arXiv: arXiv:0000.0000

Convergence Rates for Residual Branching Particle Filters*

Michael A. Kouritzin ,

Department of Mathematical and Statistical Sciences
University of Alberta, Edmonton, AB T6G 2G1 Canada
e-mail: michaelk@ualberta.ca

Abstract: A large class of proven discrete-time branching particle filters with Bayesian model selection capabilities and effective resampling is analyzed mathematically. The particles interact weakly in the branching procedure through the total mass process in such a way that the expected number of particles can remain constant. The weighted particle filter, which has no resampling, and the fully-resampled branching particle filter are included in the class as extreme points. Otherwise, selective residual branching is used allowing any number of offspring. Each particle filter in the class is coupled to a McKean-Vlasov particle system, corresponding to a reduced, unimplementable branching particle filter, for which Marcinkiewicz strong laws of large numbers (Mllns) and the central limit theorem (clt) can be written down. Coupling arguments are used to show the reduced system can be used to predict performance of and to transfer the Mllns to the real weakly-interacting residual branching particle filter. This clt is also shown transferable when (a few) extra particles are used.

MSC 2010 subject classifications: Primary 60F05, 60G35; secondary 62M20, 60G09, 62E20, 60J80.
Keywords and phrases: Branching Process, Central Limit Theorem, Coupling, McKean-Vlasov, Sequential Monte Carlo, Marcinkiewicz strong laws of large numbers.

1. Introduction

Sequential Monte Carlo (SMC) algorithms are used in diverse problems like tracking, prediction, paramter estimation, model calibration, classification, Bayesian model selection and imaging (see e.g. [18], [17], [13] and [9] for sample applications). Branching SMC algorithms have the advantage that offspring generation only depends upon the parent not the whole population and the disadvantage of having randomly-varying populations (i.e. particle numbers). Recently, Kouritzin [10] introduced four new classes of branching sequential Monte Carlo algorithms that were designed to limit wide particle variations. The tracking and model selection performance of all four algorithms was shown experimentally to be superior to a collection of popular resampled particle algorithms and these four branching algorithms have even greater advantages when it comes to distributed implementations (see Kouritzin and Wang [12]). However, there is little theory to back up these experimental findings. Theoretical rate-of-convergence results are desired to understand why these algorithms perform so well and what their weaknesses might be. Unfortunately, the branching algorithms lack the independence and fixed particle numbers of many resampled algorithms so their analysis is necessarily difficult and the desired convergence results hard to come by. Herein, we start the theoretical study by establishing Marcinkiewicz strong laws of large numbers (Mllns) and a central limit theorem (clt) for the residual branching algorithm, which is the simplest of the four branching algorithms introduced in [10]. We get around the lack of independence by using exchangeability techniques and by introducing an

*Supported in part by an NSERC Discovery Grant.

unrealizable approximate McKean-Vlasov particle system (originally motivated by the work of McKean [16]) which has independence.

The weighted particle filter, largely credited to Handschin [6] as well as Handschin and Mayne [7], approximates the unnormalized filter, denoted σ_n below. This weighted particle filter is the most basic particle filter and is embarassingly computer parallelizable. However, it is well know to suffer particle spread issues that have to be corrected by branching or resampling. Branching particle filters, like those introduced by Crisan and Lyons [3], can have effective resampling yet still be be highly parallelizable. Nonetheless, these early branching particle filters generally have very unstable particle numbers, which affects performance adversely. Recently, Kouritzin [10] introduced four successively more refined branching particle filters with the aim of reducing particle number fluctuations and thereby improving performance and reliability. Even the simplest of these four, the Residual Branching Particle Filter, was shown in [10], [12] to avoid wild particle swings and to outperform many popular sequential Monte Carlo methods by a large amount. Herein, we analyze this Residual Branching Particle Filter by way of Marcinkiewicz strong laws of large numbers (Mllns) and the central limit theorem (clt). As a consequence, we also layout a framework for further analysis of the Residual Branching filter as well as the three more-complicated improvements of this filter given in [10].

The *bootstrap particle filter algorithm* was introduced in 1993 by Gordon, Salmond and Smith [5]. It has been improved by using residuals and stratified random variables. This collection *resampled* particle filters is one of the big breakthroughs in big data sequential estimation and their convergence properties have been thoroughly studied by many authors (see e.g. Douc et. al. [4]). In particular, Chopin [2] obtained a clt for the residual improvement of the bootstrap algorithm. However, these particle filters approximate the actual filter π_n not the unnormalized σ_n, do not have the (same degree of) ancestrial dependence as the Residual Branching filter and base their resampling decisions upon the (locations of the) whole population. Hence, their analysis is quite different from what is required for the Residual Branching particle filter.

In terms of convergence results for branching filters to the unnormalized filter, Kouritzin and Sun [11] obtain L_2-rates of convergence for a partially-resampled branching algorithm. However, no other results were attained and their results are in a specific setting. From a mathematical perspective our work might be closest to Kurtz and Xiong [14], [15]. Their work applies to a more general setting than nonlinear filtering but in the non-linear filtering setting it only considers the weighted particle filter. Consequently, substantially new methods are required herein. We make use of classical exchangeability works like Weber [19] and McKean-Vlasov equations as in [16]. However, several new (at least to particle filtering) ideas including branching particle filter coupling, use of infinite branching particle systems, use of tracking systems and Hoeffding-inequality-based particle system bounding are also utilized.

For motivational purposes, we consider tracking a non-observable, random, dynamic signal X given the history of a distorted, corrupted partial observation process Y living on the same probability space $(\Omega, \mathcal{F}, P)$ as X. For many practical problems the signal is a time-homogeneous discrete-time Markov process $\{X_n,\ n = 0, 1, 2, ...\}$, living on some complete, separable metric space (E, ρ), with initial distribution π_0 and transition probability kernel K. The observation process takes the form ($Y_0 = 0$ and) $Y_n = h\left(X_{n-1}\right) + V_n$ for $n \in \mathbb{N}$, where $\{V_n\}_{n=1}^{\infty}$ are independent random vectors with common *strictly positive, bounded* density g that are independent of X, and the sensor function h is a measurable mapping from E to $\mathbb{R}^d$. (Such g still allows popular observation noise like Gaussian or Cauchy distributed ones.) Then, the objective of filtering is to compute the conditional expectations $\pi_n\left(f\right) = E^P\left(f\left(X_n\right)\middle|\mathcal{F}_n^Y\right)$ for all bounded, measurable functions $f : E \to \mathbb{R}$, where $\mathcal{F}_n^Y \doteq \sigma\{Y_l, l = 1, ..., n\}$ is the information obtained from the back observations.

Suppose without loss of generality that $\Omega = (E \times \mathbb{R}^d)^\infty$ and $\mathcal{F} = \mathcal{B}((E \times \mathbb{R}^d)^\infty)$ until later extended. Moreover, suppose hereafter $\mathcal{F}_{-1}^\xi \doteq \{\emptyset, \Omega\}$, $\mathcal{F}_n^\xi \doteq \sigma\{\xi_l^k, k \in \mathcal{K}, l \leq n\}$ when $n \in \mathbb{N}_0$ and $\mathcal{F}_\infty^\xi \doteq \sigma\{\xi_l^k, k \in$

$\mathcal{K}, l < \infty\}$ for random variables $\{\xi_n^k, k \in \mathcal{K}, n \in \{0,1,...\}\}$ on $(\Omega, \mathcal{F})$. (This is consistent with $\mathcal{F}_n^Y$ defined above if $\mathcal{K}$ has one element.) Unnormalized filters transfer the information contained in the observations to a likelihood process by measure change. In this method, a reference probability measure Q is introduced under which the signal, observation process $\{(X_n, Y_{n+1}),\ n = 0,1,...\}$ has the same distribution as the signal, noise process $\{(X_n, V_{n+1}),\ n = 0,1,...\}$ does under P. Hence, the observations are i.i.d. random vectors with strictly positive bounded density g and are independent of X under measure Q. All the observation information is absorbed into the likelihood process $\{L_n,\ n = 1,2,...\}$ transforming Q back to P, which in our case has the form

$$\frac{dP}{dQ}\Big|_{\mathcal{F}_\infty^X \vee \mathcal{F}_n^Y} = L_n = \prod_{j=1}^n \alpha_j(X_{j-1}), \text{ with } \alpha_j(x) = \frac{g\left(Y_j - h\left(x\right)\right)}{g\left(Y_j\right)}, \tag{1.1}$$

so $L_n = \alpha_n(X_{n-1})L_{n-1}$ and $L_0 = 1$. The following (well-known) discrete Girsanov's theorem constructs the real probability P from the fictitious one Q.

Theorem 1.1. *Suppose under probability Q that $\{X_n,\ n = 0,1,...\}$ and $\{Y_n,\ n = 1,2,...\}$ are independent processes on $(\Omega, \mathcal{F})$, the $\{Y_n\}$ are i.i.d. with strictly-positive, bounded density g on $\mathbb{R}^d$ and $V_n \doteq Y_n - h(X_{n-1})$ for all $n = 1,2,...$ Then, there exists a probability measure P such that (1.1) holds, $\{V_n,\ n = 1,2,...\}$ are i.i.d. on $(\Omega, \mathcal{F}, P)$ with density g and $\{X_n\}$ is independent of $\{V_n\}$ with the same law as on $(\Omega, \mathcal{F}, Q)$.*

Using this reference probability Q and Bayes' rule, one finds the filter satisfies $\pi_n(f) = \frac{\sigma_n(f)}{\sigma_n(1)}$, where σ_n is the *unnormalized filters*

$$\sigma_n(f) = E^Q\left(L_n f\left(X_n\right)\big|\mathcal{F}_n^Y\right) \tag{1.2}$$

so $\sigma_0 = \pi_0$, as $L_0 = 1$ and $\mathcal{F}_0^Y = \{\emptyset, \Omega\}$. Hence, the normalized filter π_n can be estimated by constructing approximations (denoted $\mathcal{S}_n^N$ and $\mathbb{S}_n^N$ below) to the unnormalized filter model. (It well known that Bayes factor can also be obtained from the unnormalized filter [13].)

Our algorithm is given in the next section and our mathematical notation in Section 3. To state our results, we let $\mathbb{S}_n^N(f)$ be our branching particle approximation to the unnormalized filter $\sigma_n(f)$. Our main result, Theorem 5.1 in Section 5, states that, for almost all observation paths, $\mathbb{S}_n^N(f)$ satisfies the Mllns (with all possible rates) and the normalized difference $\sqrt{N}(\mathbb{S}_n^N(f) - \sigma_n(f))$ satisfies the clt (with variance characterized by the resampling employed). Taken together these results say the same polynomial rates of almost-sure convergence in number of particles N hold for the Residual Branching particle filters as for other particle filters (like the weighted) even when *no extra* particles are used. Moreover, under the extra particle condition $\frac{N}{m_N} \to 0$, the random weak particle interactions in our algorithm average out enough to characterize the optimal convergence with a clt. To obtain these results, we couple our algorithm to a reduced particle system, introduced in Section 4, which is unimplementable but mathematically simpler. Conceptually, our Residual Branching particle filter is a weakly-interacting particle system and the reduced system is a more-tractable McKean-Vlasov-type limit (with average weight $\mathbb{A}_n$ replaced by $\sigma_n(1)$), which can be used to predict performance of the Residual Branching particle filter. We also introduce tracking systems in Section 6, which run as weighted filters but indicate where the Residual and reduced filters would resample (at least initially). These tracking systems are introduced for purely analytical reasons to help us divide the Residual and weighted particle filters into comparable pieces. They also have to be coupled to the Residual and reduced particle filters. The actual coupling and its ramifications are contained in Section 7. The first appendix contains the derivation of the clt variance for the McKean-Vlasov and Residual

Branching filter. The second appendix contains a technical total-mass ergodic theorem for the Residual Branching filter using the coupling.

2. Residual Branching Algorithm

To analyze our branching particle filter, we introduce it using two initial particle types: $N \in \mathbb{N}$ filter particles and $m_N - N \in \mathbb{N}$ extra particles. (A more implementable version without the extra particles and the $\{\mathbb{V}_n^{k,i}\}$ is given in [10].) The purpose of the extra particles is to allow enough asymptotic independence for the central limit theorem (clt) to hold. (Extra particles are not necessary for the Mllns to hold.) We define the following branching Markov process $\{\mathbb{S}_n^N, n = 0, 1, ...\}$ approximation to $\{\sigma_n, n = 0, 1, ...\}$ in terms of the observations and signal transition kernel K introduced in the introduction as follows:

Initialize: $\left\{\mathbb{X}_0^{k,1}\right\}_{k=1}^{m_N}$ are independent (initial particle) samples from π_0, $\left\{\mathbb{V}_n^{k,i}\right\}_{n,i,k=1}^{\infty,\infty,m_N}$ are zero-mean i.i.d. random variables, and $\mathbb{N}_0^k = 1$, $\mathbb{L}_0^{k,1} = 1$ for $k = 1, ..., m_N$.
To handle possible degeneracy, we also preset $\mathbb{N}_n^{k,i} = 0$ for all $i, k, n \in \mathbb{N}$.

Repeat: for $n = 0, 1, 2, ...$ do

1. Weight by Observation: $\widehat{\mathbb{L}}_n^{k,i} = \alpha_{n+1}\left(\mathbb{X}_n^{k,i}\right)\mathbb{L}_n^{k,i}$ for $i = 1, 2, ..., \mathbb{N}_n^k$, $k = 1, 2, ..., m_N$
2. Average Weight: $\mathbb{A}_{n+1} = \frac{1}{m_N}\sum\limits_{k=1}^{m_N}\sum\limits_{i=1}^{\mathbb{N}_n^k}\widehat{\mathbb{L}}_n^{k,i}$

Repeat (3-5): for $k = 1, 2, ..., m_N$ do

Repeat (3-5): for $i = 1, 2, ..., \mathbb{N}_n^k$ do

3. Resampled Case: If $\widehat{\mathbb{L}}_n^{k,i} + \mathbb{V}_{n+1}^{k,i} \notin (a_n\mathbb{A}_{n+1}, b_n\mathbb{A}_{n+1})$ then
   (a) Offspring Number: $\mathbb{N}_{n+1}^{k,i} = \left\lfloor \frac{\widehat{\mathbb{L}}_n^{k,i}}{\mathbb{A}_{n+1}}\right\rfloor + \rho_n^{k,i}$, with $\rho_n^{k,i}$ a $\left(\frac{\widehat{\mathbb{L}}_n^{k,i}}{\mathbb{A}_{n+1}} - \left\lfloor \frac{\widehat{\mathbb{L}}_n^{k,i}}{\mathbb{A}_{n+1}}\right\rfloor\right)$-Bernoulli independent of everything
   (b) Resampled Weight: $\overline{\mathbb{L}}_n^{k,i} = \mathbb{A}_{n+1}$
4. Non-resample Case: If $\widehat{\mathbb{L}}_n^{k,i} + \mathbb{V}_{n+1}^{k,i} \in (a_n\mathbb{A}_{n+1}, b_n\mathbb{A}_{n+1})$ then
   $\overline{\mathbb{L}}_n^{k,i} = \widehat{\mathbb{L}}_n^{k,i}$, $\mathbb{N}_{n+1}^{k,i} = 1$
5. Combine: $\widehat{\mathbb{X}}_n^{k,j} \doteq \mathbb{X}_n^{k,i}$, $\mathbb{L}_{n+1}^{k,j} \doteq \overline{\mathbb{L}}_n^{k,i}$ for

$$j \in \left\{\mathbb{N}_{n+1}^{k,1} + \cdots + \mathbb{N}_{n+1}^{k,i-1} + 1, ..., \mathbb{N}_{n+1}^{k,1} + \cdots + \mathbb{N}_{n+1}^{k,i}\right\}$$

6. Evolve Independently:

$$P(\mathbb{X}_{n+1}^{k,j} \in \Gamma_{k,j}\ \forall\ k,j | \mathcal{F}_n^{\mathbb{X}} \vee \mathcal{F}_n^Y \vee \mathcal{F}_\infty^{\mathbb{U},\mathbb{V}}) = \prod_{k=1}^{m_N}\prod_{j=1}^{\mathbb{N}_{n+1}^k} K(\widehat{\mathbb{X}}_n^{k,j}, \Gamma_{k,j})$$

   for all $\Gamma_{k,j} \in \mathcal{B}(E)$, where $\mathbb{N}_{n+1}^k = \mathbb{N}_{n+1}^{k,1} + \cdots + \mathbb{N}_{n+1}^{k,\mathbb{N}_n^k}$

7. Estimate σ_{n+1} by: $\mathbb{S}_{n+1}^N = \dfrac{1}{N}\displaystyle\sum_{k=1}^{N}\sum_{j=1}^{\mathbb{N}_{n+1}^k} \mathbb{L}_{n+1}^{k,j}\delta_{\mathbb{X}_{n+1}^{k,j}}$.

Remark 2.1. *(1) weights particles by their odds of producing the last observation. (3-5) resample the particles without bias, killing unlikely particles and duplicating likely ones while keeping the expected number of particles and total mass of all the particles constant. Parameter a_n, b_n in (3,4) control the amount of resampling. $a_n = -\infty, b_n = \infty$ turns off resampling and results in the weighted particle system. $a_n = b_n$ ensure complete resampling.*

Remark 2.2. *The $\{\mathbb{V}_n^{k,i}\}$ are required for analytical reasons. They provide enough smoothness that we can compare this resampled branching particle filter to a reduced McKean-Vlasov particle system. Without these $\mathbb{V}$'s the resampling events would be abrupt in the weight values.*

Remark 2.3. *The algorithm can fail (as all finite particle filters can) even though it is designed to have a constant expected number of particles. During resampling, there is a possibility of immediately killing all particles if $\max\limits_{j\le \mathbb{N}_{n-1}^k, k\le m_N} \dfrac{m_N \widehat{\mathbb{L}}_n^{k,j}}{\sum_{k=1}^{m_N}\sum_{i=1}^{\mathbb{N}_n^k} \widehat{\mathbb{L}}_n^{k,j}} < 1$. Ironically, this can only happen if there are more particles than at start. However, it may still be possible to degenerate immediately to one particle when $\displaystyle\sum_{k=1}^{m_N} \mathbb{N}_n^k \le m_N$. Conversely, it is not possible to increase by more than $m_N - 1$ particles in one step. The weight variation is a big concern: $\mathbb{L}_n^{k,j}$ can become very uneven as m_N increases. Some regularity results are required to ensure that there are enough effective particles and moment bounds to justify the anticipation of the clt as $m_N \to \infty$.*

To rationalize the use of $m_N - N$ extra particles, we quote the clt (see Weber [19]) for triangular sequences of exchangeable random variables:

Theorem 2.1. *Suppose $\{X_{N,j} : j = 1, ..., m_N\}$ are exchangeable random variables for all $N = 2, 3, ...$ and: (i) $\frac{N}{m_N} \to 0$, (ii) $NE[X_{N,1}^2] \to 1$, (iii) $\displaystyle\sum_{j=1}^{N} X_{N,j}^2 \to^P 1$, (iv) $N^2 E[X_{N,1}X_{N,2}] \to 0$, and (v) $\max_{j\le N} |X_{N,j}| \to^P 0$. Then, $\displaystyle\sum_{j=1}^{N} X_{N,j} \stackrel{\mathcal{D}}{\to} \mathcal{N}(0,1)$.*

Notice $m_N - N$ extra random variables are required for the desired central limit theorem. Moreover, when using our resampled branching particle filter in practice, you can take m_N to be something like $m_N = N(1 + \log\log\log N)$ (for large enough N) so you may not add many extra particles until N is very large. Finally, the Mlln rates of convergence hold even for $m_N = N$ so the extra particles are really only for *characterizing* performance.

3. Notation, Unnormalized Filter, Weighted Approximation

3.1. Basic Notation and Convention

Recall E is a Polish space and let $B(E)$, $B(E)_+$, $C(E)_{++}$, $\overline{C}(E)$ and $\overline{C}(E)_+$ be the bounded measurable, non-negative bounded, strictly-positive continuous, continuous bounded, and non-negative continuous bounded $\mathbb{R}$-valued functions respectively and define $|f|_\infty = \sup_{x\in E}|f(x)|$. Next, let $\mathcal{B}(E)$ be the Borel σ-algebra on E.

For a finite set A, we use $\#A$ to denote its number of elements.

We use the extended Vinogradov symbol (introduced in [8]): Suppose $a(n,m)$, $b(n,m)$ are expressions depending upon two sets of variables n,m. Then, $a(n,m) \overset{n}{\ll} b(n,m)$ means there exists a $c_m > 0$, depending only on m, such that $a(n,m) \le c_m b(n,m)$ for all n,m.

The main mathematical difficulty of this paper is due to the weak interaction of the particles in the residual branching system introduced in the previous section. In particular, the i^{th} particle $\mathbb{X}_n^{k,i}$ depends upon $\mathbb{A}_n$, which in turn depends on the other particles. Our approach to dealing with this weak interation is to couple this system to another (McKean-Vlasov) reduced particle system, where there is no interaction, using the fact that $A_n^N \to \sigma_n$ so the weak interactions die out as the initial number of particles increases. To do this coupling, we need to introduce tracking particle systems for the residual and reduced McKean-Vlasov particle systems that do not branch but tell us when each would. We also need an infinite particle system into which each finite particle system will be imbedded, forming our coupling between residual and reduced McKean-Vlasov particle systems. We reserve the following nomenclature for these systems:

- Blackboard bold will be used for all residual particle system objects ($\mathbb{X}_n^{k,i}$, $\mathbb{L}_n^{k,i}$, $\mathbb{A}_n$).
- Caligrahpic will be used for all reduced (McKean-Vlasov) particle system objects ($\mathcal{X}_n^{k,i}$, $\mathcal{L}_n^{k,i}$,). (σ_n replaces $\mathbb{A}_n$ in the McKean-Vlasov system.)
- The tracking system objects will be distinguished from the branching object by using an underline so ($\underline{\mathbb{X}}_n^k, \underline{\mathbb{L}}_n^k$) and ($\underline{\mathcal{L}}_n^k, \underline{\mathcal{L}}_n^k$) will respectively be the k^{th} (particle, weight) at time n in the tracking system for the residual and reduced systems.
- Normal font will be used for the infinite particle system but it will be indexed with a multi-index with the length of the multi-index being the time at which the particle is alive so X_κ is a particle in the infinite system that is alive at time $|\kappa|$. The infinite particle system does not need to have its own weights but rather the weights of the residual and reduced particle systems will be redefined on it.
- Mathfrak $\mathfrak{X}_n^k$ will be used once immediately below to introduce the weighted particle filter but will not re-appear again.

The rest of the infinite particle system notation will be introduced just prior to its use in Section 7.

3.2. Kernels, Measures and Unnormalized Filter

Since $Q\big(X_{n+1} \in A\,\big|\mathcal{F}_n^X\big) = K(X_n, A)$, one has $E^Q\big[f(X_n)\,\big|\mathcal{F}_{n-1}^X\big] = E^P\big[f(X_n)\,\big|\mathcal{F}_{n-1}^X\big] = Kf(X_{n-1})$. Clearly, $Kf \in B(E)_+$ if $f \in B(E)_+$. For any finite measure μ and integrable function f, we define

$$\mu f = \int_E f(x)\,\mu(dx)\,,\;\; K^n(y,dx) = \int_E K^{n-1}(z,dx)K(y,dz)$$
$$\mu K^n(dx) = \int_E K^n(z,dx)\,\mu(dz) \;\text{ and }\; K^n f(x) = \int_E f(z)\,K^n(x,dz)$$

for all $n = 2, 3, ...$ with $K^1 = K$.

Lastly, we define the (observation-dependent) operators A_n and $A_{i,n}$ as

$$A_n f(x) = \begin{cases} \alpha_n(x) K f(x)\,, & n \in \mathbb{N} \\ f(x), & n = 0 \end{cases} \text{ and} \tag{3.1}$$

$$A_{i,n} f(x) = \begin{cases} A_i\left(A_{i+1}\cdots(A_n f)\right)(x)\,, & i \le n \\ f(x), & i = n+1 \end{cases} . \tag{3.2}$$

Then, $\sigma_0 = \pi_0$ and, using (1.1,1.2), we have the following recursion for σ_n:

$$\sigma_n(f) \quad = \quad \sigma_{n-1}(A_n f) \ \ \forall n = 1, 2, ..., \tag{3.3}$$

Applying this recursion repeatedly, we have that

$$\sigma_n(f) = \pi_0(A_{1,n} f). \tag{3.4}$$

Bayes' rule implies that $\pi_n(f) = \frac{\sigma_{n-1}(A_n f)}{\sigma_{n-1}(A_n 1)} = \frac{\pi_0(A_{1,n} f)}{\pi_0(A_{1,n} 1)}$.

