The Correlated Pseudo-Marginal Method

George Deligiannidis[†], Arnaud Doucet[‡] and Michael K. Pitt^{†*}

July 20, 2017

Abstract

The pseudo-marginal algorithm is a Metropolis–Hastings-type scheme which samples asymptotically from a target probability density when we are only able to estimate unbiasedly an unnormalised version of it. In a Bayesian context, it is a state-of-the-art posterior simulation technique when the likelihood function is intractable but can be estimated unbiasedly using Monte Carlo samples. However, for the performance of this scheme not to degrade as the number T of data points increases, it is typically necessary for the number N of Monte Carlo samples to be proportional to T to control the relative variance of the likelihood ratio estimator appearing in the acceptance probability of this algorithm. The correlated pseudo-marginal algorithm is a modification of the pseudo-marginal method using a likelihood ratio estimator computed using two correlated likelihood estimators. For random effects models, we show under regularity conditions that the parameters of this scheme can be selected such that the relative variance of this likelihood ratio estimator is controlled when N increases sublinearly with T and we provide guidelines on how to optimise the parameters of the algorithm based on a non-standard weak convergence analysis. The efficiency of computations for Bayesian inference relative to the pseudo-marginal method empirically increases with T and is higher than two orders of magnitude in some of our examples.

Keywords: Asymptotic posterior normality; Correlated random numbers; Intractable likelihood; Metropolis—Hastings algorithm; Particle filter; Random effects model; State-space model; Weak convergence.

1 Introduction

Consider a Bayesian model where the likelihood of the observations y is denoted by $p(y \mid \theta)$ and the prior for the parameter $\theta \in \Theta \subseteq \mathbb{R}^d$ admits a density $p(\theta)$ with respect to Lebesgue measure $d\theta$. Then the posterior density of interest is $\pi(\theta) \propto p(y \mid \theta)p(\theta)$. We slightly abuse notation by using the same symbols for distributions and densities.

A standard approach to compute expectations with respect to $\pi\left(\theta\right)$ is to use the Metropolis–Hastings (MH) algorithm to generate an ergodic Markov chain of invariant density $\pi\left(\theta\right)$. Given the current state θ of the Markov chain, one samples at each iteration a candidate θ' which is accepted with a probability which depends on the likelihood ratio $p(y\mid\theta')/p(y\mid\theta)$. For many latent variable models, the likelihood is intractable and it is thus impossible to implement the MH algorithm. In this context, Markov chain Monte Carlo (MCMC) schemes targeting the joint posterior density of the parameter and latent variables are often inefficient as the parameter and latent variables can be strongly correlated under the posterior, or cannot even be used if only forward simulation of the latent variables is feasible; see [28], [31], and [1, Section 2.3] for a detailed discussion.

[†]Department of Mathematics, King's College London, UK.

[‡]Department of Statistics, University of Oxford, UK.

^{*}corresponding author

Contrary to these approaches, the pseudo-marginal (PM) algorithm directly mimicks the MH scheme targeting the marginal $\pi(\theta)$ by substituting an estimator of the likelihood ratio $p(y \mid \theta')/p(y \mid \theta)$ for the true likelihood ratio in the MH acceptance probability [36], [5], [3]. This estimator is obtained by computing a non-negative unbiased estimator of $p(y \mid \theta')$ and dividing it by the estimator of $p(y \mid \theta)$ computed when θ was accepted. This simple yet powerful idea has become very popular as it is often possible to obtain a non-negative unbiased estimator of intractable likelihoods and it provides state-of-the-art performance in many scenarios; see, e.g., [1], [21]. Qualitative convergence results for this procedure have been obtained in [3] and [4].

Assuming that the likelihood estimator is evaluated using importance sampling or particle filters for state-space models with N particles, it has also been shown under various assumptions in [45], [18] and [48] that N should be selected such that the variance of the loglikelihood ratio estimator should take a value between 1.0 and 2.0 in regions of high probability mass to minimise the computational resources necessary to achieve a specific asymptotic variance for a particular PM average. As the number T of data $y = (y_1, ..., y_T)$ increases, this implies that N should increase linearly with T [6, Theorem 1] and the computational cost of PM is thus of order T^2 at each iteration. This can be prohibitive for large datasets.

The reason for this is that the PM algorithm is based on an estimator of $p(y \mid \theta')/p(y \mid \theta)$ obtained by dividing estimators of $p(y \mid \theta')$ and $p(y \mid \theta)$ which are independent given θ and θ' . In contrast, the correlated pseudo-marginal (CPM) method correlates the estimators of $p(y \mid \theta')$ and $p(y \mid \theta)$ so as to reduce the variance of the resulting ratio. Correlation between these estimators is introduced by correlating the auxiliary random variates used to obtain these estimators. Two implementations of this generic idea are detailed. We show how to correlate importance sampling estimators for random effects models and how to correlate particle filter estimators for state-space models using the Hilbert sort procedure proposed in [23].

We study in detail large sample properties of the CPM scheme for random effects models. In this scenario, the loglikelihood ratio estimator based on our correlation scheme is shown to satisfy a conditional Central Limit Theorem (CLT) whenever N grows to infinity sublinearly with T and the Euclidean distance between θ and θ' is of order $1/\sqrt{T}$. When the posterior concentrates towards a normal of standard deviation $1/\sqrt{T}$, this CLT can be used to show that a space-rescaled version of the CPM chain converges weakly to a discrete-time Markov chain on the parameter space. The Integrated Autocorrelation Time (IACT) of the weak limit is not impacted by how fast N goes to infinity with T. However the lower this growth rate is, the more correlated the auxiliary variables need to be to control the variance of this estimator. We provide results suggesting we need N to grow at least at rate \sqrt{T} for the IACT of the original CPM chain to remain finite as $T \to \infty$. We use these results to provide practical guidelines on how to optimise performance of the algorithm for large data sets which are validated experimentally. In our numerical examples on random effects models and state-space models, CPM always outperforms PM and the improvement increases with T from 20 to 50 times when T is a few hundreds to more than 100 times when T is a few thousands.

The rest of the paper is organised as follows. In Section 2, we introduce the CPM algorithm and detail its implementation for random effects and state-space models. In Section 3, we present various CLTs for the loglikelihood estimator and loglikelihood ratio estimators used by PM and CPM. In Section 4, we exploit these results to analyze and optimize the CPM kernel in the large sample regime. We demonstrate experimentally the efficiency of this methodology in Section 5 and discuss various potential extensions in Section 6. All the proofs are given in the Supplementary Material.

2 Metropolis–Hastings and correlated pseudo-marginal schemes

2.1 Metropolis-Hastings algorithm

The transition kernel Q_{EX} of the MH algorithm targeting $\pi\left(\theta\right)$ using a proposal distribution $q\left(\theta, d\theta'\right) = q\left(\theta, \theta'\right) d\theta'$ is given by

$$Q_{\rm EX}(\theta, d\theta') = q(\theta, d\theta') \alpha_{\rm EX}(\theta, \theta') + \{1 - \varrho_{\rm EX}(\theta)\} \delta_{\theta}(d\theta'), \qquad (1)$$

where

$$r_{\text{EX}}(\theta, \theta') = \frac{\pi(\theta')q(\theta', \theta)}{\pi(\theta)q(\theta, \theta')} = \frac{p(y \mid \theta')p(\theta')q(\theta', \theta)}{p(y \mid \theta)p(\theta)q(\theta, \theta')},$$
(2)

and

$$\alpha_{\rm EX}(\theta, \theta') = \min\{1, r_{\rm EX}(\theta, \theta')\}, \quad \varrho_{\rm EX}(\theta) = \int q(\theta, d\theta') \,\alpha_{\rm EX}(\theta, \theta'). \tag{3}$$

Implementing this MH scheme requires being able to evaluate the likelihood ratio $p(y \mid \theta')/p(y \mid \theta)$ appearing in the expression of $r_{\text{ex}}(\theta, \theta')$.

2.2 The correlated pseudo-marginal algorithm

Assume $\widehat{p}(y \mid \theta, U)$ is a non-negative unbiased estimator of the intractable likelihood $p(y \mid \theta)$ when $U \sim m$. Here U corresponds to the \mathcal{U} -valued auxiliary random variables used to obtain the estimator. We assume that m(du) = m(u) du and introduce the joint density $\overline{\pi}(\theta, u)$ on $\Theta \times \mathcal{U}$, where

$$\overline{\pi}(\theta, u) = \pi(\theta) m(u) \,\widehat{p}(y \mid \theta, u) / p(y \mid \theta). \tag{4}$$

As $\widehat{p}(y \mid \theta, U)$ is unbiased, $\overline{\pi}(\theta, u)$ admits $\pi(\theta)$ as marginal density. The CPM algorithm is a MH scheme targeting (4) with proposal density $q(\theta, d\theta') K(u, du')$ where K admits an m-reversible Markov transition density, i.e.

$$m(u) K(u, u') = m(u') K(u', u).$$
 (5)

This yields the acceptance probability

$$\alpha_{Q}\left\{\left(\theta,u\right),\left(\theta',u'\right)\right\} = \min\left\{1, r_{\text{EX}}(\theta,\theta') \frac{\widehat{p}(y\mid\theta',u')/p(y\mid\theta')}{\widehat{p}(y\mid\theta,u)/p(y\mid\theta)}\right\}. \tag{6}$$

Hence, the CPM algorithm admits $\overline{\pi}(\theta, u)$ as an invariant density by construction and its transition kernel Q is given by

$$Q\{(\theta, u), (d\theta', du')\} = q(\theta, d\theta') K(u, du') \alpha_Q\{(\theta, u), (\theta', u')\} + \{1 - \varrho_Q(\theta, u)\} \delta_{(\theta, u)} (d\theta', du'),$$
(7)

where $1 - \varrho_Q(\theta, u)$ is the corresponding rejection probability. For K(u, u') = m(u'), we recover the PM algorithm.

Let $\varphi(z; \mu, \Sigma)$ be the multivariate normal density of argument z, mean μ and covariance matrix Σ and let $X \sim \mathcal{N}(\mu, \Sigma)$ denote a sample from this distribution. Henceforth, we assume the likelihood estimator is computed using $M \geq 1$ standard normal random variables so

$$m(u) = \varphi(u; 0_M, I_M)$$
 and $K_\rho(u, u') = \varphi(u'; \rho u, (1 - \rho^2) I_M)$, (8)

where $\rho \in (-1,1)$, 0_M is the $M \times 1$ vector with zero entries and I_M the $M \times M$ identity matrix. It is straightforward to check that K_ρ is m-reversible. There is no loss of generality to select m as a normal density since inversion techniques can be used to form any random variable of interest¹. In addition the kernel K_ρ has the advantage that it can be regarded as a discretized Ornstein-Uhlenbeck process. This property is exploited to establish the main result of Section 3.

Algorithm 1 summarizes how one samples from $Q\{(\theta, U), \cdot\}$.

Algorithm 1 Correlated Pseudo-Marginal Algorithm

- 1. Sample $\theta' \sim q(\theta, \cdot)$.
- 2. Sample $\varepsilon \sim \mathcal{N}\left(0_M, I_M\right)$ and set $U' = \rho U + \sqrt{1 \rho^2} \varepsilon$.
- 3. Compute the estimator $\widehat{p}(y \mid \theta', U')$ of $p(y \mid \theta')$.
- 4. With probability

$$\alpha_{Q}\left\{\left(\theta, U\right), \left(\theta', U'\right)\right\} = \min\left\{1, \frac{\widehat{p}(y \mid \theta', U')}{\widehat{p}(y \mid \theta, U)} \frac{p(\theta')}{p(\theta)} \frac{q\left(\theta', \theta\right)}{q\left(\theta, \theta'\right)}\right\},\tag{9}$$

output (θ', U') . Otherwise, output (θ, U) .

¹For example, in Section 2.3.2, it is necessary to generate uniform random variates and these may be constructed as $\Phi(u_i)$ where u_i is a scalar element of u and Φ the cumulative distribution function of the standard normal.

Contrary to the PM method corresponding to $\rho = 0$, we need to store the vector u instead of $\hat{p}(y \mid \theta, u)$ to implement the algorithm when $\rho \neq 0$. In the applications considered, this overhead is mild.

The rationale behind the CPM scheme is that if $(\theta, u) \longmapsto \widehat{p}(y \mid \theta, u)$ is a regular enough function and (θ, U) and (θ', U') are "close" enough then we expect the ratio estimator $\widehat{p}(y \mid \theta', U')/\widehat{p}(y \mid \theta, U)$ to have small relative variance and therefore to better mimick the noiseless MH scheme Q_{EX} . In many situations, the posterior $\pi(\theta)$ will be approximately normal for large data sets with a covariance scaling in $1/\sqrt{T}$ so an appropriately scaled MH random walk or autoregressive proposal $q(\theta, d\theta')$ will ensure that θ and θ' are "close". We explain in Section 3 how ρ can be selected as a function of T to ensure that U and U' are "close" enough so that the loglikelihood ratio estimator $\log\{\widehat{p}(y \mid \theta', U')/\widehat{p}(y \mid \theta, U)\}$ satisfies a conditional CLT at stationarity. As alluded to in the introduction, properties of this estimator and in particular its asymptotic distribution and variance at stationarity are critical to our analysis of the CPM scheme in the large sample regime detailed in Section 4.

2.3 Application to latent variable models

2.3.1 Random effects models

Consider the model

$$X_t \stackrel{\text{i.i.d.}}{\sim} f_{\theta}(\cdot), \qquad Y_t | X_t \sim g_{\theta}(\cdot | X_t),$$
 (10)

where $\{X_t; t \geq 1\}$ are \mathbb{R}^k -valued latent variables and $\{Y_t; t \geq 1\}$ are Y-valued observations. For any i < j, let $i: j = \{i, i+1, ..., j\}$. For a realization $Y_{1:T} = y_{1:T}$, the likelihood satisfies

$$p(y_{1:T} \mid \theta) = \prod_{t=1}^{T} p(y_t \mid \theta), \qquad p(y_t \mid \theta) = \int g_{\theta}(y_t \mid x_t) f_{\theta}(x_t) dx_t.$$
 (11)

If the T integrals appearing in (11) are intractable, we can estimate them using importance sampling

$$\widehat{p}(y_{1:T} \mid \theta, U) = \prod_{t=1}^{T} \widehat{p}(y_t \mid \theta, U_t), \qquad \widehat{p}(y_t \mid \theta, U_t) = \frac{1}{N} \sum_{i=1}^{N} \omega(y_t, U_{t,i}; \theta), \qquad (12)$$

where the importance weights $\omega(y, U_{t,i}; \theta)$ are given by

$$\omega(y_t, U_{t,i}; \theta) = \frac{g_{\theta}(y_t \mid X_{t,i}) f_{\theta}(X_{t,i})}{q_{\theta}(X_{t,i} \mid y_t)}, \tag{13}$$

assuming that there exists a deterministic map $\Xi_t : \mathbb{R}^p \times \Theta \to \mathbb{R}^k$ such that $X_{t,i} = \Xi_t(U_{t,i};\theta) \sim q_\theta(\cdot \mid y_t)$ for $U_{t,i} \sim \mathcal{N}\left(0_p, I_p\right)$. In this case, we have $U = (U_1, \dots, U_T)$, $U_t = (U_{t,1}, \dots, U_{t,N})$ so $U \sim \mathcal{N}\left(0_M, I_M\right)$ where M = TNp.

2.3.2 State-space models

Consider a generalization of the model (10) where the latent variables $\{X_t; t \geq 1\}$ now arise from a homogeneous \mathbb{R}^k -valued Markov process of initial density ν_{θ} and Markov transition density f_{θ} , i.e. for $t \geq 1$

$$X_1 \sim \nu_{\theta}, \quad X_{t+1} | X_t \sim f_{\theta}(\cdot | X_t), \qquad Y_t | X_t \sim g_{\theta}(\cdot | X_t).$$
 (14)

For a realization $Y_{1:T} = y_{1:T}$, the likelihood satisfies the predictive decomposition

$$p(y_{1:T} \mid \theta) = p(y_1 \mid \theta) \prod_{t=1}^{T} p(y_t \mid y_{1:t-1}, \theta),$$
(15)

with

$$p(y_t \mid y_{1:t-1}, \theta) = \int g_{\theta}(y_t \mid x_t) . p_{\theta}(x_t \mid y_{1:t-1}) dx_t,$$
(16)

where $p_{\theta}(x_1 \mid y_{1:0}) = \nu_{\theta}(x_1)$ and $p_{\theta}(x_t \mid y_{1:t-1})$ denotes the posterior density of X_t given $Y_{1:t-1} = y_{1:t-1}$ for $t \geq 2$. Importance sampling estimators of the likelihood have relative variance typically increasing exponentially with T so the likelihood is usually estimated using particle filters.

Particle filters propagate N random samples, termed particles, over time using a sequence of resampling steps and importance sampling steps using the importance densities $q_{\theta}\left(x_{1}|y_{1}\right)$ at time 1 and $q_{\theta}\left(x_{t}|y_{t},x_{t-1}\right)$ at times $t\geq 2$. Let $\Xi_{1}:\mathbb{R}^{p}\times\Theta\to\mathbb{R}^{k}$ and $\Xi_{t}:\mathbb{R}^{k}\times\mathbb{R}^{p}\times\Theta\to\mathbb{R}^{k}$ for $t\geq 2$ be deterministic maps such that $X_{1}=\Xi_{1}(V;\theta)\sim q_{\theta}(\cdot\mid y_{1})$ and $X_{t}=\Xi_{t}(x_{t-1},V;\theta)\sim q_{\theta}(\cdot\mid y_{t},x_{t-1})$ for $t\geq 2$ if $V\sim\mathcal{N}\left(0_{p},I_{p}\right)$. If we use these representations to sample the particles and normal random variables to obtain uniform random variables to sample the categorical distributions appearing in the resampling steps then we can obtain an unbiased estimator $\widehat{p}(y_{t}\mid\theta,U)$ of the likelihood where U follows a multivariate normal [15]. When this estimator is used within a PM scheme, the resulting algorithm is known as the particle marginal MH [1]. However if this likelihood estimator is used in the CPM context, the likelihood ratio estimator $\widehat{p}(y_{1:T}\mid\theta',u')/\widehat{p}(y_{1:T}\mid\theta,u)$ can significantly deviate from 1 even when (θ,u) is close to (θ',u') and the true likelihood is continuous at θ . This is because the resampling steps introduce discontinuities in the particles that are selected when θ and u are modified, even slightly.

To reduce the variability of this likelihood ratio estimator, we use a resampling scheme based on the Hilbert space-filling curve which is a continuous fractal map $H:[0,1]\to [0,1]^k$ whose image is $[0,1]^k$. It admits a pseudo-inverse $h:[0,1]^k\to [0,1]$, that is $H\circ h(x)=x$ for all $x\in [0,1]^k$. For most points x,x' that are close in $[0,1]^k$, their images h(x) and h(x') tend to be close. This property can be used to build a "sorted" resampling procedure which will ensure that when the parameter or auxiliary variables change only slightly the particles that are selected remain close. Practically, this resampling procedure proceeds as follows: 1) the \mathbb{R}^k -valued particles are projected in the hypercube $[0,1]^k$ using a bijection $\varkappa:\mathbb{R}^k\to [0,1]^k$, 2) The resulting $[0,1]^k$ -valued particles are projected on [0,1] using the pseudo-inverse h, 3) These projected [0,1]-valued particles are sorted, 4) The systematic resampling scheme proposed in [9] is used on the sorted points.

Let us introduce the importance weights $\omega_1(u_1; \theta) = \nu_{\theta}(x_1) g_{\theta}(y_1 \mid x_1) / q_{\theta}(x_1 \mid y_1)$ and $\omega_t(x_{t-1}, u_t; \theta) = f_{\theta}(x_t \mid x_{t-1}) g_{\theta}(y_t \mid x_t) / q_{\theta}(x_t \mid y_t, x_{t-1})$ for $t \geq 2$. The only difference between the resulting particle filter presented below and the algorithm proposed in [23] is that we use normal random variates instead of randomized quasi-Monte Carlo points in $[0, 1]^p$. For the mapping \varkappa , we adopt the logistic transform used in [23].

Algorithm 2 Particle filter using Hilbert sort

- 1. Sample $U_{1,i} \sim \mathcal{N}(0_p, I_p)$ and set $X_{1,i} = \Xi_1(U_{1,i}; \theta)$ for $i \in 1: N$.
- 2. For t = 1, ..., T 1
 - (a) Find the permutation σ_t such that $h \circ \varkappa \left(X_{t,\sigma_t(1)} \right) \leq \ldots \leq h \circ \varkappa \left(X_{t,\sigma_t(N)} \right)$ if $t \geq 2$, or $X_{t,\sigma_t(1)} \leq \ldots \leq X_{t,\sigma_t(N)}$ if t = 1.
 - (b) Sample $U_t^R \sim \mathcal{N}\left(0,1\right)$, set $\overline{U}_{t,i} = (i-1)/N + \Phi(U_t^R)/N$ for $i \in 1:N$.
 - (c) Sample $A_{t,i} \sim F_t^{-1}\left(\overline{U}_{t,i}\right)$ for $i \in 1:N$ where F_t^{-1} is the generalized inverse distribution function of the categorical distribution with weights $\{\omega_1(U_{1,\sigma_1(i)};\theta); i \in 1:N\}$ if t=1 and $\{\omega_t(X_{t-1,\sigma_{t-1}\left(A_{t-1,\sigma_t(i)}\right)}, U_{t,\sigma_t(i)};\theta); i \in 1:N\}$ for $t \geq 2$.
 - $\text{(d) Sample } U_{t+1,i} \sim \mathcal{N}\left(0_p, I_p\right) \text{ and set } X_{t+1,i} = \Xi_{t+1}(X_{t,\sigma_t(A_{t,i})}, U_{t+1,i}; \theta) \text{ for } i \in 1:N.$

If we denote by $U = (U_{1,1},...,U_{T,N},U_1^R,...,U_{T-1}^R)$ the vector of all the standard normal variables used within this particle filter, the corresponding unbiased likelihood estimator is given by

$$\widehat{p}(y_{1:T} \mid \theta, U) = \left\{ \frac{1}{N} \sum_{i=1}^{N} \omega_1(U_{1,i}; \theta) \right\} \prod_{t=2}^{T} \left\{ \frac{1}{N} \sum_{i=1}^{N} \omega_t(X_{t-1,\sigma_{t-1}(A_{t-1,i})}, U_{t,i}; \theta) \right\}.$$
(17)

In this case, we have M = TNp + T - 1. We can now directly use this estimator within the CPM method.

