mathsf{F}}_s\hat{\pi}_{s-1}^x - \tilde{\mathsf{F}}_{s+1}\hat{\mathsf{F}}_s\hat{\pi}_{s-1}^x\|_K$ rather than $|||\tilde{\mathsf{F}}_s\hat{\pi}_{s-1}^x - \hat{\mathsf{F}}_s\hat{\pi}_{s-1}^x|||_K$), which allows us to exploit the fact that the dynamics P has a density.

Let us begin with the most trivial result: a one-step bound in the $|||\cdot|||_J$-norm. This estimate will be used to bound the error in the last time step $s = n$.

LEMMA 4.19 (Sampling error, $s = n$). *Suppose there exists $\kappa > 0$ such that*

$$\kappa \le g^v(x^v, y^v) \le \kappa^{-1} \qquad \text{for all } v \in V, x \in \mathbb{X}, y \in \mathbb{Y}.$$

Then

$$\max_{K \in \mathcal{K}} \|\!| \tilde{\mathsf{F}}_n \hat{\pi}^\mu_{n-1} - \hat{\mathsf{F}}_n \hat{\pi}^\mu_{n-1} |\!\|_K \le \frac{2\kappa^{-2|\mathcal{K}|_\infty}}{\sqrt{N}}.$$

PROOF. Note that

$$\begin{aligned}\|\!| \tilde{\mathsf{F}}_n \hat{\pi}^\mu_{n-1} - \hat{\mathsf{F}}_n \hat{\pi}^\mu_{n-1} |\!\|_K &= \|\!| \mathsf{C}_n^K \mathsf{B}^K \mathsf{P} \hat{\pi}^\mu_{n-1} - \mathsf{C}_n^K \mathsf{B}^K \mathsf{S}^N \mathsf{P} \hat{\pi}^\mu_{n-1} |\!\| \\ &\le 2\kappa^{-2\operatorname{card} K} \|\!| \mathsf{P} \hat{\pi}^\mu_{n-1} - \mathsf{S}^N \mathsf{P} \hat{\pi}^\mu_{n-1} |\!\| \le \frac{2\kappa^{-2\operatorname{card} K}}{\sqrt{N}},\end{aligned}$$

where the first inequality is Lemma 4.2 and the second inequality follows from the simple estimate $\|\!| \mu - \mathsf{S}^N \mu |\!\| \le 1/\sqrt{N}$ that holds for any probability μ. □

For the error in steps $s < n$, the requisite one-step bound (Proposition 4.22) is more involved. Before we prove it, we must first introduce an elementary lemma about products of empirical measures that will be needed below.

LEMMA 4.20. *For any probability measure μ, we have*

$$\|\!| \mu^{\otimes d} - \hat{\mu}^{\otimes d} |\!\| \le \frac{4d}{\sqrt{N}},$$

where $\hat{\mu} = \frac{1}{N} \sum_{k=1}^N \delta_{X_k}$ and $X_1, \ldots, X_N$ are i.i.d. $\sim \mu$.

PROOF. We assume throughout that $N \ge d^2$ without loss of generality (otherwise the bound is trivial). Let $|f| \le 1$ be a measurable function. Then

$$\hat{\mu}^{\otimes d}(f) = \frac{1}{N^d} \sum_{k_1, \ldots, k_d = 1}^N f(X_{k_1}, \ldots, X_{k_d}).$$

We begin by bounding

$$\operatorname{Var}[\hat{\mu}^{\otimes d}(f)] = \frac{1}{N^{2d}} \sum_{k_1, \ldots, k_d = 1}^N \sum_{k'_1, \ldots, k'_d = 1}^N \mathbf{E}[F_{k_1, \ldots, k_d} F_{k'_1, \ldots, k'_d}],$$

where

$$F_{k_1, \ldots, k_d} := f(X_{k_1}, \ldots, X_{k_d}) - \mathbf{E}[f(X_{k_1}, \ldots, X_{k_d})].$$