3.3. Weighted Particle Filter

Weighted particle filters approximate the unnormalized filter σ_n without resampling. The conditional expectation $\sigma_n(f) = E^Q[L_n f(X_n) \big| \mathcal{F}_n^Y]$ with respect to reference probability Q is replaced with an independent sample average to obtain

$$\sigma_n^N(f) = \frac{1}{N} \sum_{k=1}^{N} L_n^k f\left(\mathfrak{X}_n^k\right), \tag{3.5}$$

our weighted-particle estimator of $\sigma_n(f)$, where the *particles* $\left\{\mathfrak{X}^k\right\}_{k=1}^{\infty}$ are independent (π_0, K)-Markov processes that are independent of Y and the *weights* satisfy $L_n^k = \prod_{j=1}^{n} \alpha_j(\mathfrak{X}_{j-1}^k)$.

In the sequel, we will fix an observation path, set $Q^Y(\cdot) = Q(\cdot|\mathcal{F}_\infty^Y)$ and let $E^Y[Z]$ denote expectation with respect to Q^Y.

4. Reduced McKean-Vlasov Particle System

The problem with the weighted particle system is, due to randomness, most particles do not behave like the signal so their weights become small compared to the weights of very few good particles. This results in a particle filter that effectively consists of an average over only a very small portion of the particles. This problem manifests itself theoretically in the large expected variance of the central limit theorem and practically in the need to use a huge number of particles in most applications. Indeed, the weighted particle filter might not work regardless of the number of particles. To combate these problems, one introduces resampling. Initially, we pretend herein that we have access to the actual unnormalized filter total mass $\{\sigma_n(1),\ n = 0, 1, 2, ...\}$ and consider an unimplementable reduced system of McKean-Vlasov type. In particular, we use the algorithm given in Section 2 with $\mathbb{A}_n$ replaced with $\sigma_n(1)$. To facilitate analysis, we make explicit reference to the random variables that drive the particle system. Suppose we have enlarged $(\Omega, \mathcal{F}, Q)$ to support the following random variables:

1. $\left\{\chi^k\right\}_{k=1}^{\infty}$ are independent samples from π_0,
2. $\{\mathcal{Z}_n^{k,i,x} : n, k, i \in \mathbb{N}, x \in E\}$ are independent with $\mathcal{Z}_n^{k,i,x}$ having distribution $K(x, \cdot)$,
3. $\{\mathcal{U}_n^{k,i} : n, k, i \in \mathbb{N}\}$ are independent and Uniform$[0, 1]$,
4. $\{\mathcal{V}_n^{k,i} : n, k, i \in \mathbb{N}\}$ are zero mean, i.i.d. with common pdf f_V,

which are mutually independent and independent of X, Y. The actual pdf f_V does not matter for this section but has to be bounded in the next section. k is used to denote the first ancestor of each particle. Then, our reduced particle filter will be the average of N *i.i.d.* weighted branching Markov processes $\{\mathcal{B}_n^k,\ n = 0, 1, ...\}$ each starting from an independent sample δ_{χ^k}. All particles evolve independently of each other only interacting with $\{\sigma_n(1)\}$, which is deterministic with respect to Q^Y. At any time, many of the $\mathcal{B}^k$ may have died out while others have branched into multiple particles. For clarity, the particles at time n (if any) that are offspring of the initial particle χ^k will be denoted $\{\mathcal{X}_n^{k,i}\}_{i=1}^{\mathcal{N}_n^k}$ and the weight of such a particle after resampling will be denoted $\mathcal{L}_n^{k,i}$. Then, the branching Markov process corresponding to the k^{th} original particle and the complete filter estimate will be

$$\mathcal{B}_n^k = \sum_{i=1}^{\mathcal{N}_n^k} \mathcal{L}_n^{k,i} \delta_{\mathcal{X}_n^{k,i}} \text{ and } \mathcal{S}_n^N = \frac{1}{N} \sum_{k=1}^{N} \mathcal{B}_n^k \tag{4.1}$$

respectively. We define the branching Markov processes $\{\mathcal{B}^k\}$ as follows:

Initialize: $\mathcal{X}_0^{k,1} = \chi^k$, $\mathcal{N}_0^k = \mathcal{L}_0^{k,1} = 1\ \forall k = 1, ..., m_N$; $\mathcal{N}_n^{k,i} = 0\ \forall i, k, n \in \mathbb{N}$.

Repeat: for $n = 0, 1, 2, ...$ do

Repeat (1-6): for $k = 1, 2, ..., m_N$ do

Repeat (1-5): for $i = 1, 2, ..., \mathcal{N}_n^k$ do

1. Weight:

$$\widehat{\mathcal{L}}_n^{k,i} = \alpha_{n+1}(\mathcal{X}_n^{k,i}) \mathcal{L}_n^{k,i} \tag{4.2}$$

2. Resample Case: If $\widehat{\mathcal{L}}_n^{k,i} + \mathcal{V}_{n+1}^{k,i} \notin (a_n \sigma_{n+1}(1), b_n \sigma_{n+1}(1))$ then

$$\mathcal{N}_{n+1}^{k,i} = \left\lfloor \frac{\widehat{\mathcal{L}}_n^{k,i}}{\sigma_{n+1}(1)} \right\rfloor + 1_{\mathcal{U}_{n+1}^{k,i} + \left\lfloor \frac{\widehat{\mathcal{L}}_n^{k,i}}{\sigma_{n+1}(1)} \right\rfloor \le \frac{\widehat{\mathcal{L}}_n^{k,i}}{\sigma_{n+1}(1)}}, \overline{\mathcal{L}}_n^{k,i} = \sigma_{n+1}(1) \tag{4.3}$$

3. Non-resample Case: If $\widehat{\mathcal{L}}_n^{k,i} + \mathcal{V}_{n+1}^{k,i} \in (a_n \sigma_{n+1}(1), b_n \sigma_{n+1}(1))$ then $\overline{\mathcal{L}}_n^{k,i} = \widehat{\mathcal{L}}_n^{k,i}$, $\mathcal{N}_{n+1}^{k,i} = 1$
4. Combine: $\widehat{\mathcal{X}}_n^{k,j} \doteq \mathcal{X}_n^{k,i}$, $\mathcal{L}_{n+1}^{k,j} \doteq \overline{\mathcal{L}}_n^{k,i}$ for $j \in \left\{\overline{\mathcal{N}}_{n+1}^{k,i-1} + 1, ..., \overline{\mathcal{N}}_{n+1}^{k,i}\right\}$, where

$$\overline{\mathcal{N}}_{n+1}^{k,i-1} = \sum_{j=1}^{i-1} \mathcal{N}_{n+1}^{k,j} \tag{4.4}$$

5. Evolve Independently: $\mathcal{X}_{n+1}^{k,j} = \mathcal{Z}_{n+1}^{k,j,\widehat{\mathcal{X}}_n^{k,j}}$ for $j \in \left\{\overline{\mathcal{N}}_{n+1}^{k,i-1} + 1, ..., \overline{\mathcal{N}}_{n+1}^{k,i}\right\}$
6. Estimate: $\mathcal{B}_{n+1}^k = \sum_{j=1}^{\mathcal{N}_{n+1}^k} \mathcal{L}_{n+1}^{k,j} \delta_{\mathcal{X}_{n+1}^{k,j}}$, where $\mathcal{N}_{n+1}^k = \mathcal{N}_{n+1}^{k,1} + \cdots + \mathcal{N}_{n+1}^{k,\mathcal{N}_n^k}$.

Remark 4.1. *Notice that a particle in this reduced particle system behaves exactly as the same particle in residual system introduced earlier if neither branches. Conversely, if they both first branch at exactly the* *same time and produce the same number of offspring, then their offspring locations can be the same but the weights will generally differ as $\mathbb{A}_n$ will not be exactly σ_n. We will later capture these two cases in our coupling proofs. Also, situations where one system branches and the other does not or a different number of offspring are produced by the two systems will be shown to have limited occurrences.*

Remark 4.2. *This reduced filter can plunge into a zero particle trap if $\max_{j\le\mathcal{N}_{n-1}^k, k\le m_N} \frac{\widehat{\mathcal{L}}_n^{k,j}}{\sigma_{n+1}(1)} < 1$. The weights can also become very uneven. We defined an extra $m_N - N$ particles that were independent of the other particles and not used in the estimate. This was entirely for comparison with the resampled system (given in Section 2), where the extra particles are required to establish the central limit theorem.*

Remark 4.3. *To handle the index change in Step 5, we use the parent operators*

$$p_{n+1}(j) = i \quad \text{such that } j \in \{\overline{\mathcal{N}}_{n+1}^{k,i-1} + 1, ..., \overline{\mathcal{N}}_{n+1}^{k,i}\}. \tag{4.5}$$

This i is unique. p_{n+1} is defined explicitly in a slightly different context in (7.45) to follow.

After Step (4), we have $\mathcal{N}_{n+1}^{k,i}$ particles at location $\mathcal{X}_n^{k,i}$ each with weight $\overline{\mathcal{L}}_n^{k,i}$. Hence, the expected weight at location $\mathcal{X}_n^{k,i}$ after possible resampling satisfies:

$$E^Y\left[\overline{\mathcal{L}}_n^{k,i}\mathcal{N}_{n+1}^{k,i}\,\middle|\mathcal{F}_n^{\mathcal{U}\mathcal{X}} \vee \mathcal{F}_{n+1}^{\mathcal{V}}\right] = \widehat{\mathcal{L}}_n^{k,i} \quad \forall i = 1, 2, ..., \mathcal{N}_n^k, \tag{4.6}$$

which is the weight in (1) prior to resampling, so the system is *unbiased*. However, we need to go further and establish a martingale property. First, averaging over the $\mathcal{U}_n^{k,i}$, one has

$$\begin{aligned} &E^Y\left[\sum_{j=\overline{\mathcal{N}}_n^{k,i-1}+1}^{\overline{\mathcal{N}}_n^{k,i}} f(\mathcal{X}_n^{k,j})\,\middle|\mathcal{F}_{n-1}^{\mathcal{U}^{k,i}} \vee \mathcal{F}_n^{\mathcal{V}\mathcal{X}}\right] \qquad (4.7)\\ &= E^Y\left[\sum_{j=\overline{\mathcal{N}}_n^{k,i-1}+1}^{\widehat{\mathcal{N}}_n^{k,i}} f(\mathcal{X}_n^{k,j}) + \left|\frac{\widehat{\mathcal{L}}_{n-1}^{k,i}}{\overline{\mathcal{L}}_{n-1}^{k,i}} - \left\lfloor\frac{\widehat{\mathcal{L}}_{n-1}^{k,i}}{\overline{\mathcal{L}}_{n-1}^{k,i}}\right\rfloor\right| f(\mathcal{X}_n^{k,\widehat{\mathcal{N}}_n^{k,i}+1})\,\middle|\mathcal{F}_{n-1}^{\mathcal{U}^{k,i}} \vee \mathcal{F}_n^{\mathcal{V}\mathcal{X}}\right], \end{aligned}$$

where $\widehat{\mathcal{N}}_n^{k,i} = \overline{\mathcal{N}}_n^{k,i-1} + \left\lfloor\frac{\widehat{\mathcal{L}}_{n-1}^{k,i}}{\overline{\mathcal{L}}_{n-1}^{k,i}}\right\rfloor$ and $\mathcal{F}_{n-1}^{\mathcal{U}^{k,i}} = \sigma\{\mathcal{U}_m^{l,j} : m \le n, (l,j,m) \ne (k,i,n)\}$. (Notice (4.7) holds whether we resample or not.) Using (4.7) plus the facts $\mathcal{N}_{n-1}^k \in \mathcal{F}_{n-1}^{\mathcal{U}\mathcal{V}X}$ and $(\mathcal{L}_n^{k,j}, E^Y[f(\mathcal{X}_n^{k,j})|\mathcal{F}_{n-1}^{\mathcal{X}} \vee$

$\mathcal{F}_n^{\mathcal{U}\mathcal{V}}\vee]) = (\overline{\mathcal{L}}_{n-1}^{k,i}, Kf(\mathcal{X}_{n-1}^{k,i}))$ for $j \in \left\{\overline{\mathcal{N}}_n^{k,i-1}+1, ..., \overline{\mathcal{N}}_n^{k,i}\right\}$, one finds by (4.1,4.4,3.1) that

$$\begin{aligned} E^Y[\mathcal{B}_n^k(f)\big|\mathcal{F}_{n-1}^{\mathcal{U}\mathcal{V}\mathcal{X}}] &= E^Y\left[\sum_{j=1}^{\mathcal{N}_n^k}\mathcal{L}_n^{k,j}f(\mathcal{X}_n^{k,j})\Big|\mathcal{F}_{n-1}^{\mathcal{U}\mathcal{V}\mathcal{X}}\right] \\ &= \sum_{i=1}^{\mathcal{N}_{n-1}^k} E^Y\left[\sum_{j=\overline{\mathcal{N}}_n^{k,i-1}+1}^{\overline{\mathcal{N}}_n^{k,i}} \mathcal{L}_n^{k,j}f(\mathcal{X}_n^{k,j})\Big|\mathcal{F}_{n-1}^{\mathcal{U}\mathcal{V}\mathcal{X}}\right] \\ &= \sum_{i=1}^{\mathcal{N}_{n-1}^k} E^Y\left[\frac{\widehat{\mathcal{L}}_{n-1}^{k,i}}{\overline{\mathcal{L}}_{n-1}^{k,i}}\overline{\mathcal{L}}_{n-1}^{k,i}Kf(\mathcal{X}_{n-1}^{k,i})\Big|\mathcal{F}_{n-1}^{\mathcal{U}\mathcal{V}\mathcal{X}}\right] \\ &= \sum_{i=1}^{\mathcal{N}_{n-1}^k}\alpha_n(\mathcal{X}_{n-1}^{k,i})\mathcal{L}_{n-1}^{k,i}Kf(\mathcal{X}_{n-1}^{k,i}) \\ &= \mathcal{B}_{n-1}^k(A_nf) \ \text{ subject to } \mathcal{B}_0^k(f) = f(\chi^k). \end{aligned} \tag{4.8}$$

(One can check this equation in the two cases: $\mathcal{N}_{n-1}^k = 0$ and $\mathcal{N}_{n-1}^k \neq 0$.) Using (4.8) recursively, one finds by (3.2,3.4) that

$$E^Y[\mathcal{B}_n^k(f)] = E^Y\left[A_{1,n}f(\chi^k)\right] = \sigma_n(f) \tag{4.9}$$

so by (4.8,4.9)

$$\mathcal{B}_n^k(f) - \sigma_n(f) = M_n^{\mathcal{B}^k}(f), \quad \text{where} \tag{4.10}$$

$$M_n^{\mathcal{B}^k}(f) = \sum_{l=0}^n \left[\mathcal{B}_l^k(A_{l+1,n}f) - E^Y\left[\mathcal{B}_l^k(A_{l+1,n}f)\left|\mathcal{F}_{l-1}^{\mathcal{U}\mathcal{V}\mathcal{X}}\right.\right]\right]. \tag{4.11}$$

$\{M_n^{\mathcal{B}^k}(f),\ n = 0,1,...\}$ is a zero-mean $\{\mathcal{F}_n^{\mathcal{U}\mathcal{V}\mathcal{X}}\}_{n=0}^\infty$-martingale with respect to Q^Y. Averaging over the initial ancestrial branches k, one finds by (4.1,4.8,4.9,4.10,4.11) that

$$E^Y[\mathcal{S}_n^N(f)\big|\mathcal{F}_{n-1}^{\mathcal{U}\mathcal{V}\mathcal{X}}] = \mathcal{S}_{n-1}^N(A_nf) \ \text{ subject to } \mathcal{S}_0^N(f) = \frac{1}{N}\sum_{k=1}^N f(\chi^k) \tag{4.12}$$

$$E^Y[\mathcal{S}_n^N(f)] = \sigma_n(f) \tag{4.13}$$

$$\mathcal{S}_n^N(f) = \sigma_n(f) + \mathcal{M}_n^N(f) \tag{4.14}$$

with

$$\begin{aligned} \mathcal{M}_n^N(f) &= \frac{1}{N}\sum_{k=1}^N M_n^{\mathcal{B}^k}(f) \\ &= \sum_{l=0}^n\left[\mathcal{S}_l^N(A_{l+1,n}f) - E^Y\left[\mathcal{S}_l^N(A_{l+1,n}f)\left|\mathcal{F}_{l-1}^{\mathcal{U}\mathcal{V}\mathcal{X}}\right.\right]\right]. \end{aligned} \tag{4.15}$$

Now, we define the $\mathcal{F}^Y_\infty$-measurable random variance

$$\gamma^P_n(f) = E^Y[|M^{\mathcal{B}^1}_n(f)|^2]. \tag{4.16}$$

Recall $\sigma_n(f)$, α_n from (1.2),(1.1) respectively. To find an expression for the variance $\gamma^P_n(f)$ of this reduced system and the resampled system to follow, we define the *resampling* function:

$$r(x) = x - \lfloor x \rfloor - (x - \lfloor x \rfloor)^2, \tag{4.17}$$

which is an artifact of resampling and is clearly bounded by $\frac{1}{4}$. Now, let

$$\nu_n(l) = \int 1_{(a_{n-1}\sigma_n(1), b_{n-1}\sigma_n(1))}(s) f_V(l-s)ds,\ \overline{\nu}_n(l) = 1 - \nu_n(l). \tag{4.18}$$

For notational simplicity, we recall $\sigma_0(1) = \pi_0(1) = 1$ and define

$$\alpha_{i,m}(x_i, ..., x_{m-1}) = \alpha_m(x_{m-1}) \cdots \alpha_{i+2}(x_{i+1})\alpha_{i+1}(x_i)\sigma_i(1) \tag{4.19}$$
$$\nu_{i,m}(x_i, ..., x_{m-1}) = \nu_m(\alpha_{i,m}(x_i, ..., x_{m-1})) \cdots \nu_{i+1}(\alpha_{i,i+1}(x_i)) \tag{4.20}$$
$$\overline{\nu}_{i,m}(x_i, ..., x_{m-1}) = \overline{\nu}_m(\alpha_{i,m}(x_i, ..., x_{m-1})) \tag{4.21}$$

so $\alpha_{i,i}(x) = \sigma_i(1)$ and $\nu_{i,i}(x) = 1$. The following proposition gives the clt variance for the reduced McKean-Vlasov particle system in terms of the resampling used. The proof is necessarily technical, and hence delayed until Appendix 1.

Proposition 4.1. *Let h be bounded and $\sum\limits_{\substack{i_1<\cdots<i_j\\ j<l}}$ denote the sum over $1 \le i_1 < \cdots < i_j < l$ and $0 \le j < l \le n$. Then,*

$$\begin{aligned}&\gamma^P_n(f) = \pi_0((A_{1,n}f)^2) - (\pi_0(A_{1,n}f))^2 \\ &+ \sum_{\substack{i_1<\cdots<i_j\\ j<l}} \sigma_l(1)\pi_0[A_{1,l-1}\{A_l(A_{l+1,n}f)^2 - \alpha_l(KA_{l+1,n}f)^2\}\overline{\nu}_{i_j,l}\nu_{i_j,l-1}\overline{\nu}_{i_1,i_2,...,i_j}] \\ &+ \sum_{\substack{i_1<\cdots<i_j\\ j<l}} \pi_0[A_{1,l-1}\alpha_{i_j,l}\{A_l(A_{l+1,n}f)^2 - \alpha_l(KA_{l+1,n}f)^2\}\nu_{i_j,l}\overline{\nu}_{i_1,i_2,...,i_j}] \\ &+ \sum_{\substack{i_1<\cdots<i_j\\ j<l}} \sigma_l^2(1)\pi_0[A_{1,l-1}\frac{\overline{\nu}_{i_j,l}}{\alpha_{i_j,l-1}} r\left(\frac{\alpha_{i_j,l}}{\sigma_l(1)}\right)(KA_{l+1,n}f)^2\nu_{i_j,l-1}\overline{\nu}_{i_1,i_2,...,i_j}]\end{aligned} \tag{4.22}$$

for all $f \in B(E)_+$, where

$$\overline{\nu}_{i_1,i_2,...,i_j} \doteq \overline{\nu}_{i_{j-1},i_j} \cdots \nu_{i_1,i_2-1}\overline{\nu}_{0,i_1}\nu_{0,i_1-1} \tag{4.23}$$

$A_{1,m}$ is defined in (3.2) and operator A_i applies to the last argument of $A_{i+1,m}\phi_m(x_0, x_1, ..., x_{i-1}, x_i)$.

Remark 4.4. *We will later show a central limit theorem for the residual branching particle filter with this same variance. This might initially seem surprising since it uses $\mathbb{A}^N_n$ instead of the unnormalized filter σ_n. However, under our conditions $\mathbb{A}^N_n \to \sigma_n$ fast enough as $N \to \infty$ that a central limit theorem with the same variance results. Of course, we needed the extra $m_N - N$ particles for this to happen.*

Remark 4.5. *The first term on the right of (4.22) represents the error variance of introducing an independent particle system. The remaining terms incorporate the resampling scheme used. To understand this formula, we can think of $j \in \{0, 1, ..., l-1\}$ as a number of resampling events up to $l-1$ and $i_1, i_2, ..., i_j$ as possible resample times up to $l-1$ so the system would run without resampling between these times. $\nu_{i_j,l-1}\overline{\nu}_{i_{j-1},i_j}\cdots\nu_{i_1,i_2-1}\overline{\nu}_{0,i_1}\nu_{0,i_1-1}$ is then the joint probability that these are the resample times. In particular, ν_{0,i_1-1} is the probability of not resampling before i_1 and $\overline{\nu}_{0,i_1}$ is the conditional probability of resampling at i_1 given no prior resampling. Under our conditions, each σ_l is a finite measure and $\frac{\overline{\nu}_{i_j,l}}{\alpha_{i_j,l-1}}$, α_l, $A_{l,n}f$ are bounded for each fixed $Y_1, ..., Y_n$, $f \in B(E)_+$ so $\gamma_n^P(f)$ is an $\mathbb{R}$-valued random variable.*

To facilitate the discussion to follow, we break the final two terms of (4.22) into the cases of resampling at time $l-1$ and not, which yields:

$$\begin{aligned}
&\gamma_n^P(f) = \pi_0((A_{1,n}f)^2) - (\pi_0(A_{1,n}f))^2 \qquad (4.24)\\
&+ \sum_{l=1}^n \sigma_l(1)\sum_{j=0}^{l-1}\sum_{1\le i_1<\cdots<i_j<l} \pi_0[A_{1,l-1}\{f_{l,n}\}\overline{\nu}_{i_j,l}\nu_{i_j,l-1}\overline{\nu}_{i_1,i_2,...,i_j}]\\
&+ \sum_{l=2}^n\sum_{j=0}^{l-2}\sum_{1\le i_1<\cdots<i_j<l-1} \pi_0[A_{1,l-1}\alpha_{i_j,l}\{f_{l,n}\}\nu_{i_j,l}\overline{\nu}_{i_1,i_2,...,i_j}]\\
&+ \sum_{l=1}^n \sigma_{l-1}(1)\sum_{j=1}^{l-1}\sum_{1\le i_1<\cdots<i_j=l-1} \pi_0[A_{1,l-1}\alpha_l\{f_{l,n}\}\nu_{i_j,l}\overline{\nu}_{i_1,i_2,...,i_j}]\\
&+ \sum_{l=2}^n \sigma_l^2(1)\sum_{j=0}^{l-2}\sum_{1\le i_1<\cdots<i_j<l-1} \pi_0\left[A_{1,l-1}\frac{\overline{\nu}_{i_j,l}}{\alpha_{i_j,l-1}}r\left(\frac{\alpha_{i_j,l}}{\sigma_l(1)}\right)f^{l,n}\nu_{i_j,l-1}\overline{\nu}_{i_1,i_2,...,i_j}\right]\\
&+ \sum_{l=1}^n \sigma_l^2(1)\sum_{j=1}^{l-1}\sum_{1\le i_1<\cdots<i_j=l-1} \pi_0\left[A_{1,l-1}\frac{\overline{\nu}_{i_j,l}}{\sigma_{l-1}(1)}r\left(\frac{\alpha_l\sigma_{l-1}(1)}{\sigma_l(1)}\right)f^{l,n}\overline{\nu}_{i_1,i_2,...,i_j}\right]
\end{aligned}$$

for all $f \in B(E)_+$, where $f_{l,n} = A_l(A_{l+1,n}f)^2 - \alpha_l(KA_{l+1,n}f)^2$ and $f^{l,n} = (KA_{l+1,n}f)^2$.