2.4 Discussion

Ideas related to the CPM scheme have previously been proposed: [35] suggest combining PM steps with updates where U is held fixed and only θ is updated but this scheme will scale poorly with T as it

still uses PM steps. In [2], the authors propose combining PM steps with steps where θ is held fixed and correlation between $\widehat{p}(y \mid \theta, U)$ and $\widehat{p}(y \mid \theta, U')$ is introduced by sampling U' using a m-reversible Markov kernel K. However, the crucial selection of K was not discussed in [2]. After the first version of this work was made available, [14] proposed independently to use the correlation scheme (8) but their guidelines for the correlation parameter ρ differ from the ones we give in subsequent sections and do not ensure that the variance of the loglikelihood ratio estimator is controlled as T increases. They also use a standard particle filter which will .

As the density m of U is independent of θ , it might be argued that a Gibbs algorithm sampling alternately from the full conditional densities $\overline{\pi}(\theta|u)$ and $\overline{\pi}(u|\theta)$ of $\overline{\pi}(\theta,u)$ could mix well and that it is unnecessary to update θ and U jointly as in the CPM scheme. Related ideas have been explored in [42]. However, such a Gibbs strategy is usually not implementable in the applications considered here. Sophisticated particle Gibbs samplers have been proposed to mimick it but their computational complexity is of order T^2N per iteration for state-space models when using such a parameterisation [38, Section 6.2]. Thus they are not even competitive to the standard PM algorithm whose cost per iteration is of order T^2 .

3 Asymptotics of the loglikelihood ratio estimators

To understand the quantitative properties of the CPM scheme, it is key to establish the statistical properties of the likelihood ratio estimator appearing in its acceptance probability (6). For the random effects models introduced in Section 2.3.1, we establish conditional CLTs for the loglikelihood estimator (12) and the corresponding loglikelihood ratio estimators used by the PM and the CPM algorithms when $N, T \to \infty$. Here N will be a deterministic function of T denoted by N_T . We show that these estimators exhibit very different behaviours, underlining the benefits of the CPM over the PM.

Consider a sequence of random variables $\{M^T; T \geq 1\}$ defined on a common probability space (Ω, \mathcal{G}, P) and sub- σ -algebras $\{\mathcal{G}^T; T \geq 1\}$ of \mathcal{G} and write $\stackrel{P}{\to}$ to denote convergence in probability. We write subsequently $M^T \mid \mathcal{G}^T \Rightarrow \lambda$ if $M \sim \lambda$ and $\mathbb{E}[f(M^T) \mid \mathcal{G}^T] \stackrel{P}{\to} \mathbb{E}[f(M)]$ as $T \to \infty$ for any bounded continuous function f.

Henceforth, we will make the assumption that $Y_t \overset{\text{i.i.d.}}{\sim} \mu$ and denote by \mathcal{Y}^T the σ -field spanned by $Y_{1:T}$. When additionally $U \sim m$, we denote the associated probability measure, expectation and variance by \mathbb{P} , \mathbb{E} and \mathbb{V} . As our limit theorems consider the asymptotic regime where $N_T, T \to \infty$, we should write m_T, π_T instead m, π and similarly U^T , U_t^T and $U_{t,i}^T$ instead of U, U_t and $U_{t,i}$. The probability space is defined precisely in Supplementary Material A.1. We do not emphasise here this dependency on T for notational simplicity but it should be kept in mind that we deal with triangular arrays of random variables. We can write unambiguously $\mathbb{E}(\psi(Y_1, U_{1,1}^T; \theta))$ as $\mathbb{E}(\psi(Y_1, U_{1,1}; \theta))$ because $U_{1,1}^T \sim \mathcal{N}(0_p, I_p)$ under \mathbb{P} for any T > 1.

3.1 Asymptotic distribution of the loglikelihood error

Let $\gamma(y_1;\theta)^2 = \mathbb{V}(\varpi(y_1,U_{1,1};\theta))$ be the variance conditional upon $Y_1 = y_1$ and $\gamma(\theta)^2 = \mathbb{V}(\varpi(Y_1,U_{1,1};\theta)) = \mathbb{E}(\gamma(Y_1;\theta)^2)$ the unconditional variance of the normalized importance weight

$$\varpi(Y_t, U_{1,1}; \theta) = \frac{\omega(Y_t, U_{1,1}; \theta)}{p(Y_t \mid \theta)},\tag{18}$$

where $\omega(Y_t, U_{1,1}; \theta)$ is defined in (13).

We present conditional CLTs for the loglikelihood error

$$Z_T(\theta) = \log \widehat{p}(Y_{1:T} \mid \theta, U) - \log p(Y_{1:T} \mid \theta), \tag{19}$$

when U arises from the proposal m or from

$$\overline{\pi}(u|\theta) = \frac{\overline{\pi}(\theta, u)}{\pi(\theta)} = \prod_{t=1}^{T} \frac{\widehat{p}(Y_t \mid \theta, u_t)}{p(Y_t \mid \theta)} \varphi(u_t; 0_{pN_T}, I_{pN_T}), \tag{20}$$

 $\overline{\pi}(\theta, u)$ being given in (4). The density $\overline{\pi}(u|\theta)$ depends upon N_T as the estimator of $p(Y_t \mid \theta)$ is obtained using N_T samples.

Theorem 1. Let $N_T = \lceil \beta T^{\alpha} \rceil$ with $1/3 < \alpha \le 1$, $\beta > 0$ and $Y_t \stackrel{i.i.d.}{\sim} \mu$.

1. If $\mathbb{E}\left(\varpi(Y, U_{1,1}; \theta)^8\right) < \infty$ and $U \sim m$ then

$$T^{(\alpha-1)/2}Z_T(\theta) + \frac{1}{2}T^{(1-\alpha)/2}\beta^{-1}\gamma(\theta)^2 \left| \mathcal{Y}^T \Rightarrow \mathcal{N}\left(0, \beta^{-1}\gamma(\theta)^2\right).$$
 (21)

2. If $\mathbb{E}\left(\varpi(Y_1, U_{1,1}; \theta)^9\right) + \mathbb{E}\left(\gamma(Y_1; \theta)^4\right) < \infty \text{ and } U \sim \overline{\pi}(\cdot | \theta) \text{ then }$

$$T^{(\alpha-1)/2}Z_{T}\left(\theta\right) - \frac{1}{2}T^{(1-\alpha)/2}\beta^{-1}\gamma\left(\theta\right)^{2} \left| \mathcal{Y}^{T} \Rightarrow \mathcal{N}\left(0, \beta^{-1}\gamma\left(\theta\right)^{2}\right).$$
 (22)

Remark. To establish (21), respectively (22), for $1/2 < \alpha \le 1$, the condition $\mathbb{E}\left(\varpi(Y_1, U_{1,1}; \theta)^4\right) < \infty$, respectively $\mathbb{E}\left(\varpi(Y_1, U_{1,1}; \theta)^5\right) < \infty$, is sufficient.

For particle filters, a CLT for $Z_T(\theta)$ of the form (21) has already been established for the case $\alpha=1$ in [6] when using multinomial resampling under strong mixing assumptions. We conjecture that both (21) and (22) hold under weaker assumptions for $1/3 < \alpha < 1$ and the Hilbert sort resampling scheme. However, it is technically very challenging to establish this result².

The results (21) and (22) imply that for large T we expect that, under the proposal, $Z_T(\theta)$ is approximately normal with mean $-\beta^{-1}T^{1-\alpha}\gamma(\theta)^2/2$ and variance $\beta^{-1}T^{1-\alpha}\gamma(\theta)^2$, whereas at equilibrium it is also approximately normal with the same variance but opposite mean. The empirical distribution of $Z_T(\theta)$ is examined for random effects models and state-space models in Section 5 and shown to closely match these limiting distributions.

3.2 Asymptotic distribution of the loglikelihood ratio error

Assume that we are at state (θ, U) and propose (θ', U') using $\theta' \sim q(\theta, \cdot)$, $U' \sim m$ as in the PM algorithm or $\theta' \sim q(\theta, \cdot)$, $U' \sim K_{\rho}(U, \cdot)$ as in the CPM algorithm. In both cases, the acceptance ratio (6) depends on the loglikelihood ratio error

$$R_T(\theta, \theta') = \log \frac{\widehat{p}(Y_{1:T} \mid \theta', U')}{\widehat{p}(Y_{1:T} \mid \theta, U)} - \log \frac{p(Y_{1:T} \mid \theta')}{p(Y_{1:T} \mid \theta)}.$$

$$(23)$$

We examine here the limiting distribution of $R_T(\theta, \theta + \xi/\sqrt{T})$ for fixed θ and ξ . The rationale for examining this ratio is that the posterior typically concentrates at rate $1/\sqrt{T}$ when T increases so a correctly scaled random walk proposal for a MH algorithm will be of the form $\theta' = \theta + \xi/\sqrt{T}$ for ξ a random variable of distribution independent of T.

For the PM algorithm, we have the following conditional CLT.

Theorem 2. Let θ, ξ be fixed. Assume that $\vartheta \mapsto \varpi(y_1, u_{1,1}; \vartheta)$ and $\vartheta \mapsto \mathbb{E}\left(\varpi(Y_1, U_{1,1}; \vartheta)^9\right)$ are continuous at $\vartheta = \theta$ for any $(y_1, u_{1,1}) \in \mathsf{Y} \times \mathbb{R}^p$, $\vartheta \mapsto \gamma(\vartheta)$ is continuously differentiable at $\vartheta = \theta$ and $\mathbb{E}\left(\varpi(Y_1, U_{1,1}; \vartheta)^9\right) + \mathbb{E}\left(\gamma(Y_1; \theta)^4\right) < \infty$. For $N_T = \lceil \beta T^\alpha \rceil$ with $1/3 < \alpha \leq 1$, $\beta > 0$, $Y_t \stackrel{i.i.d.}{\sim} \mu$, $U \sim \overline{\pi}(\cdot \mid \theta)$ and $U' \sim m$ where U and U' are independent, we have

$$T^{(\alpha-1)/2}R_{T}(\theta,\theta+\xi/\sqrt{T}) + T^{(1-\alpha)/2}\beta^{-1}\gamma(\theta)^{2} | \mathcal{Y}^{T} \Rightarrow \mathcal{N}\left(0,2\beta^{-1}\gamma(\theta)^{2}\right). \tag{24}$$

This result shows that the loglikelihood ratio error in the PM case can only have a limiting variance of order 1 if N_T is proportional to T. The loglikelihood ratio estimator used by the CPM exhibits a markedly different behaviour if we consider $U' \sim K_{\rho_T}(U,\cdot)$ with

$$\rho_T = \exp\left(-\psi \frac{N_T}{T}\right),\tag{25}$$

for some $\psi > 0$. Let us denote by \mathcal{F}^T the σ -field spanned by $\{Y_t; t \in 1 : T\}$ and $\{U_{t,i}; t \in 1 : T, i \in 1 : N\}$. We also denote the Euclidean norm by $\|\cdot\|$ and we write $\nabla_u f = (\partial_{u^1} f, ..., \partial_{u^p} f)'$ for a real-valued function $f : \mathbb{R}^p \to \mathbb{R}$ where $u = (u^1, ..., u^p)$.

²In the simpler scenario where one uses systematic resampling, such a CLT has not yet been established. Some of the technical problems arising when attempting to carry out such an analysis are detailed in [22].

Theorem 3. Let θ, ξ be fixed. Let $Y_t \overset{i.i.d.}{\sim} \mu$, $U \sim \overline{\pi}(\cdot | \theta)$ and $U' \sim K_{\rho_T}(U, \cdot)$ where ρ_T is given by (25) then if Assumptions 4-9 in Supplementary Material A.5 hold and if $N_T \to \infty$ as $T \to \infty$ with $N_T/T \to 0$, we have

$$R_T(\theta, \theta + \xi/\sqrt{T}) | \mathcal{F}^T \Rightarrow \mathcal{N}\left(-\kappa(\theta)^2/2, \kappa(\theta)^2\right),$$
 (26)

where

$$\kappa (\theta)^{2} = 2\psi \mathbb{E} \left(\left\| \nabla_{u} \overline{\omega}(Y_{1}, U_{1,1}; \theta) \right\|^{2} \right). \tag{27}$$

We do not make any structural assumption on $\varpi(y, u; \theta)$ to establish Theorem 3. Assumptions 4-9 are differentiability and integrability assumptions of this quantity with respect to y, u and θ . For CPM, this result states that the limiting variance of the loglikelihood ratio is of order 1 when N_T grows sublinearly with T. Moreover, it shows that the distribution of the loglikelihood ratio error becomes asymptotically independent of U, suggesting that the CPM chain is less prone to sticking than the PM at stationarity.

This conditional CLT has not been established for particle filters. For univariate state-space models, i.e. k = 1, we have observed experimentally on various stationary state-space models that a similar conditional CLT appears to hold. For multivariate state-space models, the CLT only appears to hold conditional upon \mathcal{Y}^T when N_T grows at least at rate $T^{k/(k+1)}$; see Section 5.

4 Analysis and optimisation

4.1 Weak convergence in the large sample regime

The use of weak convergence techniques to analyse and optimise MCMC schemes was pioneered in [46] and has found numerous applications ever since; see, e.g., [48] for a recent application to PM. To the best of our knowledge, all these contributions consider the asymptotic regime where the parameter dimension $d \to \infty$ while T is fixed. In these scenarios, a time rescaling is introduced and the limiting Markov process is usually a diffusion process. We analyze here the CPM scheme for random effects models after space rescaling under the large sample regime standard in asymptotic statistics; i.e., d is fixed while $T \to \infty$. Our analysis assumes the statistical model is regular enough to ensure that the posteriors $\{\pi_T(\theta); T \geq 1\}$ can be approximated by normal densities which concentrate. Here $\pi_T(\theta)$ is a random probability density dependent on $Y_{1:T}$ assumed to be measurable w.r.t. \mathcal{Y}^T ; see, e.g., [7, 13] for a formal definition. We write $\stackrel{\mathbb{P}^Y}{\to}$ to denote convergence in probability with respect to the law of $\{Y_t; t \geq 1\}$.

Assumption 1. The sequence of random probability densities $\{\pi_T(\theta); T \geq 1\}$ satisfies at $T \to \infty$

$$\int \left| \pi_T(\theta) - \varphi(\theta; \widehat{\theta}_T, \overline{\Sigma}/T) \right| d\theta \stackrel{\mathbb{P}^Y}{\to} 0$$

where $\{\widehat{\theta}_T; T \geq 1\}$ is a random sequence such that $\widehat{\theta}_T$ is \mathcal{Y}^T -measurable, $\widehat{\theta}_T \to_{\mathbb{P}^Y} \overline{\theta}$ and $\overline{\Sigma}$ is a positive definite matrix.

This assumption will be satisfied if a Berstein-von Mises theorem holds; see [53, Section 10.2] for sufficient conditions.

Consider the stationary CPM chain $\{(\vartheta_n^T, \mathsf{U}_n^T); n \geq 0\}$ of proposal $q_T(\theta, \theta')$ targeting the random measure (4) $\overline{\pi}_T(\mathrm{d}\theta, \mathrm{d}u) = \pi_T(\mathrm{d}\theta)\overline{\pi}_T(\mathrm{d}u|\theta)$ associated with $Y_{1:T}$. By rescaling the parameter component of the CPM chain using $\widetilde{\vartheta}_n^T := \sqrt{T}(\vartheta_n^T - \widehat{\theta}_T)$, we obtain the stationary Markov chain $\{(\widetilde{\vartheta}_n^T, \mathsf{U}_n^T); n \geq 0\}$ of initial distribution $(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T) \sim \widetilde{\pi}_T$ where

$$\widetilde{\pi}_{T}(\widetilde{\theta}, u) = \widetilde{\pi}_{T}(\widetilde{\theta})\widetilde{\pi}_{T}(u|\widetilde{\theta}), \quad \widetilde{\pi}_{T}(\widetilde{\theta}) = \pi_{T}(\widehat{\theta}_{T} + \widetilde{\theta}/\sqrt{T})/\sqrt{T}, \quad \widetilde{\pi}_{T}(u|\widetilde{\theta}) = \overline{\pi}_{T}(u|\widehat{\theta}_{T} + \widetilde{\theta}/\sqrt{T}), \quad (28)$$

and the associated proposal density for the parameter becomes

$$\widetilde{q}_T(\widetilde{\theta}, \widetilde{\theta}') = q_T(\widehat{\theta}_T + \widetilde{\theta}/\sqrt{T}, \widehat{\theta}_T + \widetilde{\theta}'/\sqrt{T})/\sqrt{T}.$$
(29)

We will assume here that we use a random walk proposal scaled appropriately.

Assumption 2. The proposal density is of the form

$$q_T(\theta, \theta') = \sqrt{T} v(\sqrt{T}(\theta' - \theta)), \tag{30}$$

where v is a probability density on \mathbb{R}^d ; that is $\theta' \sim q_T(\theta, \cdot)$ when $\theta' = \theta + \xi/\sqrt{T}$ with $\xi \sim v$.

Finally, we assume that we can control uniformly the rate at which convergence of the CLT in Theorem 3 holds in a neighbourhood of $\overline{\theta}$ specified in Assumption 1. We denote by $d_{\rm BL}(\mu, \nu)$ the bounded Lipschitz metric between probability measures μ, ν ; see, e.g., [53, p. 332] or Supplementary Material A.9.

Assumption 3. There exists a neighbourhood $N(\bar{\theta})$ of $\bar{\theta}$ such that the loglikelihood ratio error considered in Theorem 3 with $\xi \sim v(\cdot)$ satisfies as $T \to \infty$

$$\sup_{\theta \in N(\bar{\theta})} \widetilde{\mathbb{E}} \left[\left. d_{\mathrm{BL}} \left\{ \mathcal{L}\mathrm{aw} \left(\left. R_T(\theta, \theta + \xi / \sqrt{T}) \right| \mathcal{F}^T \right), \mathcal{N} \left(-\kappa \left(\theta \right)^2 / 2, \kappa \left(\theta \right)^2 \right) \right\} \right| \mathcal{Y}^T \right] \stackrel{\mathbb{P}^Y}{\to} 0.$$

We prove that Assumption 3 holds under regularity conditions in Supplementary Material A.6.

Under Assumption 2, the proposal $\widetilde{q}_T(\widetilde{\theta}, \widetilde{\theta}')$ defined in (29) satisfies $\widetilde{q}_T(\widetilde{\theta}, \widetilde{\theta}') = \upsilon(\widetilde{\theta}' - \widetilde{\theta})$ and will be denoted $\widetilde{q}(\widetilde{\theta}, \widetilde{\theta}')$. In this case, the corresponding transition kernel of the rescaled CPM chain is given by

$$Q_T\{(\widetilde{\theta}, u), (\mathrm{d}\widetilde{\theta}', \mathrm{d}u')\} = \widetilde{q}(\widetilde{\theta}, \mathrm{d}\widetilde{\theta}') K_{\rho_T}(u, \mathrm{d}u') \alpha_{Q_T}\{(\widetilde{\theta}, u), (\widetilde{\theta}', u')\} + \{1 - \varrho_{Q_T}(\widetilde{\theta}, u)\} \delta_{(\widetilde{\theta}, u)}(\mathrm{d}\widetilde{\theta}', \mathrm{d}u') \quad (31)$$

with acceptance probability

$$\alpha_{Q_T}\{(\widetilde{\theta},u),(\widetilde{\theta}',u')\} = \min\left\{1,\frac{\widetilde{\pi}_T(\widetilde{\theta}',u')\widetilde{q}(\widetilde{\theta}',\widetilde{\theta})K_{\rho_T}\left(u',u\right)}{\widetilde{\pi}_T(\widetilde{\theta},u)\widetilde{q}(\widetilde{\theta},\widetilde{\theta}')K_{\rho_T}\left(u,u'\right)}\right\},$$

and corresponding rejection probability $1 - \varrho_{Q_T}(\widetilde{\theta}, u)$. The kernel Q_T depends on $Y_{1:T}$ and is assumed to be measurable w.r.t. \mathcal{Y}^T . Let $\Theta_T = \{\widetilde{\vartheta}_n^T; n \geq 0\}$. The following result shows that the non-Markov stationary sequences $\{\Theta_T; T \geq 1\}$ converge weakly as $T \to \infty$ to a stationary Markov chain corresponding to the Penalty method—an "ideal" Monte Carlo technique which cannot be practically implemented [10], [41, p. 7].