Note that $\mathbf{E}[F_{k_1, \ldots, k_d} F_{k'_1, \ldots, k'_d}] = 0$ when $\{k_1, \ldots, k_d\} \cap \{k'_1, \ldots, k'_d\} = \varnothing$. Thus

$$\operatorname{Var}[\hat{\mu}^{\otimes d}(f)] \le \frac{4}{N^{2d}} \sum_{k_1, \ldots, k_d = 1}^N \sum_{k'_1, \ldots, k'_d = 1}^N \mathbf{1}_{\{k_1, \ldots, k_d\} \cap \{k'_1, \ldots, k'_d\} \ne \varnothing},$$

where we use $|F_{k_1,\ldots,k_d}| \le 2$. But for each choice of $k_1,\ldots,k_d$, there are at least $(N-d)^d$ choices of $k'_1,\ldots,k'_d$ such that $\{k_1,\ldots,k_d\} \cap \{k'_1,\ldots,k'_d\} = \varnothing$, so

$$\mathrm{Var}[\hat\mu^{\otimes d}(f)] \le 4\left(1 - \frac{N^d(N-d)^d}{N^{2d}}\right) = 4\left(1 - \left(1-\frac{d}{N}\right)^d\right) \le \frac{4d^2}{N}.$$

We can therefore estimate

$$\begin{aligned}\|\!|\mu^{\otimes d} - \hat\mu^{\otimes d}|\!\| &\le \|\mu^{\otimes d} - \mathbf{E}[\hat\mu^{\otimes d}]\| + \|\!|\mathbf{E}[\hat\mu^{\otimes d}] - \hat\mu^{\otimes d}|\!\| \\ &\le \|\mu^{\otimes d} - \mathbf{E}[\hat\mu^{\otimes d}]\| + \frac{2d}{\sqrt{N}}.\end{aligned}$$

It remains to estimate the first term. To this end, note that $\mathbf{E}[f(X_{k_1},\ldots,X_{k_d})] = \mu^{\otimes d}(f)$ whenever $k_1 \ne \cdots \ne k_d$. Therefore, we evidently have

$$\begin{aligned}|\mathbf{E}[\hat\mu^{\otimes d}(f)] - \mu^{\otimes d}(f)| &\le \frac{1}{N^d}\sum_{k_1,\ldots,k_d=1}^{N} |\mathbf{E}[f(X_{k_1},\ldots,X_{k_d})] - \mu^{\otimes d}(f)| \\ &\le 2\left(1 - \frac{1}{N^d}\frac{N!}{(N-d)!}\right) \le 2\left(1-\left(1-\frac{d}{N}\right)^d\right) \le \frac{2d^2}{N}.\end{aligned}$$

But as $N \ge d^2$, we have $d^2/N \le d/\sqrt{N}$. The result follows. □

This result will be used in the following form.

COROLLARY 4.21. *For any subset of blocks $\mathcal{L} \subseteq \mathcal{K}$, we have*

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

for every probability measure μ on $\mathbb{X}$ and $s \ge 1$.

PROOF. Write $\hat\mu := \mathsf{S}^N\mu$ and $d = \mathrm{card}\,\mathcal{L}$, and let us enumerate the blocks $\mathcal{L} = \{K_1,\ldots,K_d\}$. Then for any bounded function $f : \mathbb{X}^{\cup\mathcal{L}} \to \mathbb{R}$, we can write

$$\left(\bigotimes_{K\in\mathcal{L}} \mathsf{B}^K\mu\right)(f) = \int f(x_1^{K_1},\ldots,x_d^{K_d})\mu(dx_1)\cdots\mu(dx_d),$$

$$\left(\bigotimes_{K\in\mathcal{L}} \mathsf{B}^K\mathsf{S}^N\mu\right)(f) = \int f(x_1^{K_1},\ldots,x_d^{K_d})\hat\mu(dx_1)\cdots\hat\mu(dx_d).$$

Thus, evidently

$$\left\|\!\!\left|\bigotimes_{K\in\mathcal{L}} \mathsf{B}^K\mu - \bigotimes_{K\in\mathcal{L}} \mathsf{B}^K\mathsf{S}^N\mu\right|\!\!\right\| \le \|\!|\mu^{\otimes d} - \hat\mu^{\otimes d}|\!\|,$$

and the result follows from Lemma 4.20. □

We now proceed to prove a one-step error bound for time steps $s < n$.