Remark 4.6. *Notice, there are no $j = 0$ cases in the fourth and sixth terms of (4.24). For the second, third and fifth terms, the multiple sum over the i's degenerates to just one item,*

$$\sigma_l(1)\pi_0[A_{1,l-1}\left\{A_l(A_{l+1,n}f)^2 - \alpha_l(KA_{l+1,n}f)^2\right\}\overline{\nu}_{0,l}\nu_{0,l-1}], \qquad (4.25)$$

$$\pi_0[A_{1,l-1}\alpha_{0,l}\left\{A_l(A_{l+1,n}f)^2 - \alpha_l(KA_{l+1,n}f)^2\right\}\nu_{0,l}] \text{ and} \qquad (4.26)$$

$$\sigma_l^2(1)\pi_0[A_{1,l-1}\frac{1}{\alpha_{0,l-1}}r\left(\frac{\alpha_{0,l}}{\sigma_l(1)}\right)(KA_{l+1,n}f)^2\overline{\nu}_{0,l}\nu_{0,l-1}] \qquad (4.27)$$

respectively, when $j = 0$. Furthermore, in the non-resampled case where $a_i = -\infty$ and $b_i = \infty$ so $\nu_i = 1$, we have this $j = 0$ case only but also we do not resample at time l either so terms (4.25) and (4.27) also disappear. Then, we can incorporate the α_j into the operators by letting

$$A_j^{(2)}f(x) = \begin{cases} \alpha_j^2(x)Kf(x) & j = 1, 2, ... \\ f(x) & j = 0 \end{cases} \quad \text{and} \qquad (4.28)$$

$$A_{i,n}^{(2)}f = \begin{cases} A_i^{(2)}\left(A_{i+1}^{(2)}\cdots\left(A_n^{(2)}f\right)\right) & \forall i \le n \\ f & i = n+1 \end{cases}, \qquad (4.29)$$

and note that $\nu_{0,l}=1$ in this non-resampled case. Hence, the non-resampled case variance is

$$\begin{aligned}\gamma_n^W(f) &= \pi_0((A_{1,n}f)^2)-(\pi_0(A_{1,n}f))^2 \\ &+ \sum_{l=1}^n \pi_0 A_{1,l-1}^{(2)}\left[A_l^{(2)}(A_{l+1,n}f)^2-(A_{l,n}f)^2\right]\ \forall f\in B(E)_+,\end{aligned} \tag{4.30}$$

which is the variance for the weighted particle filter.

Remark 4.7. *Full resampling occurs if all $a_i=b_i$ so $\overline{\nu}_i=1$ so only the $j=l-1$ terms remain. The multiple sums over the i's in the second, fourth and sixth terms of (4.24) reduce to*

$$\sigma_l(1)\pi_0[A_{1,l-1}\left\{A_l(A_{l+1,n}f)^2-\alpha_l(KA_{l+1,n}f)^2\right\}\overline{\nu}_{l-1,l}], \tag{4.31}$$

$$\sigma_{l-1}(1)\pi_0[A_{1,l-1}\alpha_l\left\{A_l(A_{l+1,n}f)^2-\alpha_l(KA_{l+1,n}f)^2\right\}\nu_{l-1,l}], \tag{4.32}$$

$$\sigma_l^2(1)\pi_0\Big[\frac{A_{1,l-1}}{\sigma_{l-1}(1)}r\left(\frac{\alpha_l\sigma_{l-1}(1)}{\sigma_l(1)}\right)(KA_{l+1,n}f)^2\overline{\nu}_{l-1,l}] \tag{4.33}$$

respectively since $\overline{\nu}_{l-2,l-1}\cdots\overline{\nu}_{0,1}=1$ in this case. However, (4.32) also vanishes since $\nu_{l-1,l}=0$. Therefore, the variance of the fully-resampled McKean-Vlasov system is by (1.1), (1.2) and (3.4)

$$\begin{aligned}\gamma_n^R(f) &= \pi_0((A_{1,n}f)^2)-(\pi_0(A_{1,n}f))^2 \\ &+ \sum_{l=1}^n \sigma_{l-1}(\alpha_l)\sigma_{l-1}\left(A_l(A_{l+1,n}f)^2-\alpha_l(KA_{l+1,n}f)^2\right) \\ &+ \sum_{l=1}^n \sigma_{l-1}(\alpha_l)\sigma_{l-1}\left(\frac{\sigma_l(1)}{\sigma_{l-1}(1)}r\left(\frac{\alpha_l\sigma_{l-1}(1)}{\sigma_l(1)}\right)(KA_{l+1,n}f)^2\right)\end{aligned} \tag{4.34}$$

for all $f\in B(E)_+$. Comparing γ^W and the non-remainder part of γ^R (i.e. ignoring the last term of γ^R), we see that the main difference is that the former uses $A^{(2)}$ while the later uses A, so the function α_l is not squared in γ^R. Roughly speaking, this means that the errors are not compounded to the same degrees. The remaining last term in (4.34) does not have a corresponding term in (4.30) and should be thought of as the resampling noise variance. It is also compunded back in the same mild manner as the second term in (4.34). However, the existence of this term shows that one should not over resample so $a_n=b_n$ is rarely a good choice.

Remark 4.8. *By the above expressions and the proof (in the first appendix), we see that there is no need for h to be bounded in either the non-resampled (i.e. weighted) or fully-resampled case.*

This leads us to our main results of this section, which are laws of large numbers, rates of L^p-convergence and a quenched central limit theorem.

Theorem 4.1. *Let h be bounded and g be positive and continuous. Then, for Q-a.a. Y, the reduced particle system satisfies:*

slln: $\mathcal{S}_n^N\Rightarrow\sigma_n$ *(i.e. weak convergence) a.s. $[Q^Y]$;*

Mlln: $\left|\mathcal{S}_n^N(f)-\sigma_n(f)\right|\overset{N}{\ll}N^{-\beta}$ *a.s. $[Q^Y]$ for all $f\in\overline{C}(E)_+$, $0\le\beta<\frac12$;*

L^2-rates: $E^Y\left|\mathcal{S}_n^N(f)-\sigma_n(f)\right|^2=\frac{\gamma_n^P(f)}{N}$ *for all $f\in\overline{C}(E)_+$;*

L^p-rates: $E^Y\left|\mathcal{S}_n^N(f)-\sigma_n(f)\right|^p\overset{N}{\ll}N^{-\frac{p}{2}}$ *for all $f\in\overline{C}(E)_+$, $p\ge1$;*

clt: $\sqrt{N}\left(\mathcal{S}_n^N(f)-\sigma_n(f)\right)\Rightarrow\mathcal{N}\left(0,\sqrt{\gamma_n^P(f)}\right)$ *for all* $f\in\overline{C}(E)_+$.

Proof. $\mathcal{S}_n^N(f)-\sigma_n(f)=\frac{1}{N}\sum_{k=1}^{N}M_n^{\mathcal{B}^k}(f)$ is an average of i.i.d. random variables (see (4.15)) so the theorem follows by (4.22), the classical laws of large numbers, L^p bounds and central limit theorem. Note: 1) $M_n^{\mathcal{B}^k}(f)$ is bounded for fixed $Y_1,...,Y_n$ by the following Lemma. 2) $\mathcal{S}_n^N(f_i)\to\sigma_n(f_i)$ a.s. $[Q^Y]$ for all i implies $\mathcal{S}_n^N\Rightarrow\sigma_n$ a.s. $[Q^Y]$, where

$$\{f_i\}_{i=1}^{\infty}=\left\{\prod_{j=1}^{l}(1-\rho(\cdot,x_j))\vee 0: l\in\{0,1,2,...\}, x_j\in\{y_k\}_{k=1}^{\infty}\right\},\tag{4.35}$$

for some dense collection $\{y_k\}\subset E$. (See Blount and Kouritzin [1] and note the product over zero functions is taken to be the constant function 1.) □

The boundedness of $M_n^{\mathcal{B}^k}(f)$, required above follows from (4.11,4.8,4.1) and the following lemma.

Lemma 4.1. *Suppose* h *is bounded while* g *is positive and continuous. Then, there is a function* $C_n:\mathbb{R}^{dn}\to(0,\infty)$ *such that the reduced system particle numbers and weights satisfy:*

$$\mathcal{N}_l^k,\max_{i\in\{1,...,\mathcal{N}_l^k\}}\mathcal{L}_l^{k,i}\le C_n(Y_1,...,Y_n)\quad\forall k\in\{1,...,m_N\}, l\in\{0,...,n\}\ \textit{on}\ \Omega.$$

Proof. Let $\mathcal{W}_l^{k,i}=\alpha_l(\mathcal{X}_{l-1}^{k,i})$ with α_l defined in (1.1). Since

$$0<\inf_{x\in E}\frac{g(Y_l-h(x))}{g(Y_l)}<\sup_{x\in E}\frac{g(Y_l-h(x))}{g(Y_l)}<\infty$$

and σ_l is a positive finite measure for each $l\in\mathbb{N}$, there is a $C=C(Y_1,...,Y_n)>1$ such that

$$\begin{aligned}&\frac{1}{C}\le\sigma_l(1),\mathcal{W}_l^{k,i}\le C\\&\forall i=1,...,\mathcal{N}_{l-1}^k; l=1,...,n; k=1,...,m_N; N=1,2,....\end{aligned}\tag{4.36}$$

Now, recall from the reduced system algorithm (given above) that

$$\mathcal{L}_{l+1}^{k,j}\ \le\ \sigma_{l+1}(1)\vee\mathcal{W}_{l+1}^{k,p_{l+1}(j)}\mathcal{L}_l^{k,p_{l+1}(j)}\tag{4.37}$$

$$\mathcal{N}_{l+1}^k\ =\ \sum_{i_l=1}^{\mathcal{N}_l^k}\mathcal{N}_{l+1}^{k,i_l}\le\sum_{i_1=1}^{\mathcal{N}_1^k}\ \sum_{i_2=\overline{\mathcal{N}}_2^{k,i_1-1}+1}^{\overline{\mathcal{N}}_2^{k,i_1}}\cdots\sum_{i_l=\overline{\mathcal{N}}_l^{k,i_{l-1}-1}+1}^{\overline{\mathcal{N}}_l^{k,i_{l-1}}}\left[\frac{\mathcal{L}_l^{k,i_l}\mathcal{W}_{l+1}^{k,i_l}}{\sigma_{l+1}(1)}+1\right]\tag{4.38}$$

for $j=1,...,N_{l+1}^k; k=1,2,...,m_N$, where the parent operator p is defined in (4.5). Now, the stated bounds follow from (4.36,4.37,4.38), the fact $\mathcal{N}_0^k=\mathcal{L}_0^{k,1}=1$ and induction. □

$\gamma_n^P(f)$ is $\gamma_n^W(f)$ or $\gamma_n^R(f)$ when there is no resampling or full resampling respectively, where $\gamma_n^W(f)$, $\gamma_n^R(f)$ are defined in Remarks 4.6, 4.7. h need not be bounded in these two cases.

Bounded regularity for the residual branching system will not be so easy to come by but is handled in the next section.

5. Residual Branching Particle System

The reduced system uses $\sigma_n(1)$, which is usually unrepresentable on a finite computer, so we use the particle filter algorithm in the introduction, expressed now in terms of random variables $\{\chi^k\}$, $\{\mathbb{U}_n^{k,i}\}$, $\{\mathbb{V}_n^{k,i}\}$ and $\{\mathbb{Z}_n^{k,i,x}\}$ analogous to those of the previous section. Particles can now interact weakly through an average weight process $\{\mathbb{A}_n^{m_N}, n=0,1,...\}$. However, we still break up the system by the first ancestor of each particle so our resampled particle filter will be the average of N *exchangeable* branching Markov processes $\{\mathbb{B}_n^k,\ n=0,1,...\}$, each starting from an independent sample δ_{χ^k}. For clarity, the particles at time n that are offspring from the original particle χ^k will be denoted $\{\mathbb{X}_n^{k,i}\}_{i=1}^{\mathbb{N}_n^k}$ and the weight of such a particle after resampling will be denoted $\mathbb{L}_n^{k,i}$. Then, the branching Markov process corresponding to this original particle and the complete Residual Branching particle filter are:

$$\mathbb{B}_n^k = \sum_{i=1}^{\mathbb{N}_n^k} \mathbb{L}_n^{k,i}\delta_{\mathbb{X}_n^{k,i}} \text{ and } \mathbb{S}_n^N = \frac{1}{N}\sum_{k=1}^{N}\mathbb{B}_n^k. \tag{5.1}$$

The branching Markov processes $\{\mathbb{B}_n^k\}$ are defined by:

Initialize: $\mathbb{X}_0^{k,1} = \chi^k$, $\mathbb{N}_0^k = 1 = \mathbb{L}_0^{k,1}\ \forall k=1,2,...,m_N$. $\mathbb{N}_n^{k,i} = 0\ \forall i,k,n\in\mathbb{N}$.
Repeat: for $n=0,1,2,...$ do

1. Weight: $\widehat{\mathbb{L}}_n^{k,i} = \alpha_{n+1}\left(\mathbb{X}_n^{k,i}\right)\mathbb{L}_n^{k,i}$ for $i=1,2,...,\mathbb{N}_n^k$, $k=1,2,...,m_{\mathbb{N}}$
2. Average Weight:

$$\mathbb{A}_{n+1} = \frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i=1}^{\mathbb{N}_n^k}\widehat{\mathbb{L}}_n^{k,i} \tag{5.2}$$

Repeat (3-7): for $k=1,2,...,m_N$ do
Repeat (3-6): for $i=1,2,...,\mathbb{N}_n^k$ do

3. Resampled Case: If $\widehat{\mathbb{L}}_n^{k,i} + \mathbb{V}_{n+1}^{k,i} \notin (a_n\mathbb{A}_{n+1}, b_n\mathbb{A}_{n+1})$ then
   (a) Offspring Numbers: $\mathbb{N}_{n+1}^{k,i} = \left\lfloor \frac{\widehat{\mathbb{L}}_n^{k,i}}{\mathbb{A}_{n+1}}\right\rfloor + 1_{\mathbb{U}_{n+1}^{k,i} + \left\lfloor \frac{\widehat{\mathbb{L}}_n^{k,i}}{\mathbb{A}_{n+1}}\right\rfloor \le \frac{\widehat{\mathbb{L}}_n^{k,i}}{\mathbb{A}_{n+1}}}$,
   (b) Resampled Weight: $\overline{\mathbb{L}}_n^{k,i} = \mathbb{A}_{n+1}$
4. Non-resample Case: If $\widehat{\mathbb{L}}_n^{k,i} + \mathbb{V}_{n+1}^{k,i} \in (a_n\mathbb{A}_{n+1}, b_n\mathbb{A}_{n+1})$ then $\overline{\mathbb{L}}_n^{k,i} = \widehat{\mathbb{L}}_n^{k,i}$, $\mathbb{N}_{n+1}^{k,i} = 1$
5. Combine: $\widehat{\mathbb{X}}_n^{k,j} \doteq \mathbb{X}_n^{k,i}$, $\mathbb{L}_{n+1}^{k,j} \doteq \overline{\mathbb{L}}_n^{k,i}$ for $j \in \left\{\overline{\mathbb{N}}_{n+1}^{k,i-1}+1,...,\overline{\mathbb{N}}_{n+1}^{k,i}\right\}$, where

$$\overline{\mathbb{N}}_{n+1}^{k,i} = \sum_{m=1}^{i}\mathbb{N}_{n+1}^{k,m}. \tag{5.3}$$

6. Evolve Independently: $\mathbb{X}_{n+1}^{k,j} = \mathbb{Z}_{n+1}^{k,j,\widehat{\mathbb{X}}_n^{k,j}}$ for all $j \in \left\{\overline{\mathbb{N}}_{n+1}^{k,i-1}+1,...,\overline{\mathbb{N}}_{n+1}^{k,i}\right\}$
7. Estimate: $\mathbb{B}_{n+1}^k = \sum_{j=1}^{\mathbb{N}_{n+1}^k}\mathbb{L}_{n+1}^{k,j}\delta_{\mathbb{X}_{n+1}^{k,j}}$, where $\mathbb{N}_{n+1}^k = \mathbb{N}_{n+1}^{k,1} + \cdots + \mathbb{N}_{n+1}^{k,\mathbb{N}_n^k}$.

Remark 5.1. *For the index change in Step 5, we re-use the parent operator*

$$p_{n+1}(j) = i \quad \text{such that } j \in \{\overline{\mathbb{N}}_{n+1}^{k,i-1} + 1, ..., \overline{\mathbb{N}}_{n+1}^{k,i}\}, \tag{5.4}$$

defined now in terms of $\overline{\mathbb{N}}_{n+1}^{k,i}$ *instead of* $\overline{\mathcal{N}}_{n+1}^{k,i}$. *The context will make it clear for which system* p_n *is operating on.*

Remark 5.2. *The distinguishing feature between the Residual Branching and reduced particle filters is the resampling events. The resample sets for these systems are respectively*

$$\mathbb{H}_m^{k,i} = \left\{ \frac{\widehat{\mathbb{L}}_{m-1}^{k,i} + \mathbb{V}_m^{k,i}}{\mathbb{A}_m^{m_N}} \notin (a_{m-1}, b_{m-1}) \right\}, \tag{5.5}$$

$$\mathcal{H}_m^{k,i} = \left\{ \frac{\widehat{\mathcal{L}}_{m-1}^{k,i} + \mathcal{V}_m^{k,i}}{\sigma_m(1)} \notin (a_{m-1}, b_{m-1}) \right\}. \tag{5.6}$$

The conditionally expected effective weight of resampled filter particle $\mathbb{X}_n^i$ after resampling is:

$$E^Y \left[\overline{\mathbb{L}}_n^{k,i} \mathbb{N}_{n+1}^{k,i} \middle| \mathcal{F}_n^{\mathbb{U}\mathbb{X}} \vee \mathcal{F}_{n+1}^{\mathbb{V}} \right] = \widehat{\mathbb{L}}_n^{k,i},$$

which is the weight before resampling so the system is *unbiased*. Moreover, noting $\widehat{\mathbb{L}}_n^{k,i} \in \mathcal{F}_{n-1}^{\mathbb{U}\mathbb{V}\mathbb{X}}$, one finds as in (4.7-4.8) that

$$E^Y [\mathbb{B}_n^k(f) \,|\mathcal{F}_{n-1}^{\mathbb{U}\mathbb{V}\mathbb{X}}] = \mathbb{B}_{n-1}^k(A_n f) \text{ subject to } \mathbb{B}_0^k(f) = f(\chi^k). \tag{5.7}$$

Using (5.7) recursively with (3.2) and (3.4), one finds that

$$E^Y [\mathbb{B}_n^k(f)] = \sigma_n(f) \quad \text{and } \mathbb{B}_n^k(f) - \sigma_n(f) = M_n^{\mathbb{B}^k}(f), \tag{5.8}$$

with

$$\begin{aligned} M_n^{\mathbb{B}^k}(f) &= \sum_{l=0}^n \left[\mathbb{B}_l^k(A_{l+1,n} f) - E^Y \left[\mathbb{B}_l^k(A_{l+1,n} f) \,|\mathcal{F}_{l-1}^{\mathbb{U}\mathbb{V}\mathbb{X}} \right] \right] \\ &= \sum_{l=0}^n \left[\mathbb{B}_l^k(A_{l+1,n} f) - \mathbb{B}_{l-1}^k(A_{l,n} f) \right] \quad \text{if } \mathbb{B}_{-1}^k = \pi_0. \end{aligned} \tag{5.9}$$

Hence, $E^Y[M_n^{\mathbb{B}^k}(f)] = 0$ by (5.9). Moreover, $\{M_n^{\mathbb{B}^k}(f),\ n = 0, 1, ...\}$ is a $\{\mathcal{F}_n^{\mathbb{U}\mathbb{V}\mathbb{X}}\}$-martingale with respect to Q^Y. Averaging over the first N ancestrial branches, one finds that

$$E^Y [\mathbb{S}_n^N(f) | \mathcal{F}_{n-1}^{\mathbb{U}\mathbb{V}\mathbb{X}}] = \mathbb{S}_{n-1}^N(A_n f) \quad \text{subject to } \mathbb{S}_0^N(f) = \frac{1}{N} \sum_{k=1}^N f(\chi^k) \tag{5.10}$$

$$E^Y [\mathbb{S}_n^N(f)] = \sigma_n(f) \tag{5.11}$$

$$\mathbb{S}_n^N(f) = \sigma_n(f) + \mathbb{M}_n^N(f) \tag{5.12}$$

with

$$\mathbb{M}_n^N(f) = \frac{1}{N}\sum_{k=1}^{N} M_n^{\mathbb{B}^k} = \sum_{l=0}^{n}\left[\mathbb{S}_l^N(A_{l+1,n}f) - E^Y\left[\mathbb{S}_l^N(A_{l+1,n}f)\,\middle|\mathcal{F}_{l-1}^{\mathbb{UVX}}\right]\right]. \tag{5.13}$$

This leads to our main result, laws of large numbers and a quenched clt for our Residual Branching particle filter.

Theorem 5.1. *Suppose $m_N \ge N$; h and f_V are bounded; and g is strictly positive and continuous. Then, for any $n \in \mathbb{N}$ and Q-almost all Y, the Residual Branching particle filter satisfies:*

slln: $\mathbb{S}_n^N \Rightarrow \sigma_n$ *(i.e. weak convergence) a.s. $[Q^Y]$;*

Mlln: $\left|\mathbb{S}_n^N(f) - \sigma_n(f)\right| \overset{N}{\ll} N^{-\beta}$ *a.s. $[Q^Y]$ $\forall f \in \overline{C}(E)_+$, $0 \le \beta < \frac{1}{2}$;*

clt: $\sqrt{N}\left(\mathbb{S}_n^N(f) - \sigma_n(f)\right) \Rightarrow \mathcal{N}\left(0, \sqrt{\gamma_n^P(f)}\right)$ $\forall f \in \overline{C}(E)_+$ *if* $\frac{N}{m_N} \to 0$.

Remark 5.3. *The L^p rate results of Theorem 4.1 were not transferred here due to a missing bound. The proof of these L^p bounds might follow a similar line as our Mlln proof, which breaks the estimates over a "large" good set $\mathbb{D}_n^N$ (defined in Theorem 5.3 below) and "small" bad set $(\mathbb{D}_n^N)^C$. Indeed, we do provide L^p bounds on $\mathbb{D}_n^N$ but only (fast decaying) probability bounds on its complement. This is fine for our Mlln but not immediately good enough for L^p rates.*

Remark 5.4. *1) This clt requires exactly the same "extra particle" condition $\frac{N}{m_N} \to 0$ as the clt for exchangeable random variables in Theorem 2.1. 2) $\gamma_n^P(f) = \gamma_n^W(f)$, given in (4.30), when there is no resampling and $\gamma_n^P(f) = \gamma_n^R(f)$, given in (4.34), when there is full resampling.*

We use the following theorem to prove Theorem 5.1.

Theorem 5.2. *Suppose $\rho \in [0,1]$, $N_0 \in \mathbb{N}$, $m_N \ge N + N^\rho - 1$ for all $N \ge N_0$ and $\{\psi_{N,k}\}_{k=1}^{m_N}$ are exchangeable random variables such that: i) $N^{1-\rho}E\left[\psi_{N,1}^2\right] \to 0$, and ii) $NE\left[\psi_{N,1}\psi_{N,2}\right] \to 0$. Then, $\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\psi_{N,k} \to^P 0$.*

Proof. Define $\mathcal{F}_{N,i} = \sigma\left\{\psi_{N,1}, ..., \psi_{N,i}, \sum_{j=i+1}^{m_N}\psi_{N,j}\right\}$ and let $\Theta_{N,i} = \psi_{N,i} - E\left[\psi_{N,m_N}\middle|\mathcal{F}_{N,i-1}\right]$. Then, using the exchangeability, one has that

$$\begin{aligned}\lim_{N\to\infty} E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Theta_{N,i}\right|^2 &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[\Theta_{N,i}^2\right] \\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[\psi_{N,i}^2\right] \\ &- \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[E^2\left[\psi_{N,i}\middle|\mathcal{F}_{N,i-1}\right]\right] \\ &\le \lim_{N\to\infty}E\left[\psi_{N,1}^2\right] = 0 \text{ by i).}\end{aligned} \tag{5.14}$$

By exhangeability, linearity and the definition of $\mathcal{F}_{N,i}$, we find that

$$E\left[\psi_{N,m_N}\middle|\mathcal{F}_{N,i-1}\right] = \frac{1}{m_N-i+1}\sum_{j=i}^{m_N} E\left[\psi_{N,j}\middle|\mathcal{F}_{N,i-1}\right] = (m_N-i+1)^{-1}\sum_{j=i}^{m_N}\psi_{N,j}. \tag{5.15}$$

Therefore, it follows by Jensen's inequality that

$$\begin{aligned} &\lim_{N\to\infty} E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} E\left[\psi_{N,m_N}\middle|\mathcal{F}_{N,i-1}\right]\right| \\ \le\ &\lim_{N\to\infty}\sum_{i=1}^{N}\sqrt{\frac{\sum_{j=i}^{m_N} E\left[\psi_{N,j}^2\right]+\sum_{j\ne k=i}^{m_N} E\left[\psi_{N,j}\psi_{N,k}\right]}{N\left(m_N-i+1\right)^2}} \\ \le\ &\lim_{N\to\infty}\sqrt{\frac{N}{m_N-N+1}E\left[\psi_{N,1}^2\right]+NE\left[\psi_{N,1}\psi_{N,2}\right]}=0 \end{aligned} \tag{5.16}$$

by i) and ii). □

As noted in Remark 2.3, our resampled filter can degenerate to few particles or grossly uneven weights. The following *bounds*, used to prove Theorem 5.1, ensure the risk of such system irregularity decreases exponentially in the initial number of particles.