Theorem 4. If Assumptions 1, 2 and 3 hold and $\vartheta \mapsto \kappa(\vartheta)$ is continuous at $\vartheta = \overline{\theta}$ then the random probability measures on $(\mathbb{R}^d)^{\infty}$ given by the laws of $\{\Theta_T; T \geq 1\}$ converge weakly in probability \mathbb{P}^Y as $T \to \infty$ to the law of a stationary Markov chain $\{\widetilde{\vartheta}_n; n \geq 0\}$ defined by $\widetilde{\vartheta}_0 \sim \mathcal{N}(0, \overline{\Sigma})$ and $\widetilde{\vartheta}_n \sim P(\widetilde{\vartheta}_n, \cdot)$ for $n \geq 1$ with

$$P(\widetilde{\theta}, d\widetilde{\theta}') = \widetilde{q}(\widetilde{\theta}, d\widetilde{\theta}')\alpha_P(\widetilde{\theta}, \widetilde{\theta}') + \{1 - \varrho_P(\widetilde{\theta})\}\delta_{\widetilde{\theta}}(d\widetilde{\theta}'), \tag{32}$$

and

$$\alpha_P(\widetilde{\theta}, \widetilde{\theta}') = \int \varphi\left(\mathrm{d}w; -\kappa^2/2, \kappa^2\right) \min\left\{1, \frac{\varphi(\widetilde{\theta}'; 0, \overline{\Sigma})\widetilde{q}(\widetilde{\theta}', \widetilde{\theta})}{\varphi(\widetilde{\theta}; 0, \overline{\Sigma})\widetilde{q}(\widetilde{\theta}, \widetilde{\theta}')} \exp\left(w\right)\right\},\,$$

 $1-\varrho_P(\widetilde{\theta})$ being the corresponding rejection probability and $\kappa:=\kappa(\overline{\theta}).$

The consequence of this result is that, as $T \to \infty$, only the asymptotic distribution of the loglikelihood ratio error at the central parameter value $\bar{\theta}$ impacts the acceptance probability of the limiting chain. For large T and a proposal of the form specified in Assumption 2, we thus expect some of the quantitative properties of the CPM kernel Q, where we now omit T from notation, to be captured by the Markov kernel

$$\widehat{Q}(\theta, d\theta') = q(\theta, d\theta') \alpha_{\widehat{Q}}(\theta, \theta') + \{1 - \varrho_{\widehat{Q}}(\theta)\} \delta_{\theta}(d\theta'), \qquad (33)$$

where

$$\alpha_{\widehat{Q}}\left(\theta,\theta'\right) = \int \varphi(\mathrm{d}w; -\kappa^2/2, \kappa^2) \min\left\{1, r_{\mathrm{EX}}(\theta,\theta') \exp\left(w\right)\right\},\,$$

 $1 - \varrho_{\widehat{Q}}(\theta)$ being the corresponding rejection probability. We have obtained (33) by using the change of variables $\theta = \hat{\theta}_T + \widetilde{\theta}/\sqrt{T}$ and substituting the true target for its normal approximation in (32), hence removing a level of approximation.

4.2 A bounding Markov chain

We analyse here the stationary Markov chain of transition kernel \widehat{Q} arising from our weak convergence analysis. To state our results, we need the following notation. For any real-valued measurable function h, probability measure and Markov kernel K on a measurable space (E,\mathcal{E}) , we write $\mu(h) = \int_E h(x) \, \mu(\mathrm{d}x)$, $Kh(x) = \int_E K(x,\mathrm{d}x') \, h(x')$ and $K^nh(x) = \int_E K^{n-1}(x,\mathrm{d}z) \, K(z,\mathrm{d}x') \, h(x')$ for $n \geq 2$ with $K^1 = K$. We also introduce the Hilbert space $L^2(\mu) = \{h: E \to \mathbb{R}: \mu(h^2) < \infty\}$ equipped with the inner product $\langle g,h\rangle_{\mu} = \int_E g(x) \, h(x) \, \mu(\mathrm{d}x)$. For any $h \in L^2(\mu)$, the autocorrelation at lag $n \geq 0$ is $\phi_n(h,K) = \langle \overline{h}, K^nh\rangle_{\mu}/\mu(\overline{h}^2)$ where $\overline{h} = h - \mu(h)$. The IACT associated with a function h under a Markov kernel K is given by IF $(h,K) = 1 + 2\sum_{n=1}^{\infty} \phi_n(h,K)$ and will be referred to subsequently as the inefficiency. For $\mu(\mathrm{d}x) = \mu(\mathrm{d}x_1,\mathrm{d}x_2)$, we will slightly abuse notation and write IF(h,K) instead of IF(g,K) when $g(x_1,x_2) = h(x_1)$ or $g(x_1,x_2) = h(x_2)$. When estimating $\mu(h)$, nIF(h,K) samples from a stationary Markov chain of μ -invariant transition kernel K are necessary to obtain approximately an estimator of the same precision as an average of n independent draws from μ ; see, e.g., [24].

We provide an upper bound on $\mathrm{IF}(h,\widehat{Q})$ which we exploit to provide guidelines on how to optimise the performance of the CPM scheme in Subsection 4.4. The inefficiency $\mathrm{IF}(h,\widehat{Q})$ is difficult to work with but we give an upper bound that only depends on $\mathrm{IF}(h,Q_{\mathrm{EX}})$ and κ . To proceed, we introduce an auxiliary Markov kernel Q^* given by

$$Q^{*}(\theta, d\theta') = \varrho_{U}(\kappa) Q_{EX}(\theta, d\theta') + \{1 - \varrho_{U}(\kappa)\} \delta_{\theta}(d\theta'), \qquad (34)$$

where

$$\varrho_{\mathcal{U}}(\kappa) = \int \varphi(\mathrm{d}w; -\kappa^2/2, \kappa^2) \min\{1, \exp(w)\} = 2\Phi(-\kappa/2). \tag{35}$$

We denote by $\bar{\varrho}_{Q^*}(\kappa)$, respectively $\bar{\varrho}_{\widehat{Q}}(\kappa)$, the average acceptance probability of Q^* , respectively \widehat{Q} , at stationarity. The kernel Q^* is a "lazy" version of Q_{EX} which satisfies the following properties.

Proposition 5. The kernel Q^* is reversible w.r.t. π and $\mathrm{IF}(h,\widehat{Q}) \leq \mathrm{IF}(h,Q^*)$ for any $h \in L^2(\pi)$ where

IF
$$(h, Q^*) = \{1 + IF(h, Q_{EX})\}/\rho_{U}(\kappa) - 1,$$
 (36)

with equality when $\varrho_{\text{EX}}\left(\theta\right)=1$ for all $\theta\in\Theta$ and

$$\bar{\varrho}_{\mathbf{Q}^*}(\kappa) = \varrho_{\mathbf{U}}(\kappa) \pi(\varrho_{\mathbf{E}\mathbf{X}}) \leq \bar{\varrho}_{\widehat{\mathbf{Q}}}(\kappa).$$
(37)

Moreover, Q^* is geometrically ergodic if Q_{EX} is geometrically ergodic.

For any transition kernel K admitting an invariant distribution given by π , or admitting π as a marginal, we define the relative inefficiency RIF (h, K) of the kernel K with respect to the known likelihood kernel Q_{EX} and the auxiliary relative computing time ARCT (h, K) by

$$RIF(h,K) := \frac{IF(h,K)}{IF(h,Q_{EX})}, \qquad ARCT(h,K) := \sqrt{\frac{RIF(h,K)}{\kappa^2 \varrho_{U}(\kappa)}}.$$
 (38)

We next minimise ARCT (h, Q^*) , an upper bound on ARCT (h, \widehat{Q}) , w.r.t. κ - this quantity is a component of the function we need to minimize in order to optimize the performance of the CPM algorithm; see Section 4.4.

Proposition 6. The following results hold:

1. If $IF(h, Q_{EX}) = 1$, then

RIF
$$(h, Q^*) = \{2 - \rho_{\mathrm{U}}(\kappa)\}/\rho_{\mathrm{U}}(\kappa),$$

and ARCT (h, Q^*) is minimised at $\kappa = 1.35$, at which point $\varrho_{\rm U}(\kappa) = 0.50$, RIF $(h, Q^*) = 2.99$ and ARCT $(h, Q^*) = 1.81$.

2. As IF $(h, Q_{\text{ex}}) \longrightarrow \infty$,

RIF
$$(h, Q^*) = 1/\rho_{\rm U}(\kappa)$$
,

and ARCT (h, Q^*) is minimised at $\kappa = 1.50$, at which point $\varrho_U(\kappa) = 0.43$, RIF $(h, Q^*) = 2.20$ and ARCT $(h, Q^*) = 1.47$.

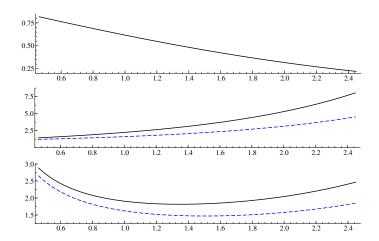


Figure 1: Illustrations of Proposition 6. Top: Acceptance probability $\varrho_{\rm U}(\kappa)$ against κ . Middle: Relative inefficiency RIF (h,Q^*) against κ (solid line IF $(h,Q_{\rm EX})=1$, dashed line IF $(h,Q_{\rm EX})\to\infty$). Bottom: Auxiliary relative computing time ARCT (h,Q^*) against κ (solid line IF $(h,Q_{\rm EX})=1$, dashed line IF $(h,Q_{\rm EX})\to\infty$).

3. RIF (h, Q^*) and ARCT (h, Q^*) are decreasing functions of IF (h, Q_{EX}) . The minimising argument rises monotonically from 1.35 to 1.50 as IF (h, Q_{EX}) increases from 1 to ∞ .

Figure 1 displays $\varrho_{\rm U}(\kappa)$, RIF (h,Q^*) and ARCT (h,Q^*) against κ . The two scenarios displayed are for IF $(h,Q_{\rm EX})=1$, corresponding to the "perfect" proposal case where $q(\theta,\theta')=\pi(\theta')$, and for the limiting case where IF $(h,Q_{\rm EX})\longrightarrow\infty$. These are parts (i) and (ii) of Proposition 6. From Figure 1, it is also clear that ARCT (h,Q^*) , for both scenarios, is fairly flat as a function of κ . The function only approximately doubles relative to the minimum at $\kappa=1$ or 4.

4.3 A lower bound on the integrated autocorrelation time

The weak convergence result presented in Theorem 4 does not imply that $\operatorname{IF}(h,Q) \stackrel{\mathbb{P}^Y}{\to} \operatorname{IF}(h,P)$ as $T \to \infty$ for any test function $h:\Theta \to \mathbb{R}$ such that $\operatorname{IF}(h,Q_T)+\operatorname{IF}(h,P)<\infty$. Whereas our weak convergence result holds whenever $N_T \to \infty$ as $T \to \infty$ and $N_T/T \to 0$, we establish here a result suggesting that we need N_T to increase with T at least at rate \sqrt{T} for $\operatorname{IF}(h,Q_T)$ to remain controlled. To simplify the presentation in this section, we assume further on that d=1.

In the CPM context, the auxiliary variables sequence $\{U_n; n \geq 0\}$ evolve at a much slower scale than $\{\vartheta_n; n \geq 0\}$ according to the kernel K_{ρ_T} where ρ_T is given by (25). We expect and observe empirically that the inefficiency IF (h, Q_T) is of the same order as the inefficiency of $\{\mathbb{E}[h(\vartheta_n)|\mathsf{U}_n]; n \geq 0\}$ when N_T increases too slowly to infinity with T. Moreover, under large T, we have under regularity conditions, see, e.g., [17, Lemma 2]

$$\mathbb{E}\left[h(\vartheta_n)|\mathsf{U}_n\right] = h(\widehat{\theta}_T) + \frac{\overline{\Sigma}}{2T}\nabla_{\vartheta,\vartheta}h(\widehat{\theta}_T) + \frac{\overline{\Sigma}}{T}\nabla_{\vartheta}h(\widehat{\theta}_T)\ \Psi(\widehat{\theta}_T,\mathsf{U}_n) + O_{\widetilde{\mathbb{P}}}\left(T^{-2}\right),\tag{39}$$

where

$$\Psi(\widehat{\theta}_T, U) = \nabla_{\vartheta} \log\{\widehat{p}(Y_{1:T} \mid \widehat{\theta}_T, U) / p(Y_{1:T} \mid \widehat{\theta}_T)\}$$
(40)

is the error in the simulated score at $\widehat{\theta}_T$, which we will refer to as the score error. As a first step, we compute the inefficiency IF(Ψ , Q_T) of the score error.

Proposition 7. Under regularity conditions given in Section A.10, there exists C > 0 such that $IF(\Psi, Q_T) \ge C \mathbb{V}_{\overline{\pi}}(\Psi)$ \mathbb{P}^Y – a.s.

It follows from calculations similar to Supplementary Material A.11, see also [37, Proposition 3], that under regularity conditions there exists A > 0 such that $\mathbb{V}_{\overline{\pi}}(\Psi) \sim AT/N \mathbb{P}^Y$ – a.s. By combining (39)

and Proposition 7, we thus expect the inefficiency of $\{\mathbb{E}[h(\vartheta_n)|\mathsf{U}_n]; n \geq 0\}$ to be lower bounded by a term of order

$$\frac{\mathrm{IF}(\Psi,Q_T) \ \mathbb{V}_{\overline{\pi}}\left(\Psi/T\right)}{\mathbb{V}_{\pi}\left(h\right)} \gtrsim B \frac{T}{N_T} \frac{T^{1-\alpha}}{T^2} T = B T^{1-2\alpha}$$

for $N_T = \lceil \beta T^{\alpha} \rceil$ and a constant B > 0. This result suggests that a necessary condition for IF (h, Q_T) to remain finite as $T \to \infty$ is to have N_T growing at least at rate \sqrt{T} . This is validated by our experimental results of Section 5 which also suggest that this rate is sufficient.

4.4 Optimization

We provide a heuristic to select the parameters of the CPM so as to optimise its performance which is validated by experimental results in Section 5. Again, we assume here that d = 1. For a test function $h: \Theta \to \mathbb{R}$, we want to minimize

$$CT(h, Q_T) = N_T \times IF(h, Q_T), \tag{41}$$

where the factor N_T arises from the fact that the computational cost of obtaining the likelihood estimator is proportional to N_T for random effects models. The results of Section 4.3 suggest that we should choose the number of Monte Carlo samples to scale as $N_T = \beta T^{1/2}$ so that $\rho_T = \exp\left(-\psi \beta T^{-1/2}\right)$. It remains to determine both ψ and β .

To evaluate (41), we first decompose the functional of interest evaluated at the parameter at the n-th iteration as

$$h(\vartheta_n) = f(\mathsf{U}_n) + g(\vartheta_n, \mathsf{U}_n).$$

where

$$f(\mathsf{U}) := \mathbb{E}_{\bar{\pi}_T} \left[h(\vartheta) | \mathsf{U} \right], \qquad g(\vartheta, \mathsf{U}) := h(\vartheta) - \mathbb{E}_{\bar{\pi}_T} \left[h(\vartheta) | \mathsf{U} \right]. \tag{42}$$

From [32], we have that

$$\mathbb{V}_{\pi_T}(h)\operatorname{IF}(h, Q_T) \leq 2\mathbb{V}_{\bar{\pi}_T}(f)\operatorname{IF}(f, Q_T) + 2\mathbb{V}_{\bar{\pi}_T}(g)\operatorname{IF}(g, Q_T).$$

Assumption 1 combined to mild regularity assumptions on h and integrability conditions shows that $\mathbb{V}_{\bar{\pi}_T}(h(\vartheta_n)) \approx \overline{\Sigma}_h/T$, where $\overline{\Sigma}_h = |h'(\bar{\theta})|^2 \overline{\Sigma}$. Since $f(\vartheta_n, \mathsf{U}_n)$ and $g(\vartheta_n, \mathsf{U}_n)$ are clearly uncorrelated, this implies that $\mathbb{V}_{\bar{\pi}_T}(h) = \mathbb{V}_{\bar{\pi}_T}(f) + \mathbb{V}_{\bar{\pi}_T}(g)$. From (39) we have $\mathbb{V}_{\bar{\pi}_T}(f) \approx \overline{\Sigma}^2 \mathbb{V}_{\bar{\pi}_T}(\Psi/T) \approx \overline{\Sigma}_f/(TN_T)$, so it follows that $\overline{\Sigma}_{-}$ $\overline{\Sigma}_{-}$ $\overline{\Sigma}_{-}$ $\overline{\Sigma}_{-}$ $\overline{\Sigma}_{-}$

$$\mathbb{V}_{\bar{\pi}_T}(g) \approx \frac{\overline{\Sigma}_h}{T} - \frac{\overline{\Sigma}_f}{TN_T} \approx \frac{\overline{\Sigma}_h}{T}.$$

Using the reasoning of Section 4.3 and the calculations above we obtain

$$\operatorname{IF}(h, Q_{T}) \leq \frac{2}{\overline{\Sigma}_{h}} \left(\mathbb{V}_{\overline{\pi}_{T}} \left(\sqrt{T} f \right) \operatorname{IF}(f, Q_{T}) + \mathbb{V}_{\overline{\pi}_{T}} \left(\sqrt{T} g \right) \operatorname{IF}(g, Q_{T}) \right) \\
\approx \frac{2}{\overline{\Sigma}_{h}} \left(\frac{\overline{\Sigma}_{f}}{N_{T}} \operatorname{IF}(\Psi, Q_{T}) + \overline{\Sigma}_{h} \operatorname{IF}(g, Q_{T}) \right). \tag{43}$$

Proposition 7 states that IF(Ψ , Q_T) is of order at least T/N_T in probability as $T \to \infty$. Numerical results suggest that in fact we have IF(Ψ , Q_T) $\approx A/(\delta_T \varrho_U(\kappa))$ where $\delta_T = \psi N_T/T = -\log \rho_T$ as detailed in Section 5.1, Figure 5. Hence, by substituting this expression of IF(Ψ , Q_T) in (43), it follows that

IF
$$(h, Q_T) \lesssim \frac{2}{\overline{\Sigma}_h} \left(\frac{\overline{\Sigma}_f}{N_T} \frac{A}{\delta_T \varrho_{\mathrm{U}}(\kappa)} + \overline{\Sigma}_h \operatorname{IF}(g, Q_T) \right).$$

It can also be observed empirically from Figure 4, described in Section 5.1, that the autocorrelations of $g(\vartheta_n, \mathsf{U}_n)$ decay exponentially, at a rate independent of T. Thus we expect that, at least approximately, we have IF $(g, Q_T) \approx \mathrm{IF}(h, \widehat{Q}_T)$ in probability. Therefore overall we have that for some constant B

IF
$$(h, Q_T) \lesssim 2 \left(\frac{B}{\varrho_{\mathrm{U}}(\kappa) \delta_T N_T} + \mathrm{IF} \left(h, \widehat{Q}_T \right) \right)$$
.

We are interested in optimizing $\mathrm{CT}(h,Q_T) \simeq N_T \times \mathrm{IF}(h,Q_T)$ w.r.t. ψ and β where we recall from (27) that $\delta_T = \psi N_T/T = \psi \beta/\sqrt{T} = (\kappa^2 \beta)/(\gamma^2 \sqrt{T})$ as $\kappa^2 = \psi \gamma^2$. Therefore

$$\operatorname{CT}(h, Q_T) \lesssim 2T^{1/2} \left(\frac{C}{\beta \varrho_{\mathrm{U}}(\kappa)\kappa^2} + \beta \operatorname{IF}\left(h, \widehat{Q}_T\right) \right),$$
 (44)

where $C = B\gamma^2$, and thus the upper bound on $CT(h, Q_T)$ is minimized for

$$\beta = \sqrt{\frac{C}{\varrho_{\mathrm{U}}(\kappa)\kappa^{2}\mathrm{IF}\left(h,\widehat{Q}_{T}\right)}}.$$

Substituting this expression in the upper bound on IF (h, \widehat{Q}_T) , we obtain IF $(h, Q_T) \lesssim 4$ IF (h, \widehat{Q}_T) . Then by further substituting the last expression in the resulting upper bound on CT (h, Q_T) , we obtain

$$CT(h, Q_T) \lesssim 4\sqrt{C}T^{1/2}ARCT\left(h, \widehat{Q}_T\right) \lesssim 4\sqrt{C}T^{1/2}ARCT(h, Q_T^*)$$
 (45)

where ARCT is introduced in (38). Therefore in practice we will minimize $\operatorname{ARCT}(h,Q_T^*)$ w.r.t. κ , which has already been addressed in Proposition 6. The minimiser $\hat{\kappa}$ is a function of $\operatorname{IF}(h,Q_{\operatorname{EX}})$ which varies only slightly as $\operatorname{IF}(h,Q_{\operatorname{EX}})$ varies from 1 to ∞ as observed in Figure 1. Consequently, we propose the following procedure to optimize the performance of CPM. Let T be fixed and large enough for the asymptotic assumptions to hold approximately. First, we choose a candidate value for N and determine $\hat{\psi}$ such that the standard deviation of the log-likelihood ratio estimator around the mode of the posterior satisfies $\hat{\kappa} \approx 1.4$. Second, fixing ψ to $\hat{\psi}$, we evaluate for several values of β the computational time $\operatorname{CT}(h,Q_T)$ which we assume is of the form of the upper bound (44), i.e.,

$$CT(h, Q_T) = C_0/\beta + C_1\beta, \tag{46}$$

with κ and T kept constant; see Figure 6 in Section 5.1 for empirical results. This function is minimized for $\beta = \sqrt{C_0/C_1}$. Practically we evaluate $\mathrm{CT}(h,Q_T)$ only on a subset of the data. We then estimate through regression the constants C_0 and C_1 by \hat{C}_0 and \hat{C}_1 which in turn provide the following estimate of β

$$\hat{\beta} = \sqrt{\hat{C}_0/\hat{C}_1}.\tag{47}$$

We examine in Section 5.1 the assumptions made here, illustrate this procedure and demonstrate its robustness.