PROPOSITION 4.22 (Sampling error, $s < n$). *Suppose there exist* $\varepsilon, \kappa > 0$ *with*

$$\varepsilon \le p^v(x, z^v) \le \varepsilon^{-1}, \qquad \kappa \le g^v(x^v, y^v) \le \kappa^{-1} \qquad \forall v \in V, x, z \in \mathbb{X}, y \in \mathbb{Y}.$$

Then

$$\max_{K \in \mathcal{K}} \mathbf{E}\big[\|\tilde{\mathsf{F}}_{s+1}\tilde{\mathsf{F}}_s\hat{\pi}^\mu_{s-1} - \tilde{\mathsf{F}}_{s+1}\hat{\mathsf{F}}_s\hat{\pi}^\mu_{s-1}\|^2_K\big]^{1/2} \le \frac{16\Delta_{\mathcal{K}}\varepsilon^{-2|\mathcal{K}|_\infty}\kappa^{-4|\mathcal{K}|_\infty\Delta_{\mathcal{K}}}}{\sqrt{N}}$$

for every $0 < s < n$.

PROOF. We begin by bounding using Lemma 4.2

$$\begin{aligned}\|\tilde{\mathsf{F}}_{s+1}\tilde{\mathsf{F}}_s\hat{\pi}^\mu_{s-1} - \tilde{\mathsf{F}}_{s+1}\hat{\mathsf{F}}_s\hat{\pi}^\mu_{s-1}\|_K &= \|\mathsf{C}^K_{s+1}\mathsf{B}^K\mathsf{P}\tilde{\mathsf{F}}_s\hat{\pi}^\mu_{s-1} - \mathsf{C}^K_{s+1}\mathsf{B}^K\mathsf{P}\hat{\mathsf{F}}_s\hat{\pi}^\mu_{s-1}\| \\ &\le 2\kappa^{-2|\mathcal{K}|_\infty}\|\mathsf{B}^K\mathsf{P}\tilde{\mathsf{F}}_s\hat{\pi}^\mu_{s-1} - \mathsf{B}^K\mathsf{P}\hat{\mathsf{F}}_s\hat{\pi}^\mu_{s-1}\|.\end{aligned}$$

Now note that

$$\begin{aligned}&\frac{(\mathsf{B}^K\mathsf{P}\tilde{\mathsf{F}}_s\hat{\pi}^\mu_{s-1})(dx^K)}{\psi^K(dx^K)} \\ &\quad= \frac{\int\prod_{v\in K}p^v(z,x^v)\prod_{K'\in N(K)}\prod_{v'\in K'}g^{v'}(z^{v'},Y^{v'}_s)(\mathsf{B}^{K'}\mathsf{P}\hat{\pi}^\mu_{s-1})(dz^{K'})}{\int\prod_{K'\in N(K)}\prod_{v'\in K'}g^{v'}(z^{v'},Y^{v'}_s)(\mathsf{B}^{K'}\mathsf{P}\hat{\pi}^\mu_{s-1})(dz^{K'})}, \\ &\frac{(\mathsf{B}^K\mathsf{P}\hat{\mathsf{F}}_s\hat{\pi}^\mu_{s-1})(dx^K)}{\psi^K(dx^K)} \\ &\quad= \frac{\int\prod_{v\in K}p^v(z,x^v)\prod_{K'\in N(K)}\prod_{v'\in K'}g^{v'}(z^{v'},Y^{v'}_s)(\mathsf{B}^{K'}\mathsf{S}^N\mathsf{P}\hat{\pi}^\mu_{s-1})(dz^{K'})}{\int\prod_{K'\in N(K)}\prod_{v'\in K'}g^{v'}(z^{v'},Y^{v'}_s)(\mathsf{B}^{K'}\mathsf{S}^N\mathsf{P}\hat{\pi}^\mu_{s-1})(dz^{K'})},\end{aligned}$$

where $\psi^K(dx^K) := \prod_{v\in K}\psi^v(dx^v)$, and we can write

$$\begin{aligned}&\|\mathsf{B}^K\mathsf{P}\tilde{\mathsf{F}}_s\hat{\pi}^\mu_{s-1} - \mathsf{B}^K\mathsf{P}\hat{\mathsf{F}}_s\hat{\pi}^\mu_{s-1}\| \\ &\qquad= \int\left|\frac{(\mathsf{B}^K\mathsf{P}\tilde{\mathsf{F}}_s\hat{\pi}^\mu_{s-1})(dx^K)}{\psi^K(dx^K)} - \frac{(\mathsf{B}^K\mathsf{P}\hat{\mathsf{F}}_s\hat{\pi}^\mu_{s-1})(dx^K)}{\psi^K(dx^K)}\right|\psi^K(dx^K).\end{aligned}$$