Theorem 5.3. *Suppose $n\in\mathbb{N}$; $\{m_N\}_{N=1}^\infty$ satisfies $m_1\ge 2$, $m_N\nearrow\infty$; $h\in B(\mathbb{R}^d)$; $f_V\in B(\mathbb{R})$; and $g\in C_{++}(\mathbb{R}^d)$. Then, there are $\epsilon_n>0$, $C_n>1$ and $\mathbb{D}_n^N\in\sigma\left\{\sum_{k=1}^{m_N}\mathbb{N}_l^k,\ l\le n\right\}$ such that $\mathbb{D}_{n+1}^N\subset\mathbb{D}_n^N$ for all $n=0,1,2...$; $Q^Y\left(\mathbb{D}_n^N\right)\ge 1-2ne^{-\epsilon_n m_N}$ for $N\ge1$; and*

$$\mathbb{N}_l^k,\ \max_{i\in\{1,...,\mathbb{N}_l^k\}}\mathbb{L}_l^{k,i},\mathbb{A}_l^{m_N}\le C_n\ \forall k\in\{1,...,m_N\}, l\in\{0,...,n\}\ \text{on}\ \mathbb{D}_{n-1}^N.$$

Remark 5.5. *This result says that the algorithms are well behaved up until at least one step in the future on $\mathbb{D}_n^N$, which allows comparsion of the Residual and reduced branching filters on $\mathbb{D}_n^N$.*

Remark 5.6. *$g\in C_{++}(\mathbb{R}^d)$ still allows quite general noise, including Gaussian, Cauchy, Laplace etc.*

Proof. **Initial Setup:** Let $\mathbb{W}_l^{k,i}=\alpha_l(\mathbb{X}_{l-1}^{k,i})$. Since

$$0<\inf_{x\in E}\frac{g(Y_l-h(x))}{g(Y_l)},\ \sup_{x\in E}\frac{g(Y_l-h(x))}{g(Y_l)}<\infty$$

there is a $C=C(Y_1,...,Y_n)>1$ such that

$$\frac{1}{C}\le\mathbb{W}_l^{k,i}\le C\ \forall 1\le i\le N_{l-1}^k; 1\le l\le n; 1\le k\le m_N; N\ge 1. \tag{5.17}$$

For $l \geq 1$, we define $\upsilon_C(l), \tau_C(l), \mathbb{D}_l^N$ recursively by

$$\upsilon_C(l) = C\upsilon_C(l-1)\tau_C(l-1), \quad \text{subject to } \upsilon_C(0) = 1, \tag{5.18}$$

$$\tau_C(l) = 2\tau_C(l-1)(C\upsilon_C(l)\upsilon_C(l-1)+1) \quad \text{subject to } \tau_C(0) = 1, \tag{5.19}$$

$$\mathbb{D}_l^N = \left\{\frac{1}{\tau_C(l)} \leq \frac{1}{m_N}\sum_{k=1}^{m_N}\mathbb{N}_l^k \leq \tau_C(l)\right\} \cap \mathbb{D}_{l-1}^N \text{ subject to } \mathbb{D}_0^N = \Omega. \tag{5.20}$$

Clearly, $\mathbb{D}_l^N \in \mathcal{F}_l^{\mathbb{X}\mathbb{U}\mathbb{V}}$. Now, recall from (5.2) and the Residual Branching algorithm that

$$\mathbb{A}_{l+1}^{m_N} = \frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i=1}^{\mathbb{N}_l^k}\mathbb{W}_{l+1}^{k,i}\mathbb{L}_l^{k,i} \tag{5.21}$$

$$\mathbb{A}_{l+1}^{m_N} \wedge \mathbb{W}_{l+1}^{k,p_{l+1}(j)}\mathbb{L}_l^{k,p_{l+1}(j)} \leq \mathbb{L}_{l+1}^{k,j} \leq \mathbb{A}_{l+1}^{m_N} \vee \mathbb{W}_{l+1}^{k,p_{l+1}(j)}\mathbb{L}_l^{k,p_{l+1}(j)} \tag{5.22}$$

$$\mathbb{N}_{l+1}^k = \sum_{i_l=1}^{\mathbb{N}_l^k}\mathbb{N}_{l+1}^{k,i_l} \leq \sum_{i_1=1}^{\mathbb{N}_1^k}\sum_{i_2=\overline{\mathbb{N}}_2^{k,i_1-1}+1}^{\overline{\mathbb{N}}_2^{k,i_1}}\cdots\sum_{i_l=\overline{\mathbb{N}}_l^{k,i_{l-1}-1}+1}^{\overline{\mathbb{N}}_l^{k,i_{l-1}}}\left\lceil\frac{\mathbb{L}_l^{k,i_l}\mathbb{W}_{l+1}^{k,i_l}}{\mathbb{A}_{l+1}^{m_N}}+1\right\rceil \tag{5.23}$$

for $j = 1, ..., N_{l+1}^k; k = 1, 2, ..., m_N$. These imply that

$$\frac{1}{\upsilon_C(l+1)} \leq \mathbb{A}_{l+1}^{m_N} \leq \upsilon_C(l+1) \tag{5.24}$$

$$\frac{1}{\upsilon_C(l+1)} \leq \mathbb{L}_{l+1}^{k,i} \leq \upsilon_C(l+1) \ \forall k \in \{1,2,...,m_N\},\ i \in \{1,...,\mathbb{N}_{l+1}^k\} \tag{5.25}$$

$$\frac{1}{C\upsilon_C(l)} \leq \mathbb{W}_{l+1}^{k,i}\mathbb{L}_l^{k,i} \leq C\upsilon_C(l) \ \forall k \in \{1,2,...,m_N\},\ i \in \{1,...,\mathbb{N}_{l+1}^k\} \tag{5.26}$$

$$\mathbb{N}_{l+1}^k \leq \prod_{i=0}^{l}(\upsilon_C(i+1)\upsilon_C(i)C+1) \doteq M_C(l+1) \ \forall k \in \{1,2,...,m_N\} \tag{5.27}$$

on $\mathbb{D}_l^N$ for all $l = 0, 1, 2..., n$ by induction and (5.17).

Base Case: $\{\mathbb{N}_1^k\}$ are bounded by $M_C(1)$ (since $\mathbb{D}_0^N = \Omega$) and conditionally independent so Hoeffding's inequality applies to find

$$Q^Y\left(\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\left[\mathbb{N}_1^k - \left[\frac{\mathbb{W}_1^{k,1}}{\mathbb{A}_1^{m_N}}1_{\mathbb{H}_1^{k,1}} + 1_{(\mathbb{H}_1^{k,1})^C}\right]\right]\right| > t\middle|\mathcal{F}_0^X \vee \mathcal{F}_1^V\right) \tag{5.28}$$

$$\leq \quad 2\exp\left(-\frac{2m_N t^2}{M_C^2(1)}\right) \text{ a.s.,}$$

where resample set $\mathbb{H}_1^{k,i}$ is defined in (5.5). Next, by (5.17), (5.24), (5.18), (5.19) and (5.28)

$$
\begin{aligned}
& Q^Y\left(\left\{\frac{1}{\tau_C(1)} \le \frac{1}{m_N}\sum_{k=1}^{m_N} \mathbb{N}_1^k \le \tau_C(1)\right\}\right) \qquad (5.29)\\
\ge\ & Q^Y\left(\left\{\frac{2}{\tau_C(1)} \le \frac{1}{m_N}\sum_{k=1}^{m_N}\left[\frac{\mathbb{W}_1^{k,1}}{\mathbb{A}_1^{m_N}} 1_{\mathbb{H}_1^{k,1}} + 1_{(\mathbb{H}_1^{k,1})^C}\right] \le \frac{\tau_C(1)}{2}\right\}\right)\\
-\ & E^Y\left[Q^Y\left(\left|\sum_{k=1}^{m_N}\left[\mathbb{N}_1^k - \left[\frac{\mathbb{W}_1^{k,1}}{\mathbb{A}_1^{m_N}} 1_{\mathbb{H}_1^{k,1}} + 1_{(\mathbb{H}_1^{k,1})^C}\right]\right]\right| > \frac{m_N}{\tau_C(1)}\middle| \mathcal{F}_0^{\mathbb{X}} \vee \mathcal{F}_1^{\mathbb{V}}\right)\right]\\
\ge\ & 1 - 2\exp\left(-\frac{2m_N}{M_C^2(1)\tau_C^2(1)}\right).
\end{aligned}
$$

Inductive Step: Suppose that

$$
Q^Y(\mathbb{D}_l^N) \quad \ge \quad 1 - 2l\exp\left(-\frac{2m_N}{M_C^2(l)\tau_C^2(l)}\right), \qquad (5.30)
$$

which is true when $l = 1$, and let

$$
\rho_l^{k,i} = \frac{\mathbb{W}_{l+1}^{k,i}\mathbb{L}_l^{k,i}}{\mathbb{A}_{l+1}^{m_N}} 1_{\mathbb{H}_{l+1}^{k,i}} + 1_{(\mathbb{H}_{l+1}^{k,i})^C}. \qquad (5.31)
$$

Then, it follows by (5.26), (5.24) (5.19) and (5.18) that

$$
\begin{aligned}
& Q^Y\left(\left\{\frac{1}{\tau_C(l+1)} \le \frac{1}{m_N}\sum_{k=1}^{m_N} \mathbb{N}_{l+1}^k \le \tau_C(l+1)\right\} \cap \mathbb{D}_l^N\right) \qquad (5.32)\\
\ge\ & Q^Y\left(\left\{\frac{2}{\tau_C(l+1)} \le \frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i=1}^{\mathbb{N}_l^k} \rho_l^{k,i} \le \frac{\tau_C(l+1)}{2}\right\} \cap \mathbb{D}_l^N\right)\\
-\ & Q^Y\left(\left\{\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\left[\mathbb{N}_{l+1}^k - \sum_{i=1}^{\mathbb{N}_l^k}\rho_l^{k,i}\right]\right| > \frac{1}{\tau_C(l+1)}\right\} \cap \mathbb{D}_l^N\right)\\
\ge\ & Q^Y(\mathbb{D}_l^N) - Q^Y\left(\left\{\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\left[\mathbb{N}_{l+1}^k - \sum_{i=1}^{\mathbb{N}_l^k}\rho_l^{k,i}\right]\right| > \frac{1}{\tau_C(l+1)}\right\} \cap \mathbb{D}_l^N\right).
\end{aligned}
$$

However, we have by the independence of the U's, (5.27) and Hoeffding's inequality that

$$
\begin{aligned}
& Q^Y\left(\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\left[\mathbb{N}_{l+1}^k - \sum_{i=1}^{\mathbb{N}_l^k}\rho_l^{k,i}\right]\right| > t \middle| \mathcal{F}_\infty^{\mathbb{X},\mathbb{V}} \vee \mathcal{F}_l^{\mathbb{U}}\right) \qquad (5.33)\\
\le\ & 2\exp\left(-\frac{2m_N t^2}{M_C^2(l+1)}\right) \quad \text{on } \mathbb{D}_l^N,
\end{aligned}
$$

where resample set $\mathbb{H}_l^{k,i}$ is defined in (5.5), so by (5.32), (5.30) and (5.33) with $t=\frac{1}{\tau_C(l+1)}$

$$\begin{aligned} & Q^Y\Bigg(\Bigg\{\tau_C(l+1)\le \frac{1}{m_N}\sum_{k=1}^{m_N}\mathbb{N}_{l+1}^k\le \tau_C(l+1)\Bigg\}\cap \mathbb{D}_l^N\Bigg) \\ \ge\;\; & 1-2(l+1)\exp\Bigg(-\frac{2m_N}{M_C^2(l+1)\tau_C^2(l+1)}\Bigg). \end{aligned} \tag{5.34}$$

Conclusion: The result follows by induction, (5.20), (5.24) and (5.25). □

The proof of Theorem 5.1 also relies on a coupling of our systems as well as tracking systems that run as weighted particle filters but signal the resampling events for the Residual and reduced branching filters.

6. Tracking Systems

For analytical reasons, we define tracking systems corresponding to the Residual and reduced systems. These systems do not resample but do track where resampling would occur (at least initially). They are used in Appendix 2 to establish the "closeness" of the Residual and weighted filter total masses. However, they are introduced now in order that we can couple these tracking systems with the Residual and reduced systems on the same probability space.

The *reduced tracking* system is defined as follows:

Initialize: $\underline{\mathcal{X}}_0^k=\chi^k$ and $\underline{\mathcal{L}}_0^k=1$ for $k=1,2,...,m_N$;

Repeat: for $n=0,1,2,...$ do

For $k=1,2,...,m_N$ **do:**

$$\widehat{\underline{\mathcal{L}}}_n^k \;=\; \alpha_{n+1}(\underline{\mathcal{X}}_n^k)\underline{\mathcal{L}}_n^k \tag{6.1}$$

$$\underline{\mathcal{L}}_{n+1}^k \;=\; \begin{cases} \sigma_{n+1}(1), & \widehat{\underline{\mathcal{L}}}_n^k+\mathcal{V}_{n+1}^{k,1}\notin(a_n\sigma_{n+1}(1),b_n\sigma_{n+1}(1)) \\ \widehat{\underline{\mathcal{L}}}_n^k, & \widehat{\underline{\mathcal{L}}}_n^k+\mathcal{V}_{n+1}^{k,1}\in(a_n\sigma_{n+1}(1),b_n\sigma_{n+1}(1)) \end{cases} \tag{6.2}$$

$$\underline{\mathcal{X}}_{n+1}^k \;=\; \mathcal{Z}_{n+1}^{k,1,\underline{\mathcal{X}}_n^k} \tag{6.3}$$

while the *Residual tracking* system is:

Initialize: $\underline{\mathbb{X}}_0^k=\chi^k$ and $\underline{\mathbb{L}}_0^k=1$ for $k=1,2,...,m_N$;

Repeat: for $n=0,1,2,...$ do

For $k=1,2,...,m_N$ **do:**

$$\widehat{\underline{\mathbb{L}}}_n^k \;=\; \alpha_{n+1}(\underline{\mathbb{X}}_n^k)\underline{\mathbb{L}}_n^k \tag{6.4}$$

$$\underline{\mathbb{L}}_{n+1}^k \;=\; \begin{cases} \mathbb{A}_{n+1}, & \widehat{\underline{\mathbb{L}}}_n^k+\mathbb{V}_{n+1}^{k,1}\notin(a_n\mathbb{A}_{n+1},b_n\mathbb{A}_{n+1}) \\ \widehat{\underline{\mathbb{L}}}_n^k, & \widehat{\underline{\mathbb{L}}}_n^k+\mathbb{V}_{n+1}^{k,1}\in(a_n\mathbb{A}_{n+1},b_n\mathbb{A}_{n+1}) \end{cases} \tag{6.5}$$

$$\underline{\mathbb{X}}_{n+1}^k \;=\; \mathbb{Z}_{n+1}^{k,1,\underline{\mathbb{X}}_n^k}. \tag{6.6}$$

In the above algorithms, $\{\mathbb{V}_n^{k,1};n,k=1,2,...\}$ and $\{\mathbb{Z}_n^{k,1,x};n,k=1,2,...,x\in E\}$ are the random variables used in the Residual system while $\{\mathcal{V}_n^{k,1};n,k=1,2,...\}$ and $\{\mathcal{Z}_n^{k,1,x};n,k=1,2,...,x\in E\}$ are the random variables used in the reduced system. $\{\mathbb{A}_n, n = 0, 1, 2, ...\}$ is also from the Residual system. Hence, the Residual and reduced tracking systems have been defined on the same probability space as the Residual and reduced particle filters respectively.

One would never implement these tracking systems. Roughly speaking, they run as weighted filters but indicate (at least initially) where resampling for the reduced and Residual particle filter would have taken place. Their importance is solely to ease the analysis by facilitating a break up of the weighted and Residual particle filters over certain resampling events. In particular, the resample sets of the tracking systems:

$$\underline{\mathbb{H}}_m^k = \left\{ \frac{\widehat{\underline{\mathbb{L}}}_{m-1}^k + \mathbb{V}_m^{k,1}}{\mathbb{A}_m^{m_N}} \notin (a_{m-1}, b_{m-1}) \right\}, \tag{6.7}$$

$$\underline{\mathcal{H}}_m^k = \left\{ \frac{\widehat{\underline{\mathcal{L}}}_{m-1}^k + \mathcal{V}_m^{k,1}}{\sigma_m(1)} \notin (a_{m-1}, b_{m-1}) \right\} \tag{6.8}$$

are important to break up the weighted and reduced particle systems into comparable pieces once we have coupled all systems together on the same probability space.

7. Coupling

To obtain "nearness" estimates between the resampled, tracking and reduced filters, we couple them through an infinite particle system. Suppose $\mathbb{N}^0 = \{\emptyset\}$, $\mathbb{M} = \bigcup_{n=0}^{\infty} \mathbb{N}^n$, $|\kappa| = n$ if multi-index $\kappa \in \mathbb{N}^n$ and we enlarge $(\Omega, \mathcal{F}, Q)$ to support the following random variables:

1. $\{\chi^k\}_{k=1}^{\infty}$ are independent samples from π_0,
2. $\left\{Z_\kappa^{k;x} : \kappa \in \bigcup_{n=1}^{\infty} \mathbb{N}^n, k \in \mathbb{N}; x \in E\right\}$ are independent, distribution $K(x, \cdot)$,
3. $\{U_\kappa^k : \kappa \in \bigcup_{n=1}^{\infty} \mathbb{N}^n, k \in \mathbb{N}\}$ are independent and Uniform$[0, 1]$,
4. $\{V_\kappa^k : \kappa \in \bigcup_{n=1}^{\infty} \mathbb{N}^n, k \in \mathbb{N}\}$ are independent and zero mean with common pdf f_V,

which are mutually independent and independent of X, Y. Then, at time n, there is a particle X_κ^k corresponding to each initial particle k and multi-index κ with $|\kappa| = n$ that satisfies:

$$X_\emptyset^k = \chi^k,\ X_{(\kappa,i)}^k = Z_{(\kappa,i)}^{k;X_\kappa^k} \quad \forall \kappa \in \mathbb{M}; k, i \in \mathbb{N}. \tag{7.1}$$

$(\mathbb{N}_\kappa^k,\mathbb{L}_\kappa^k)_{\kappa\in\mathbb{M},k\in\mathbb{N}}$, $(\mathcal{N}_\kappa^k,\mathcal{L}_\kappa^k)_{\kappa\in\mathbb{M},k\in\mathbb{N}}$, $(\underline{\mathbb{L}}_\kappa^k)_{\kappa\in\mathbb{M},k\in\mathbb{N}}$ and $(\underline{\mathcal{L}}_\kappa^k)_{\kappa\in\mathbb{M},k\in\mathbb{N}}$ then extend the notion of offspring numbers and likelihood for the finite systems to the infinite system, where

$$(\mathbb{N}_\emptyset^k,\mathbb{L}_\emptyset^k) = (\mathcal{N}_\emptyset^k,\mathcal{L}_\emptyset^k)=(1,1),\ \underline{\mathbb{L}}_\emptyset^k=\underline{\mathcal{L}}_\emptyset^k=1 \tag{7.2}$$

$$(\mathbb{N}_{(\kappa,i)}^k,\mathbb{L}_{(\kappa,i)}^k) = \begin{cases}\left(\mathbb{L}U_{(\kappa,i)}^k,\mathbb{A}_{n+1}\right), & \frac{\widehat{\mathbb{L}}_\kappa^k+V_{(\kappa,i)}^k}{\mathbb{A}_{n+1}}\notin(a_n,b_n),\, i\le\mathbb{N}_\kappa^k\\ (1,\widehat{\mathbb{L}}_\kappa^k), & \frac{\widehat{\mathbb{L}}_\kappa^k+V_{(\kappa,i)}^k}{\mathbb{A}_{n+1}}\in(a_n,b_n),\, i\le\mathbb{N}_\kappa^k\\ (0,0), & i>\mathbb{N}_\kappa^k\end{cases} \tag{7.3}$$

$$(\mathcal{N}_{(\kappa,i)}^k,\mathcal{L}_{(\kappa,i)}^k) = \begin{cases}\left(\mathcal{L}U_{(\kappa,i)}^k,\sigma_{n+1}(1)\right), & \frac{\widehat{\mathcal{L}}_\kappa^k+V_{(\kappa,i)}^k}{\sigma_{n+1}(1)}\notin(a_n,b_n),\, i\le\mathcal{N}_\kappa^k\\ (1,\widehat{\mathcal{L}}_\kappa^k), & \frac{\widehat{\mathcal{L}}_\kappa^k+V_{(\kappa,i)}^k}{\sigma_{n+1}(1)}\in(a_n,b_n),\, i\le\mathcal{N}_\kappa^k\\ (0,0), & i>\mathcal{N}_\kappa^k\end{cases} \tag{7.4}$$

$$\underline{\mathbb{L}}_{(\kappa,i)}^k = \begin{cases}\mathbb{A}_{n+1}, & \frac{\widehat{\mathbb{L}}_\kappa^k+V_{(\kappa,i)}^k}{\mathbb{A}_{n+1}}\notin(a_n,b_n),\, i=1\\ \widehat{\underline{\mathbb{L}}}_\kappa^k, & \frac{\widehat{\mathbb{L}}_\kappa^k+V_{(\kappa,i)}^k}{\mathbb{A}_{n+1}}\in(a_n,b_n),\, i=1\\ 0, & i>1\end{cases} \tag{7.5}$$

$$\underline{\mathcal{L}}_{(\kappa,i)}^k = \begin{cases}\sigma_{n+1}(1), & \frac{\widehat{\mathcal{L}}_\kappa^k+V_{(\kappa,i)}^k}{\sigma_{n+1}(1)}\notin(a_n,b_n),\, i=1\\ \widehat{\underline{\mathcal{L}}}_\kappa^k, & \frac{\widehat{\mathcal{L}}_\kappa^k+V_{(\kappa,i)}^k}{\sigma_{n+1}(1)}\in(a_n,b_n),\, i=1\\ 0, & i>1\end{cases} \tag{7.6}$$

for all $k,i\in\mathbb{N},\ |\kappa|=n,\ n=0,1,2,....$ Here,

$$\mathbb{L}U_{(\kappa,i)}^k = \left\lfloor\frac{\widehat{\mathbb{L}}_\kappa^k}{\mathbb{A}_{n+1}}\right\rfloor+1_{U_{(\kappa,i)}^k+\left\lfloor\frac{\widehat{\mathbb{L}}_\kappa^k}{\mathbb{A}_{n+1}}\right\rfloor\le\frac{\widehat{\mathbb{L}}_\kappa^k}{\mathbb{A}_{n+1}}} \tag{7.7}$$

$$\mathcal{L}U_{(\kappa,i)}^k = \left\lfloor\frac{\widehat{\mathcal{L}}_\kappa^k}{\sigma_{n+1}(1)}\right\rfloor+1_{U_{(\kappa,i)}^k+\left\lfloor\frac{\widehat{\mathcal{L}}_\kappa^k}{\sigma_{n+1}(1)}\right\rfloor\le\frac{\widehat{\mathcal{L}}_\kappa^k}{\sigma_{n+1}(1)}} \tag{7.8}$$

$$\mathbb{A}_{n+1} = \frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{\kappa:|\kappa|=n}\widehat{\mathbb{L}}_\kappa^k \text{ for } n=0,1,...;\mathbb{A}_0=1; \tag{7.9}$$

$$\widehat{\mathbb{L}}_\kappa^k = \alpha_{|\kappa|+1}(X_\kappa^k)\mathbb{L}_\kappa^k,\qquad \widehat{\mathcal{L}}_\kappa^k=\alpha_{|\kappa|+1}(X_\kappa^k)\mathcal{L}_\kappa^k, \tag{7.10}$$

$$\widehat{\underline{\mathbb{L}}}_\kappa^k = \alpha_{|\kappa|+1}(X_\kappa^k)\underline{\mathbb{L}}_\kappa^k \text{ and } \widehat{\underline{\mathcal{L}}}_\kappa^k=\alpha_{|\kappa|+1}(X_\kappa^k)\underline{\mathcal{L}}_\kappa^k \text{ for } \kappa\in\mathbb{M},k\in\mathbb{N}. \tag{7.11}$$

Next, we introduce a partial order on $\mathbb{M}$: $\kappa\prec\widehat{\kappa}$ if $|\kappa|=|\widehat{\kappa}|$ and $\min\{i:\kappa_i<\widehat{\kappa}_i\}<\min\{i:\widehat{\kappa}_i<\kappa_i\}$. To make *room* for live particles from all finite systems, we let

$$N_\kappa^k=\mathbb{N}_\kappa^k\vee\mathcal{N}_\kappa^k\vee 1\ \ \forall k\in\mathbb{N},\kappa\in\mathbb{M} \tag{7.12}$$

and define the subset of *alive multi-indices* $\mathbb{M}^A$ by $\kappa\in\mathbb{M}^A$ if $\kappa\in\mathbb{M}$ and either

$$\kappa=\emptyset \text{ or } \kappa=(\kappa_1,...,\kappa_n) \text{ with } \kappa_l\in\{1,...,N_{(\kappa_1,...,\kappa_{l-1})}^k\}\ \forall l=1,...,n, \tag{7.13}$$

so particles $X^k_{(\kappa_1,...,\kappa_n)}$ with ($n\ge 1$ and) some $\kappa_l > N^k_{(\kappa_1,...,\kappa_{l-1})}$ are not in any finite system. To recover the finite systems, we drop explicit reference to the ancesteral chain and set:

$$X_n^{k,j} = X_\kappa^k,\ U_n^{k,j}=U_\kappa^k,\ V_n^{k,j}=V_\kappa^k,\ Z_n^{k,j,x}=Z_\kappa^{k;x}, \tag{7.14}$$
$$\mathcal{K}_n^{k,j} = \mathcal{L}_\kappa^k,\ \mathbb{K}_n^{k,j}=\mathbb{L}_\kappa^k,\ \widehat{\mathcal{K}}_n^{k,j}=\widehat{\mathcal{L}}_\kappa^k,\ \widehat{\mathbb{K}}_n^{k,j}=\widehat{\mathbb{L}}_\kappa^k, \tag{7.15}$$
$$N_n^{k,j} = N_\kappa^k, \overline{N}_l^{k,i}=\sum_{m=1}^{i} N_l^{k,m} \text{ and } N_l^k=\sum_{m=1}^{N_{l-1}^k} N_l^{k,m} \text{ with } N_0^k=1, \tag{7.16}$$

where κ is the unique alive multi-index such that $|\kappa|=n$ and

$$j=\eta(\kappa)\doteq \#\{\widehat{\kappa}\in\mathbb{M}^A:\widehat{\kappa}\prec\kappa\}+1. \tag{7.17}$$