5 Applications

5.1 Random effects model

We illustrate the performance of the PM and CPM schemes on a simple Gaussian random effects model where

$$X_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, 1), \qquad Y_t | X_t \sim \mathcal{N}(X_t, 1).$$
 (48)

We are interested in estimating θ (which has a true value of 0.5) to which we assign a zero-mean Gaussian prior with large variance. In this scenario, the likelihood is known as $Y_t \sim \mathcal{N}(\theta, 2)$. This allows for detailed experimental analysis of the loglikelihood error and the loglikelihood ratio error. This also allows us to implement the MH algorithm with the true likelihood. The same normal random walk proposal is used for all three schemes (MH, PM and CPM) and the following unbiased estimator of the likelihood is used for the PM and CPM schemes:

$$\widehat{p}(y_{1:T} \mid \theta, U) = \prod_{t=1}^{T} \widehat{p}(y_t \mid \theta, U_t), \qquad \widehat{p}(y_t \mid \theta, U_t) = \frac{1}{N} \sum_{i=1}^{N} \varphi(y_t; \theta + U_{t,i}, 1), \qquad U_{t,i} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

The inefficiency is estimated for all three schemes for $h(\theta) = \theta$ using $1 + 2\sum_{n=1}^{L} \widehat{\phi}_n$ where $\widehat{\phi}_n$ is the estimated correlation for θ at lag n and L is a suitable cutoff value. We use the notation Z = 0

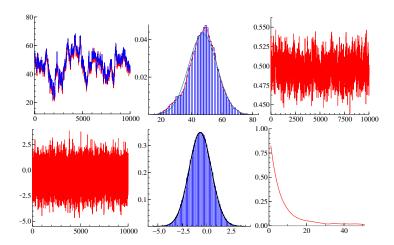


Figure 2: Random effects model using CPM: $T=8192, N=80, \rho=0.9963$. Left: the first 10,000 iterations of W (blue) and Z (red) (top), the difference R (bottom). Middle: Histograms of Z (top) and R (bottom) and the theoretical Gaussian densities. Right: draws of θ (top) and the corresponding correlogram (bottom).

 $\log \left\{ \widehat{p}(y_{1:T} \mid \theta, U) / p(y_{1:T} \mid \theta) \right\} \text{ and } W = \log \left\{ \widehat{p}(y_{1:T} \mid \theta', U') / p(y_{1:T} \mid \theta') \right\} \text{ where } \theta' \sim q(\theta, \cdot), U' \sim K_{\rho}(U, \cdot) \text{ and write } R = W - Z \text{ for } R_{T}(\theta, \theta') \text{ defined in } (23).$

As discussed in Section 4, for large datasets, the relative inefficiency RIF = IF/IF_{MH} and associated relative computing time RCT = $N \times \text{RIF}$ of the CPM scheme depend on the standard deviation κ of R at stationarity and the correlation parameter ρ . To validate experimentally the results of Section 3, we first analyze the case where T=8192 in more detail. We run CPM using a random walk proposal for N=80 and $\rho=0.9963$, so that $\kappa=1.145$. The draws of W and Z at equilibrium, together with R, are displayed in Figure 2. The draws of Z are approximately distributed according to $\mathcal{N}(\sigma^2/2, \sigma^2)$ (middle left), where the variance σ^2 is high. The draws of R appear uncorrelated (in unreported tests) and their histogram is indistinguishable from the expected theoretical distribution $\mathcal{N}(-\kappa^2/2, \kappa^2)$ established in Theorem 3 (middle right). This is in agreement with Theorem 1, equation (22), the posterior of θ being concentrated. The resulting draws and correlogram (bottom panel) of θ demonstrate low persistence.

For the PM scheme, it is necessary to take N=5000 samples to ensure that the variance of Z evaluated at a central value $\hat{\theta}$ is approximately one [18]. We next validate experimentally the theoretical results of Section 4 by investigating the performance of CPM for this dataset, varying N, and thus also hence $\kappa^2 = \mathbb{V}(R)$, while keeping $\rho = 0.9963$. Figure 3 displays the values of RIF and RCT against κ as well as the marginal acceptance probabilities, showing that RCT is approximately minimized around $\kappa = 1.6$ close to the minimizing argument of ARCT (h, Q_T^*) established in Proposition 6 which satisfies (45). The bottom two plots show that $\log \kappa^2$ decreases linearly with $\log N$ as expected (bottom right) and that the marginal probability of acceptance in the CPM scheme is close to the asymptotic lower bound (bottom left) given by (37). From these experimental results, it is clear that for all values of N considered, the gains of the CPM scheme over the PM method in terms of RCT are very significant. The optimal value of N for the CPM scheme is 35 ($\kappa = 1.6$) which gives RCT = 61 against a value of RCT = 14100 for the PM scheme. As a consequence, PM would take more than 200 times as long in computational time to produce an estimate of the posterior mean of θ of the same accuracy.

We next investigate the performance of CPM when T and $N=\beta\sqrt{T}$ vary while ψ , equivalently ρ , is scaled such that κ is approximately constant. The results are recorded in Table 1. They suggest that the scaling $N=\beta\sqrt{T}$ is successful as IF_{CPM} appears to stabilize whereas the scaling $N=\beta T$ would be necessary for IF_{PM} to stabilize. Experimental results not reported here confirm that if N grows at a slower rate than \sqrt{T} , then IF_{CPM} increases without bound with T.

We now justify empirically some of the assumptions made in Section 4 on how to select the parameters ψ and β . First, we show that the CPM process can be thought of as a combination of two different processes: a 'slow' moving component $f(\mathsf{U}_n) \approx \hat{f}(\mathsf{U}_n) = \hat{\theta}_T + \overline{\Sigma} T^{-1} \Psi(\hat{\theta}_T, \mathsf{U}_n)$, the modified score error associated to the score error $\Psi(\hat{\theta}_T, \mathsf{U}_n)$ defined in (40), and a 'fast' component $g(\vartheta_n, \mathsf{U}_n) = \vartheta_n - f(\mathsf{U}_n) \approx \hat{g}(\vartheta_n, \mathsf{U}_n) = \vartheta_n - \hat{f}(\mathsf{U}_n)$. We display these components for a CPM run and the associated correlograms in Figure 4 for

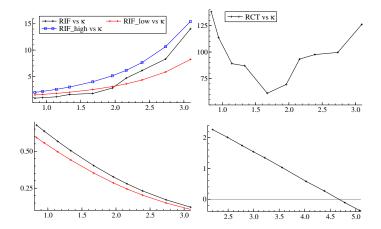


Figure 3: Random effects model using CPM: T=8192, ρ fixed and various N. RIF_{CPM} (top left) and RIF_{Q*} for IF $(h,Q_{\rm EX})=1$ and IF $(h,Q_{\rm EX})=\infty$ against κ , see Corollary 6. RCT_{CPM} against κ (top right). The acceptance probability of the CPM and the theoretical lower bound, of (37), against κ (bottom left). $\log(\kappa^2)$ against $\log(N)$ (bottom right).

T	N	ρ	κ^2	$ar{arrho}_{ m MH}$	$\mathrm{IF}_{\mathrm{MH}}$	$\bar{\varrho}_{\mathrm{CPM}}$	IF_{CPM}	RIF_{CPM}
1024	19	0.9894	2.0	0.71	10.71	0.48	43.26	4.04
2048	28	0.9925	1.9	0.69	8.21	0.49	38.50	4.61
4096	39	0.9947	1.7	0.72	11.75	0.51	21.01	1.79
8192	56	0.9962	1.8	0.81	15.61	0.50	24.25	1.55
16384	79	0.9974	1.8	0.70	9.37	0.50	20.05	2.14

Table 1: Random effects model. Inefficiency and acceptance probabilities for MH and CPM, $N = \beta \sqrt{T}$ and ρ selected such that κ^2 is approximately constant.

fixed κ . We also illustrate in Figure 5 that IF(Ψ,Q_T) $\approx A/(\delta_T \varrho_U(\kappa))$ where $\delta_T = \psi N_T/T = -\log \rho_T$. The mixing of the score error deteriorates as ρ approaches one. For fixed κ , as N increases then ρ decreases so the autocorrelation function of the score decays faster and, additionally, the variability of the modified score error decreases. Hence its contribution to the autocorrelation function of the CPM scheme decreases. The optimization scheme developed in Section 4.4 essentially selects β such that the asymptotic variances of both the slow and fast components $\hat{f}(U_n)$ and $\hat{g}(\vartheta_n, U_n)$ are of the same order.

To apply the optimization procedure, we first run the algorithm for N=20 and tune ψ to get $\hat{\kappa}\approx 1.4$. For the resulting value $\hat{\psi}$, we then evaluate ${\rm CT_{CPM}}=N\times {\rm IF_{CPM}}$ for various values of β and perform a regression based on (46)-(47). Practically, we can only use a subset of the data to perform this optimization to speed up computation. The results are fairly insensitive to the size of this subset as illustrated in Figure 6 and suggest selecting β around 0.25.

5.2 Heston stochastic volatility model

We investigate here the empirical performance of CPM on the Heston model [27, 11], a popular stochastic volatility model with leverage which is a partially observed diffusion model. The logarithm of observed price P(t) evolves according to

$$d \log P(t) = \sigma(t) dB(t),$$

$$d\sigma^{2}(t) = \upsilon \left\{ \mu - \sigma^{2}(t) \right\} dt + \omega \sigma(t) dW(t),$$

where $\sigma(t)$ is a stationary latent spot stochastic volatility process such that $\sigma^2(t) \sim \mathcal{G}(\alpha, \beta)$ where $\mathcal{G}(\alpha, \beta)$ is the gamma distribution of shape $\alpha = 2\mu v/\omega^2$ and rate $\beta = 2v/\omega^2$. The Brownian motions B(t) and W(t) are correlated with $\chi = \text{corr}\{B(t), W(t)\}$. We shall suppose that the log prices are observed at equally spaced times $\tau_0 < \cdots < \tau_T$, where $\Delta = \tau_s - \tau_{s-1}$ for all s and we denote $Y_s = t$

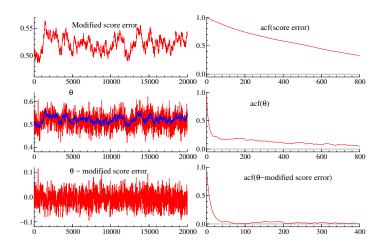


Figure 4: Random effects model using CPM: $T=2,560,~\beta=0.12,~N=6,~\rho=0.9977$. Top: modified score error $\hat{f}(\mathsf{U}_n)$ (left) and its correlogram (right). Middle: parameter ϑ_n (red) and modified score error (blue) (left) and correlogram ϑ_n (right). Bottom: residual $\hat{g}(\vartheta_n,\mathsf{U}_n)$ (left) and correlogram (right).

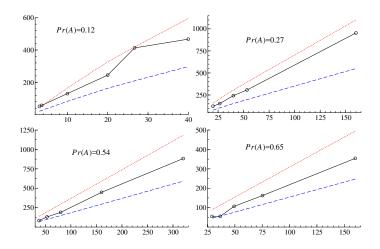


Figure 5: Random effects model using CPM: T=320. Inefficiency of the score error (black line) plotted against $1/\delta$ for four different values of $\kappa^2=9.5, 4.9, 1.42, 0.75$ from top left to bottom right clockwise and corresponding acceptance probability $\bar{\varrho}_{\text{CPM}}$. Upper bound $2/(\delta\bar{\varrho}_{\text{CPM}})$ (dotted red) and lower bound $1/(\delta\bar{\varrho}_{\text{CPM}})$ (dotted blue).

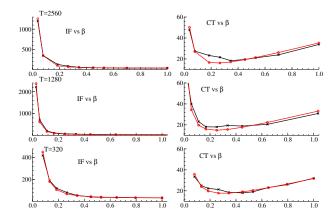


Figure 6: Random effects model. IF and CT as a function of β . Top to bottom: T=2560, 1280, 320. Left: IF = IACT vs β . Right: CT = IF × β vs β . The regression fit based upon estimated CT is included in red.

 $\log P(\tau_s) - \log P(\tau_{s-1})$ for s = 1, ..., T. Conditional on the volatility and driving processes $\sigma^2(t)$ and W(t), the distribution of these returns is given by

$$Y_s \sim \mathcal{N}\left\{\chi \gamma_s; (1 - \chi^2)\sigma_s^{2*}\right\},\tag{50}$$

$$\sigma_s^{2*} = \int_{\tau_{s-1}}^{\tau_s} \sigma^2(t) dt, \quad \gamma_s = \int_{\tau_{s-1}}^{\tau_s} \sigma(t) dW(t). \tag{51}$$

To perform inference, we first reparameterise the model in terms of $x(t) = \log \sigma^2(t)$. We apply Itô's lemma to x(t) and discretize the resulting diffusion using an Euler scheme. We denote by $x_k^s = x (\tau_s + \epsilon i)$ where $\epsilon = \triangle/K$ for i = 0, ..., I so that $x_I^s = x_0^{s+1}$. The evolution of these latent variables is given by

$$x_{i+1}^{s} = x_{i}^{s} + \epsilon \left[v \left\{ \mu e^{-x_{i}^{s}} - 1 \right\} - \frac{\omega^{2}}{2} e^{-x_{i}^{s}} \right] + \sqrt{\epsilon} \omega e^{-x_{i}^{s}/2} \eta_{i},$$

where $\eta_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}\left(0,1\right)$ for i=0,...,I-1. Under the Euler scheme, the distribution of the returns is given by

$$Y_s \sim \mathcal{N}\left\{\chi \widehat{\gamma}_s; (1-\chi^2)\widehat{\sigma}_s^{2*}\right\},\tag{52}$$

$$\widehat{\sigma}_s^{2*} = \epsilon \sum_{i=1}^{I} \exp(x_i^s), \quad \widehat{\gamma}_t = \sqrt{\epsilon} \sum_{i=1}^{I} \exp(x_i^s/2) \eta_i.$$
 (53)

where $\hat{\sigma}_t^{2*}$ and $\hat{\gamma}_t$ are the Euler approximations to the corresponding expressions in (51). We are interested in inferring $\theta = (\mu, v, \omega, \chi)$ given T = 4,000 daily returns $y_{1:T}$ from the S&P 500 index from 15/08/1990 to 03/07/2006. We use here I = 10. Although the state is scalar, it is very difficult to perform inference using standard MCMC techniques as this involves $T \times I = 40000$ highly correlated latent variates.

We first run the CPM by keeping the parameter fixed at the posterior mean $\hat{\theta}$, estimated from a full CPM run, and only updating the auxiliary variables. We display the histograms of $Z = \log \hat{p}(y_{1:T} \mid \hat{\theta}, U)$, $W = \log \hat{p}(y_{1:T} \mid \hat{\theta}, U')$ and $R = \log \{\hat{p}(y_{1:T} \mid \hat{\theta}, U')/\hat{p}(y_{1:T} \mid \hat{\theta}, U)\}$ in Figure 7 for N = 80 and N = 300 using the parameters given in Table 2. We observe that R is approximately distributed according to $\mathcal{N}(-\kappa^2/2, \kappa^2)$ for $\kappa = 1.35$ in both cases. Additionally the sequence of estimates is almost uncorrelated across CPM iterations.

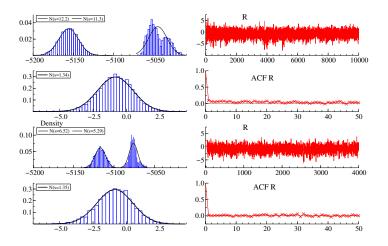


Figure 7: Histograms of W, Z for N=80 (1st left), N=300 (3rd left), histograms of R for N=80 (2nd left), N=300 (4th left). R across CPM iterations and associated correlograms for N=80 (1st right, 2nd right), N=300 (3rd right, 4th right).

We then run the CPM using a random walk proposal. Using N=300, we first select $\psi=0.125$ to get the standard deviation of the loglikelihood ratio estimator at $\hat{\theta}$ around $\kappa=1.4$. We then run the CPM schemes for other values of N, $N=\beta\sqrt{T}$, and compute $\mathrm{CT}=N\times\mathrm{IF}$. These results are summarized in Table 2. The posterior estimates are in very close agreement across the different values of N. In unreported results, we observe empirically that the dependence of CT on of β for parameters $(\mu,\phi:=e^{-v},\omega,\chi)$, matches (46) which can be optimized, suggesting that an optimal value of N around 70-80. As in the random effect scenario, we also observe on datasets of increasing length that the scaling

 $N = \beta \sqrt{T}$ is successful as IF_{CPM} appears to stabilize. In this context, the PM procedure is extremely expensive computationally as we need approximately N = 20000 to obtain a standard deviation of Z around one [18], our implementation taking 7 minutes per iteration to run on a standard desktop. In terms of CT, CPM is approximately 100 times more efficient than PM.

$\mathbb{E}(\theta) (\mathrm{SD}(\theta))$	μ	φ	ω	χ	CPM ρ
N = 80	1.258 (0.098)	0.981 (0.0027)	0.142 (0.0099)	-0.676 (0.027)	0.9975
N = 150	1.253 (0.098)	$0.981 \ (0.0028)$	$0.142\ (0.0105)$	-0.672 (0.034)	0.9953
N = 300	1.255 (0.099)	$0.981 \ (0.0028)$	$0.142\ (0.0110)$	-0.671 (0.032)	0.9907
$CT(\theta)$	μ	ϕ	ω	χ	$\bar{\varrho}_{\mathrm{CPM}}$
N = 80	9,995	12,555	13,571	33,794	0.276
N = 150	19,691	$20,\!256$	17,931	$32,\!588$	0.272
N = 300	32,970	30,432	35,103	35,505	0.281

Table 2: Heston model. Posterior means and standard deviations over 10,000 iterations (top). CT = IF $\times N$ for the CPM scheme for $N = \beta \sqrt{T}$ and ρ selected such that $\kappa \approx 1.4$ at $\hat{\theta}$.

5.3 Linear Gaussian state-space model

We examine empirically the performance of the CPM for multivariate state-space models using the particle filter with Hilbert sort described in Algorithm 2 and compare it to the PM procedure. Attention is restricted to a linear Gaussian state-space model which allows exact calculation of the likelihood and of the loglikelihood error $Z_T(\theta, U) = \log \hat{p}(Y_{1:T} \mid \theta, U) - \log p(Y_{1:T} \mid \theta)$. Very similar empirical results for non-linear non-Gaussian state-space models were observed.

We consider the model discussed in [26, 29] where $\{X_t; t \geq 1\}$ and $\{Y_t; t \geq 1\}$ are \mathbb{R}^k -valued with

$$X_1 \sim \mathcal{N}(0, I_n), \qquad X_{t+1} = A_{\theta} X_t + V_{t+1}, \qquad Y_t = X_t + W_t,$$
 (54)

where $V_{t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0_{k}, I_{k}\right), W_{t} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0_{k}, I_{k}\right) \text{ and } A_{\theta}^{i,j} = \theta^{|i-j|+1}.$

We use for the proposal density within the particle filter the transition density of $\{X_t; t \geq 1\}$. We first examine the achieved correlation between successive draws of $Z = \log \{\widehat{p}(y_{1:T} \mid \theta, U)/p(y_{1:T} \mid \theta)\}$ by running the CPM procedure holding the parameter fixed and equal to its true value $\theta = 0.4$. This allows for the examination of the variance of $R = \log \{\widehat{p}(y_{1:T} \mid \theta', U')/p(y_{1:T} \mid \theta)\} - Z$ where $U' \sim K_{\rho}(U, \cdot)$ is the proposal when $\theta' = \theta$. We examine the results for various values of T, with $N = \lceil \beta T^{\alpha} \rceil$ and $\rho = \exp(-\psi N/T)$ for $k \in \{2, 3, 4\}$.

We will now discuss the choice of α for state-space models. In sharp contrast to random effects models, we found empirically that there are dimension dependent limitations to the realized correlation that can be achieved through the particle filter with Hilbert sort. In particular we found that, due to resampling, the realized correlation is limited by min $\{1-c_1N^{-1/k}, 1-c_2\delta\}$ for some constants c_1, c_2 , unless we set δ extremely small. A back of the envelope calculation suggests that to overcome this limitation we would need to choose δ around $N^{-2/k}$. Since $\delta \propto N/T$ this would result in choosing $N \propto T^{k/(k+2)}$ in which case empirical results suggest that the number of particles is too small to control κ^2 , the variance of the loglikelihood ratio error. Since the inefficiency tends to increase if we set δ too small, the above considerations suggest that we set $\delta = N^{-1/k}$, or equivalently scaling $N \propto T^{k/(k+1)}$. Thus for the following examples we set $\alpha = k/k + 1$.

We run the simulated chain for 1000 iterations recording $\kappa^2 = \mathbb{V}(R)$ and $\sigma^2 = \mathbb{V}(Z)$. The values of β and ψ have been chosen so that they result in a particular target value of κ^2 as will be evident from the following tables. The asymptotic acceptance probability of the CPM scheme is thus in this case given by $\varrho_{\text{CPM}}(\kappa) := \varrho_{\text{U}}(\kappa) = 2\Phi(-\kappa/2)$ while it is $\varrho_{\text{PM}}(\sigma) = 2\Phi(-\sigma/\sqrt{2})$ for the PM [18].