We therefore have by Minkowski's integral inequality

$$\begin{aligned}&\mathbf{E}[\|\mathsf{B}^K\mathsf{P}\tilde{\mathsf{F}}_s\hat{\pi}^\mu_{s-1} - \mathsf{B}^K\mathsf{P}\hat{\mathsf{F}}_s\hat{\pi}^\mu_{s-1}\|^2]^{1/2} \\ &\quad\le \int\mathbf{E}\left[\left|\frac{(\mathsf{B}^K\mathsf{P}\tilde{\mathsf{F}}_s\hat{\pi}^\mu_{s-1})(dx^K)}{\psi^K(dx^K)} - \frac{(\mathsf{B}^K\mathsf{P}\hat{\mathsf{F}}_s\hat{\pi}^\mu_{s-1})(dx^K)}{\psi^K(dx^K)}\right|^2\right]^{1/2}\psi^K(dx^K) \\ &\quad\le \psi^K(\mathbb{X}^K)\sup_{x^K\in\mathbb{X}^K}\mathbf{E}\left[\left|\frac{(\mathsf{B}^K\mathsf{P}\tilde{\mathsf{F}}_s\hat{\pi}^\mu_{s-1})(dx^K)}{\psi^K(dx^K)} - \frac{(\mathsf{B}^K\mathsf{P}\hat{\mathsf{F}}_s\hat{\pi}^\mu_{s-1})(dx^K)}{\psi^K(dx^K)}\right|^2\right]^{1/2}.\end{aligned}$$

As we have

$$\varepsilon\psi^v(\mathbb{X}^v) \le \int p^v(x,z^v)\psi^v(dz^v) = 1, \qquad \prod_{v\in K} p^v(z,x^v) \le \varepsilon^{-|\mathcal{K}|_\infty}$$

and

$$\kappa^{|\mathcal{K}|_\infty \Delta_{\mathcal{K}}} \le \prod_{K'\in N(K)} \prod_{v'\in K'} g^{v'}(z^{v'}, Y_s^{v'}) \le \kappa^{-|\mathcal{K}|_\infty \Delta_{\mathcal{K}}},$$

we can apply Lemma 4.2 to estimate

$$\begin{aligned}&\mathbf{E}\big[\big\|\mathsf{B}^K \mathsf{P}\tilde{\mathsf{F}}_s \hat\pi^\mu_{s-1} - \mathsf{B}^K \mathsf{P}\hat{\mathsf{F}}_s \hat\pi^\mu_{s-1}\big\|^2\big]^{1/2} \\ &\quad\le 2\varepsilon^{-2|\mathcal{K}|_\infty}\kappa^{-2|\mathcal{K}|_\infty\Delta_{\mathcal{K}}} \Big\| \bigotimes_{K'\in N(K)} \mathsf{B}^{K'}\mathsf{P}\hat\pi^\mu_{s-1} - \bigotimes_{K'\in N(K)} \mathsf{B}^{K'}\mathsf{S}^N\mathsf{P}\hat\pi^\mu_{s-1}\Big\|.\end{aligned}$$

By Corollary 4.21 (applied conditionally given $\hat\pi^\mu_{s-1}$), we obtain

$$\mathbf{E}\big[\big\|\mathsf{B}^K \mathsf{P}\tilde{\mathsf{F}}_s \hat\pi^\mu_{s-1} - \mathsf{B}^K \mathsf{P}\hat{\mathsf{F}}_s \hat\pi^\mu_{s-1}\big\|^2\big]^{1/2} \le \frac{8\Delta_{\mathcal{K}}\varepsilon^{-2|\mathcal{K}|_\infty}\kappa^{-2|\mathcal{K}|_\infty\Delta_{\mathcal{K}}}}{\sqrt{N}}.$$

The result follows immediately. □

We finally put everything together.