(Many $\mathbb{K}$, $\mathcal{K}$ could be zero.) For the tracking systems, we define

$$\underline{\mathcal{K}}_n^k = \underline{\mathcal{L}}_\kappa^k,\ \underline{\mathbb{K}}_n^k=\underline{\mathbb{L}}_\kappa^k,\ \widehat{\underline{\mathcal{K}}}_n^k=\widehat{\underline{\mathcal{L}}}_\kappa^k,\ \widehat{\underline{\mathbb{K}}}_n^k=\widehat{\underline{\mathbb{L}}}_\kappa^k, \tag{7.18}$$

for $\kappa=(1,1,...,1)$ with $|\kappa|=n$. Now, it follows by (7.9), (7.3), (7.12), (7.10), (7.1) and (7.14-7.16) that

$$\mathbb{A}_{n+1} = \frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{\kappa:|\kappa|=n}\widehat{\mathbb{L}}_\kappa^k=\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{j=1}^{N_n^k}\widehat{\mathbb{K}}_n^{k,j} \text{ and} \tag{7.19}$$
$$X_n^{k,j} = Z_n^{k,j,X_{n-1}^{k,i}} \text{ for } j\in\{\overline{N}_n^{k,i-1}+1,...,\overline{N}_n^{k,i}\}. \tag{7.20}$$

For convenience, let $\mathbb{I}_n^k=\{i:\mathbb{K}_n^{k,i}\neq 0\}$ and $\mathcal{I}_n^k=\{i:\mathcal{K}_n^{k,i}\neq 0\}$ be the Residual and reduced particles at time n that started from the k^{th} inital particle and $|\mathbb{I}_n^k|$ denote the cardinality of $\mathbb{I}_n^k$. Redefine the resample and non-resample sets (previously defined in (5.5,5.6,6.7, 6.8))

$$\mathbb{R}_m^{k,i} = \left\{\frac{\widehat{\mathbb{K}}_{m-1}^{k,i}+V_m^{k,i}}{\mathbb{A}_m^{m_N}}\notin(a_{m-1},b_{m-1}), i\in\mathbb{I}_{m-1}^k\right\}, \tag{7.21}$$
$$\mathbb{S}_m^{k,i} = \{i\in\mathbb{I}_{m-1}^k\}\setminus\mathbb{R}_m^{k,i}, \tag{7.22}$$
$$\mathcal{R}_m^{k,i} = \left\{\frac{\widehat{\mathcal{K}}_{m-1}^{k,i}+V_m^{k,i}}{\sigma_m(1)}\notin(a_{m-1},b_{m-1}), i\in\mathcal{I}_{m-1}^k\right\}, \tag{7.23}$$
$$\mathcal{S}_m^{k,i} = \{i\in\mathcal{I}_{m-1}^k\}\setminus\mathcal{R}_m^{k,i}, \tag{7.24}$$
$$\underline{\mathbb{R}}_m^k = \left\{\frac{\widehat{\underline{\mathbb{K}}}_{m-1}^k+V_m^{k,1}}{\mathbb{A}_m^{m_N}}\notin(a_{m-1},b_{m-1})\right\}, \tag{7.25}$$
$$\underline{\mathcal{R}}_m^k = \left\{\frac{\widehat{\underline{\mathcal{K}}}_{m-1}^k+V_m^{k,1}}{\sigma_m(1)}\notin(a_{m-1},b_{m-1})\right\}. \tag{7.26}$$

The following combinations of resample and non-resample events will be useful in comparing our Residual particle filter total mass to the weighted total mass in Appendix 2:

$$\mathbb{RSI}_{l,n}^{k,i_{l-1},i_l,...,i_n} = \mathbb{R}_l^{k,i_{l-1}}\cap\mathbb{S}_{l+1}^{k,i_l}\cap\cdots\cap\mathbb{S}_n^{k,i_{n-1}}\cap\{i_n\in\mathbb{I}_n^k\}, \tag{7.27}$$
$$\mathcal{RSI}_{l,n}^{k,i_{l-1},i_l,...,i_n} = \mathcal{R}_l^{k,i_{l-1}}\cap\mathcal{S}_{l+1}^{k,i_l}\cap\cdots\cap\mathcal{S}_n^{k,i_{n-1}}\cap\{i_n\in\mathcal{I}_n^k\}. \tag{7.28}$$

Our coupling of the finite systems on common probability space $(\Omega, \mathcal{F}, Q^Y)$ is complete. We use this coupling to transfer the bounds of Theorem 5.3 to the infinite particle system, to prove Theorem 5.1 and to ease notation about (9.14) of Appendix 2. For these uses, we need the following result.

Theorem 7.1. *Suppose $\mathbb{B}_n^k$, $\mathcal{B}_n^k$ are the Residual, reduced Markov branching processes defined in (5.1), (4.1) and $\mathbb{N}_n^k$, $\mathcal{N}_n^k$ are the corresponding particle numbers. Then,*

$$\left\{\mathbb{N}_n^k, \mathbb{B}_n^k\right\}_{1\le k\le m_N, n\in\mathbb{N}_0, N\in\mathbb{N}} \overset{D}{=} \left\{|\mathbb{I}_n^k|, \sum_{i=1}^{N_n^k} \mathbb{K}_n^{k,i} \delta_{X_n^{k,i}}\right\}_{1\le k\le m_N, n\in\mathbb{N}_0, N\in\mathbb{N}} \tag{7.29}$$

$$\left\{\mathcal{N}_n^k, \mathcal{B}_n^k\right\}_{1\le k\le m_N, n\in\mathbb{N}_0, N\in\mathbb{N}} \overset{D}{=} \left\{|\mathcal{I}_n^k|, \sum_{i=1}^{N_n^k} \mathcal{K}_n^{k,i} \delta_{X_n^{k,i}}\right\}_{1\le k\le m_N, n\in\mathbb{N}_0, N\in\mathbb{N}} \tag{7.30}$$

and

$$\left\{\sum_{i_{l-1}=1}^{N_{l-1}^k} \sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}} W_{l+1,n+1}^{k,i_l,\dots,i_n} \sigma_l(1) 1_{\mathcal{RSI}_{l,n}^{k,i_{l-1},i_l,\dots,i_n}}\right\}_{l,k,n,N} \tag{7.31}$$

$$\overset{D}{=} \left\{\sum_{i_{l-1}=1}^{\mathcal{N}_{l-1}^k} \sum_{i_l=\overline{\mathcal{N}}_l^{k,i_{l-1}-1}+1}^{\overline{\mathcal{N}}_l^{k,i_{l-1}}} \mathcal{W}_{l+1,n+1}^{k,i_l,\dots,i_n} \sigma_l(1) 1_{\mathcal{H}_l^{k,i_{l-1}}\left(\mathcal{H}_{l+1}^{k,i_l}\right)^C \cdots \left(\mathcal{H}_n^{k,i_{n-1}}\right)^C}\right\}_{l,k,n,N}$$

where

$$W_{l+1,n+1}^{k,i_l,\dots,i_n} = W_{n+1}^{k,i_n} \cdots W_{l+2}^{k,i_{l+1}} W_{l+1}^{k,i_l} \tag{7.32}$$

$$\mathcal{W}_{l+1,n+1}^{k,i_l,\dots,i_n} = \mathcal{W}_{n+1}^{k,i_n} \cdots \mathcal{W}_{l+2}^{k,i_{l+1}} \mathcal{W}_{l+1}^{k,i_l} \tag{7.33}$$

$$W_l^{k,i} = \alpha_l(X_{l-1}^{k,i}) \text{ and } \mathcal{W}_l^{k,i} = \alpha_l(\mathcal{X}_{l-1}^{k,i}). \tag{7.34}$$

Moreover, there are $\epsilon_n, C_n > 0$ and $\mathbb{D}_n^N \in \sigma\left\{\sum_{k=1}^{m_N} |\mathbb{I}_l^k|,\ l \le n\right\}$, such that $\mathbb{D}_{n+1}^N \subset \mathbb{D}_n^N$,

$$Q^Y(\mathbb{D}_n^N) \ge 1 - 2ne^{-\epsilon_n m_N} \tag{7.35}$$

$$\max_{i\in\{1,\dots,N_l^k\}} \mathbb{K}_l^{k,i} \vee |\mathbb{I}_l^k| \vee \mathbb{A}_l \le C_n \ \forall 1 \le k \le m_N; 0 \le l \le n \text{ on } \mathbb{D}_{n-1}^N, \tag{7.36}$$

$$\max_{i\in\{1,\dots,N_l^k\}} \mathcal{K}_l^{k,i} \vee |\mathcal{I}_l^k| \le C_n \ \forall 1 \le k \le m_N; 0 \le l \le n \text{ on } \Omega \tag{7.37}$$

for all $n = 0, 1, 2...$

Note: For notational simplicity, we take $\mathbb{D}_{-1}^N = \Omega$ in the sequel.

Proof. Suppose (temporarily) the alive multi-indices $\mathbb{M}^A$ were $\kappa \in \mathbb{M}^A$ if $\kappa \in \mathbb{M}$ and either

$$\kappa = \emptyset \text{ or } \kappa = (\kappa_1, \dots, \kappa_n) \text{ with } \kappa_l \in \{1, \dots, \mathbb{N}_{(\kappa_1,\dots,\kappa_{l-1})}^k\} \ \forall l = 1, \dots, n, \tag{7.38}$$

replacing $N^k_{(\kappa_1,...,\kappa_{l-1})}$ with $\mathbb{N}^k_{(\kappa_1,...,\kappa_{l-1})}$ in (7.13), and we defined

$$\mathbb{X}^{k,j}_n = X^k_\kappa,\ \mathbb{U}^{k,j}_n = U^k_\kappa,\ \mathbb{V}^{k,j}_n = V^k_\kappa, \tag{7.39}$$

$$\mathbb{Z}^{k,j,x}_n = Z^{k;x}_\kappa,\ \mathbb{L}^{k,j}_n = \mathbb{L}^k_\kappa,\ \widehat{\mathbb{L}}^{k,j}_n = \widehat{\mathbb{L}}^k_\kappa, \tag{7.40}$$

$$\mathbb{N}^{k,j}_n = \mathbb{N}^k_\kappa,\ \overline{\mathbb{N}}^{k,i}_l = \sum_{m=1}^{i} \mathbb{N}^{k,m}_l \text{ and } \mathbb{N}^k_l = \sum_{m=1}^{N^k_{l-1}} \mathbb{N}^{k,m}_l \text{ with } \mathbb{N}^k_0 = 1, \tag{7.41}$$

where κ is the unique alive multi-index such that $|\kappa| = n$ and

$$j = \eta(\kappa) \mathrel{\hat{=}} \#\{\widehat{\kappa} \in \mathbb{M}^A : \widehat{\kappa} \prec \kappa\} + 1. \tag{7.42}$$

Then, the Residual particle system algorithm is recovered by (7.1-7.3), (7.9), (7.10) with these definitions. Moreover, the process distribution of the Residual estimates $\mathbb{B}^k_n$ and particle numbers $\mathbb{N}^k_n$ do not change if we select from the (independent) $\{Z^{k;x}_\kappa\}$, $\{U^k_\kappa\}$ and $\{V^k_\kappa\}$ differently nor if we add in zero weights and zero offspring numbers. Therefore, examining the equations (7.1-7.42) and concentrating on this Residual particle algorithm, we find

$$\{\mathbb{N}^k_n, \mathbb{B}^k_n\}_{k,n,N} \stackrel{D}{=} \left\{ |\mathbb{I}^k_n|, \sum_{i=1}^{N^k_n} \mathbb{K}^{k,i}_n \delta_{X^{k,i}_n} \right\}_{k,n,N}. \tag{7.43}$$

(7.30,7.31) are handled similarly. (7.35-7.37) now follow from Lemma 4.1, Theorem 5.3 and (7.29,7.30). □
For notational convenience, we define the (exchangeable random) signed measures $\{B^{N,k}_n\}_{k=1}^{m_N}$ and the parent operators (with respect to κ and η defined in (7.42) but with $\mathbb{M}^A$ reset to (7.13)) by:

$$B^{N,k}_n = 1_{\mathbb{D}^N_{n-1}} \sum_{i=1}^{N^k_n} K^{k,i}_n \delta_{X^{k,i}_n} \text{ with } K^{k,i}_n = \mathbb{K}^{k,i}_n - \mathcal{K}^{k,i}_n \tag{7.44}$$

$$p_l(i) = \eta(\kappa_1, ..., \kappa_{l-1}) \text{ when } i = \eta(\kappa_1, ..., \kappa_l) \tag{7.45}$$

$$p_{l,m}(i) = \begin{cases} p_l(\cdots p_{m-1}(p_m(i))) & \text{for } l \le m \\ i & \text{for } l > m \end{cases} \tag{7.46}$$

(so $i \in \{\overline{N}^{k,p_l(i)-1}_l + 1, ..., \overline{N}^{k,p_l(i)}_l\}$). Finally, by the argument in (5.17) there is a $c > 1$ so that

$$W^{k,i}_l \le c\ \forall i, k, l \in \mathbb{N}. \tag{7.47}$$

Now that we have redefined the algorithms on the same (infinite particle system and) probability space $(\Omega, \mathcal{F}, Q^Y)$ (for each fixed Y), we can compare their particle systems.

Theorem 7.2. *Suppose $p \in \mathbb{N}$ as well as the conditions and setting of Theorem 7.1 with all algorithms* *defined on* $(\Omega, \mathcal{F}, Q^Y)$. *Then, there are* $C_n = C_n^{p,Y} > 0$ *such that*

$$E^Y\left[|\mathbb{A}_n^{m_N} - \sigma_n(1)|^p 1_{\mathbb{D}_{n-1}^N}\right] \le C_n m_N^{-\frac{p}{2}}, \tag{7.48}$$

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_n^k\cup\mathcal{I}_n^k}|\mathbb{K}_n^{k,i} - \mathcal{K}_n^{k,i}|\right|^p 1_{\mathbb{D}_{n-1}^N}\right] \le C_n m_N^{-\frac{p}{2}} \tag{7.49}$$

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_n^k\cup\mathcal{I}_n^k}|\widehat{\mathbb{K}}_n^{k,i} - \widehat{\mathcal{K}}_n^{k,i}|\right|^p 1_{\mathbb{D}_{n-1}^N}\right] \le C_n m_N^{-\frac{p}{2}} \qquad \textit{and} \tag{7.50}$$

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}|\underline{\widehat{\mathbb{K}}}_n^k - \underline{\widehat{\mathcal{K}}}_n^k|\right|^p 1_{\mathbb{D}_{n-1}^N}\right] \le C_n m_N^{-\frac{p}{2}} \tag{7.51}$$

for all $m_N = p+1, p+2, \dots$ *and* $n = 1, 2 \dots$, *where* $\mathbb{D}_{n-1}^N$ *is as in Theorem 7.1.*

As these are bounds in N, we highlighted the previously-suppressed N-dependence in $\mathbb{A}_n^{m_N}$. The following lemma is used (with induction) to prove (7.50) implies (7.48) in Theorem 7.2.

Lemma 7.1. *Suppose* $n \in \mathbb{N}_0$ *and* $E^Y\left[|\mathbb{A}_l^{m_N} - \sigma_l(1)|^p 1_{\mathbb{D}_{l-1}^N}\right] \overset{N}{\ll} m_N^{-\frac{p}{2}}$ *for all* $l \le n$. *Then,*

$$E^Y\left[|\mathbb{A}_{n+1}^{m_N} - \sigma_{n+1}(1)|^p 1_{\mathbb{D}_n^N}\right] \overset{N}{\ll} \sum_{j=1}^{n-1} E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}|\underline{\widehat{\mathcal{K}}}_j^k - \underline{\widehat{\mathbb{K}}}_j^k|\right|^p 1_{\mathbb{D}_{j-1}^N}\right]$$
$$+ \quad m_N^{-\frac{p}{2}} + \sum_{j=1}^{n-1} E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_j^k\cup\mathcal{I}_j^k}|\widehat{\mathbb{K}}_j^{k,i} - \widehat{\mathcal{K}}_j^{k,i}|\right|^p 1_{\mathbb{D}_{j-1}^N}\right].$$

The proof of this lemma is involved and hence delayed to Appendix 2.

Proof of Theorem 7.2. Set Up: Using the independence of the V's, letting

$$\mathcal{G}_k^l = \sigma\{V_m^{j,i} : i, j \in \mathbb{N}, m \ne l\} \vee \sigma\{V_l^{j,i} : j \le k, i \in \mathbb{N}\} \vee \mathcal{F}_\infty^{UZ}, \tag{7.52}$$

noting the boundedness of f_V and considering $-\infty \le a_l \le b_l \le \infty$, one has by (7.21,7.23,7.25) that

$$\begin{aligned}
& E^Y\left[\sum_{i\in\mathbb{I}_{l-1}^k\cup\mathcal{I}_{l-1}^k} 1_{\mathbb{R}_l^{k,i}\triangle\mathcal{R}_l^{k,i}}\Big|\mathcal{G}_{k-1}^l\right] && (7.53)\\
\le & \sum_{i\in\mathbb{I}_{l-1}^k\cup\mathcal{I}_{l-1}^k}\left[\left|\int_{a_{l-1}\mathbb{A}_l^{m_N}-\widehat{\mathbb{K}}_{l-1}^{k,i}}^{a_{l-1}\sigma_l(1)-\widehat{\mathcal{K}}_{l-1}^{k,i}} f_V(v)dv\right| + \left|\int_{b_{l-1}\mathbb{A}_l^{m_N}-\widehat{\mathbb{K}}_{l-1}^{k,i}}^{b_{l-1}\sigma_l(1)-\widehat{\mathcal{K}}_{l-1}^{k,i}} f_V(v)dv\right|\right]\\
\overset{N}{\ll} & 1_{(\mathbb{D}_{l-1}^N)^C} + |\mathbb{A}_l^{m_N}-\sigma_l(1)|1_{\mathbb{D}_{l-1}^N} + \sum_{i\in\mathbb{I}_{l-1}^k\cup\mathcal{I}_{l-1}^k}|\widehat{\mathcal{K}}_{l-1}^{k,i}-\widehat{\mathbb{K}}_{l-1}^{k,i}|1_{\mathbb{D}_{l-1}^N}\\
& E^Y\left[\sum_{i\in\mathbb{I}_{l-1}^k\cup\mathcal{I}_{l-1}^k} 1_{\mathbb{S}_l^{k,i}\triangle\mathcal{S}_l^{k,i}}\Big|\mathcal{G}_{k-1}^l\right] && (7.54)\\
\overset{N}{\ll} & 1_{(\mathbb{D}_{l-1}^N)^C} + |\mathbb{A}_l^{m_N}-\sigma_l(1)|1_{\mathbb{D}_{l-1}^N} + \sum_{i\in\mathbb{I}_{l-1}^k\cup\mathcal{I}_{l-1}^k}|\widehat{\mathcal{K}}_{l-1}^{k,i}-\widehat{\mathbb{K}}_{l-1}^{k,i}|1_{\mathbb{D}_{l-1}^N}\\
& E^Y\left[1_{\underline{\mathbb{R}}_l^k\triangle\underline{\mathcal{R}}_l^k}\Big|\mathcal{G}_0^l\right] && (7.55)\\
\overset{N}{\ll} & 1_{(\mathbb{D}_{l-1}^N)^C} + |\mathbb{A}_l^{m_N}-\sigma_l(1)|1_{\mathbb{D}_{l-1}^N} + |\underline{\widehat{\mathcal{K}}}_{l-1}^k-\underline{\widehat{\mathbb{K}}}_{l-1}^k|1_{\mathbb{D}_{l-1}^N}\\
& E^Y\left[1_{\underline{\mathbb{S}}_l^k\triangle\underline{\mathcal{S}}_l^k}\Big|\mathcal{G}_0^l\right] && (7.56)\\
\overset{N}{\ll} & 1_{(\mathbb{D}_{l-1}^N)^C} + |\mathbb{A}_l^{m_N}-\sigma_l(1)|1_{\mathbb{D}_{l-1}^N} + |\underline{\widehat{\mathcal{K}}}_{l-1}^k-\underline{\widehat{\mathbb{K}}}_{l-1}^k|1_{\mathbb{D}_{l-1}^N}
\end{aligned}$$

almost surely for all $l=1,2,...$ Now, set $S_l^{k,j}=\mathbb{S}_l^{k,j}\cap\mathcal{S}_l^{k,j}$, with $\mathbb{S}_l^{k,j}$, $\mathcal{S}_l^{k,j}$ defined in (7.22,7.24). If $i\in\mathbb{I}_n^k\cup\mathcal{I}_n^k$, then either $i\in\mathbb{I}_n^k\triangle\mathcal{I}_n^k$ so there is a time $l\ge 1$ when only one algorithm ancestor was resampled or $i\in\mathbb{I}_n^k\mathcal{I}_n^k$ so the Residual and reduced particles have the same ancestorial chains. Hence, by (7.45,7.46)

$$\begin{aligned}
|\mathbb{K}_n^{k,i}-\mathcal{K}_n^{k,i}| \le \sum_{l=1}^n\Bigg|&\left\{|\mathbb{K}_n^{k,i}|+|\mathcal{K}_n^{k,i}|\right\}1_{\mathbb{R}_l^{k,p_{l,n}(i)}\triangle\mathcal{R}_l^{k,p_{l,n}(i)}} + && (7.57)\\
&1_{S_n^{k,p_{n,n}(i)}S_{n-1}^{k,p_{n-1,n}(i)}\cdots S_{l+1}^{k,p_{l+1,n}(i)}}\left|\prod_{j=l+1}^n W_j^{k,p_{j,n}(i)}[\mathbb{A}_l^{m_N}-\sigma_l(1)]\right|1_{\mathbb{R}_l^{k,p_{l,n}(i)}\mathcal{R}_l^{k,p_{l,n}(i)}}\Bigg|,\\
|\widehat{\mathbb{K}}_n^{k,i}-\widehat{\mathcal{K}}_n^{k,i}| \le \sum_{l=1}^n\Bigg|&\left\{|\widehat{\mathbb{K}}_n^{k,i}|+|\widehat{\mathcal{K}}_n^{k,i}|\right\}1_{\mathbb{R}_l^{k,p_{l,n}(i)}\triangle\mathcal{R}_l^{k,p_{l,n}(i)}} + && (7.58)\\
&1_{S_n^{k,p_{n,n}(i)}S_{n-1}^{k,p_{n-1,n}(i)}\cdots S_{l+1}^{k,p_{l+1,n}(i)}}\left|\prod_{j=l+1}^{n+1} W_j^{k,p_{j,n}(i)}[\mathbb{A}_l^{m_N}-\sigma_l(1)]\right|1_{\mathbb{R}_l^{k,p_{l,n}(i)}\mathcal{R}_l^{k,p_{l,n}(i)}}\Bigg|.
\end{aligned}$$

For the tracking systems, we let $\underline{S}_l^k = (\underline{\mathbb{R}}_l^k)^C \cap (\underline{\mathcal{R}}_l^k)^C$ and find by (7.5,7.6,7.11,7.18) that

$$|\underline{\widehat{\mathbb{K}}}_n^k - \underline{\widehat{\mathcal{K}}}_n^k| \leq \sum_{l=1}^{n} \left\{ |\underline{\widehat{\mathbb{K}}}_n^k| + |\underline{\widehat{\mathcal{K}}}_n^k| \right\} 1_{\underline{\mathbb{R}}_l^k \triangle \underline{\mathcal{R}}_l^k} + \sum_{l=1}^{n} 1_{\underline{S}_n^k \underline{S}_{n-1}^k \cdots \underline{S}_{l+1}^k} \left| \prod_{j=l+1}^{n+1} W_j^{k,1} [\mathbb{A}_l^{m_N} - \sigma_l(1)] \right| 1_{\underline{\mathbb{R}}_l^k \underline{\mathcal{R}}_l^k}. \tag{7.59}$$