The results for k=2 are reported in Table 3, where the two eigenvalues of A_{θ} are 0.56 and 0.24. It is clear that the proposed scaling rules result in values of κ^2 which are approximately constant, remaining at values of around 2 for $T \geq 1600$. The implied acceptance probability of the CPM scheme $\varrho_{\text{CPM}}(\kappa)$ therefore settles at a value just below 0.5. By contrast the marginal variance σ^2 increases at the expected rate $T^{1-\alpha}$ and accordingly the implied acceptance probability $\varrho_{\text{PM}}(\sigma)$ deteriorates rapidly,

even for T=100. Similar results are found for the case k=3, reported in Table 4, where the eigenvalues of Λ are (0.6605,0.3360,0.2035) resulting in a model with moderately high persistence. In this case we set $\alpha=3/4$. Although less dramatic, the implied gain of the CPM method over the PM is substantial even for T=100 and increases as T goes up. The variance κ^2 appears again to stabilize at a value less than 3

State dimension $k=2$ with $\beta=0.854,\psi=0.12,\alpha=2/3$							
\overline{T}	N	$\delta = -\log \rho$	κ^2	σ^2	$\varrho_{\mathrm{CPM}}\left(\kappa\right)$	$\varrho_{\mathrm{PM}}\left(\sigma\right)$	
100	18	0.0216	2.59	16.3	0.42	0.004	
400	46	0.0138	2.71	20.5	0.41	0.0013	
1600	116	0.0087	2.01	34.1	0.48	3.6×10^{-5}	
6400	294	0.0055	2.07	49.7	0.47	6.0×10^{-7}	
25600	742	0.0034	1.97	105.9	0.48	3.4×10^{-13}	

Table 3: Linear state-space model. Results for k = 2 for varying T.

State dimension $k = 3$ with $\beta = 1.57$, $\psi = 0.042$, $\alpha = 3/4$							
\overline{T}	N	$\delta = -\log \rho$	κ^2	σ^2	$\varrho_{\mathrm{CPM}}\left(\kappa\right)$	$\varrho_{\mathrm{PM}}\left(\sigma\right)$	
100	49	0.0205	3.15	13.7	0.37	0.0089	
400	140	0.0147	2.97	16.6	0.39	0.0039	
1600	397	0.0104	3.44	26.7	0.35	0.00025	
6400	1124	0.0074	3.03	34.1	0.38	3.66×10^{-5}	
25600	3181	0.0052	2.69	49.4	0.41	6.74×10^{-7}	

Table 4: Linear state-space model. Results for k = 3 for varying T.

The full CPM procedure is now implemented for T = 400 and T = 6400 when k = 2 and k = 3 using the parameters of Tables 3 and 4. An autoregressive proposal in the Metropolis algorithm is employed for θ which is based on the posterior mode and the second derivative at this point [52].

The results for k=3 and T=6400 are shown in Figure 8. The mixing for θ is fairly rapid from the achieved value of $\kappa=2.26$. The empirical distributions of Z under m and $\bar{\pi}$ are plotted (middle left) and are close to the theoretical distributions $\mathcal{N}\left(-\sigma^2/2,\sigma^2\right)$ and $\mathcal{N}\left(\sigma^2/2,\sigma^2\right)$ respectively, where $\sigma=7.5$. The middle right plot and the third row show the draws of R, the empirical distribution and the associated correlogram arising from the CPM scheme. It is clear that R is approximately distributed according to $\mathcal{N}\left(-\kappa^2/2,\kappa^2\right)$ for some κ , which is overlaid, but the correlations across iterations vanish slowler than for random effect models and one-dimensional state-space models. The gain over the PM method is around σ^2 meaning we need around 50 times as many particles in the PM method to achieve similar results to the CPM scheme. When T=400, we obtained $\kappa=1.92$ and $\sigma=4.30$ resulting in gains over the PM of approximately 18 fold. When k=2, the gains are more impressive and are around 25 fold for T=400 and 80 fold when T=6400.

6 Discussion

The CPM method is a generic extension of the PM method based on an estimator of the likelihood ratio appearing in its acceptance probability which is obtained by correlating the estimators of its numerator and denominator. We have detailed two implementations of this idea for random effects and state-space models. For random effects models, we have provided theory to perform an efficient implementation of the methodology and have verified empirically that this methodology is also useful for state-space models. In our examples, the efficiency of computations using the CPM relative to the PM method increases with T and is improved by more than two order orders of magnitude for large data sets. This methodology is particularly beneficial in scenarios such as partially observed diffusions where sophisticated MCMC alternatives, such as particle Gibbs techniques, are inefficient.

From a theoretical point of view, we have obtained for random effects models a result suggesting that a necessary condition to ensure finiteness of the IACT of the CPM parameter sequence as T increases is to have N_T growing at least at rate \sqrt{T} . Our experimental results suggest that this condition is also sufficient for a large class of functions and thus that the computational complexity of CPM for random

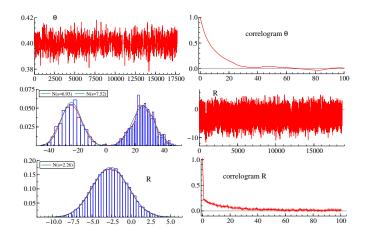


Figure 8: The CPM results for the 3-dimensional state space model with T=6400. Top: parameter samples (left) and corresponding correlogram (right). Middle: Histograms of Z arising from m and $\bar{\pi}$ (left), draws of R (right). Bottom: Histogram of R (left) and correlogram (right).

effects models is $O(T^{\frac{3}{2}})$ versus $O(T^2)$ for PM. For state-space models, our empirical results indicate that this scaling degrades with the state dimension k and that we need N_T to grow at rate $T^{\frac{k}{k+1}}$, suggesting that the computational complexity of CPM in this context is thus $O(T^{\frac{2k+1}{k+1}})$ up to a logarithmic factor³ versus $O(T^2)$ for PM. It would be of interest but technically very involved to establish these results rigorously.

From a methodological point of view, it is possible in the state-space context to use alternatives to the Hilbert resampling sort to implement the CPM algorithm [40, Section 6], [33] and several such methods have been proposed (for example [29], [47]) following the first version of this work (arXiv:1511.04992). Our empirical results suggest that all these procedures provide roughly similar improvements over the PM. It could also be beneficial to use the sequential randomised Quasi Monte Carlo (QMC) algorithm proposed in [23], [11] within the CPM scheme by correlating the single uniform used to randomize the QMC grid. In a random effects context, it has already be demonstrated that this can provide significant improvements [51]. Finally, a sequential extension of the particle marginal Metropolis–Hastings algorithm [1], a PM method, has been proposed in [12] and it would be interesting to develop an efficient sequential version of CPM.

7 Acknowledgments

Arnaud Doucet's research is partially supported by the Engineering and Physical Sciences Research Council, grant EP/K000276/1. We thank Sebastian Schmon for his comments.

References

- [1] Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society, Series B* **72**, 269–342.
- [2] Andrieu, C., Doucet, A. and Lee, A. (2012). Discussion of "Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation" by P. Fearnhead and D. Prangle. *Journal of the Royal Statistical Society, Series* B **72**, 451–452.
- [3] Andrieu, C. and Roberts G.O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics* **37**, 697–725.
- [4] Andrieu, C. and Vihola, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *The Annals of Applied Probability* **25**, 1030–1077.

³The particle filter with Hilbert sort has computational complexity $N_T \log N_T$ per observation.

- [5] Beaumont, M. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164**, 1139–1160.
- [6] Bérard, J., Del Moral, P. and Doucet, A. (2014). A lognormal central limit theorem for particle approximations of normalizing constants. *Electronic Journal of Probability* **19**, 1–28.
- [7] Berti, P., Pratelli, L. and Rigo, P. (2006). Almost sure weak convergence of random probability measures. *Stochastics* **78**, 91–97.
- [8] Billingsley, P. (1968). Convergence of Probability Measures. Wiley: New York.
- [9] Carpenter, J., P. Clifford and Fearnhead, P. (1999). Improved particle filter for nonlinear problems. *IEE Proceedings-F* **146**, 2–7.
- [10] Ceperley, D.M. and Dewing, M. (1999). The penalty method for random walks with uncertain energies. *Journal of Chemical Physics* **110**, 9812–9820.
- [11] Chopin, N. and Gerber, M. (2017). Sequential quasi-Monte Carlo: Introduction for non-experts, dimension reduction, application to partly observed diffusion processes. arXiv preprint arXiv:1706.05305v1.
- [12] Chopin, N., Jacob, P.E. and Papaspiliopoulos, O. (2013). SMC²: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society, Series* B **75**, 397–426.
- [13] Crauel, H. (2003). Random Probability Measures on Polish Spaces. CRC Press.
- [14] Dalhin, J., Lindsten, F., Kronander, J. and Schön, T.B. (2015). Accelerating pseudo-marginal Metropolis-Hastings by correlating auxiliary variables. arXiv preprint arXiv:1511.05483.
- [15] Del Moral, P. (2004). Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications. Springer-Verlag: New York.
- [16] Douc, R., Moulines, E. and Stoffer, D.S. (2014). Nonlinear Time Series. CRC Press.
- [17] Doucet, A., Jacob, P.E. and Rubenthaler, S. (2013). Derivative-free estimation of the score vector and observed information matrix with applications to state-space models. arXiv:1304:5768.
- [18] Doucet, A., Pitt, M.K., Deligiannidis, G. and Kohn, R. (2015). Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika* **102**, 295-313.
- [19] Durrett, R. (2010). Probability: Theory and Examples. Cambridge University Press.
- [20] Ethier, S.N. and Kurtz, T.G. (2005). Markov Processes: Characterization and Convergence. Wiley: New York.
- [21] Flury, T. and Shephard, N. (2011). Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. *Econometric Theory* 27, 933–956.
- [22] Gentil, I. and Rémillard, B. (2008). Using systematic sampling selection for Monte Carlo solutions of Feynman-Kac equations. Advances in Applied Probability 40, 454-472.
- [23] Gerber, M. and Chopin, N. (2015). Sequential quasi Monte Carlo (with discussion). *Journal of the Royal Statistical Society, Series B* 77, 509-579.
- [24] Geyer, C.J. (1992). Practical Markov chain Monte Carlo. Statistical Science 7, 473-483.
- [25] Gordon, N. J., Salmond, D. and Smith, A.F.M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F*, **140**, 107–113.
- [26] Guarniero, P., Johansen, A.M., and Lee, A. (2017). The iterated auxiliary particle filter. *Journal of the American Statistical Association*, to appear.
- [27] Heston, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bound and currency options. *The Review of Financial Studies*, **6**, 327–343.
- [28] Ionides, E.L., Breto, C. and King, A.A. (2006). Inference for nonlinear dynamical systems. Proceedings of the National Academy of Science USA 103, 18438–18443.
- [29] Jacob, P.E., Lindsten, F. and Schön, T.B. (2016). Coupling of particle filters. arXiv preprint arXiv:1606.01156.
- [30] Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40**:2, 633-643.
- [31] Johndrow, J.E., Smith, A., Pillai, N.S. and Dunson, D.B. (2016). Inefficiency of data augmentation for large sample imbalanced data. arXiv preprint arXiv:1605.05798.
- [32] Kipnis, C. and Varadhan S. R. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics* **104**, 1-19.
- [33] L'Ecuyer, P., Munger, D., Lécot, C. and Tuffin, B. (2016). Sorting methods and convergence rates for array-RQMC: Some empirical comparisons. *Mathematics and Computers in Simulation*, to appear.
- [34] Lee, A. (2008). Towards smooth particle filters for likelihood estimation with multivariate latent

- variables. M.Sc. Thesis, Department of Computer Science, University of British Columbia.
- [35] Lee, A. and Holmes, C.C. (2010). Discussion of "Particle Markov chain Monte Carlo methods" by C. Andrieu, A. Doucet and R. Holenstein. *Journal of the Royal Statistical Society, Series B* 72, 327.
- [36] Lin, L., Liu, K.F., Sloan, J. (2000). A noisy Monte Carlo algorithm. Physical Review D 61, 074505.
- [37] Lindsten, F. and Doucet, A. (2016). Pseudo-marginal Hamiltonian Monte Carlo. arXiv preprint arXiv:1607.02516.
- [38] Lindsten, F., Jordan, M.I. and Schön, T.B. (2014). Particle Gibbs with ancestor sampling. *Journal of Machine Learning Research* 15, 2145–2184.
- [39] Liu, J.S. (2001). Monte Carlo Strategies in Scientific Computing. Springer-Verlag, New York.
- [40] Malik, S. and Pitt, M.K. (2011). Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics* **165**, 190-209.
- [41] Nicholls, G.K., Fox, C. and Watt, A.M. (2012). Coupled MCMC with a randomized acceptance probability. arXiv preprint arXiv:1205.6857.
- [42] Papaspiliopoulos, O., Roberts, G.O. and Sköld, M. (2007). A general framework for the parametrization of hierarchical models. *Statistical Science* **22**, 59–73.
- [43] Peskun, P.H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60**, 607–612.
- [44] Pitt, M.K. and Shephard, N. (1999). Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association* **94**, 590-599.
- [45] Pitt, M.K., Silva, R., Giordani, P. and Kohn, R. (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics* **171**, 134–151.
- [46] Roberts, G.O., Gelman, A. and Gilks, W.R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* 7, 110–120.
- [47] Sen, D., Thiery, A.H. and Jasra, A. (2017). On coupling particle filter trajectories. arXiv preprint arXiv:1606.01016, to appear in *Statistics and Computing*.
- [48] Sherlock, C., Thiery, A., Roberts, G.O. and Rosenthal, J.S. (2015). On the efficiency of pseudomarginal random walk Metropolis algorithms. *The Annals of Statistics* **43**, 238–275.
- [49] Stein, C.M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics* **9**, 1135–1151.
- [50] Tierney, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *The Annals of Applied Probability* **8**, 1–9.
- [51] Tran, M-N., Kohn, R., Quiroz, M. and Villani, M. (2017). The block pseudo-marginal sampler. arXiv preprint arXiv:1603.024845v4.
- [52] Tran, M-N., Pitt, M.K. and Kohn, R. (2016). Adaptive Metropolis-Hastings sampling using reversible dependent mixture proposals. *Statistics and Computing* **26**, 361–381.
- [53] van der Vaart, A.W. (2000). Asymptotic Statistics. Cambridge University Press.
- [54] Zakai, M. (1967). Some moment inequalities for stochastic integrals and for solutions of stochastic differential equations. *Israel Journal of Mathematics* 5, 170–176.

A Supplementary Material

A.1 Notation

We define a reference probability space $(\Omega_T, \mathcal{G}_T, \mathbb{P}_T)$ which supports the following random variables:

- 1. $\theta^T \sim \pi_T$ where π_T denote the posterior distribution associated to observations $y_{1:T}$,
- 2. $\{U_{t,i}^T: t \in 1: T, i \in 1: N\}$ independent and identically distributed $\mathcal{N}(0_p, I_p)$ random variables,
- 3. $\{B_{t,i}^T(\cdot): t \in 1: T, i \in 1: N\}$ where the $B_{t,i}^T(\cdot)$ are mutually independent, p-dimensional standard Brownian motions.

We set

$$\Omega_T := \Theta \times \mathbb{R}^{pNT} \times C^p[0, \infty)^{NT},$$

and

$$\mathbb{P}_T\left(\mathrm{d}\theta^T, \left\{\mathrm{d}u_{t,i}^T\right\}_{t,i}, \left\{\mathrm{d}\left(\beta_{t,i}^T(\cdot)\right)\right\}_{t,i}\right) = \pi_T(\mathrm{d}\theta^T) \prod_{t=1}^T \prod_{i=1}^N \varphi(\mathrm{d}u_{t,i}^T; 0_p, I_p) \prod_{t=1}^T \prod_{i=1}^N \mathbb{W}^p\left(\mathrm{d}\beta_{t,i}^T(\cdot)\right),$$

where $\mathbb{W}^p(\mathbf{d}\cdot)$ denotes the Wiener measure on $C^p[0,\infty)$, $C^p[0,\infty)$ being the space of \mathbb{R}^p - valued continuous paths on $[0,\infty)$.

Let $\mathfrak{Y} = \bigotimes_{t=1}^{\infty} \mathsf{Y}$ where Y is a topological space, y_t is Y -valued and $\mathcal{B}(\mathfrak{Y})$ the associated Borel σ -algebra. Then we consider the product space $(\Omega, \mathcal{G}, \mathbb{P})$ where

$$\Omega = \mathfrak{Y} \times \prod_{T} \Omega_{T}, \ \mathcal{G} = B(\mathfrak{Y}) \otimes (\otimes_{T} \mathcal{G}_{T}),$$

and

$$\mathbb{P} = \left(\prod_{t=1}^{\infty} \mu(\mathrm{d}y_t)\right) \otimes (\otimes_T \mathbb{P}_T).$$

In most cases we will be working with the probability measure $\widetilde{\mathbb{P}}$ capturing the scenario when the CPM algorithm is in the stationary regime, which is defined as follows. For every $T \geq 1$ and sequence of observations $y_{1:T}$, we define the probability measure $\widetilde{\mathbb{P}}_T^{y_{1:T}}$ by

$$\frac{\mathrm{d}\widetilde{\mathbb{P}}_{T}^{y_{1:T}}}{\mathrm{d}\mathbb{P}_{T}}\left(\theta^{T},\left\{u_{t,i}^{T}\right\}_{i,t},\left\{\beta_{t,i}^{T}(\cdot)\right\}_{t,i}\right) = \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} \varpi(y_{t};\theta^{T},u_{t,i}^{T}),$$

and let

$$\widetilde{\mathbb{P}} = \left(\prod_{t=1}^{\infty} \mu(\mathrm{d}y_t) \right) \otimes \left(\otimes_T \widetilde{\mathbb{P}}_T^{y_{1:T}} \right).$$

We will denote by \mathbb{E} , \mathbb{V} and $\widetilde{\mathbb{E}}$, $\widetilde{\mathbb{V}}$ the expectation and variance under \mathbb{P} and $\widetilde{\mathbb{P}}$ respectively.

When T and θ^T are understood fixed, allowing some abuse of notation, we will write \mathbb{P} to denote the measure

$$\mathbb{P}\left(\mathrm{d}y_1,\ldots,\mathrm{d}y_T,\left\{\mathrm{d}u_{t,i}^T\right\}_{t,i},\left\{\mathrm{d}\left(\beta_{t,i}^T(\cdot)\right)\right\}_{t,i}\right) = \prod_{t=1}^T \mu(\mathrm{d}y_t) \prod_{t=1}^T \prod_{i=1}^N \varphi(\mathrm{d}u_{t,i}^T;0_p,I_p) \prod_{t=1}^T \prod_{i=1}^N \mathbb{W}^p\left(\mathrm{d}\beta_{t,i}^T(\cdot)\right),$$

and similarly

$$\begin{split} &\widetilde{\mathbb{P}}\left(\mathrm{d}y_{1},\ldots,\mathrm{d}y_{T},\left\{\mathrm{d}u_{t,i}^{T}\right\}_{t,i},\left\{\mathrm{d}\left(\beta_{t,i}^{T}(\cdot)\right)\right\}_{t,i}\right) \\ &= \bar{\pi}_{T}\left(\left\{\mathrm{d}u_{t,i}^{T}\right\} \mid \theta^{T}\right)\prod_{t=1}^{T}\mu(\mathrm{d}y_{t})\prod_{t=1}^{T}\prod_{i=1}^{N}\mathbb{W}^{p}\left(\mathrm{d}\beta_{t,i}^{T}(\cdot)\right) \\ &= \prod_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\varpi(y_{t};\theta^{T},u_{t,i}^{T})\prod_{t=1}^{T}\prod_{i=1}^{N}\varphi(\mathrm{d}u_{t,i}^{T};0_{p},I_{p})\prod_{t=1}^{T}\mu(\mathrm{d}y_{t})\prod_{t=1}^{T}\prod_{i=1}^{N}\mathbb{W}^{p}\left(\mathrm{d}\beta_{t,i}^{T}(\cdot)\right). \end{split}$$

To ease notation, we will often drop the superscript T, since we will always be considering variables belonging to the same row. In addition we will write N for N_T in the proofs, omitting the explicit dependence of N_T on T. In the proofs of Theorem 1, Theorem 2 and Theorem 3, we also write m, $\bar{\pi} \left(\mathrm{d}u \mid \theta \right)$, $B_{t,i}$, $U_{t,i}$ instead of m_T , $\bar{\pi}_T \left(\mathrm{d}u^T \mid \theta^T \right)$, $B_{t,i}^T$ and $U_{t,i}^T$. Notice that $\mathbb{E} \left(\varpi \left(Y_1, U_{1,1}^T; \theta \right)^j \right)$ is independent of T for any j as $U_{1,1}^T \sim \mathcal{N}(0_p, I_p)$ under \mathbb{P} .

A.2 Proof of Part 1 of Theorem 1

The starting point of our analysis is the following decomposition

$$\log \widehat{p}(Y_{1:T}|\theta) - \log p(Y_{1:T}|\theta) = \sum_{t=1}^{T} \log \left\{ 1 + \frac{\varepsilon_N(Y_t;\theta)}{\sqrt{N}} \right\}$$
 (55)

with

$$\varepsilon_{N}(Y_{t};\theta) := \sqrt{N} \frac{\widehat{p}(Y_{t}|\theta) - p(Y_{t}|\theta)}{p(Y_{t}|\theta)} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{ \varpi(Y_{t}, U_{t,i};\theta) - 1 \right\}.$$

We will denote by $\rho_i(\theta)$ the i^{th} order cumulant of the normalized importance weight $\varpi(Y_1, U_{1,1}; \theta)$ given in (18) under \mathbb{P} and by $\gamma(\theta)^2$ its variance, so that $\rho_2(\theta) = \gamma(\theta)^2$.

We first establish three preliminary lemmas.