THEOREM 4.23 (Variance term). *Suppose there exist $\varepsilon,\kappa>0$ with*

$$\varepsilon \le p^v(x,z^v) \le \varepsilon^{-1}, \qquad \kappa \le g^v(x^v,y^v) \le \kappa^{-1} \qquad \forall v\in V, x,z\in\mathbb{X}, y\in\mathbb{Y}$$

such that

$$\varepsilon > \varepsilon_0 = \left(1 - \frac{1}{6\Delta_{\mathcal{K}}\Delta^2}\right)^{1/2\Delta}.$$

Let $\beta = -\log 6\Delta_{\mathcal{K}}\Delta^2(1-\varepsilon^{2\Delta}) > 0$. Then

$$\big|\!\big|\!\big|\tilde\pi^x_n - \hat\pi^x_n\big|\!\big|\!\big|_J \le \operatorname{card} J \,\frac{64\Delta_{\mathcal{K}}e^\beta}{1-e^{-\beta}}\,\frac{\varepsilon^{-4|\mathcal{K}|_\infty}\kappa^{-4|\mathcal{K}|_\infty\Delta_{\mathcal{K}}}}{\sqrt{N}}$$

for every $n\ge 0$, $x\in\mathbb{X}$, $K\in\mathcal{K}$ and $J\subseteq K$.

PROOF. We begin with the elementary error decomposition

$$\big|\!\big|\!\big|\tilde\pi^x_n - \hat\pi^x_n\big|\!\big|\!\big|_J \le \sum_{s=1}^n \big|\!\big|\!\big|\tilde{\mathsf{F}}_n\cdots\tilde{\mathsf{F}}_{s+1}\tilde{\mathsf{F}}_s\hat\pi^x_{s-1} - \tilde{\mathsf{F}}_n\cdots\tilde{\mathsf{F}}_{s+1}\hat{\mathsf{F}}_s\hat\pi^x_{s-1}\big|\!\big|\!\big|_J.$$

The term $s=n$ in this sum is bounded in Lemma 4.19:

$$\big|\!\big|\!\big|\tilde{\mathsf{F}}_n\hat\pi^x_{n-1} - \hat{\mathsf{F}}_n\hat\pi^x_{n-1}\big|\!\big|\!\big|_J \le \frac{2\kappa^{-2|\mathcal{K}|_\infty}}{\sqrt{N}}.$$

The term $s = n - 1$ is bounded in Proposition 4.22:

$$\left|\!\left|\!\left| \tilde{\mathsf{F}}_n \tilde{\mathsf{F}}_{n-1} \hat{\pi}^x_{s-1} - \tilde{\mathsf{F}}_n \hat{\mathsf{F}}_{n-1} \hat{\pi}^x_{s-1} \right|\!\right|\!\right|_J \leq \frac{16 \Delta_{\mathcal{K}} \varepsilon^{-2|\mathcal{K}|_\infty} \kappa^{-4|\mathcal{K}|_\infty \Delta_{\mathcal{K}}}}{\sqrt{N}}.$$

Now suppose $s < n - 1$. Then we can estimate using Corollary 4.18

$$\begin{aligned} &\left|\!\left|\!\left| \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1} \tilde{\mathsf{F}}_s \hat{\pi}^x_{s-1} - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1} \hat{\mathsf{F}}_s \hat{\pi}^x_{s-1} \right|\!\right|\!\right|_J \\ &\quad \leq \frac{4}{\varepsilon^{2|\mathcal{K}|_\infty}} \operatorname{card} J e^{-\beta(n-s-1)} \max_{K \in \mathcal{K}} \mathbf{E}\big[\big\| \tilde{\mathsf{F}}_{s+1} \tilde{\mathsf{F}}_s \hat{\pi}^x_{s-1} - \tilde{\mathsf{F}}_{s+1} \hat{\mathsf{F}}_s \hat{\pi}^x_{s-1} \big\|^2_K \big]^{1/2}. \end{aligned}$$

Applying Proposition 4.22 yields

$$\begin{aligned} &\left|\!\left|\!\left| \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1} \tilde{\mathsf{F}}_s \hat{\pi}^x_{s-1} - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1} \hat{\mathsf{F}}_s \hat{\pi}^x_{s-1} \right|\!\right|\!\right|_J \\ &\quad \leq \operatorname{card} J e^{-\beta(n-s-1)} \frac{64 \Delta_{\mathcal{K}} \varepsilon^{-4|\mathcal{K}|_\infty} \kappa^{-4|\mathcal{K}|_\infty \Delta_{\mathcal{K}}}}{\sqrt{N}}. \end{aligned}$$

Substituting the above three cases into the error decomposition and summing the geometric series yields the statement of the theorem. □

Theorems 4.14 and 4.23 now immediately yield Theorem 2.1.