Base Case: Clearly, (7.48-7.51) hold with $n = 0$ and $C_0 = 0$, even though this trivial case is not claimed in the theorem statement.
Inductive Step: Suppose

$$E^Y \left[|\mathbb{A}_l^{m_N} - \sigma_l(1)|^p 1_{\mathbb{D}_{l-1}^N} \right] \overset{N}{\ll} m_N^{-\frac{p}{2}} \tag{7.60}$$

$$E^Y \left[\left| \frac{1}{m_N} \sum_{k=1}^{m_N} \sum_{i \in \mathbb{I}_l^k \cup \mathcal{I}_l^k} |\mathbb{K}_l^{k,i} - \mathcal{K}_l^{k,i}| \right|^p 1_{\mathbb{D}_{l-1}^N} \right] \overset{N}{\ll} m_N^{-\frac{p}{2}} \tag{7.61}$$

$$E^Y \left[\left| \frac{1}{m_N} \sum_{k=1}^{m_N} \sum_{i \in \mathbb{I}_l^k \cup \mathcal{I}_l^k} |\widehat{\mathbb{K}}_l^{k,i} - \widehat{\mathcal{K}}_l^{k,i}| \right|^p 1_{\mathbb{D}_{l-1}^N} \right] \overset{N}{\ll} m_N^{-\frac{p}{2}} \tag{7.62}$$

and

$$E^Y \left[\left| \frac{1}{m_N} \sum_{k=1}^{m_N} |\underline{\widehat{\mathbb{K}}}_l^k - \underline{\widehat{\mathcal{K}}}_l^k| \right|^p 1_{\mathbb{D}_{l-1}^N} \right] \overset{N}{\ll} m_N^{-\frac{p}{2}} \tag{7.63}$$

hold for all $l \leq n$, which are true when $n = 0$. Then, it follows by Lemma 7.1 that

$$E^Y \left[|\mathbb{A}_l^{m_N} - \sigma_l(1)|^p 1_{\mathbb{D}_{l-1}^N} \right] \overset{N}{\ll} m_N^{-\frac{p}{2}} \quad \forall\, l \leq n+1. \tag{7.64}$$

Recalling (7.58), noting (7.36), (7.37), (7.47), (7.64) and using exchangeability, one finds that

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_{n+1}^k\cup\mathcal{I}_{n+1}^k}|\widehat{\mathbb{K}}_{n+1}^{k,i}-\widehat{\mathcal{K}}_{n+1}^{k,i}|\right|^p 1_{\mathbb{D}_n^N}\right] \tag{7.65}$$

$$\overset{N}{\ll} E^Y\left[\left|\sum_{j=1}^{n+1}\left\{|\mathbb{A}_j-\sigma_j(1)|+\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}\mathcal{I}_{n+1}^k}1_{\mathbb{R}_j^{k,p_{j,n+1}(i)}\triangle\mathcal{R}_j^{k,p_{j,n+1}(i)}}\right\}\right|^p 1_{\mathbb{D}_n^N}\right]$$

$$\overset{N}{\ll}\sum_{j=1}^{n+1}\left\{E^Y\Big[|\mathbb{A}_j-\sigma_j(1)|^p 1_{\mathbb{D}_{j-1}^N}\Big]+E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}\mathcal{I}_{j-1}^k}1_{\mathbb{R}_j^{k,i}\triangle\mathcal{R}_j^{k,i}}\right|^p 1_{\mathbb{D}_{j-1}^N}\right]\right\}$$

$$\overset{N}{\ll} m_N^{-\frac{p}{2}}+\sum_{\substack{k_1\neq k_2\neq\cdots\neq k_q\\ q\le p}}\sum_{j=1}^{n+1}\frac{E^Y\left[\sum_{i_1}1_{\mathbb{R}\mathcal{R}_j^{k_1,i_1}}\sum_{i_2}1_{\mathbb{R}\mathcal{R}_j^{k_2,i_2}}\cdots\sum_{i_q}1_{\mathbb{R}\mathcal{R}_j^{k_q,i_q}}1_{\mathbb{D}_{j-1}^N}\right]}{m_N^p}$$

$$\overset{N}{\ll} m_N^{-\frac{p}{2}}+\sum_{q=1}^{p}\sum_{j=1}^{n+1}\frac{E^Y\left[\sum_{i_1\in\mathbb{I}\mathcal{I}_{j-1}^1}1_{\mathbb{R}\mathcal{R}_j^{1,i_1}}\sum_{i_2\in\mathbb{I}\mathcal{I}_{j-1}^2}1_{\mathbb{R}\mathcal{R}_j^{2,i_2}}\cdots\sum_{i_q\in\mathbb{I}\mathcal{I}_{j-1}^q}1_{\mathbb{R}\mathcal{R}_j^{q,i_q}}1_{\mathbb{D}_{j-1}^N}\right]}{m_N^{p-q}},$$

where

$$\mathbb{I}\mathcal{I}_{j-1}^q=\mathbb{I}_{j-1}^q\cup\mathcal{I}_{j-1}^q \text{ and } \mathbb{R}\mathcal{R}_j^{k,i}=\mathbb{R}_j^{k,i}\triangle\mathcal{R}_j^{k,i}. \tag{7.66}$$

In exactly the same way, we also get from (7.57), (7.36,7.37), (7.47) and (7.64)

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_{n+1}^k\cup\mathcal{I}_{n+1}^k}|\mathbb{K}_{n+1}^{k,i}-\mathcal{K}_{n+1}^{k,i}|\right|^p 1_{\mathbb{D}_n^N}\right] \tag{7.67}$$

$$\overset{N}{\ll} m_N^{-\frac{p}{2}}+\sum_{q=1}^{p}\sum_{j=1}^{n+1}\frac{E^Y\left[\sum_{i_1\in\mathbb{I}\mathcal{I}_{j-1}^1}1_{\mathbb{R}\mathcal{R}_j^{1,i_1}}\sum_{i_2\in\mathbb{I}\mathcal{I}_{j-1}^2}1_{\mathbb{R}\mathcal{R}_j^{2,i_2}}\cdots\sum_{i_q\in\mathbb{I}\mathcal{I}_{j-1}^q}1_{\mathbb{R}\mathcal{R}_j^{q,i_q}}1_{\mathbb{D}_{j-1}^N}\right]}{m_N^{p-q}}$$

and from (7.59,7.47) and (7.64)

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}|\underline{\widehat{\mathbb{K}}}_{n+1}^k-\underline{\widehat{\mathcal{K}}}_{n+1}^k|\right|^p 1_{\mathbb{D}_n^N}\right] \tag{7.68}$$

$$\overset{N}{\ll}\quad m_N^{-\frac{p}{2}}+\sum_{q=1}^{p}\sum_{j=1}^{n+1}\frac{E^Y\Big[1_{\underline{\mathbb{R}\mathcal{R}}_j^1}1_{\underline{\mathbb{R}\mathcal{R}}_j^2}\cdots 1_{\underline{\mathbb{R}\mathcal{R}}_j^q}1_{\mathbb{D}_{j-1}^N}\Big]}{m_N^{p-q}}$$

where

$$\underline{\mathbb{R}\mathcal{R}}_j^k = \underline{\mathbb{R}}_j^k \triangle \underline{\mathcal{R}}_j^k. \tag{7.69}$$

However, letting $\widehat{\mathbb{K}}\widehat{\mathcal{K}}_j^{k,i} = |\widehat{\mathbb{K}}_j^{k,i} - \widehat{\mathcal{K}}_j^{k,i}|$, we find by (7.53,7.36,7.37), exchangeability and (7.64) that

$$\begin{aligned}
& E^Y\left[\sum_{i_1\in\mathbb{I}_{j-1}^1\cup\mathcal{I}_{j-1}^1} 1_{\mathbb{R}\mathcal{R}_j^{1,i_1}} \sum_{i_2\in\mathbb{I}_{j-1}^2\cup\mathcal{I}_{j-1}^2} 1_{\mathbb{R}\mathcal{R}_j^{2,i_2}} \cdots \sum_{i_q\in\mathbb{I}_{j-1}^q\cup\mathcal{I}_{j-1}^q} 1_{\mathbb{R}\mathcal{R}_j^{q,i_q}} 1_{\mathbb{D}_{j-1}^N}\right] \qquad (7.70) \\
\overset{N}{\ll}\ & E^Y\left[\left(|\mathbb{A}_j - \sigma_j(1)| + \sum_{i_1} \widehat{\mathbb{K}}\widehat{\mathcal{K}}_{j-1}^{1,i_1}\right)\cdots\left(|\mathbb{A}_j - \sigma_j(1)| + \sum_{i_q}\widehat{\mathbb{K}}\widehat{\mathcal{K}}_{j-1}^{q,i_q}\right) 1_{\mathbb{D}_{j-1}^N}\right] \\
\overset{N}{\ll}\ & \sum_{r=0}^q E^Y\left[|\mathbb{A}_j - \sigma_j(1)|^{q-r} \sum_{i_1}\widehat{\mathbb{K}}\widehat{\mathcal{K}}_{j-1}^{1,i_1} \sum_{i_2}\widehat{\mathbb{K}}\widehat{\mathcal{K}}_{j-1}^{2,i_2}\cdots\sum_{i_r}\widehat{\mathbb{K}}\widehat{\mathcal{K}}_{j-1}^{r,i_r}\ 1_{\mathbb{D}_{j-1}^N}\right] \\
\overset{N}{\ll}\ & \sum_{r=0}^q E^Y\left[|\mathbb{A}_j - \sigma_j(1)|^{q-r}\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_{j-1}^k\cup\mathcal{I}_{j-1}^k}\widehat{\mathbb{K}}\widehat{\mathcal{K}}_{j-1}^{k,i}\right|^r 1_{\mathbb{D}_{j-1}^N}\right] \\
\overset{N}{\ll}\ & m_N^{-\frac{q}{2}} + E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_{j-1}^k\cup\mathcal{I}_{j-1}^k}|\widehat{\mathbb{K}}_{j-1}^{k,i} - \widehat{\mathcal{K}}_{j-1}^{k,i}|\right|^q 1_{\mathbb{D}_{j-1}^N}\right].
\end{aligned}$$

Substituting (7.70,7.62) into (7.65) and using Hölder's inequality, we find

$$\begin{aligned}
& E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_{n+1}^k\cup\mathcal{I}_{n+1}^k}|\widehat{\mathbb{K}}_{n+1}^{k,i} - \widehat{\mathcal{K}}_{n+1}^{k,i}|\right|^p 1_{\mathbb{D}_n^N}\right] \qquad (7.71) \\
\overset{N}{\ll}\quad & m_N^{-\frac{p}{2}} + \sum_{q=1}^p\sum_{j=1}^{n+1}\frac{m_N^{-\frac{q}{2}} + \left(E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_i|\widehat{\mathbb{K}}_{j-1}^{k,i} - \widehat{\mathcal{K}}_{j-1}^{k,i}|\right|^p 1_{\mathbb{D}_{j-2}^N}\right]\right)^{\frac{q}{p}}}{m_N^{p-q}} \\
\overset{N}{\ll}\quad & m_N^{-\frac{p}{2}}
\end{aligned}$$

so

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_l^k\cup\mathcal{I}_l^k}|\widehat{\mathbb{K}}_l^{k,i} - \widehat{\mathcal{K}}_l^{k,i}|\right|^p 1_{\mathbb{D}_{l-1}^N}\right] \overset{N}{\ll} m_N^{-\frac{p}{2}} \quad \forall l \le n+1. \tag{7.72}$$

Similarly, replacing (7.65) with (7.67), we have

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_l^k\cup\mathcal{I}_l^k}|\mathbb{K}_l^{k,i} - \mathcal{K}_l^{k,i}|\right|^p 1_{\mathbb{D}_{l-1}^N}\right] \overset{N}{\ll} m_N^{-\frac{p}{2}} \quad \forall l \le n+1. \tag{7.73}$$

Turning to the tracking system and following (7.70), we find by (7.55,7.64) and exchangeability that

$$\begin{aligned} & E^Y\Big[1_{\underline{\mathbb{R}\mathcal{R}}_j^1} 1_{\underline{\mathbb{R}\mathcal{R}}_j^2} \cdots 1_{\underline{\mathbb{R}\mathcal{R}}_j^q} 1_{\mathbb{D}_{j-1}^N}\Big] \\ \overset{N}{\ll} \;& m_N^{-\frac{q}{2}} + E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_i |\underline{\widehat{\mathbb{K}}}_{j-1}^k - \underline{\widehat{\mathcal{K}}}_{j-1}^k|\right|^q 1_{\mathbb{D}_{j-1}^N}\right] \end{aligned} \tag{7.74}$$

so by (7.68) and the method of (7.71-7.72) one has that

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}|\underline{\widehat{\mathbb{K}}}_l^k - \underline{\widehat{\mathcal{K}}}_l^k|\right|^p 1_{\mathbb{D}_{l-1}^N}\right] \overset{N}{\ll} m_N^{-\frac{p}{2}} \quad \forall l \le n+1. \tag{7.75}$$

□

With this coupling and prior preliminary results, we can establish our main result.

Proof of Theorem 5.1.

We can work directly on the coupled algorithms by (7.29,7.30).

Mllns: Taking $p > \frac{2}{1-2\beta}$, we then find by (7.44), Theorem 7.2 and Fubini's theorem that

$$\begin{aligned} & E^Y\left[\sum_{N=1}^\infty \left|\frac{N^\beta}{m_N}\sum_{k=1}^{m_N} B_n^{N,k}(f)\right|^p\right] \\ \ll \;& \sum_{N=1}^\infty E^Y\left[\left|\frac{N^\beta}{m_N}\sum_{k=1}^{m_N}\sum_i |\mathbb{K}_n^{k,i} - \mathcal{K}_n^{k,i}|\right|^p 1_{\mathbb{D}_{n-1}^N}\right] \\ \ll \;& \sum_{N=1}^\infty N^{p\beta} m_N^{-\frac{p}{2}} \ll \sum_{N=1}^\infty N^{(\beta-\frac{1}{2})p} < \infty \end{aligned} \tag{7.76}$$

and it follows by N^{th}-term divergence that

$$\sum_{N=1}^\infty \left|\frac{N^\beta}{m_N}\sum_{k=1}^{m_N} B_n^{N,k}(f)\right|^p < \infty \Rightarrow \frac{N^\beta}{m_N}\sum_{k=1}^{m_N} B_n^{N,k}(f) \to 0 \text{ a.s. } [Q^Y]. \tag{7.77}$$

Moreover, by Borel-Cantelli and (7.35)

$$\begin{aligned} \sum_{N=1}^\infty Q((\mathbb{D}_{n-1}^N)^C) \;&\le\; \sum_{N=1}^\infty 2(n-1)e^{-\epsilon_{n-1}m_N} < \infty \\ &\Rightarrow\; Q((\mathbb{D}_{n-1}^N)^C \text{ i.o.}) = 0. \end{aligned} \tag{7.78}$$

Finally, we know

$$\frac{1}{m_N}\sum_{k=1}^{m_N}\mathcal{K}_n^{k,i} f(\mathcal{X}_n^{k,i}) \overset{N}{\ll} N^{-\beta} \text{ a.s. } [Q^Y]. \tag{7.79}$$

by (4.1) as well as Theorems 4.1 and 7.1 so this part follows by (7.44).

slln: This part follows from the Mllns, using the same $\{f_i\} \subset C(E)$ as in the proof of Theorem 4.1.

Establish i),ii) of Theorem 5.2: It follows by (7.44) and (7.36,7.37) that

$$E^Y|B_n^{N,1}(f)|^2 \overset{N}{\ll} E^Y\left[\sum_{i\in\mathbb{I}_n^1\cup\mathcal{I}_n^1}|\mathbb{K}_n^{1,i}-\mathcal{K}_n^{1,i}|^2 1_{\mathbb{D}_{n-1}^N}\right] \overset{N}{\ll} E^Y\left[\sum_{i\in\mathbb{I}_n^1\cup\mathcal{I}_n^1}|\mathbb{K}_n^{1,i}-\mathcal{K}_n^{1,i}|1_{\mathbb{D}_{n-1}^N}\right] \tag{7.80}$$

for $f\in\overline{C}(E)_+$. Hence, by (7.80), exchangeability and (7.49) with $p=1$

$$N^{\frac{1}{2}}E^Y|B_n^{N,1}(f)|^2 \overset{N}{\ll} \left(\frac{N}{m_N}\right)^{\frac{1}{2}}\to 0. \tag{7.81}$$

and Theorem 5.2 i) is true with $\rho=\frac{1}{2}$ and $\psi_{N,k}=B_n^{N,k}(f)$.
It follows by (7.44) and (7.49) with $p=2$ that

$$\begin{aligned} & NE^Y|B_n^{N,1}(f)B_n^{N,2}(f)| \\ \le\ & N|f|_\infty^2 E^Y\left|\sum_{i\in\mathbb{I}_n^1\cup\mathcal{I}_n^1}\sum_{j\in\mathbb{I}_n^2\cup\mathcal{I}_n^2}|\mathbb{K}_n^{1,i}-\mathcal{K}_n^{1,i}||\mathbb{K}_n^{2,j}-\mathcal{K}_n^{2,j}|1_{\mathbb{D}_{n-1}^N}\right| \\ \le\ & \frac{N}{m_N(m_N-1)}|f|_\infty^2 E^Y\left[\left|\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_n^k\cup\mathcal{I}_n^k}|\mathbb{K}_n^{k,i}-\mathcal{K}_n^{k,i}|\right|^2 1_{\mathbb{D}_{n-1}^N}\right] \\ \overset{N}{\ll}\ & \frac{N}{m_N}\to 0. \end{aligned} \tag{7.82}$$

Apply Exchangeability Result, Reduced System clt:
$\frac{1}{N^{\frac{1}{2}}}\sum_{k=1}^{N}B_n^{N,k}(f)\to^P 0$ by (7.82), (7.81) and Theorem 5.2 with $\rho=\frac{1}{2}$, with $E=E^Y$ and $\psi_{N,k}=B_n^{N,k}(f)$.
Therefore, it follows by (7.44), (7.35) that for any $\epsilon>0$

$$\begin{aligned} & Q^Y(N^{-\frac{1}{2}}\sum_{k=1}^{N}\sum_i(\mathbb{K}_n^{k,i}f(X_n^{k,i})-\mathcal{K}_n^{k,i}f(X_n^{k,i}))>\epsilon) \\ \le\ & Q^Y(N^{-\frac{1}{2}}\sum_{k=1}^{N}B_n^{N,k}(f)>\epsilon)+Q^Y((\mathbb{D}_{n-1}^N)^C)\to 0. \end{aligned} \tag{7.83}$$

The clt in Theorem 5.1 now follows from the clt in Theorem 4.1 and Theorem 7.1. □

8. Appendix I: Proof of Proposition 4.1, variance calculation

Abbreviating $M_n^k=M_n^{\mathcal{B}^k}(f)$, one notes from (4.11) and (4.8) that

$$M_n^k = \sum_{l=0}^{n}\left[\mathcal{B}_l^k\left(A_{l+1,n}f\right)-\mathcal{B}_{l-1}^k\left(A_{l,n}f\right)\right], \tag{8.1}$$

where $\mathcal{B}^k_{-1}=\pi_0$. The variance of the '$l=0$' term is

$$E^Y\left|\mathcal{B}^k_0(A_{1,n}f)-\mathcal{B}^k_{-1}(A_{0,n}f)\right|^2 = E^Y\left|(A_{1,n}f(\chi^k))-\pi_0(A_{1,n}f)\right|^2 = \pi_0\left((A_{1,n}f)^2\right)-\left(\pi_0(A_{1,n}f)\right)^2. \tag{8.2}$$

The martingale differences for $l\geq 1$ are by (4.1), (4.4), (4.2), (3.2), Step 4 of the reduced algorithm and (4.7)

$$\begin{aligned}
&\mathcal{B}^k_l(A_{l+1,n}f)-\mathcal{B}^k_{l-1}(A_{l,n}f) \\
&= \sum_{i=1}^{\mathcal{N}^k_{l-1}}\left\{\sum_{j=\overline{\mathcal{N}}^{k,i-1}_l+1}^{\overline{\mathcal{N}}^{k,i}_l}\mathcal{L}^{k,j}_l A_{l+1,n}f(\mathcal{X}^{k,j}_l)-\widehat{\mathcal{L}}^{k,i}_{l-1}KA_{l+1,n}f(\mathcal{X}^{k,i}_{l-1})\right\} \\
&= \sum_{i=1}^{\mathcal{N}^k_{l-1}}\overline{\mathcal{L}}^{k,i}_{l-1}\left\{\sum_{j=\overline{\mathcal{N}}^{k,i-1}_l+1}^{\overline{\mathcal{N}}^{k,i}_l}A_{l+1,n}f(\mathcal{X}^{k,j}_l)\right. \\
&\qquad\left.-E^Y\left[\sum_{j=\overline{\mathcal{N}}^{k,i-1}_l+1}^{\overline{\mathcal{N}}^{k,i}_l}A_{l+1,n}f(\mathcal{X}^{k,j}_l)\,\middle|\,\mathcal{F}^{\mathcal{UX}}_{l-1}\vee\mathcal{F}^{\mathcal{V}}_l\right]\right\}.
\end{aligned} \tag{8.3}$$

Therefore, by the independence of the $\{\mathcal{U},\mathcal{V},\mathcal{Z}\}$

$$\begin{aligned}
&E^Y\left[(\mathcal{B}^k_l(A_{l+1,n}f)-\mathcal{B}^k_{l-1}(A_{l,n}f))^2\,\middle|\,\mathcal{F}^{\mathcal{UX}}_{l-1}\vee\mathcal{F}^{\mathcal{V}}_l\right] \\
&= \sum_{i_1,i_2=1}^{\mathcal{N}^k_{l-1}}\overline{\mathcal{L}}^{k,i_1}_{l-1}\overline{\mathcal{L}}^{k,i_2}_{l-1} \\
&\left\{E^Y\left[\sum_{j_1=\overline{\mathcal{N}}^{k,i_1-1}_l+1}^{\overline{\mathcal{N}}^{k,i_1}_l}A_{l+1,n}f(\mathcal{X}^{k,j_1}_l)\sum_{j_2=\overline{\mathcal{N}}^{k,i_2-1}_l+1}^{\overline{\mathcal{N}}^{k,i_2}_l}A_{l+1,n}f(\mathcal{X}^{k,j_2}_l)\middle|\mathcal{F}^{\mathcal{UX}}_{l-1}\vee\mathcal{F}^{\mathcal{V}}_l\right]\right. \\
&- E^Y\left[\sum_{j_1=\overline{\mathcal{N}}^{k,i_1-1}_l+1}^{\overline{\mathcal{N}}^{k,i_1}_l}A_{l+1,n}f(\mathcal{X}^{k,j_1}_l)\middle|\mathcal{F}^{\mathcal{UX}}_{l-1}\vee\mathcal{F}^{\mathcal{V}}_l\right] \\
&\quad\left.\times E^Y\left[\sum_{j_2=\overline{\mathcal{N}}^{k,i_2-1}_l+1}^{\overline{\mathcal{N}}^{k,i_2}_l}A_{l+1,n}f(\mathcal{X}^{k,j_2}_l)\middle|\mathcal{F}^{\mathcal{UX}}_{l-1}\vee\mathcal{F}^{\mathcal{V}}_l\right]\right\} \\
&= \sum_{i=1}^{\mathcal{N}^k_{l-1}}\left|\overline{\mathcal{L}}^{k,i}_{l-1}\right|^2\left\{E^Y\left[\left|\sum_{j=\overline{\mathcal{N}}^{k,i-1}_l+1}^{\overline{\mathcal{N}}^{k,i}_l}A_{l+1,n}f(\mathcal{X}^{k,j}_l)\right|^2\middle|\mathcal{F}^{\mathcal{UX}}_{l-1}\vee\mathcal{F}^{\mathcal{V}}_l\right]\right. \\
&- \left.\left|E^Y\left[\sum_{j=\overline{\mathcal{N}}^{k,i-1}_l+1}^{\overline{\mathcal{N}}^{k,i}_l}A_{l+1,n}f(\mathcal{X}^{k,j}_l)\middle|\mathcal{F}^{\mathcal{UX}}_{l-1}\vee\mathcal{F}^{\mathcal{V}}_l\right]\right|^2\right\}.
\end{aligned} \tag{8.4}$$