Lemma 8. The terms $\{\varepsilon_N(y_t;\theta)\}_{t=1}^T$ are independent with for any $y \in Y$, $\mathbb{E}(\varepsilon_N(y_t;\theta)) = 0$ and

$$\mathbb{E}\left(\varepsilon_{N}\left(y;\theta\right)^{2}\right) = \mathbb{V}\left(\varpi(y,U_{1,1};\theta)\right) = \rho_{2}\left(y;\theta\right) := \gamma\left(y;\theta\right)^{2},\tag{56}$$

$$\mathbb{E}\left(\varepsilon_N\left(y;\theta\right)^3\right) = \frac{\rho_3\left(y;\theta\right)}{\sqrt{N}},\tag{57}$$

$$\mathbb{E}\left(\varepsilon_{N}\left(y;\theta\right)^{4}\right) = 3\gamma\left(y;\theta\right)^{4} + \frac{\rho_{4}\left(y;\theta\right)}{N},$$

$$\mathbb{E}\left(\varepsilon_{N}\left(y;\theta\right)^{5}\right) = \frac{10\rho_{2}\left(y;\theta\right)\rho_{3}\left(y;\theta\right)}{\sqrt{N}} + \frac{\rho_{5}\left(y;\theta\right)}{N\sqrt{N}},$$
(58)

where $\rho_i(y;\theta)$ denotes the ith-order cumulant of $\varpi(y,U_{1,1};\theta)$ and $\rho_i(\theta) = \mathbb{E}(\rho_i(Y;\theta)) = \int \rho_i(y;\theta) \, \mu(\mathrm{d}y)$.

The proof of Lemma 8 follows from direct calculations so it is omitted.

Lemma 9. For any
$$k \geq 2$$
, if $\mathbb{E}\left(\varpi(Y_1, U_{1,1}; \theta)^k\right) < \infty$ then $\limsup_{T \to \infty} \mathbb{E}\left(\left|\varepsilon_N(Y_1; \theta)\right|^k\right) < \infty$.

Proof of Lemma 9. It follows from a successive application of Marcinkiewicz-Zygmund, Jensen and C_p inequalities that for any $k \geq 2$, there exist $b(k), c(k) < \infty$ such that

$$\mathbb{E}\left(\left|\varepsilon_{N}(Y_{1};\theta)\right|^{k}\right) = \mathbb{E}\left(\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\varpi(Y_{1},U_{1,i};\theta)-1\right\}\right|^{k}\right)$$

$$\leq b\left(k\right)\mathbb{E}\left(\left|\frac{1}{N}\sum_{i=1}^{N}\left\{\varpi(Y_{1},U_{1,i};\theta)-1\right\}^{2}\right|^{k/2}\right)$$

$$\leq b\left(k\right)\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(\left|\varpi(Y_{1},U_{1,i};\theta)-1\right|^{k}\right)$$

$$= b\left(k\right)\left(\mathbb{E}\left|\varpi(Y_{1},U_{1,1};\theta)-1\right|^{k}\right)$$

$$\leq b\left(k\right)c\left(k\right)\left(\mathbb{E}\left(\varpi(Y_{1},U_{1,1};\theta)^{k}\right)+1\right).$$

This concludes the proof.

Lemma 10. Consider the triangular array $\{\varepsilon_N(Y_t;\theta)\}$ and let $k \geq 2$. If there exists $\delta > 0$ such that $\mathbb{E}\left(\varpi(Y_1,U_{1,1};\theta)^{k+\delta}\right) < \infty$ then

$$T^{-1} \sum_{t=1}^{T} \varepsilon_N(Y_t; \theta)^k - \mathbb{E}\left(\varepsilon_N(Y_1; \theta)^k\right) \stackrel{\mathbb{P}}{\to} 0.$$
 (59)

If $\mathbb{E}\left(\varpi(Y_1,U_{1,1};\theta)^{2k}\right)$ then we have for any $\lambda>0$

$$T^{-\frac{(1+\lambda)}{2}} \sum_{t=1}^{T} \varepsilon_N(Y_t; \theta)^k - T^{\frac{(1-\lambda)}{2}} \mathbb{E}\left(\varepsilon_N(Y_1; \theta)^k\right) \stackrel{\mathbb{P}}{\to} 0.$$
 (60)

Proof of Lemma 10. The results (59) follows directly from a weak law of large numbers (WLLN) applied to the triangular array $\varepsilon_N(Y_t;\theta)^k - \mathbb{E}\left(\varepsilon_N(Y_1;\theta)^k\right)$; see, e.g., [16, Theorem B.18]. This results holds as $\mathbb{E}\left(\varpi(Y_1,U_{1,1};\theta)^{k+\delta}\right) < \infty$ so $\limsup_{T\to\infty} \mathbb{E}\left(\left|\varepsilon_N(Y_1;\theta)\right|^{k+\delta}\right) < \infty$ from Lemma 9. For the second result (60), we have for any $\epsilon > 0$

$$\mathbb{P}\left\{\left|T^{-\frac{(1+\lambda)}{2}}\sum_{t=1}^{T}\left\{\varepsilon_{N}(Y_{t};\theta)^{k}-\mathbb{E}\left(\varepsilon_{N}(Y_{1};\theta)^{k}\right)\right\}\right| \geq \epsilon\right\} \leq \frac{\mathbb{E}\left[\left(\sum_{t=1}^{T}\left[\varepsilon_{N}(Y_{t};\theta)^{k}-\mathbb{E}\left(\varepsilon_{N}(Y_{1};\theta)^{k}\right)\right]\right)^{2}\right]}{T^{(1+\lambda)}\epsilon^{2}}$$

$$=\frac{\mathbb{E}\left(\left[\varepsilon_{N}(Y_{1};\theta)^{k}-\mathbb{E}\left(\varepsilon_{N}(Y_{1};\theta)^{k}\right)\right]^{2}\right)}{T^{\lambda}\epsilon^{2}}$$

$$\Rightarrow 0.$$

The result follows. \Box

We can now give the proof of Theorem 1 Part 1.

Proof of Part 1 of Theorem 1. We first perform a fourth order Taylor expansion of each term appearing in (55), i.e.

$$\log\left\{1 + \frac{\varepsilon_N(Y_t;\theta)}{\sqrt{N}}\right\} = \frac{\varepsilon_N(Y_t;\theta)}{\sqrt{N}} - \frac{\varepsilon_N(Y_t;\theta)^2}{2N} + \frac{\varepsilon_N(Y_t;\theta)^3}{3N\sqrt{N}} - \frac{\varepsilon_N(Y_t;\theta)^4}{4N^2} + R_{t,N}(Y_t;\theta)$$
(61)

where

$$R_{t,N}(Y_t;\theta) = \frac{1}{5} \frac{1}{(1+\xi_N(Y_t;\theta))^5} \left\{ \frac{\varepsilon_N(Y_t;\theta)}{\sqrt{N}} \right\}^5$$
(62)

with $|\xi_N(Y_t;\theta)| \leq \left|\frac{\varepsilon_N(Y_t;\theta)}{\sqrt{N}}\right|$. We need to ensure that these Taylor expansions are valid for $t \in 1: T$ so we control the probability of the event $B\left(Y^T,\epsilon\right) = \left\{\max_{t \leq T} \left|\frac{\varepsilon_N(Y_t;\theta)}{\sqrt{N}}\right| > \epsilon\right\}$. We have for any $\epsilon > 0$

$$\mathbb{P}\left\{B\left(Y^{T},\epsilon\right)\right\} \leq \sum_{t=1}^{T} \mathbb{P}\left(\left|\frac{\varepsilon_{N}\left(Y_{t};\theta\right)}{\sqrt{N}}\right| > \epsilon\right)$$

$$= T\mathbb{P}\left(\left|\frac{\varepsilon_{N}\left(Y_{1};\theta\right)}{\sqrt{N}}\right| > \epsilon\right)$$

$$\leq T\frac{\mathbb{E}\left(\varepsilon_{N}\left(Y_{1};\theta\right)^{8}\right)}{\epsilon^{8}N^{4}}$$

$$\leq \frac{\mathbb{E}\left(\varepsilon_{N}\left(Y_{1};\theta\right)^{8}\right)}{\epsilon^{8}\beta^{4}T^{4\alpha-1}}.$$

As $\mathbb{E}\left(\varpi\left(Y_1,U_{1,1}^T;\theta\right)^8\right)<\infty$ under assumption, the complementary event satisfies for $\alpha>1/4$

$$\lim_{T \to \infty} \mathbb{P}\left(\left(B\left(Y^{T}, \epsilon\right)\right)^{\mathsf{c}}\right) = 1. \tag{63}$$

On the event $(B(Y^T, \epsilon))^{\mathsf{C}}$, the Taylor expansion (61) holds for all $t \in 1: T$ so we can write

$$\frac{\log \widehat{p}\left(Y_{1:T}|\theta\right) - \log p\left(Y_{1:T}|\theta\right)}{T^{(1-\alpha)/2}} = \frac{1}{\beta^{1/2}T^{1/2}} \sum_{t=1}^{T} \varepsilon_N\left(Y_t;\theta\right)$$

$$\tag{64}$$

$$-\frac{1}{2\beta T^{(1+\alpha)/2}} \sum_{t=1}^{T} \varepsilon_N \left(Y_t; \theta \right)^2 \tag{65}$$

+
$$\frac{1}{3\beta^{3/2}T^{(1+2\alpha)/2}} \sum_{t=1}^{T} \varepsilon_N (Y_t; \theta)^3$$
 (66)

$$-\frac{1}{4\beta^2 T^{(1+3\alpha)/2}} \sum_{t=1}^{T} \varepsilon_N \left(Y_t; \theta \right)^4 \tag{67}$$

$$+\frac{1}{T^{(1-\alpha)/2}} \sum_{t=1}^{T} R_{t,N} (Y_t; \theta)$$

$$+ o_{\mathbb{P}} (1)$$

$$(68)$$

where the $o_{\mathbb{P}}(1)$ arises from substituting βT^{α} to $N = \lceil \beta T^{\alpha} \rceil$.

We first control the remainder (68), using the fact that (62) can be controlled on the event $B^{\mathbb{C}}(Y^T, \epsilon)$, as follows

$$\frac{1}{T^{(1-\alpha)/2}} \left| \sum_{t=1}^{T} R_{t,N} (Y_t; \theta) \right| \leq \frac{1}{5\beta^{5/2}} \frac{1}{(1-\epsilon)^5} \frac{1}{T^{(1-\alpha)/2} N^{5/2}} \sum_{t=1}^{T} \left| \varepsilon_N (Y_t; \theta) \right|^5 \\
\leq \frac{1}{5\beta^{5/2}} \frac{1}{(1-\epsilon)^5} \frac{1}{T^{(4\alpha-1)/2}} \frac{1}{T} \sum_{t=1}^{T} \left| \varepsilon_N (Y_t; \theta) \right|^5.$$

The WLLN for triangular arrays holds by a similar argument to Lemma 10 so we have

$$\frac{1}{T} \sum_{t=1}^{T} \left| \varepsilon_{N} \left(Y_{t}; \theta \right) \right|^{5} - \mathbb{E} \left(\left| \varepsilon_{N} \left(Y_{1}; \theta \right) \right|^{5} \right) \stackrel{\mathbb{P}}{\to} 0.$$

Hence as $\alpha > 1/4$, we have

$$\frac{1}{T^{(1-\alpha)/2}} \left| \sum_{t=1}^{T} R_{t,N} \left(Y_t; \theta \right) \right| \stackrel{\mathbb{P}}{\to} 0. \tag{69}$$

The term on the r.h.s. of (64) satisfies a conditional CLT for triangular arrays; see Lemma 27 in Section A.9. Indeed, we have for any $\epsilon > 0$

$$\mathbb{E}\left[T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{N}\left(Y_{t};\theta\right)^{2}\mathbb{I}_{\left\{\left|\varepsilon_{N}\left(Y_{t};\theta\right)\right|\geq\sqrt{T}\epsilon\right\}}\middle|\mathcal{Y}^{T}\right)\right] = \mathbb{E}\left[\epsilon^{2}\sum_{t=1}^{T}\mathbb{E}\left(\frac{\varepsilon_{N}\left(Y_{t};\theta\right)^{2}}{\epsilon^{2}T}\mathbb{I}_{\left\{\left|\varepsilon_{N}\left(Y_{t};\theta\right)\right|\geq\sqrt{T}\epsilon\right\}}\middle|\mathcal{Y}^{T}\right)\right]\right]$$

$$\leq \epsilon^{2}\sum_{t=1}^{T}\mathbb{E}\left(\frac{\varepsilon_{N}\left(Y_{t};\theta\right)^{4}}{\epsilon^{4}T^{2}}\right)$$

$$= \frac{1}{T\epsilon^{2}}\mathbb{E}\left(\varepsilon_{N}\left(Y_{t};\theta\right)^{4}\right)$$

$$= \frac{1}{T\epsilon^{2}}\left\{3\gamma\left(\theta\right)^{4} + \frac{\rho_{4}\left(\theta\right)}{N}\right\}$$

$$\Rightarrow 0$$

so the following conditional Lindeberg condition holds

$$T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(\varepsilon_{N}\left(Y_{t};\theta\right)^{2} \mathbb{I}_{\left\{\left|\varepsilon_{N}\left(Y_{t};\theta\right)\right| \geq \sqrt{T}\epsilon\right\}} \middle| \mathcal{Y}^{T}\right) \stackrel{\mathbb{P}}{\to} 0.$$

As (56) holds, by the strong law of large numbers (SLLN), the limiting variance is given by

$$\lim_{T \to \infty} \frac{1}{\beta T} \sum_{t=1}^{T} \mathbb{E}\left(\varepsilon_{N}\left(Y_{t};\theta\right)^{2} \middle| \mathcal{Y}^{T}\right) = \lim_{T \to \infty} \frac{1}{\beta T} \sum_{t=1}^{T} \gamma\left(Y_{t};\theta\right)^{2} = \beta^{-1} \gamma\left(\theta\right)^{2}.$$

Lemma 10 shows that the second term (65) satisfies

$$\frac{1}{T^{(1+\alpha)/2}} \sum_{t=1}^{T} \varepsilon_N \left(Y_t; \theta \right)^2 - T^{(1-\alpha)/2} \gamma \left(\theta \right)^2 \stackrel{\mathbb{P}}{\to} 0, \tag{71}$$

while the third term (66) satisfies

$$\frac{1}{T^{(1+2\alpha)/2}} \sum_{t=1}^{T} \varepsilon_N (Y_t; \theta)^3 - \frac{\rho_3(\theta)}{\beta^{1/2} T^{(3\alpha-1)/2}} \stackrel{\mathbb{P}}{\to} 0, \tag{72}$$

hence it vanishes for $\alpha > 1/3$. Similarly, Lemma 10 and (58) show that

$$\frac{1}{T^{(1+3\alpha)/2}} \sum_{t=1}^{T} \varepsilon_N \left(Y_t; \theta \right)^4 - \frac{3\gamma \left(\theta \right)^4}{T^{(3\alpha-1)/2}} - \frac{\rho_4 \left(\theta \right)}{\beta T^{(5\alpha-1)/2}} \stackrel{\mathbb{P}}{\to} 0 \tag{73}$$

where $\frac{\rho_4(\theta)}{\beta T^{(5\alpha-1)/2}} \to 0$ for any $\alpha > 1/5$.

The term $T^{-(1-\alpha)/2}\{\log \widehat{p}(Y_{1:T}|\theta) - \log p(Y_{1:T}|\theta)\}$ is asymptotically equivalent in distribution to the sum of the terms (64), (65), (66), (67) and (68). By combining (63) to the fact that (64) satisfies a conditional CLT, (71), (72), (73), (69) hold and Lemma 29, the result follows.

A.3 Proof of Part 2 of Theorem 1

Lemma 11. For any $y \in Y$ and integer $k \geq 1$, if $\mathbb{E}\left[\left|\varepsilon_{N}\left(y;\theta\right)\right|^{k+1}\right] < \infty$ then $\widetilde{\mathbb{E}}\left[\left|\varepsilon_{N}\left(y;\theta\right)\right|^{k}\right] < \infty$ and $\widetilde{\mathbb{E}}\left[\varepsilon_{N}\left(y;\theta\right)^{k}\right] = \mathbb{E}\left[\varepsilon_{N}\left(y;\theta\right)^{k}\right] + \frac{1}{\sqrt{N}}\mathbb{E}\left[\varepsilon_{N}\left(y;\theta\right)^{k+1}\right].$

Proof of Lemma 11. We have

$$\begin{split} \widetilde{\mathbb{E}}\left[\varepsilon_{N}\left(y;\theta\right)^{k}\right] &= \frac{1}{N^{k/2}}\int \dots \int \left[\sum_{i=1}^{N}\left\{\varpi\left(y,u_{1,i};\theta\right)-1\right\}\right]^{k}\overline{\pi}(\mathrm{d}u_{1,1:N}|\theta) \\ &= \frac{1}{N^{1+k/2}}\int \dots \int \left[\sum_{i=1}^{N}\left\{\varpi\left(y,u_{1,i};\theta\right)-1\right\}\right]^{k}\left[N+\sum_{i=1}^{N}\left\{\varpi\left(y,u_{1,i};\theta\right)-1\right\}\right]\prod_{j=1}^{N}\varphi\left(\mathrm{d}u_{1,j};0_{p},I_{p}\right) \\ &= \frac{1}{N^{k/2}}\int \dots \int \left[\sum_{i=1}^{N}\left\{\varpi\left(y,u_{1,i};\theta\right)-1\right\}\right]^{k}\prod_{j=1}^{N}\varphi\left(\mathrm{d}u_{1,j};0_{p},I_{p}\right) \\ &+ \frac{1}{N^{1+k/2}}\int \dots \int \left[\sum_{i=1}^{N}\left\{\varpi\left(y,u_{1,i};\theta\right)-1\right\}\right]^{k+1}\prod_{j=1}^{N}\varphi\left(\mathrm{d}u_{1,j};0_{p},I_{p}\right). \end{split}$$

The result follows directly.

Corollary 12. By combining Lemma 8 and Lemma 11, we obtain

$$\widetilde{\mathbb{E}}\left[\varepsilon_{N}\left(y;\theta\right)\right] = \frac{\gamma\left(y;\theta\right)^{2}}{\sqrt{N}},\tag{74}$$

$$\widetilde{\mathbb{E}}\left[\varepsilon_N\left(y;\theta\right)^2\right] = \gamma\left(y;\theta\right)^2 + \frac{\rho_3\left(y;\theta\right)}{N},\tag{75}$$

$$\widetilde{\mathbb{E}}\left[\varepsilon_{N}\left(y;\theta\right)^{3}\right] = \frac{3\gamma\left(y;\theta\right)^{4} + \rho_{3}\left(y;\theta\right)}{\sqrt{N}} + \frac{\rho_{4}\left(y;\theta\right)}{N\sqrt{N}},\tag{76}$$

$$\widetilde{\mathbb{E}}\left[\varepsilon_{N}\left(y;\theta\right)^{4}\right] = 3\gamma\left(y;\theta\right)^{4} + \frac{\rho_{4}\left(y;\theta\right) + 10\rho_{2}\left(y;\theta\right)\rho_{3}\left(y;\theta\right)}{N} + \frac{\rho_{5}\left(y;\theta\right)}{N^{2}}.$$
(77)

Similarly, we have $\widetilde{\mathbb{E}}\left[\varepsilon_{N}\left(Y_{1};\theta\right)\right] = \gamma\left(\theta\right)^{2}/\sqrt{N}, \ \widetilde{\mathbb{E}}\left[\varepsilon_{N}\left(Y_{1};\theta\right)^{2}\right] = \gamma\left(\theta\right)^{2} + \rho_{3}\left(\theta\right)/N, \ etc.$

We can now give the proof of Part 2 of Theorem 1. For $\alpha = 1$, it is possible to combine Part 1 of Theorem 1 to a uniform integrability argument to establish this result but this argument does not extend to $1/3 < \alpha < 1$.

Proof of Part 2 of Theorem 1. The proof of this CLT is very similar to the proof of Part 1 of Theorem 1 so we skip some details. We again first perform a fourth order Taylor expansion of each term appearing in (55), i.e. see (61) and (62). We also need to ensure that these Taylor expansions are valid for $t \in 1: T$ so we need to control the probability of the event $B\left(Y^T, \epsilon\right) = \left\{\max_{t \leq T} \left|N^{-1/2} \varepsilon_N\left(Y_t; \theta\right)\right| > \epsilon\right\}$. We have for any $\epsilon > 0$

$$\widetilde{\mathbb{P}}\left\{B\left(Y^{T},\epsilon\right)\right\} \leq \frac{\widetilde{\mathbb{E}}\left(\varepsilon_{N}\left(Y_{1};\theta\right)^{8}\right)}{\epsilon^{8}\beta^{4}T^{4\alpha-1}}.$$

As $\mathbb{E}\left(\varpi\left(Y_{1},U_{1,1}^{T};\theta\right)^{9}\right)<\infty$ holds, Lemma 11 ensures that $\widetilde{\mathbb{E}}\left(\varepsilon_{N}\left(Y_{1};\theta\right)^{8}\right)<\infty$ so

$$\lim_{T \to \infty} \widetilde{\mathbb{P}} \left(\left(B \left(Y^T, \epsilon \right) \right)^{c} \right) = 1 \tag{78}$$

for $\alpha > 1/4$. On the event $(B(Y^T, \epsilon))^{\mathfrak{c}}$, the Taylor expansion (61) holds for all $t \in 1 : T$ so we can similarly decompose $T^{-(1-\alpha)/2}\{\log \widehat{p}(Y_{1:T}|\theta) - \log p(Y_{1:T}|\theta)\}$ as the sum of the terms (64), (65), (66), (67), (68) and an additional $o_{\widetilde{p}}(1)$ term.