REFERENCES

[1] APTE, A., JONES, C. K. R. T., STUART, A. M. and VOSS, J. (2008). Data assimilation: Mathematical and statistical perspectives. *Internat. J. Numer. Methods Fluids* **56** 1033–1046. MR2393497

[2] BESKOS, A., CRISAN, D., JASRA, A. and WHITELEY, N. (2014). Error bounds and normalizing constants for sequential Monte Carlo in high dimensions. *Adv. Appl. Probab.* To appear.

[3] BICKEL, P., LI, B. and BENGTSSON, T. (2008). Sharp failure rates for the bootstrap particle filter in high dimensions. In *Pushing the Limits of Contemporary Statistics*: *Contributions in Honor of Jayanta K. Ghosh*. *Inst. Math. Stat. Collect.* **3** 318–329. IMS, Beachwood, OH. MR2459233

[4] BLANK, M. (1997). *Discreteness and Continuity in Problems of Chaotic Dynamics*. *Translations of Mathematical Monographs* **161**. Amer. Math. Soc., Providence, RI. MR1440853

[5] CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer, New York. MR2159833

[6] DEL MORAL, P. and GUIONNET, A. (2001). On the stability of interacting processes with applications to filtering and genetic algorithms. *Ann. Inst. Henri Poincaré Probab. Stat.* **37** 155–194. MR1819122

[7] DOBRUSHIN, R. L., KRYUKOV, V. I. and TOOM, A. L., eds. (1990). *Stochastic Cellular Systems*: *Ergodicity*, *Memory*, *Morphogenesis*. Manchester Univ. Press, Manchester.

[8] GEORGII, H.-O. (2011). *Gibbs Measures and Phase Transitions*, 2nd ed. *de Gruyter Studies in Mathematics* **9**. de Gruyter, Berlin. MR2807681

[9] LAW, K. J. H. and STUART, A. M. (2012). Evaluating data assimilation algorithms. *Mon. Weather Rev.* **140** 3757–3782.

[10] LEBOWITZ, J. L., MAES, C. and SPEER, E. R. (1990). Statistical mechanics of probabilistic cellular automata. *J. Stat. Phys.* **59** 117–170. MR1049965

[11] MAJDA, A. J. and HARLIM, J. (2012). *Filtering Complex Turbulent Systems*. Cambridge Univ. Press, Cambridge. MR2934167

[12] MARTINELLI, F. (2004). Relaxation times of Markov chains in statistical mechanics and combinatorial structures. In *Probability on Discrete Structures*. *Encyclopaedia Math. Sci.* **110** 175–262. Springer, Berlin. MR2023653

[13] REBESCHINI, P. and VAN HANDEL, R. (2014). Phase transitions in nonlinear filtering. Preprint. Available at arXiv:1401.6450.

[14] REBESCHINI, P. and VAN HANDEL, R. (2014). Comparison theorems for Gibbs measures. Preprint. Available at arXiv:1308.4117.

[15] SNYDER, C. (2011). Particle filters, the "optimal" proposal and high-dimensional systems. In *Proceedings of the ECMWF Seminar on Data Assimilation for Atmosphere and Ocean*.

[16] SNYDER, C., BENGTSSON, T., BICKEL, P. and ANDERSON, J. (2008). Obstacles to high-dimensional particle filtering. *Mon. Weather Rev.* **136** 4629–4640.

[17] TATIKONDA, S. C. and JORDAN, M. I. (2002). Loopy belief propagation and Gibbs measures. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence* 493–500. Morgan Kaufmann, San Francisco, CA.

[18] TONG, X. T. and VAN HANDEL, R. (2014). Conditional ergodicity in infinite dimension. *Ann. Probab.* **42** 2243–2313.

[19] VAN LEEUWEN, P. J. (2009). Particle filtering in geophysical systems. *Mon. Weather Rev.* **137** 4089–4114.

SHERRERD HALL
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
USA
E-MAIL: prebesch@princeton.edu
rvan@princeton.edu