However, by the independence of the $\{\mathcal{U},\mathcal{V},\mathcal{Z}\}$ again as well as (4.6)

$$\begin{aligned} & E^Y\left[\left(\sum_{j=\overline{\mathcal{N}}_l^{k,i-1}+1}^{\overline{\mathcal{N}}_l^{k,i}} A_{l+1,n}f(\mathcal{X}_l^{k,j})\right)^2 \big|\mathcal{F}_{l-1}^{\mathcal{U}\mathcal{X}}\vee\mathcal{F}_l^{\mathcal{V}}\right] \qquad (8.5) \\ = & \ E^Y\Big[\mathcal{N}_l^{k,i}\Big\{K(A_{l+1,n}f)^2(\mathcal{X}_{l-1}^{k,i})-(KA_{l+1,n}f)^2(\mathcal{X}_{l-1}^{k,i})\Big\} \\ + & \ \Big(\mathcal{N}_l^{k,i}KA_{l+1,n}f\Big)^2(\mathcal{X}_{l-1}^{k,i})\,\big|\mathcal{F}_{l-1}^{\mathcal{U}\mathcal{X}}\vee\mathcal{F}_l^{\mathcal{V}}\Big] \\ = & \ \frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}\Big\{K(A_{l+1,n}f)^2(\mathcal{X}_{l-1}^{k,i})-(KA_{l+1,n}f)^2(\mathcal{X}_{l-1}^{k,i})\Big\} \\ + & \left\{\left|\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}\right|^2+\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}-\left\lfloor\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}\right\rfloor-\left|\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}-\left\lfloor\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}\right\rfloor\right|^2\right\}(KA_{l+1,n}f)^2(\mathcal{X}_{l-1}^{k,i}), \end{aligned}$$

since

$$E^Y\Big[|\mathcal{N}_l^{k,i}|^2\,\big|\mathcal{F}_{l-1}^{\mathcal{U}\mathcal{X}}\vee\mathcal{F}_l^{\mathcal{V}}\Big]=\left\lfloor\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}\right\rfloor^2+\left|2\left\lfloor\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}\right\rfloor+1\right|Q^Y\!\left(U_l^{k,i}\le\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}-\left\lfloor\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}\right\rfloor\right),$$

and

$$\begin{aligned} & \left|E^Y\left[\sum_{j=\overline{\mathcal{N}}_l^{k,i-1}+1}^{\overline{\mathcal{N}}_l^{k,i}} A_{l+1,n}f(\mathcal{X}_l^{k,j})\,\big|\mathcal{F}_{l-1}^{\mathcal{U}\mathcal{X}}\vee\mathcal{F}_l^{\mathcal{V}}\right]\right|^2 \qquad (8.6) \\ = & \ \left|E^Y\Big[\mathcal{N}_l^{k,i}KA_{l+1,n}f(\mathcal{X}_{l-1}^{k,i})\,\big|\mathcal{F}_{l-1}^{\mathcal{U}\mathcal{X}}\vee\mathcal{F}_l^{\mathcal{V}}\Big]\right|^2 \\ = & \ \left|\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}\right|^2(KA_{l+1,n}f)^2(\mathcal{X}_{l-1}^{k,i}). \end{aligned}$$

Combining the last three equations, letting $f_{l,n}=A_l(A_{l+1,n}f)^2-\alpha_l(KA_{l+1,n}f)^2$, breaking over the resample and non-resample cases, and averaging over the $\mathcal{V}_l^{k,i}$, one finds by (4.17,4.18,4.2,4.3) that

$$
\begin{aligned}
& E^Y\left[(\mathcal{B}_l^k(A_{l+1,n}f)-\mathcal{B}_{l-1}^k(A_{l,n}f))^2\,\middle|\mathcal{F}_{l-1}^{\mathcal{UVX}}\right] \qquad (8.7)\\
= \;& E^Y\left[\sum_{i=1}^{\mathcal{N}_{l-1}^k}\overline{\mathcal{L}}_{l-1}^{k,i}\widehat{\mathcal{L}}_{l-1}^{k,i}\left\{K(A_{l+1,n}f)^2(\mathcal{X}_{l-1}^{k,i})-(KA_{l+1,n}f)^2(\mathcal{X}_{l-1}^{k,i})\right\}\right.\\
+ \;& \left.\sum_{i=1}^{\mathcal{N}_{l-1}^k}\left(\overline{\mathcal{L}}_{l-1}^{k,i}\right)^2 r\left(\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\overline{\mathcal{L}}_{l-1}^{k,i}}\right)(KA_{l+1,n}f)^2(\mathcal{X}_{l-1}^{k,i})\,\middle|\mathcal{F}_{l-1}^{\mathcal{UVX}}\right]\\
= \;& \sigma_l(1)\sum_{i=1}^{\mathcal{N}_{l-1}^k}\mathcal{L}_{l-1}^{k,i}\overline{\nu}_l(\alpha_l(\mathcal{X}_{l-1}^{k,i})\mathcal{L}_{l-1}^{k,i})\left\{f_{l,n}(\mathcal{X}_{l-1}^{k,i})\right\}\\
+ \;& \sigma_l^2(1)\sum_{i=1}^{\mathcal{N}_{l-1}^k}\overline{\nu}_l(\alpha_l(\mathcal{X}_{l-1}^{k,i})\mathcal{L}_{l-1}^{k,i})r\left(\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\sigma_l(1)}\right)(KA_{l+1,n}f)^2(\mathcal{X}_{l-1}^{k,i})\\
+ \;& \sum_{i=1}^{\mathcal{N}_{l-1}^k}\left(\mathcal{L}_{l-1}^{k,i}\right)^2\alpha_l(\mathcal{X}_{l-1}^{k,i})\nu_l(\alpha_l(\mathcal{X}_{l-1}^{k,i})\mathcal{L}_{l-1}^{k,i})\left\{f_{l,n}(\mathcal{X}_{l-1}^{k,i})\right\}
\end{aligned}
$$

since $r\left(\frac{\widehat{\mathcal{L}}_{l-1}^{k,i}}{\widehat{\mathcal{L}}_{l-1}^{k,i}}\right)=0$. Now, in the case '$l=1$' we have $\mathcal{L}_{l-1}^{k,i}=1=\mathcal{N}_{l-1}^k$ and

$$
\begin{aligned}
& E^Y\left[(\mathcal{B}_1^k(A_{2,n}f)-\mathcal{B}_0^k(A_{l,n}f))^2\right] \qquad (8.8)\\
= \;& \sigma_1(1)E^Y[\overline{\nu}_1(\alpha_1(\chi^k))\left\{A_1(A_{2,n}f)^2(\chi^k)-\alpha_1(\chi^k)(KA_{2,n}f)^2(\chi^k)\right\}]\\
+ \;& \sigma_1^2(1)E^Y[\overline{\nu}_1(\alpha_1(\chi^k))r\left(\frac{\alpha_1(\chi^k)}{\sigma_1(1)}\right)(KA_{2,n}f)^2(\chi^k)]\\
+ \;& E^Y[\alpha_1(\chi^k)\nu_1(\alpha_1(\chi^k))\left\{A_1(A_{2,n}f)^2(\chi^k)-\alpha_1(\chi^k)(KA_{2,n}f)^2(\chi^k)\right\}].
\end{aligned}
$$

Moreover, for any $l \geq 2$, $m \in \{1,2,...,l-1\}$ and bounded function ϕ_m, we have by (4.4), (4.6), (3.1), independence and division over resampled and non-resampled cases that

$$\begin{aligned} & E^Y\left[\sum_{i=1}^{\mathcal{N}_m^k} \mathcal{L}_m^{k,i}\phi_m(\mathcal{X}_m^{k,i},\mathcal{L}_m^{k,i})\middle|\mathcal{F}_{m-1}^{\mathcal{UVX}}\right] \\ =& \sum_{j=1}^{\mathcal{N}_{m-1}^k} E^Y\left[\sum_{i=\overline{\mathcal{N}}_m^{k,j-1}+1}^{\overline{\mathcal{N}}_m^{k,j}} \overline{\mathcal{L}}_{m-1}^{k,j}\phi_m(\mathcal{X}_m^{k,i},\overline{\mathcal{L}}_{m-1}^{k,j})\middle|\mathcal{F}_{m-1}^{\mathcal{UVX}}\right] \\ =& \sum_{j=1}^{\mathcal{N}_{m-1}^k} E^Y\left[\sum_{i=\overline{\mathcal{N}}_m^{k,j-1}+1}^{\overline{\mathcal{N}}_m^{k,j}} \overline{\mathcal{L}}_{m-1}^{k,j}K\phi_m(\mathcal{X}_{m-1}^{k,j},\overline{\mathcal{L}}_{m-1}^{k,j})\middle|\mathcal{F}_{m-1}^{\mathcal{UVX}}\right] \\ =& \sum_{j=1}^{\mathcal{N}_{m-1}^k} E^Y\left[\widehat{\mathcal{L}}_{m-1}^{k,j}K\phi_m(\mathcal{X}_{m-1}^{k,j},\overline{\mathcal{L}}_{m-1}^{k,j})\middle|\mathcal{F}_{m-1}^{\mathcal{UVX}}\right] \\ =& \sum_{j=1}^{\mathcal{N}_{m-1}^k} \mathcal{L}_{m-1}^{k,j}E^Y\left[A_m\phi_m(\mathcal{X}_{m-1}^{k,j},\overline{\mathcal{L}}_{m-1}^{k,j})\middle|\mathcal{F}_{m-1}^{\mathcal{UVX}}\right] \\ =& \sum_{j=1}^{\mathcal{N}_{m-1}^k} \mathcal{L}_{m-1}^{k,j}\phi_{m-1}(\mathcal{X}_{m-1}^{k,j},\mathcal{L}_{m-1}^{k,j}), \end{aligned} \tag{8.9}$$

where

$$\begin{aligned} & \phi_{m-1}(\mathcal{X},\mathcal{L}) \\ =& A_m\phi_m(\mathcal{X},\sigma_m(1))\,\overline{\nu}_m(\alpha_m(\mathcal{X})\mathcal{L}) + A_m\phi_m(\mathcal{X},L)\Big|_{L=\alpha_m(\mathcal{X})\mathcal{L}}\nu_m(\alpha_m(\mathcal{X})\mathcal{L}). \end{aligned} \tag{8.10}$$

(8.9) implies that

$$\begin{aligned} & E^Y\left[\sum_{i=1}^{\mathcal{N}_1^k} \mathcal{L}_1^{k,i}\phi_1(\mathcal{X}_1^{k,i},\mathcal{L}_1^{k,i})\right] \\ =& \pi_0\left[A_1\phi_1(\cdot,\sigma_1(1))\overline{\nu}_1(\alpha_1(\cdot)) + A_1\phi_1(\cdot,\alpha_1(\cdot))\nu_1(\alpha_1(\cdot))\right]. \end{aligned} \tag{8.11}$$

Now, recall (4.23) and suppose that

$$\begin{aligned} & E^Y\left[\sum_{i=1}^{\mathcal{N}_{m-1}^k} \mathcal{L}_{m-1}^{k,i}\phi_{m-1}(\mathcal{X}_{m-1}^{k,i},\mathcal{L}_{m-1}^{k,i})\right] \\ =& \sum_{j=0}^{m-1}\sum_{1\le i_1<i_2<\cdots<i_j\le m-1} \pi_0[A_{1,m-1}\phi_{m-1}(\cdot,\alpha_{i_j,m-1})\nu_{i_j,m-1}\overline{\nu}_{i_1,i_2,\ldots,i_j}] \end{aligned} \tag{8.12}$$

for some $m \in \{2, ..., l-1\}$, which is known when $m = 2$ by (8.11) and (4.19,4.20,4.21). (For clarity, the "$j = 0$" term on the right of (8.12) is simply $\pi_0[A_{1,m-1}\phi_{m-1}(\cdot, \alpha_{0,m-1})\nu_{0,m-1}]$.) Then, it follows from (8.9,8.10,8.12) and (4.19,4.20,4.21) by letting $r = j+1$ that

$$
\begin{aligned}
& E^Y\left[\sum_{i=1}^{\mathcal{N}_m^k} \mathcal{L}_m^{k,i}\phi_m(\mathcal{X}_m^{k,i}, \mathcal{L}_m^{k,i})\right] \qquad (8.13) \\
= & \sum_{j=0}^{m-1} \sum_{1\le i_1<i_2<\cdots<i_j\le m-1} \pi_0[A_{1,m}\phi_m(\cdot, \sigma_m(1))\overline{\nu}_{i_j,m}\nu_{i_j,m-1}\overline{\nu}_{i_1,i_2,\dots,i_j}] \\
+ & \sum_{j=0}^{m-1} \sum_{1\le i_1<i_2<\cdots<i_j\le m-1} \pi_0[A_{1,m}\phi_m(\cdot, \alpha_{i_j,m})\nu_{i_j,m}\overline{\nu}_{i_1,i_2,\dots,i_j}] \\
= & \sum_{r=1}^{m} \sum_{\substack{1\le i_1<i_2<\cdots<i_r\le m \\ i_r=m}} \pi_0[A_{1,m}\phi_m(\cdot, \alpha_{i_r,m})\nu_{i_r,m}\overline{\nu}_{i_1,i_2,\dots,i_r}] \\
+ & \sum_{j=0}^{m-1} \sum_{1\le i_1<i_2<\cdots<i_j\le m-1} \pi_0[A_{1,m}\phi_m(\cdot, \alpha_{i_j,m})\nu_{i_j,m}\overline{\nu}_{i_1,i_2,\dots,i_j}] \\
= & \sum_{j=0}^{m} \sum_{1\le i_1<i_2<\cdots<i_j\le m} \pi_0[A_{1,m}\phi_m(\cdot, \alpha_{i_j,m})\nu_{i_j,m}\overline{\nu}_{i_1,i_2,\dots,i_j}].
\end{aligned}
$$

Hence, (8.13) holds for all $m = 1, ..., l-1$ by induction and (4.22) follows by (8.1), (8.2), (8.7) and (8.13), considering the three cases:

$$\phi_{l-1}(\mathcal{X}, \mathcal{L}) = \{f_{l,n}(\mathcal{X})\}\,\overline{\nu}_l(\alpha_l(\mathcal{X})\mathcal{L}) \tag{8.14}$$
$$\phi_{l-1}(\mathcal{X}, \mathcal{L}) = \mathcal{L}\alpha_l(\mathcal{X})\,\{f_{l,n}(\mathcal{X})\}\,\nu_l(\alpha_l(\mathcal{X})\mathcal{L}) \tag{8.15}$$
$$\phi_{l-1}(\mathcal{X}, \mathcal{L}) = \frac{1}{\mathcal{L}}\overline{\nu}_l(\alpha_l(\mathcal{X})\mathcal{L})r\left(\frac{\alpha_l(\mathcal{X})\mathcal{L}}{\sigma_l(1)}\right)(KA_{l+1,n}f)^2(\mathcal{X}), \tag{8.16}$$

where $f_{l,n} = A_l(A_{l+1,n}f)^2 - \alpha_l(KA_{l+1,n}f)^2$. □

9. Appendix 2: Proof of Lemma 7.1

Essentially, we observe that this result would hold trivially for the weighted particle system and then use induction and the coupling to show the necessary differences between the Residual and weighted systems converge appropriately.

Proof. Recall $\mathbb{A}_0^{m_N} = 1$ so $\mathbb{A}_0^{m_N} = \sigma_0(1)$. It follows from (7.19,7.10,7.14,7.15) that for all $n \ge 0$

$$\mathbb{A}_{n+1}^{m_N} = \frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_n^k} \alpha_{n+1}(X_n^{k,i})\mathbb{K}_n^{k,i}. \tag{9.1}$$

Base Case: For notational reasons we consider the case $n=0$ separately. One then finds by (7.1,7.2) that (9.1) reduces to

$$\mathbb{A}_1^{m_N} \quad = \quad \frac{1}{m_N}\sum_{k=1}^{m_N}\alpha_1(\chi^k), \tag{9.2}$$

where $\{\alpha_1(\chi^k)\}_{k=1}^{\infty}$ are i.i.d., bounded and mean $\sigma_1(1)$ with respect to Q^Y. Hence, by the Marcinkiewicz-Zygmund and Jensen inequalities there is a constant $C_p>0$ such that

$$\begin{aligned} E^Y\left|\mathbb{A}_1^{m_N}-\sigma_1(1)\right|^p &\leq \frac{C_p}{m_N^p}E^Y\left|\sum_{k=1}^{m_N}\left(\alpha_1(\chi^k)-\sigma_1(1)\right)^2\right|^{\frac{p}{2}} \\ &\leq \frac{C_p}{m_N^{\frac{p}{2}}}\frac{1}{m_N}\sum_{k=1}^{m_N}E^Y\left|\alpha_1(\chi^k)-\sigma_1(1)\right|^p \overset{N}{\ll} m_N^{-\frac{p}{2}} \end{aligned} \tag{9.3}$$

for any $p\geq 1$.
Case $n\geq 1$**:** It follows from (7.3,7.2,7.15,7.21,7.22,7.45) that $\mathbb{K}_0^{k,1}=1$ and

$$\mathbb{K}_j^{k,i}=\mathbb{A}_j^{m_N}1_{\mathbb{R}_j^{k,p_j(i)}}+W_j^{k,p_j(i)}\mathbb{K}_{j-1}^{k,p_j(i)}1_{\mathbb{S}_j^{k,p_j(i)}}\ \forall i\in\mathbb{I}_j^k, j\in\mathbb{N}. \tag{9.4}$$

Using (9.1,7.34) and (9.4) recursively, one has

$$\begin{aligned} &\mathbb{A}_{n+1}^{m_N} \\ =\ &\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i_{n-1}=1}^{N_{n-1}^k}\sum_{i_n=\overline{N}_n^{k,i_{n-1}-1}+1}^{\overline{N}_n^{k,i_{n-1}}}W_{n+1}^{k,i_n}\mathbb{A}_n^{m_N}1_{\mathbb{RSI}_{n,n}^{k,i_{n-1},i_n}} \\ +\ &\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i_{n-2}=1}^{N_{n-2}^k}\sum_{i_{n-1}=\overline{N}_{n-1}^{k,i_{n-2}-1}+1}^{\overline{N}_{n-1}^{k,i_{n-2}}}W_{n+1}^{k,i_n}W_n^{k,i_{n-1}}\mathbb{A}_{n-1}^{m_N}1_{\mathbb{RSI}_{n-1,n}^{k,i_{n-2},i_{n-1},i_n}} \\ +\ &\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i_{n-3}=1}^{N_{n-3}^k}\sum_{i_{n-2}=\overline{N}_{n-2}^{k,i_{n-3}-1}+1}^{\overline{N}_{n-2}^{k,i_{n-3}}}W_{n+1}^{k,i_n}W_n^{k,i_{n-1}}W_{n-1}^{k,i_{n-2}}\mathbb{A}_{n-2}^{m_N}1_{\mathbb{RSI}_{n-2,n}^{k,i_{n-3},\ldots,i_n}} \\ +\ &\cdots+ \\ +\ &\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i_1=1}^{N_1^k}W_{n+1}^{k,i_n}W_n^{k,i_{n-1}}\cdots W_3^{k,i_2}W_2^{k,i_1}\mathbb{A}_1^{m_N}1_{\mathbb{RSI}_{1,n}^{k,i_0,\ldots,i_n}} \\ +\ &\frac{1}{m_N}\sum_{k=1}^{m_N}W_{n+1}^{k,1}W_n^{k,1}\cdots W_3^{k,1}W_2^{k,1}W_1^{k,1}1_{\mathbb{RSI}_{0,n}^{k,1,\ldots,1}}, \end{aligned} \tag{9.5}$$

where the non-summed indices satisfy $i_0=1$, $i_l=\overline{N}_l^{k,i_{l-1}-1}+1$ (since no resampling). For clarity, here and below $\mathbb{RSI}_{0,n}^{k,1,\ldots,1}=\mathbb{S}_1^{k,1}\mathbb{S}_2^{k,1}\cdots\mathbb{S}_n^{k,1}\{1\in\mathbb{I}_n^k\}$. Noting $\mathbb{S}_1^{k,1}\mathbb{S}_2^{k,1}\cdots\mathbb{S}_n^{k,1}\{1\in\mathbb{I}_n^k\}=(\underline{\mathbb{R}}_1^k)^C(\underline{\mathbb{R}}_2^k)^C\cdots(\underline{\mathbb{R}}_n^k)^C$ (by

(7.21, 7.25, 7.2, 7.3, 7.5, 7.15, 7.18)) and letting

$$\begin{aligned} W_{1,n+1}^{k,1} &\doteq W_{n+1}^{k,1} W_n^{k,1} \cdots W_2^{k,1} W_1^{k,1} \\ &= W_{1,n+1}^{k,1} 1_{\mathbb{RSI}_{0,n}^{k,1,\ldots,1}} + W_{1,n+1}^{k,1} 1_{\underline{\mathbb{R}}_1^k (\underline{\mathbb{R}}_2^k)^C \cdots (\underline{\mathbb{R}}_{n-1}^k)^C (\underline{\mathbb{R}}_n^k)^C} + \cdots \\ &+ W_{1,n+1}^{k,1} 1_{\underline{\mathbb{R}}_{n-2}^k (\underline{\mathbb{R}}_{n-1}^k)^C (\underline{\mathbb{R}}_n^k)^C} + W_{1,n+1}^{k,1} 1_{\underline{\mathbb{R}}_{n-1}^k (\underline{\mathbb{R}}_n^k)^C} + W_{1,n+1}^{k,1} 1_{\underline{\mathbb{R}}_n^k}, \end{aligned} \tag{9.6}$$

we have by (9.5-9.6) that

$$\begin{aligned} & m_N \mathbb{A}_{n+1}^{m_N} - \sigma_{n+1}(1) \\ = \; & \sum_{k=1}^{m_N} \left| W_{1,n+1}^{k,1} - \sigma_{n+1}(1) \right| \\ + \; & \sum_{k=1}^{m_N} \left| \sum_{i_{n-1}=1}^{N_{n-1}^k} \sum_{i_n = \overline{N}_n^{k,i_{n-1}-1}+1}^{\overline{N}_n^{k,i_{n-1}}} W_{n+1}^{k,i_n} \mathbb{A}_n^{m_N} 1_{\mathbb{RSI}_{n,n}^{k,i_{n-1},i_n}} - W_{1,n+1}^{k,1} 1_{\underline{\mathbb{R}}_n^k} \right| \\ + \; & \sum_{k=1}^{m_N} \Bigg| \sum_{i_{n-2}=1}^{N_{n-2}^k} \sum_{i_{n-1} = \overline{N}_{n-1}^{k,i_{n-2}-1}+1}^{\overline{N}_{n-1}^{k,i_{n-2}}} W_{n+1}^{k,i_n} W_n^{k,i_{n-1}} \mathbb{A}_{n-1}^{m_N} 1_{\mathbb{RSI}_{n-1,n}^{k,i_{n-2},i_{n-1},i_n}} \\ & \qquad - W_{1,n+1}^{k,1} 1_{\underline{\mathbb{R}}_{n-1}^k (\underline{\mathbb{R}}_n^k)^C} \Bigg| \\ + \; & \sum_{k=1}^{m_N} \Bigg| \sum_{i_{n-3}=1}^{N_{n-3}^k} \sum_{i_{n-2} = \overline{N}_{n-2}^{k,i_{n-3}-1}+1}^{\overline{N}_{n-2}^{k,i_{n-3}}} W_{n+1}^{k,i_n} W_n^{k,i_{n-1}} W_{n-1}^{k,i_{n-2}} \mathbb{A}_{n-2}^{m_N} 1_{\mathbb{RSI}_{n-2,n}^{k,i_{n-3},\ldots,i_n}} \\ & \qquad - W_{1,n+1}^{k,1} 1_{\underline{\mathbb{R}}_{n-2}^k (\underline{\mathbb{R}}_{n-1}^k)^C (\underline{\mathbb{R}}_n^k)^C} \Bigg| \\ + \; & \cdots + \\ + \; & \sum_{k=1}^{m_N} \Bigg| \sum_{i_1=1}^{N_1^k} W_{n+1}^{k,i_n} W_n^{k,i_{n-1}} \cdots W_3^{k,i_2} W_2^{k,i_1} \mathbb{A}_1^{m_N} 1_{\mathbb{RSI}_{1,n}^{k,i_0,\ldots,i_n}} \\ & \qquad - W_{1,n+1}^{k,1} 1_{\underline{\mathbb{R}}_1^k (\underline{\mathbb{R}}_2^k)^C \cdots (\underline{\mathbb{R}}_n^k)^C} \Bigg|. \end{aligned} \tag{9.7}$$