We can show that as $\alpha > 1/4$ the remainder vanishes

$$\frac{1}{T^{(1-\alpha)/2}} \left| \sum_{t=1}^{T} R_{t,N} \left(Y_t; \theta \right) \right| \stackrel{\widetilde{\mathbb{P}}}{\to} 0 \tag{79}$$

because the WLLN for triangular arrays holds so we have

$$\frac{1}{T}\sum_{t=1}^{T}\left|\varepsilon_{N}\left(Y_{t};\theta\right)\right|^{5}-\widetilde{\mathbb{E}}\left(\left|\varepsilon_{N}\left(Y_{1};\theta\right)\right|^{5}\right)\overset{\widetilde{\mathbb{P}}}{\rightarrow}0.$$

Using (74), we can rewrite the first term (64) as follows

$$\frac{1}{\beta^{1/2}T^{1/2}}\sum_{t=1}^{T}\varepsilon_{N}\left(Y_{t};\theta\right) = \frac{1}{\beta^{1/2}T^{1/2}}\sum_{t=1}^{T}\left\{\varepsilon_{N}\left(Y_{t};\theta\right) - \widetilde{\mathbb{E}}\left(\varepsilon_{N}\left(Y_{t};\theta\right)|\mathcal{Y}^{T}\right)\right\}$$
(80)

$$+\frac{1}{\beta^{1/2}T^{1/2}}\sum_{t=1}^{T}\left\{\widetilde{\mathbb{E}}\left(\varepsilon_{N}\left(Y_{t};\theta\right)|\mathcal{Y}^{T}\right)-\frac{\gamma\left(\theta\right)^{2}}{\sqrt{N}}\right\}$$
(81)

$$+\frac{T^{1/2}}{\beta^{1/2}}\frac{\gamma(\theta)^2}{\sqrt{N}}.\tag{82}$$

The r.h.s. of (80) satisfies a conditional CLT, see Lemma 27. Indeed the conditional Lindeberg condition holds using arguments similar to (70) as $T^{-1}\widetilde{\mathbb{E}}\left(\varepsilon_N^4\left(Y_t;\theta\right)\right)\to 0$. By Lemma 11 and the SLLN, the limiting variance is given by

$$\lim_{T \to \infty} \frac{1}{\beta T} \sum_{t=1}^{T} \widetilde{\mathbb{E}} \left(\varepsilon_{N} \left(Y_{t}; \theta \right)^{2} \middle| \mathcal{Y}^{T} \right) - \widetilde{\mathbb{E}} \left(\varepsilon_{N} \left(Y_{t}; \theta \right) \middle| \mathcal{Y}^{T} \right)^{2} = \lim_{T \to \infty} \frac{1}{\beta T} \sum_{t=1}^{T} \left\{ \gamma \left(Y_{t}; \theta \right)^{2} + \frac{\rho_{3} \left(Y_{t}; \theta \right)}{N} - \frac{\gamma \left(Y_{t}; \theta \right)^{4}}{N} \right\} = \beta^{-1} \gamma \left(\theta \right)^{2}$$

almost surely, (74)-(75) and using the assumption $\widetilde{\mathbb{E}}\left[\gamma\left(Y_{1};\theta\right)^{4}\right]<\infty$. The term (81) satisfies

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T} \left\{ \widetilde{\mathbb{E}} \left(\varepsilon_{N} \left(Y_{t}; \theta \right) | \mathcal{Y}^{T} \right) - \frac{\gamma \left(\theta \right)^{2}}{\sqrt{N}} \right\} = \frac{1}{T^{1/2} \sqrt{N}} \sum_{t=1}^{T} \left\{ \gamma \left(Y_{t}; \theta \right)^{2} - \gamma \left(\theta \right)^{2} \right\} \stackrel{\widetilde{\mathbb{P}}}{\to} 0$$
 (83)

by the SLLN, the assumption $\widetilde{\mathbb{E}}\left[\gamma\left(Y_1;\theta\right)^4\right]<\infty$ and Chebyshev's inequality. Finally we have for (82)

$$\frac{T^{1/2}}{\beta^{1/2}} \frac{\gamma(\theta)^2}{\sqrt{N}} - \frac{T^{(1-\alpha)/2}}{\beta} \gamma(\theta)^2 \to 0.$$
(84)

For the second term (65), using (75), we obtain using Lemma 10

$$\frac{1}{T^{(1+\alpha)/2}} \sum_{t=1}^{T} \varepsilon_N \left(Y_t; \theta \right)^2 - T^{(1-\alpha)/2} \gamma \left(\theta \right)^2 - \beta^{-1} T^{(1-3\alpha)/2} \rho_3 \left(\theta \right) \stackrel{\widetilde{\mathbb{P}}}{\to} 0, \tag{85}$$

where the third term on the l.h.s. vanishes for $\alpha > 1/3$. For the third term (66), we obtain using (76) and Lemma 10

$$\frac{1}{T^{(1+2\alpha)/2}} \sum_{t=1}^{T} \varepsilon_N (Y_t; \theta)^3 - \frac{3\gamma (\theta)^4 + \rho_3 (\theta)}{\beta^{1/2} T^{(3\alpha-1)/2}} - \frac{\rho_4 (\theta)}{\beta^{3/2} T^{(5\alpha-1)/2}} \stackrel{\tilde{\mathbb{P}}}{\to} 0.$$
 (86)

Hence, (66) vanishes for $\alpha > 1/3$. Finally for the fourth term (67), we obtain using (77) and Lemma 10

$$\frac{1}{T^{(1+3\alpha)/2}} \sum_{t=1}^{T} \varepsilon_N (Y_t; \theta)^4 - \frac{3\gamma(\theta)^4}{T^{(3\alpha-1)/2}} - \frac{\rho_4(\theta) + 10\rho_2(\theta)\rho_3(\theta)}{\beta T^{(5\alpha-1)/2}} - \frac{\rho_5(\theta)}{\beta^2 T^{(7\alpha-1)/2}} \stackrel{\widetilde{\mathbb{P}}}{\to} 0 \tag{87}$$

where $T^{-(5\alpha-1)/2}\left\{\rho_4\left(\theta\right)+10\rho_2\left(\theta\right)\rho_3\left(\theta\right)\right\}\to 0$ and $T^{-(7\alpha-1)/2}\rho_5\left(\theta\right)\to 0$ for any $\alpha>1/5$.

The term $T^{-(1-\alpha)/2}\{\log \widehat{p}(Y_{1:T}|\theta) - \log p(Y_{1:T}|\theta)\}$ is asymptotically equivalent in distribution to the sum of the terms (64), (65), (66), (67) and (68). By combining (78) to the fact that (80) satisfies a conditional CLT, (83), (84), (85), (86), (87), (79) and Lemma 29, the result follows.

Remark 13. It follows directly from our proof that for $\frac{1}{4} < \alpha \le \frac{1}{3}$

$$\frac{Z_{T}(\theta)}{T^{(1-\alpha)/2}} - \frac{T^{(1-\alpha)/2}}{2}\beta^{-1}\gamma(\theta)^{2} + T^{(1-3\alpha)/2}\beta^{-2} \left\{ \frac{\rho_{3}(\theta)}{6} - \frac{1}{4}\gamma(\theta)^{4} \right\} \left| \mathcal{Y}^{T} \Rightarrow \mathcal{N}\left\{0, \beta^{-1}\gamma(\theta)^{2}\right\}. \tag{88}$$

We also note that if we assume that higher order moments of $\varpi\left(Y_1,U_{1,1};\theta\right)$ under $\widetilde{\mathbb{P}}$ are finite then we obtain different expressions in the CLT for $\frac{1}{2k+3}<\alpha\leq\frac{1}{2k+1}$ where $k\in\mathbb{N}$.

A.4 Proof of Theorem 2

To simplify presentation, we only give the proof when θ is a scalar parameter, the multivariate extension is direct. We have $Z_T(\theta) = \log \widehat{p}(Y_{1:T} \mid \theta, U) - \log p(Y_{1:T} \mid \theta, U)$ with $U \sim \overline{\pi}(\cdot \mid \theta)$. We define $W_T(\theta + \xi/\sqrt{T}) = \log \widehat{p}(Y_{1:T} \mid \theta + \xi/\sqrt{T}, U') - \log p(Y_{1:T} \mid \theta, U')$ with $U' \sim m$.

The result will follow by the arguments used in the proof of Theorem 1, replacing

$$\epsilon_N(Y_t, U_t; \theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[\varpi(Y_t, U_{t,i}; \theta) - 1 \right],$$

with

$$\zeta_N(Y_t;\theta) = \epsilon_N(Y_t, U_t'; \theta + \xi/\sqrt{T}) - \epsilon_N(Y_t, U_t; \theta).$$

We make here the dependence of ϵ_N on U_t or U_t' explicit. We need to check that the moment conditions used for ϵ_N carry over to ζ_N . We have by the \mathcal{C}_p inequality and Lemma 11 that there exists $c<\infty$

$$\widetilde{\mathbb{E}}\left(\zeta_{N}(Y_{1};\theta)^{8}\right) \leq c\left\{\mathbb{E}\left(\epsilon_{N}\left(Y_{1},U_{1}';\theta+\xi/\sqrt{T}\right)^{8}\right) + \widetilde{\mathbb{E}}\left(\epsilon_{N}\left(Y_{1},U_{1};\theta\right)^{8}\right)\right\}$$

$$\leq c\left\{\mathbb{E}\left(\epsilon_{N}\left(Y_{1},U_{1};\theta+\xi/\sqrt{T}\right)^{8}\right) + \mathbb{E}\left(\epsilon_{N}\left(Y_{1},U_{1};\theta\right)^{8}\right) + \frac{1}{\sqrt{N}}\left|\mathbb{E}\left(\epsilon_{N}\left(Y_{1},U_{1};\theta\right)^{9}\right)\right|\right\}.$$

As $\vartheta \mapsto \varpi(Y_1, U_{1,1}; \vartheta)$ and $\vartheta \mapsto \widetilde{\mathbb{E}}(\varpi(Y_1, U_{1,1}; \vartheta)^9)$ are continuous by assumption, it is straightforward to check that lower order moments are also continuous. Therefore for T large enough

$$\widetilde{\mathbb{E}}\left(\zeta_{N}(Y_{1};\theta)^{8}\right) \leq c\left\{2\mathbb{E}\left(\epsilon_{N}\left(Y_{1},U_{1};\theta\right)^{8}\right) + \frac{1}{\sqrt{N}}\mathbb{E}\left(\left|\epsilon_{N}\left(Y_{1},U_{1};\theta\right)\right|^{9}\right)\right\},\,$$

and similar results hold for lower order moments.

We use a Taylor expansion similarly to Theorem 1 Part 1 and Part 2,

$$\begin{split} \frac{W_T\left(\theta+\xi/\sqrt{T}\right)}{T^{(1-\alpha)/2}} - \frac{Z_T\left(\theta\right)}{T^{(1-\alpha)/2}} &= \frac{1}{\beta^{1/2}T^{1/2}} \sum_{t=1}^T \left[\varepsilon_N\left(Y_t, U_t'; \theta+\xi/\sqrt{T}\right) - \varepsilon_N\left(Y_t, U_t; \theta\right) \right] \\ &- \frac{1}{2\beta T^{(1+\alpha)/2}} \sum_{t=1}^T \left[\varepsilon_N\left(Y_t, U_t'; \theta+\xi/\sqrt{T}\right)^2 - \varepsilon_N\left(Y_t, U_t; \theta\right)^2 \right] \\ &+ \frac{1}{3\beta^{3/2}T^{(1+2\alpha)/2}} \sum_{t=1}^T \left[\varepsilon_N\left(Y_t, U_t'; \theta+\xi/\sqrt{T}\right)^3 - \varepsilon_N\left(Y_t, U_t; \theta\right)^3 \right] \\ &- \frac{1}{4\beta^2T^{(1+3\alpha)/2}} \sum_{t=1}^T \left[\varepsilon_N\left(Y_t, U_t'; \theta+\xi/\sqrt{T}\right)^4 - \varepsilon_N\left(Y_t, U_t; \theta\right)^4 \right] \\ &+ \frac{1}{T^{(1-\alpha)/2}} \sum_{t=1}^T R_{t,N}'\left(Y_t; \theta, \xi\right) + o_{\mathbb{P}}(1) \,, \end{split}$$

where $\overline{\mathbb{P}}$ denotes the probability over $U \sim \overline{\pi} \left(\cdot | \theta \right)$, $U' \sim m$ and $Y_t \stackrel{\text{i.i.d.}}{\sim} \mu$ and $\overline{\mathbb{E}}$ the associated expectation. By inspecting the proofs of Parts 1 and 2 of Theorem 1, we can rewrite this as

$$\frac{W_T\left(\theta + \xi/\sqrt{T}\right)}{T^{(1-\alpha)/2}} - \frac{Z_T\left(\theta\right)}{T^{(1-\alpha)/2}} \tag{89}$$

$$= \frac{1}{\beta^{1/2} T^{1/2}} \sum_{t=1}^{T} \left[\varepsilon_N \left(Y_t, U_t'; \theta + \xi / \sqrt{T} \right) - \varepsilon_N \left(Y_t, U_t; \theta \right) \right]$$

$$(90)$$

$$-\frac{1}{2\beta T^{(1+\alpha)/2}} \sum_{t=1}^{T} \left[\varepsilon_N \left(Y_t, U_t'; \theta + \xi/\sqrt{T} \right)^2 - \varepsilon_N \left(Y_t, U_t; \theta \right)^2 \right] + o_{\mathbb{P}}(1). \tag{91}$$

The term (90) satisfies a conditional CLT for triangular arrays (Lemma 27) as the conditional Lindeberg condition is verified

$$\overline{\mathbb{E}}\left[T^{-1}\sum_{t=1}^{T}\overline{\mathbb{E}}\left(\left\{\varepsilon_{N}\left(Y_{t},U_{t}';\theta+\xi/\sqrt{T}\right)-\varepsilon_{N}\left(Y_{t},U_{t};\theta\right)\right\}^{2}\mathbb{I}_{\left\{|\varepsilon_{N}\left(Y_{t};\theta\right)-\varepsilon_{N}\left(Y_{t},U_{t};\theta\right)|\geq\sqrt{T}\epsilon\right\}}\middle|\mathcal{Y}^{T}\right)\right]$$

$$=\overline{\mathbb{E}}\left[\epsilon^{2}\sum_{t=1}^{T}\overline{\mathbb{E}}\left(\frac{\left\{\varepsilon_{N}\left(Y_{t},U_{t}';\theta+\xi/\sqrt{T}\right)-\varepsilon_{N}\left(Y_{t},U_{t};\theta\right)\right\}^{2}}{\epsilon^{2}T}\mathbb{I}_{\left\{|\varepsilon_{N}\left(Y_{t};\theta\right)-\varepsilon_{N}\left(Y_{t},U_{t};\theta\right)|\geq\sqrt{T}\epsilon\right\}}\middle|\mathcal{Y}^{T}\right)\right]$$

$$\leq\frac{1}{T\epsilon^{2}}\overline{\mathbb{E}}\left(\left\{\varepsilon_{N}\left(Y_{t},U_{t}';\theta+\xi/\sqrt{T}\right)-\varepsilon_{N}\left(Y_{t},U_{t};\theta\right)\right\}^{4}\right)$$

$$\leq\frac{c}{T\epsilon^{2}}\left\{\widetilde{\mathbb{E}}\left(\varepsilon_{N}\left(Y_{t},U_{t}';\theta+\xi/\sqrt{T}\right)-\varepsilon_{N}\left(Y_{t},U_{t}';\theta\right)\right)\right\}^{2}\left(C_{p} \text{ inequality}\right)$$

$$\to 0,$$

so

$$T^{-1} \sum_{t=1}^{T} \overline{\mathbb{E}} \left(\left\{ \varepsilon_{N} \left(Y_{t}, U_{t}'; \theta + \xi / \sqrt{T} \right) - \varepsilon_{N} \left(Y_{t}, U_{t}; \theta \right) \right\}^{2} \mathbb{I}_{\left\{ |\varepsilon_{N}(Y_{t}; \theta) - \varepsilon_{N}(Y_{t}, U_{t}; \theta)| \geq \sqrt{T}\epsilon \right\}} \middle| \mathcal{Y}^{T} \right) \xrightarrow{\overline{\mathbb{P}}} 0$$

and the limiting variance is given by

$$\lim_{T \to \infty} \frac{1}{\beta T} \sum_{t=1}^{T} \left[\mathbb{E} \left(\left\{ \varepsilon_{N} \left(Y_{t}, U_{t}'; \theta + \xi / \sqrt{T} \right) - \varepsilon_{N} \left(Y_{t}, U_{t}; \theta \right) \right\}^{2} \middle| \mathcal{Y}^{T} \right) \right.$$

$$\left. - \mathbb{E} \left(\left\{ \varepsilon_{N} \left(Y_{t}, U_{t}'; \theta + \xi / \sqrt{T} \right) - \varepsilon_{N} \left(Y_{t}, U_{t}; \theta \right) \right\} \middle| \mathcal{Y}^{T} \right)^{2} \right]$$

$$= \lim_{T \to \infty} \frac{1}{\beta T} \sum_{t=1}^{T} \mathbb{E} \left(\varepsilon_{N} \left(Y_{t}, U_{t}; \theta \right)^{2} \middle| \mathcal{Y}^{T} \right) + \mathbb{E} \left(\varepsilon_{N} \left(Y_{t}, U_{t}'; \theta + \xi / \sqrt{T} \right)^{2} \middle| \mathcal{Y}^{T} \right) - \mathbb{E} \left(\varepsilon_{N} \left(Y_{t}, U_{t}; \theta \right) \middle| \mathcal{Y}^{T} \right)^{2} \right.$$

$$= \lim_{T \to \infty} \frac{1}{\beta T} \sum_{t=1}^{T} \gamma \left(Y_{t}; \theta \right)^{2} + \frac{\rho_{3} \left(Y_{t}; \theta \right)}{N} - \frac{\gamma \left(Y_{t}; \theta \right)^{4}}{N} + \gamma \left(Y_{t}; \theta + \xi / \sqrt{T} \right)^{2}$$

as $\mathbb{E}\left[\left.\varepsilon_{N}\left(Y_{t},U_{t}';\theta+\xi/\sqrt{T}\right)\right|\mathcal{Y}^{T}\right]=0$. Now we have

$$\lim_{T \to \infty} \frac{1}{\beta T} \sum_{t=1}^{T} \gamma \left(Y_t; \theta \right)^2 + \frac{\rho_3 \left(Y_t; \theta \right)}{N} - \frac{\gamma \left(Y_t; \theta \right)^4}{N} = \beta^{-1} \gamma \left(\theta \right)^2$$

by the SLLN as $\mathbb{E}\left[\gamma\left(Y_{t};\theta\right)^{4}\right]<\infty$. We also have by the WLLN for triangular arrays that

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \gamma \left(Y_t; \theta \right)^2 - \gamma \left(Y_t; \theta + \xi / \sqrt{T} \right)^2 \stackrel{\mathbb{P}}{\to} 0 \tag{92}$$

and

$$T^{(1-\alpha)/2} \left| \gamma \left(\theta + \xi / \sqrt{T} \right)^{2} - \gamma \left(\theta \right)^{2} \right| = T^{(1-\alpha)/2} \left| \int_{\theta}^{\theta + \frac{\xi}{\sqrt{T}}} \frac{\partial \gamma \left(\vartheta \right)^{2}}{\partial \vartheta} d\vartheta \right|$$

$$\leq T^{(1-\alpha)/2} \frac{\xi}{\sqrt{T}} \sup_{\vartheta \in \left[\theta \wedge \left(\theta + \frac{\xi}{\sqrt{T}} \right), \theta \vee \left(\theta + \frac{\xi}{\sqrt{T}} \right) \right]} \left| \frac{\partial \gamma \left(\vartheta \right)^{2}}{\partial \vartheta} \right| \to 0.$$

$$(93)$$

We have already seen in the proof of Theorem 1, equation (71), that

$$\frac{1}{T^{(1+\alpha)/2}} \sum_{t=1}^{T} \varepsilon_N \left(Y_t, U_t; \theta \right)^2 - T^{(1-\alpha)/2} \gamma \left(\theta \right)^2 \stackrel{\overline{\mathbb{P}}}{\to} 0, \tag{94}$$

and using a similar argument as the one used in the proof of (22) in Theorem 1, equation (85), we have

$$\frac{1}{T^{(1+\alpha)/2}} \sum_{t=1}^{T} \varepsilon_N \left(Y_t, U_t'; \theta + \xi/\sqrt{T} \right)^2 - T^{(1-\alpha)/2} \gamma \left(\theta \right)^2 \stackrel{\overline{\mathbb{P}}}{\to} 0 \tag{95}$$

as $\alpha > 1/3$ and (93) holds.

Hence (90) minus its mean satisfies a conditional CLT of limiting variance $2\beta^{-1}\gamma(\theta)^2$ because of (92)-(93). Using (83), its mean plus $\beta^{-1}T^{(1-\alpha)/2}\gamma(\theta)^2$ converges to zero in probability and (91) vanishes in probability so the final result follows from Lemma 29.