Now, $\{W_{1,n+1}^{k,1} - \sigma_{n+1}(1)\}_{k=1}^{m_N}$ are i.i.d., zero mean and bounded with respect to Q^Y. Therefore, it follows as above by the Marcinkiewicz-Zygmund and Jensen inequalities that

$$E^Y \left| \frac{1}{m_N} \sum_{k=1}^{m_N} W_{1,n+1}^{k,1} - \sigma_{n+1}(1) \right|^p \overset{N}{\ll} m_N^{-\frac{p}{2}} \tag{9.8}$$

for any $p \geq 1$. Next, we consider a typical (non-first) term in (9.7) in terms of $l \in \{1, ..., n\}$

$$\sum_{i_{l-1}=1}^{N_{l-1}^k} \sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}} W_{n+1}^{k,i_n} \cdots W_{l+2}^{k,i_{l+1}} W_{l+1}^{k,i_l} \mathbb{A}_l^{m_N} 1_{\mathbb{RSI}_{l,n}^{k,i_{l-1},\dots,i_n}} - W_{1,n+1}^{k,1} 1_{\mathbb{R}_l^k(\mathbb{R}_{l+1}^k)^C \cdots (\mathbb{R}_n^k)^C} = \mathbb{T}_1^k + \mathbb{T}_2^k + \mathbb{T}_3^k + \mathbb{T}_4^k, \tag{9.9}$$

where

$$\mathbb{T}_1^k = \sum_{i_{l-1}=1}^{N_{l-1}^k} \sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}} W_{n+1}^{k,i_n} \cdots W_{l+2}^{k,i_{l+1}} W_{l+1}^{k,i_l} \sigma_l(1) 1_{\mathcal{RSI}_{l,n}^{k,i_{l-1},\dots,i_n}} - W_{1,n+1}^{k,1} 1_{\underline{\mathcal{R}}_l^k(\underline{\mathcal{R}}_{l+1}^k)^C \cdots (\underline{\mathcal{R}}_n^k)^C} \tag{9.10}$$

$$\mathbb{T}_2^k = \sum_{i_{l-1}=1}^{N_{l-1}^k} \sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}} W_{n+1}^{k,i_n} \cdots W_{l+2}^{k,i_{l+1}} W_{l+1}^{k,i_l} (\mathbb{A}_l^{m_N} - \sigma_l(1)) 1_{\mathbb{RSI}_{l,n}^{k,i_{l-1},\dots,i_n}} \tag{9.11}$$

$$\mathbb{T}_3^k = \sum_{i_{l-1}=1}^{N_{l-1}^k} \sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}} W_{n+1}^{k,i_n} \cdots W_{l+2}^{k,i_{l+1}} W_{l+1}^{k,i_l} \sigma_l(1) \times \left(1_{\mathbb{RSI}_{l,n}^{k,i_{l-1},\dots,i_n}} - 1_{\mathcal{RSI}_{l,n}^{k,i_{l-1},\dots,i_n}}\right) \tag{9.12}$$

and

$$\mathbb{T}_4^k = W_{1,n+1}^{k,1} 1_{\underline{\mathcal{R}}_l^k(\underline{\mathcal{R}}_{l+1}^k)^C \cdots (\underline{\mathcal{R}}_n^k)^C} - W_{1,n+1}^{k,1} 1_{\mathbb{R}_l^k(\mathbb{R}_{l+1}^k)^C \cdots (\mathbb{R}_n^k)^C}. \tag{9.13}$$

Bound $\mathbb{T}_1$: The sums in $\mathbb{T}_1$ only involve the reduced system so by Theorem 7.1 we can just work in the original (prior to coupling) reduced system setting. Now, recalling $\nu_l, \alpha_{i,m}, \overline{\nu}_{i,m}$ from (4.18,4.19,4.21) and using (7.33), one has by (7.34,5.5), independence, the fact $\sigma_l(1) 1_{\mathcal{H}_l^{k,i_{l-1}}} = \overline{\mathcal{L}}_{l-1}^{k,i_{l-1}} 1_{\mathcal{H}_l^{k,i_{l-1}}}$, (4.6), (4.4) and

(4.1) that

$$E^Y\left[\sum_{i_{l-1}=1}^{\mathcal{N}_{l-1}^k}\sum_{i_l=\overline{\mathcal{N}}_l^{k,i_{l-1}-1}+1}^{\overline{\mathcal{N}}_l^{k,i_{l-1}}}\mathcal{W}_{l+1,n+1}^{k,i_l,\dots,i_n}\sigma_l(1)1_{\mathcal{H}_l^{k,i_{l-1}}\left(\mathcal{H}_{l+1}^{k,i_l}\right)^C\cdots\left(\mathcal{H}_n^{k,i_{n-1}}\right)^C}\middle|\mathcal{F}_{l-1}^{\mathcal{UVX}}\right] \tag{9.14}$$

$$= \sum_{i_{l-1}=1}^{\mathcal{N}_{l-1}^k} E^Y\left[\sigma_l(1)\sum_{i_l=\overline{\mathcal{N}}_l^{k,i_{l-1}-1}+1}^{\overline{\mathcal{N}}_l^{k,i_{l-1}}}(\nu_n\circ\alpha_{l,n}K\alpha_{n+1})(\mathcal{X}_l^{k,i_l},\dots,\mathcal{X}_{n-1}^{k,i_{n-1}})\right.$$

$$\left.\times\,\mathcal{W}_n^{k,i_{n-1}}\cdots\mathcal{W}_{l+2}^{k,i_{l+1}}\mathcal{W}_{l+1}^{k,i_l}1_{\mathcal{H}_l^{k,i_{l-1}}\left(\mathcal{H}_{l+1}^{k,i_l}\right)^C\cdots\left(\mathcal{H}_{n-1}^{k,i_{n-2}}\right)^C}\middle|\mathcal{F}_{l-1}^{\mathcal{UVX}}\right]$$

$$= \sum_{i_{l-1}=1}^{\mathcal{N}_{l-1}^k} E^Y\left[\sigma_l(1)\sum_{i_l=\overline{\mathcal{N}}_l^{k,i_{l-1}-1}+1}^{\overline{\mathcal{N}}_l^{k,i_{l-1}}}\Gamma_{l,n+1}^k(\mathcal{X}_l^{k,i_l})1_{\mathcal{H}_l^{k,i_{l-1}}}\middle|\mathcal{F}_{l-1}^{\mathcal{UVX}}\right]$$

$$= \sum_{i_{l-1}=1}^{\mathcal{N}_{l-1}^k}\widehat{\mathcal{L}}_{l-1}^{k,i_{l-1}}E^Y\left[\Gamma_{l,n+1}^k(\mathcal{X}_l^{k,i_l})1_{\mathcal{H}_l^{k,i_{l-1}}}\middle|\mathcal{F}_{l-1}^{\mathcal{UVX}}\right]$$

$$= \sum_{i_{l-1}=1}^{\mathcal{N}_{l-1}^k}\mathcal{L}_{l-1}^{k,i_{l-1}}\left(\alpha_l\overline{\nu}_l\circ\alpha_lK\Gamma_{l,n+1}^k\right)(\mathcal{X}_{l-1}^{k,i_{l-1}})$$

$$= \mathcal{B}_{l-1}^k\left(\alpha_l\overline{\nu}_l\circ\alpha_lK\Gamma_{l,n+1}^k\right)$$

and by (7.34), (7.26, 7.18, 7.11, 7.6), (4.18,4.19,4.21) that

$$E^Y\left[W_{1,n+1}^{k,1}1_{\underline{\mathcal{R}}_l^k(\underline{\mathcal{R}}_{l+1}^k)^C\cdots(\underline{\mathcal{R}}_n^k)^C}\middle|\mathcal{F}_{l-1}^{UVX}\right] \tag{9.15}$$

$$= W_{1,l}^{k,1}\left(\overline{\nu}_l\circ\alpha_lK\Gamma_{l,n+1}^k\right)(X_{l-1}^{k,1})$$

$$= \prod_{m=1}^{l}\alpha_m(X_{m-1}^{k,1})\left(\overline{\nu}_l\circ\alpha_lK\Gamma_{l,n+1}^k\right)(X_{l-1}^{k,1})$$

for $l=1,\dots,n$, where

$$\Gamma_{l,n+1}^k(x_l) \tag{9.16}$$

$$= \left(\alpha_{l+1}\nu_{l+1}\circ\alpha_{l,l+1}K\left(\alpha_{l+2}\nu_{l+2}\circ\alpha_{l,l+2}\cdots K\left(\alpha_n\nu_n\circ\alpha_{l,n}K\alpha_{n+1}\right)\right)\right)(x_l).$$

Hence, by (7.31), (9.14), (4.1), (4.9), (1.1), (1.2) and (9.15)

$$E^Y\left[\sum_{i_{l-1}=1}^{N_{l-1}^k}\sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}}W_{n+1}^{k,i_n}\cdots W_{l+2}^{k,i_{l+1}}W_{l+1}^{k,i_l}\sigma_l(1)1_{\mathcal{RSI}_{l,n}^{k,i_{l-1},\dots,i_n}}\right] \tag{9.17}$$

$$= \sigma_{l-1}(\alpha_l\overline{\nu}_l\circ\alpha_lK\Gamma_{l,n+1}^k)$$

$$= E^Y\left[W_{1,n+1}^{k,1}1_{\underline{\mathcal{R}}_l^k(\underline{\mathcal{R}}_{l+1}^k)^C\cdots(\underline{\mathcal{R}}_n^k)^C}\right]$$

and $\{\mathbb{T}_1^k\}_{k=1}^{m_N}$, the first terms of (9.9), are i.i.d., bounded (w.r.t. E^Y) and zero mean. Therefore, it follows as above by the Marcinkiewicz-Zygmund and Jensen inequalities that

$$E^Y\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\mathbb{T}_1^k\right|^p \overset{N}{\ll} m_N^{-\frac{p}{2}} \tag{9.18}$$

for any $p\geq 1$.
Bound $\mathbb{T}_2$: One has by the induction hypothesis, (7.47,7.36,7.37) and Jensen's inequality that for any $p\geq 1$

$$E^Y\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\mathbb{T}_2^k 1_{\mathbb{D}_n^N}\right|^p \overset{N}{\ll} E^Y\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\left(\mathbb{A}_l^{m_N}-\sigma_l(1)\right)1_{\mathbb{D}_{l-1}^N}\right|^p \overset{N}{\ll} m_N^{-\frac{p}{2}}. \tag{9.19}$$

Bound $\mathbb{T}_3$: One finds by (7.47) that

$$\begin{aligned}
&\left|\sum_{k=1}^{m_N}\mathbb{T}_3^k\right| 1_{\mathbb{D}_n^N} \qquad (9.20)\\
\overset{N}{\ll}\ & \sum_{k=1}^{m_N}\sum_{i_{l-1}=1}^{N_{l-1}^k}\sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}}\left|1_{\mathbb{RSI}_{l,n}^{k,i_{l-1},\ldots,i_n}}-1_{\mathcal{RSI}_{l,n}^{k,i_{l-1},\ldots,i_n}}\right|1_{\mathbb{D}_{l-1}^N}\\
\leq\ & \sum_{j=l+1}^{n}\sum_{k=1}^{m_N}\sum_{i_{l-1}=1}^{N_{l-1}^k}\sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}} 1_{\mathbb{R}_l^{k,i_{l-1}}\mathbb{S}_{l+1}^{k,i_l}\cdots\mathbb{S}_{j-1}^{k,i_{j-2}}\mathbb{S}\triangle\mathcal{S}_j^{k,i_{j-1}}\mathcal{S}_{j+1}^{k,i_j}\cdots\mathcal{S}_n^{k,i_{n-1}}}1_{\mathbb{D}_{l-1}^N}\\
+\ & \sum_{k=1}^{m_N}\sum_{i_{l-1}=1}^{N_{l-1}^k}\sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}} 1_{\mathbb{R}\triangle\mathcal{R}_l^{k,i_{l-1}}\mathcal{S}_{l+1}^{k,i_l}\cdots\mathcal{S}_n^{k,i_{n-1}}}1_{\mathbb{D}_{l-1}^N},
\end{aligned}$$

where

$$\mathbb{S}\triangle\mathcal{S}_j^{k,i_{j-1}}=\mathbb{S}_j^{k,i_{j-1}}\triangle\mathcal{S}_j^{k,i_{j-1}} \text{ and } \mathbb{R}\triangle\mathcal{R}_j^{k,i_{j-1}}=\mathbb{R}_j^{k,i_{j-1}}\triangle\mathcal{R}_j^{k,i_{j-1}}. \tag{9.21}$$

Recalling $\mathcal{G}_k^j$ from (7.52), one has that

$$\begin{aligned}
&E^Y\left|\sum_{k=1}^{m_N}\sum_{i_{l-1}=1}^{N_{l-1}^k}\sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}} 1_{\mathbb{R}_l^{k,i_{l-1}}\mathbb{S}_{l+1}^{k,i_l}\cdots\mathbb{S}_{j-1}^{k,i_{j-2}}\mathbb{S}\triangle\mathcal{S}_j^{k,i_{j-1}}\mathcal{S}_{j+1}^{k,i_j}\cdots\mathcal{S}_n^{k,i_{n-1}}}\right|^p 1_{\mathbb{D}_{l-1}^N} \qquad (9.22)\\
\overset{N}{\ll}\ & E^Y\left|\sum_{k=1}^{m_N}\sum_{i_{l-1}}\sum_{i_l}E^Y\left[1_{\mathbb{R}_l^{k,i_{l-1}}\mathbb{S}_{l+1}^{k,i_l}\cdots\mathbb{S}_{j-1}^{k,i_{j-2}}\mathbb{S}\triangle\mathcal{S}_j^{k,i_{j-1}}\mathcal{S}_{j+1}^{k,i_j}\cdots\mathcal{S}_n^{k,i_{n-1}}}\middle|\mathcal{G}_{k-1}^j\right]\right|^p 1_{\mathbb{D}_{l-1}^N}\\
+\ & E^Y\left|\sum_{k=1}^{m_N}\Delta_j^k\right|^p,
\end{aligned}$$

for $j=l,...,n$, where

$$\Delta_j^k = \tag{9.23}$$
$$\sum_{i_{l-1}=1}^{N_{l-1}^k}\sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}} 1_{\mathbb{R}_l^{k,i_{l-1}}\mathbb{S}_{l+1}^{k,i_l}\cdots\mathbb{S}_{j-1}^{k,i_{j-2}}\mathbb{S}\triangle\mathcal{S}_j^{k,i_{j-1}}\mathcal{S}_{j+1}^{k,i_j}\cdots\mathcal{S}_n^{k,i_{n-1}}\mathbb{D}_{l-1}^N}$$
$$- \ E^Y\left[\sum_{i_{l-1}=1}^{N_{l-1}^k}\sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}} 1_{\mathbb{R}_l^{k,i_{l-1}}\mathbb{S}_{l+1}^{k,i_l}\cdots\mathbb{S}_{j-1}^{k,i_{j-2}}\mathbb{S}\triangle\mathcal{S}_j^{k,i_{j-1}}\mathcal{S}_{j+1}^{k,i_j}\cdots\mathcal{S}_n^{k,i_{n-1}}\mathbb{D}_{l-1}^N}\middle|\mathcal{G}_{k-1}^j\right]$$

are bounded $\{\mathcal{G}_k^j\}$-martingale differences (in k). Therefore, it follows by the Burkholder-Gundy-Davis inequality and Jensen's inequality as well as exchangeability that

$$E^Y\left|\sum_{k=1}^{m_N}\Delta_j^k\right|^p \ \overset{N}{\ll} \ E^Y\left|\sqrt{\sum_{k=1}^{m_N}(\Delta_j^k)^2}\right|^p \tag{9.24}$$
$$\overset{N}{\ll} \ m_N^{\frac{p}{2}-1}\sum_{k=1}^{m_N}E^Y\left[|\Delta_j^k|^p\right] = m_N^{\frac{p}{2}}E^Y\left[|\Delta_j^1|^p\right] \overset{N}{\ll} m_N^{\frac{p}{2}}$$

for $p\geq 2$. Now, by Hölder's inequality we can take p to be an integer. Moreover, by (7.36,7.37)

$$\sum_{i_{l-1}}\sum_{i_l}E^Y\left[1_{\mathbb{R}_l^{k,i_{l-1}}\mathbb{S}_{l+1}^{k,i_l}\cdots\mathbb{S}_{j-1}^{k,i_{j-2}}\mathbb{S}\triangle\mathcal{S}_j^{k,i_{j-1}}\mathcal{S}_{j+1}^{k,i_j}\cdots\mathcal{S}_n^{k,i_{n-1}}}\middle|\mathcal{G}_{k-1}^j\right]1_{\mathbb{D}_{l-1}^N} \tag{9.25}$$
$$\overset{N}{\ll} \ 1_{(\mathbb{D}_{j-1}^N)^C} + \sum_{i\in\mathbb{I}_{j-1}^k\cup\mathcal{I}_{j-1}^k}E^Y\left[1_{\mathbb{S}_j^{k,i}\triangle\mathcal{S}_j^{k,i}}\middle|\mathcal{G}_{k-1}^j\right]1_{\mathbb{D}_{j-1}^N}$$

and one has by (7.54) that

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_{j-1}^k\cup\mathcal{I}_{j-1}^k}E^Y\left[1_{\mathbb{S}_j^{k,i}\triangle\mathcal{S}_j^{k,i}}\middle|\mathcal{G}_{k-1}^j\right]\right|^p 1_{\mathbb{D}_{j-1}^N}\right] \tag{9.26}$$
$$\overset{N}{\ll} \ E^Y\left[|\mathbb{A}_j^{m_N}-\sigma_j(1)|^p 1_{\mathbb{D}_{j-1}^N}\right] + E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_{j-1}^k\cup\mathcal{I}_{j-1}^k}|\widehat{\mathcal{K}}_{j-1}^{k,i}-\widehat{\mathbb{K}}_{j-1}^{k,i}|\right|^p 1_{\mathbb{D}_{j-1}^N}\right]$$

so by (9.22), (9.24), (9.25), (9.26) and (7.35)

$$E^Y\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i_{l-1}}\sum_{i_l}1_{\mathbb{R}_l^{k,i_{l-1}}\mathbb{S}_{l+1}^{k,i_l}\cdots\mathbb{S}_{j-1}^{k,i_{j-2}}\mathbb{S}\triangle\mathcal{S}_j^{k,i_{j-1}}\mathcal{S}_{j+1}^{k,i_j}\cdots\mathcal{S}_n^{k,i_{n-1}}}\right|^p 1_{\mathbb{D}_{l-1}^N} \tag{9.27}$$

$$\overset{N}{\ll} \quad m_N^{-\frac{p}{2}}+\sum_{j=l+1}^{n}E^Y[|\mathbb{A}_j^{m_N}-\sigma_j(1)|^p 1_{\mathbb{D}_{j-1}^N}]$$

$$+\quad \sum_{j=l+1}^{n}E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_{j-1}^k\cup\mathcal{I}_{j-1}^k}|\widehat{\mathcal{K}}_{j-1}^{k,i}-\widehat{\mathbb{K}}_{j-1}^{k,i}|\right|^p 1_{\mathbb{D}_{j-1}^N}\right].$$

Similarly to (9.22-9.27), one finds that

$$E^Y\left|\sum_{k=1}^{m_N}\sum_{i_{l-1}=1}^{N_{l-1}^k}\sum_{i_l=\overline{N}_l^{k,i_{l-1}-1}+1}^{\overline{N}_l^{k,i_{l-1}}}1_{\mathbb{R}\triangle\mathcal{R}_l^{k,i_{l-1}}\mathcal{S}_{l+1}^{k,i_l}\cdots\mathcal{S}_n^{k,i_{n-1}}}\right|^p 1_{\mathbb{D}_{l-1}^N} \tag{9.28}$$

$$\overset{N}{\ll} \quad m_N^{-\frac{p}{2}}+E^Y[|\mathbb{A}_l^{m_N}-\sigma_l(1)|^p 1_{\mathbb{D}_{l-1}^N}]$$

$$+\quad E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_{l-1}^k\cup\mathcal{I}_{l-1}^k}|\widehat{\mathcal{K}}_{l-1}^{k,i}-\widehat{\mathbb{K}}_{l-1}^{k,i}|\right|^p 1_{\mathbb{D}_{l-1}^N}\right].$$

Therefore, one has by (9.20), (9.27), (9.28) and the lemma hypothesis that

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\mathbb{T}_3^k\right|^p 1_{\mathbb{D}_n^N}\right] \tag{9.29}$$

$$\overset{N}{\ll} \quad m_N^{-\frac{p}{2}}+\sum_{j=l}^{n}E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\sum_{i\in\mathbb{I}_{j-1}^k\cup\mathcal{I}_{j-1}^k}|\widehat{\mathcal{K}}_{j-1}^{k,i}-\widehat{\mathbb{K}}_{j-1}^{k,i}|\right|^p 1_{\mathbb{D}_{j-1}^N}\right]$$

for any $p\geq 1$.
Bound $\mathbb{T}_4$: We find by (7.47) and analogous to (9.20-9.25) that

$$E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\mathbb{T}_4^k\right|^p 1_{\mathbb{D}_n^N}\right] \tag{9.30}$$

$$\overset{N}{\ll} \quad m_N^{-\frac{p}{2}}+E^Y\left[\left|\sum_{j=l}^{n}\frac{1}{m_N}\sum_{k=1}^{m_N}E^Y\left[1_{\underline{\mathbb{R}}_j^k\triangle\underline{\mathcal{R}}_j^k}\middle|\mathcal{G}_0^j\right]\right|^p 1_{\mathbb{D}_{j-1}^N}\right].$$

Hence, by (7.56,7.55) and the lemma hypothesis

$$\begin{aligned} & E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}\mathbb{T}_4^k\right|^p 1_{\mathbb{D}_{l-1}^N}\right] \\ \overset{N}{\ll} \;& m_N^{-\frac{p}{2}}+\sum_{j=l}^n E^Y[|\mathbb{A}_j^{m_N}-\sigma_j(1)|^p 1_{\mathbb{D}_{j-1}^N}] \\ & +\sum_{j=l}^n E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}|\underline{\widehat{\mathcal{K}}}_{j-1}^k-\underline{\widehat{\mathbb{K}}}_{j-1}^k|\right|^p 1_{\mathbb{D}_{j-1}^N}\right] \\ \overset{N}{\ll} \;& m_N^{-\frac{p}{2}}+\sum_{j=l}^n E^Y\left[\left|\frac{1}{m_N}\sum_{k=1}^{m_N}|\underline{\widehat{\mathcal{K}}}_{j-1}^k-\underline{\widehat{\mathbb{K}}}_{j-1}^k|\right|^p 1_{\mathbb{D}_{j-1}^N}\right] \end{aligned} \tag{9.31}$$

for any $p \geq 1$. □

References

[1] Blount, D. and Kouritzin, M. A. (2010). On convergence determining and separating classes of functions. *Stoch. Proc. Appl.* **120**, 1898–1907.

[2] Chopin, N. (2004). Central Limit Theorem for Sequential Monte Carlo Methods and its Application to Bayesian Inference. *The Annals of Statistics* **32** (6), 2385-2411.

[3] Crisan, D. and Lyons, T. (1997). Nonlinear filtering and measure valued processes. *Prob. Theory Related Fields* **109**, 217-244.

[4] Douc, R., Cappé, O. and Moulines, E. (2005). Comparison of resampling schemes for particle filtering. In *Proceedings of the* 4^{th} *International Symposium on Image and Signal Processing and Analysis* 64-69.

[5] Gordon, N., Salmond, D. and Smith, A. F. M. (1993). Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *Proc. Inst. Elect. Eng., F*, **140**, 107-113.

[6] Handschin J.E. (1970). Monte Carlo Techniques for Prediction and Filtering of Non- Linear Stochastic Processes. *Automatica* **6**, 555-563.

[7] Handschin J.E. and Mayne D.Q. (1969). Monte Carlo Techniques to Estimate the Conditional Expectation in Multi-stage Non-linear Filtering. *International Journal of Control* **9**, 547-559.

[8] Kouritzin, M. A. (1996). On the interrelation of almost sure invariance principles for certain stochastic adaptive algorithms and for partial sums of random variables, *J. Theoret. Probab.*, **9**, 811-840.

[9] Kouritzin, M.A. (2015). Microstructure Models with Short-Term Inertia and Stochastic Volatility. *Mathematical Problems in Engineering* **vol. 2015**, Article ID 323475, 17 pages, doi:10.1155/2015/323475.

[10] Kouritzin, M.A. (2016). Residual and Stratified Branching Particle Filters. *Submitted*.

[11] Kouritzin, M.A. and Sun, W. (2005). Rates for Branching Particle Approximations of Continuous-Discrete Filters. *The Annals of Applied Probability* **15**, 2739-2772.

[12] Kouritzin, M.A. and Wang, X. (2016). Parallel Sequential Monte Carlo Methods. *Submitted*.

[13] Kouritzin, M.A. and Zeng, Y. (2005). Bayesian Model Selection via Filtering for a Class of Micro-movement Models of Asset Price. *International Journal of Theoretical and Applied Finance* **8**, 97-122.

[14] Kurtz, T.G., Xiong, J. (1999). Particle representations for a class of nonlinear SPDEs. *Stochastic Process. Appl.* **83**, 103-126.

[15] Kurtz, T.G. and Xiong, J. (2000). Numerical solutions for a class of SPDEs with application to filtering, Stochastics in Finite and Infinite Dimension: In Honor of Gopinath Kallianpur. Edited by T. Hida, R. Karandikar, H. Kunita, B. Rajput, S. Watanabe and J. Xiong. *Trends in Mathematics.* Birkhauser.

[16] McKean, H. P. (1966). A Class of Markov Processes Associated with Nonlinear Parabolic Equations. *Proc. Natl. Acad. Sci. USA* **56** (6): 1907-1911.

[17] Maroulas, V. and Nebenführ, A. (2015). Tracking Rapid Intracellular Movements: A Bayesian Random Set Approach. *The Annals of Applied Statistics* **9** (2): 926-949. DOI: 10.1214/15-AOAS819

[18] Sorrentino, A.; Johansen, A. M.; Aston, J.A.D.; Nichols, T. E. and Kendall, W. S. (2013) Dynamic Filtering of Static Dipoles in Magnetoencephalography. *The Annals of Applied Statistics* **7** (2): 955-988.

[19] Weber, N. C. (1980). A Martingale Approach to Central Limit Theorems for Exchangeable Random Variables. *Journal of Applied Probability* **17**: 662-673.