A.5 Proof of Theorem 3

A.5.1 Notation and continuous-time embedding

For $\delta_T = \psi \frac{N}{T}$, we have $\rho_T = \exp(-\delta_T)$ and we can write for $t \in 1: T$ and $i \in 1: N$

$$U'_{t,i} = e^{-\delta_T} U_{t,i} + \sqrt{1 - e^{-2\delta_T}} \varepsilon_{t,i}, \quad \varepsilon_{t,i} \sim \mathcal{N}\left(0_p, I_p\right). \tag{96}$$

It will prove convenient for our proof to embed this discrete-time process within the following Ornstein-Uhlenbeck process

$$dU_{t,i}(s) = -U_{t,i}(s) ds + \sqrt{2} dB_{t,i}(s),$$
(97)

where $B_{t,i}$ are independent p-dimensional standard Brownian motions for $t \in 1 : T$ and $i \in 1 : N$. It is easy to check that we can set equivalently $U'_{t,i} = U_{t,i}(\delta_T)$ as the value of the Ornstein-Uhlenbeck process at time $s = \delta_T$ which has been initialized at time s = 0 using $U_{t,i}(0) = U_{t,i}$.

Whenever it is clear, we will drop the T index to keep the notation reasonable. We define

$$\widehat{W}_{t}^{T}\left(\theta\right) = \widehat{W}_{t}^{T}(Y_{t} \mid \theta; U_{t}) = \frac{\widehat{p}\left(\left.Y_{t} \mid \theta, U_{t}\right)\right.}{p\left(\left.Y_{t} \mid \theta\right.\right)} = \frac{1}{N} \sum_{i=1}^{N} \varpi\left(\left.Y_{t}, U_{t, i}; \theta\right.\right),$$

where the full notation shall be retained when evaluating at the proposal θ', U'_t and

$$\eta_t^T = \frac{\widehat{W}_t^T(Y_t \mid \theta'; U_t') - \widehat{W}_t^T(\theta)}{\widehat{W}_t^T(\theta)}.$$
(98)

Let $\mathcal{F}^T \subset \mathcal{G}$ be the sigma-algebra spanned by $\{\mathcal{U}^T, \mathcal{Y}^T\}$ where $\mathcal{U}^T = \sigma\{U_{t,i}; t \in 1 : T, i \in 1 : N\}$ and $\mathcal{Y}^T = \sigma\{Y_t; t \in 1 : T\}$. Let $\widetilde{\mathbb{E}}\left[\cdot | \mathcal{Y}^T\right]$ denotes the expectation w.r.t $\{U_{t,i}(0); t \in 1 : T, i \in 1 : N\} \sim \overline{\pi}(\cdot | \theta)$ and the Brownian motions $\{(B_{t,i}(s); s \geq 0); t \in 1 : T, i \in 1 : N\}$ where $\overline{\pi}(\{u_{t,i}^T; t \in 1 : T, i \in 1 : N\} | \theta)$ is given by (20) whereas $\mathbb{E}\left[\cdot | \mathcal{Y}^T\right]$ denotes the expectation w.r.t $U_{t,i} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0_p, I_p)$ and the Brownian motions $\{(B_{t,i}(s); s \geq 0); t \in 1 : T, i \in 1 : N\}$. Finally, we define the Stein operator \mathcal{S} for a real-valued function $g(y, u; \theta)$

$$S\{g(y, u; \theta)\} := \langle \nabla_u, \nabla_u g(y, u; \theta) \rangle - \langle u, \nabla_u g(y, u; \theta) \rangle. \tag{99}$$

A.5.2 Assumptions

Assumption 4. There exists $\epsilon > 0$ such that

$$\limsup_{T} \mathbb{E}\left[\left(\widehat{W}_{1}^{T}\left(\theta\right)\right)^{-3-\epsilon}\right] < \infty.$$

Assumption 5. There exists $\chi : \mathsf{Y} \times \mathbb{R}^p \to \mathbb{R}^+$ such that $\vartheta \mapsto \nabla_{\vartheta} \varpi (y, u; \vartheta)$ is continuous at $\vartheta = \theta$, $\|\nabla_{\theta} \varpi (y, u; \theta)\| \le \chi(y, u)$ for all $y, u \in \mathsf{Y} \times \mathbb{R}^p$, and

$$\mathbb{E}\left[\chi\left(Y_1, U_{1,1}\right)^4\right] < \infty.$$

Assumption 6. We have

$$\mathbb{E}\left[\left|\left\langle \nabla_{u}, \nabla_{u} \varpi \left(Y_{1}, U_{1,1}; \theta\right) \right\rangle\right|\right] < \infty.$$

Assumption 7. We have

$$\mathbb{E}\left[\left(\mathcal{S}\left\{\left\|\nabla_{u}\varpi(Y_{1},U_{1,1};\theta)\right\|^{2}\right\}\right)^{2}\right]<\infty.$$

Assumption 8. There exists $\varkappa > 0$ such that

$$\mathbb{E}\left[\left\|\nabla_{u}\varpi(Y_{1},U_{1,1};\theta)\right\|^{4+\varkappa}\right]<\infty.$$

Assumption 9. We have

$$\mathbb{E}\left[\left(\mathcal{S}\left\{\varpi\left(Y_{1},U_{1,1};\theta\right)\right\}\right)^{4}\right]<\infty.$$

A.5.3 Details of the proof

To simplify the presentation of the proof, we only consider the case where θ is a scalar parameter, the dimension of $U_{t,i}$ is p=1 and $\psi=1$, the multivariate extension is straightforward although much more

tedious. Let $\theta' = \theta + \xi/\sqrt{T}$. Notice that by definition of $\widehat{W}_t^T(Y_t \mid \theta'; U_t')$, $\widehat{W}_t^T := \widehat{W}_t^T(\theta)$ and a Taylor expansion we have

$$\sum_{t=1}^{T} \log \left(\frac{\widehat{W}_{t}^{T}(Y_{t} \mid \theta'; U_{t}')}{\widehat{W}_{t}^{T}} \right) = \sum_{t=1}^{T} \log(1 + \eta_{t}^{T})$$

$$= \sum_{t=1}^{T} \eta_{t}^{T} - \frac{1}{2} \sum_{t=1}^{T} [\eta_{t}^{T}]^{2} + \sum_{t=1}^{T} h(\eta_{t}^{T}) [\eta_{t}^{T}]^{2}, \tag{100}$$

as $\log(1+x) = x - x^2/2 + h(x)x^2$ with h(x) = o(x) as $x \to 0$.

The proof proceeds through several auxiliary Lemmas in three main steps. First, we prove that $\sum_{t=1}^{T} \eta_t^T$ converges to a zero-mean normal conditional upon \mathcal{F}^T . Second, we show that $\sum_{t=1}^{T} \left(\eta_t^T\right)^2$ converges in probability towards a constant. Third, we show that high-order terms vanish in probability. The result then follows from Proposition 29.

Using Itô's formula, we decompose η_t^T as follows

$$\eta_t^T = J_t^T + L_t^T + M_t^T, (101)$$

where

$$J_{t}^{T} = \frac{1}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \{ \varpi(Y_{t}, U_{t,i}(\delta_{T}); \theta + \xi/\sqrt{T}) - \varpi(Y_{t}, U_{t,i}(\delta_{T}); \theta) \}, \tag{102}$$

$$L_{t}^{T} = \int_{0}^{\delta_{T}} \frac{1}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \left\{ -\partial_{u} \varpi \left(Y_{t}, U_{t,i} \left(s \right) ; \theta \right) U_{t,i} \left(s \right) + \partial_{u,u}^{2} \varpi \left(Y_{t}, U_{t,i} \left(s \right) ; \theta \right) \right\} ds, \tag{103}$$

$$M_{t}^{T} = \int_{0}^{\delta_{T}} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \partial_{u} \varpi \left(Y_{t}, U_{t,i} \left(s \right); \theta \right) dB_{t,i} \left(s \right).$$

$$(104)$$

The following preliminary Lemmas establish various properties of the terms J_t^T , L_t^T , M_t^T and η_t^T .

Lemma 14. The sequence $\{J_t^T; t \geq 1\}$ defined in (102) satisfies

$$\widetilde{\mathbb{E}}\left(J_{t}^{T}\right)=0, \quad \widetilde{\mathbb{V}}\left(\sum_{t=1}^{T}J_{t}^{T}\right)=T\widetilde{\mathbb{V}}\left(J_{1}^{T}\right)\rightarrow0$$

and $\sum_{t=1}^{T} J_t^T \stackrel{\widetilde{\mathbb{P}}}{\to} 0$, $\sum_{t=1}^{T} (J_t^T)^2 \stackrel{\widetilde{\mathbb{P}}}{\to} 0$.

Lemma 15. The sequence $\{L_t^T; t \geq 1\}$ defined in (103) satisfies

$$\widetilde{\mathbb{E}}\left(L_{t}^{T}\right)=0, \quad \ \widetilde{\mathbb{V}}\left(\sum_{t=1}^{T}L_{t}^{T}\right)=T\widetilde{\mathbb{V}}\left(L_{1}^{T}\right)\rightarrow0,$$

and $\sum_{t=1}^{T} L_t^T \stackrel{\widetilde{\mathbb{P}}}{\to} 0$.

Lemma 16. The sequence $\{M_t^T; t \geq 1\}$ defined in (104) satisfies

$$\widetilde{\mathbb{E}}[(M_t^T)^2] = O(1/T), \qquad \sum_{t=1}^T M_t^T \middle| \mathcal{F}^T \Rightarrow \mathcal{N}\left(0, \frac{\kappa\left(\theta\right)^2}{2}\right).$$

Lemma 17. The sequence $\{\eta_t^T; t \geq 1\}$ defined in (101) satisfies

$$\sum_{t=1}^{T} (\eta_t^T)^2 \stackrel{\widetilde{\mathbb{P}}}{\to} \kappa^2 (\theta) .$$

Armed with the above results, we can now prove Theorem 3. Combining Lemmas 14, 15, 16 and 17 with Lemma 29 from Section A.9, we immediately obtain that

$$\sum_{t=1}^{T} \eta_{t} - \frac{1}{2} \left(\eta_{t}^{T} \right)^{2} \left| \mathcal{F}^{T} \right| \Rightarrow \mathcal{N} \left(-\frac{\kappa \left(\theta \right)^{2}}{2}, \kappa \left(\theta \right)^{2} \right).$$

It remains to control the remainder from the Taylor expansion (100). We bound it using Lemma 17 as

$$\left| \sum_{t=1}^{T} h\left(\eta_{t}^{T}\right) \left(\eta_{t}^{T}\right)^{2} \right| \leq \max_{t} \left| h\left(\eta_{t}^{T}\right) \right| \sum_{t=1}^{T} \eta_{t}^{2} = \max_{t} \left| h(\eta_{t}^{T}) \right| O_{\mathbb{P}}(1).$$

Without loss of generality we can assume that $|h\left(x\right)|\leq g\left(|x|\right)$ where g is increasing on $[0,\infty)$ and $\lim_{x\to 0^+}g\left(x\right)=0$ so that

$$\max_{t} |h\left(\eta_{t}^{T}\right)| \leq g\left(\max_{t} |\eta_{t}^{T}|\right)$$

and

$$\widetilde{\mathbb{P}}\left(\max_{t}\left|\eta_{t}^{T}\right|\leq\varepsilon\right)=\prod_{t=1}^{T}\left\{1-\widetilde{\mathbb{P}}\left(\left|\eta_{t}^{T}\right|>\varepsilon\right)\right\}\geq\left(1-\varepsilon^{-2}\widetilde{\mathbb{E}}\left(\left(\eta_{1}^{T}\right)^{2}\mathbb{I}\left(\left|\eta_{1}^{T}\right|>\varepsilon\right)\right)\right)^{T}.\tag{105}$$

By using the decomposition of η_1^T , we have using the C_p inequality

$$T\widetilde{\mathbb{E}}\left(\left(\eta_{1}^{T}\right)^{2}\mathbb{I}\left(\left|\eta_{1}^{T}\right|>\varepsilon\right)\right)\leq c\ T\left(\widetilde{\mathbb{V}}(J_{1}^{T})+\widetilde{\mathbb{V}}(L_{1}^{T})+\widetilde{\mathbb{E}}\left[\left(M_{1}^{T}\right)^{2}\mathbb{I}\left(\left|\eta_{1}^{T}\right|>\varepsilon\right)\right]\right)$$
$$=o(1),$$

where we have used Lemmas 14 and 15 for the terms involving J_1^T and L_1^T . The term involving M_1^T vanishes by uniform integrability of the family $\{T(M_1^T)^2; T \geq 1\}$, the proof of which can be found in the proof of Lemma 16 where the Lindeberg condition is verified. Therefore overall we have

$$\widetilde{\mathbb{E}}\left(\left|\eta_{1}^{T}\right|^{2}\mathbb{I}\left(\left|\eta_{1}^{T}\right|>\varepsilon\right)\right)=o(1/T),$$

and thus (105) converges towards 1 as $T \to \infty$. Hence we have $g\left(\max_{t} \left|\eta_{t}^{T}\right|\right) = o_{\widetilde{\mathbb{P}}}(1)$ and the result follows.

A.5.4 Proofs of Auxiliary Results

Proof of Lemma 14. From Assumption 5, we obtain directly $\widetilde{\mathbb{E}}(J_t^T) = 0$ and we can rewrite J_t^T as follows

$$J_t^T = \frac{1}{N\widehat{W}_t^T} \sum_{i=1}^N \int_{\theta}^{\theta'} \partial_{\vartheta} \varpi(Y_t, U_{t,i}(\delta_T); \vartheta) d\vartheta,$$

where $\theta' = \theta + \xi/\sqrt{T}$. Thus we obtain

$$\begin{split} \widetilde{\mathbb{V}}\left(\sum_{t=1}^{T}J_{t}^{T}\right) &= \sum_{t=1}^{T}\widetilde{\mathbb{V}}\left(J_{t}^{T}\right) = \sum_{t=1}^{T}\widetilde{\mathbb{E}}\left(\left(J_{t}^{T}\right)^{2}\right) \\ &= \sum_{t=1}^{T}\widetilde{\mathbb{E}}\left(\left(\frac{1}{N\widehat{W}_{t}^{T}}\sum_{i=1}^{N}\int_{\theta}^{\theta'}\partial_{\vartheta}\varpi(Y_{t},U_{t,i}(\delta_{T});\vartheta)\mathrm{d}\vartheta\right)^{2}\right) \\ &= \sum_{t=1}^{T}\frac{1}{N^{2}}\mathbb{E}\left((\widehat{W}_{t}^{T})^{-1}\left(\sum_{i=1}^{N}\int_{\theta}^{\theta'}\partial_{\vartheta}\varpi(Y_{t},U_{t,i}(\delta_{T});\vartheta)\mathrm{d}\vartheta\right)^{2}\right) \\ &\leq \sum_{t=1}^{T}\frac{1}{N^{2}}\mathbb{E}\left((\widehat{W}_{t}^{T})^{-2}\right)^{1/2}\mathbb{E}\left(\left[\sum_{i=1}^{N}\int_{\theta}^{\theta'}\partial_{\vartheta}\varpi(Y_{t},U_{t,i}(\delta_{T});\vartheta)\mathrm{d}\vartheta\right]^{4}\right)^{1/2} \\ &= \frac{T}{N^{2}}\mathbb{E}\left((\widehat{W}_{1}^{T})^{-2}\right)^{1/2}\mathbb{E}\left(\left[\sum_{i=1}^{N}\int_{\theta}^{\theta'}\partial_{\vartheta}\varpi(Y_{t},U_{t,i}(\delta_{T});\vartheta)\mathrm{d}\vartheta\right]^{4}\right)^{1/2} \\ &\leq c\frac{TN}{N^{2}}\mathbb{E}\left((\widehat{W}_{1}^{T})^{-2}\right)^{1/2}\mathbb{E}\left(\left[\int_{\theta}^{\theta'}\partial_{\vartheta}\varpi(Y_{t},U_{t,i}(\delta_{T});\vartheta)\mathrm{d}\vartheta\right]^{4}\right)^{1/2} \end{split}$$

where we have interchanged derivative and integration by Assumption 5, have used the fact that the expectation of the integrals over ϑ have zero mean and that under $\mathbb P$ the terms are i.i.d. over index i. We also use the fact that for i.i.d. zero-mean random variables Z_i

$$\mathbb{E}\left[\left(\sum_{i=1}^P Z_i\right)^4\right] \leq c P^2 \mathbb{E}\left(Z_1^4\right).$$

Hence, we have using Assumptions 4 and 5 that $|\partial_{\vartheta}\varpi(Y_1,U_{1,1}(\delta_T);\vartheta)| \leq 2\chi(Y_1,U_{1,1}(\delta_T))$ for $\vartheta \in [\theta \wedge \theta',\theta \vee \theta']$ and T large enough. When θ is multidimensional, this can be established using the fundamental theorem of calculus for line integrals. It follows that

$$\widetilde{\mathbb{V}}\left(\sum_{t=1}^{T} J_{t}^{T}\right) \leq c \frac{TN}{N^{2}} \mathbb{E}\left(\left[\int_{\theta}^{\theta'} \partial_{\vartheta} \varpi(Y_{1}, U_{1,1}(\delta_{T}); \vartheta) d\vartheta\right]^{4}\right)^{1/2} \\
\leq c \frac{TN}{N^{2}} \mathbb{E}\left(\left[\int_{\theta}^{\theta'} \chi(Y_{1}, U_{1,1}(\delta_{T})) d\vartheta\right]^{4}\right)^{1/2} \\
= c \frac{TN}{N^{2}} \frac{\xi^{2}}{T} \mathbb{E}\left(\chi(Y_{1}, U_{1,1})^{4}\right)^{1/2} = O\left(\frac{1}{N}\right).$$

This concludes the proof of the lemma.

Proof of Lemma 15. We can rewrite L_t^T given by (103) as

$$L_{t}^{T} = \int_{0}^{\delta_{T}} \frac{1}{N\widehat{W}_{t}^{T}} \left(\sum_{i=1}^{N} \mathcal{S} \left\{ \varpi \left(Y_{t}, U_{t,i} \left(s \right) ; \theta \right) \right\} \right) \mathrm{d}s,$$

where S is the Stein operator defined in (99). By Assumption 6, we can apply Fubini's theorem to interchange the order of integration, and Stein's lemma [49, Lemma 1] shows that

$$\widetilde{\mathbb{E}}\left(L_{t}^{T} \middle| \mathcal{Y}^{T}\right) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left(\int_{0}^{\delta_{T}} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t, i}\left(s\right); \theta\right)\right\} ds \middle| \mathcal{Y}^{T}\right) = 0,$$

so in particular $\widetilde{\mathbb{E}}(L_t^T) = 0$. Hence, we have

$$\widetilde{\mathbb{V}}\left(L_{t}^{T}\right) = \widetilde{\mathbb{E}}\left(\left[\int_{0}^{\delta_{T}} \frac{1}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,i}\left(s\right);\theta\right)\right\} ds\right]^{2}\right)$$

$$= \widetilde{\mathbb{E}}\left(\frac{1}{\left(N\widehat{W}_{t}^{T}\right)^{2}} \int_{0}^{\delta_{T}} \int_{0}^{\delta_{T}} \left[\sum_{i=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,i}\left(s\right);\theta\right)\right\}\right] \left[\sum_{j=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,j}\left(r\right);\theta\right)\right\}\right] dr ds\right)$$

$$= \mathbb{E}\left(\frac{1}{N^{2}\widehat{W}_{t}^{T}} \int_{0}^{\delta_{T}} \int_{0}^{\delta_{T}} \left[\sum_{i=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,i}\left(s\right);\theta\right)\right\}\right] \left[\sum_{j=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,j}\left(r\right);\theta\right)\right\}\right] dr ds\right). (106)$$

The term (106) can be rewritten as

$$\begin{split} &\int_{0}^{\delta_{T}} \int_{0}^{\delta_{T}} \mathbb{E}\left[\frac{1}{\widehat{W}_{t}^{T}} \left(\frac{1}{N} \sum_{i=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,i}\left(s\right);\theta\right)\right\}\right) \left(\frac{1}{N} \sum_{j=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,j}\left(r\right);\theta\right)\right\}\right)\right] \mathrm{d}r \mathrm{d}s \\ &\leq \int_{0}^{\delta_{T}} \int_{0}^{\delta_{T}} \mathbb{E}\left[\left(\widehat{W}_{t}^{T}\right)^{-2}\right]^{1/2} \mathbb{E}\left[\left(\frac{1}{N} \sum_{i=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,i}\left(s\right);\theta\right)\right\}\right)^{2} \left(\frac{1}{N} \sum_{j=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,j}\left(r\right);\theta\right)\right\}\right)^{2}\right]^{1/2} \mathrm{d}r \mathrm{d}s \\ &\leq \int_{0}^{\delta_{T}} \int_{0}^{\delta_{T}} \mathbb{E}\left[\left(\widehat{W}_{t}^{T}\right)^{-2}\right]^{1/2} \mathbb{E}\left[\left(\frac{1}{N} \sum_{i=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,i}\left(s\right);\theta\right)\right\}\right)^{4}\right]^{1/4} \mathbb{E}\left[\left(\frac{1}{N} \sum_{j=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,j}\left(r\right);\theta\right)\right\}\right)^{4}\right]^{1/4} \\ &= \int_{0}^{\delta_{T}} \int_{0}^{\delta_{T}} \mathbb{E}\left[\left(\widehat{W}_{t}^{T}\right)^{-2}\right]^{1/2} \mathbb{E}\left[\left(\frac{1}{N} \sum_{i=1}^{N} \mathcal{S}\left\{\varpi\left(Y_{t}, U_{t,i}\left(s\right);\theta\right)\right\}\right)^{4}\right]^{1/2} \mathrm{d}r \mathrm{d}s \\ &\leq c \frac{\delta_{T}^{2}}{N} = O(\frac{N}{T^{2}}), \end{split}$$

by Assumption 9, $\mathbb{E}\left(\mathcal{S}\left\{\varpi\left(Y_{t},U_{t,i}\left(s\right);\theta\right)\right\}\right)=0$, and the fact that $U_{t,i}\left(s\right)$ are stationary and independent over i under \mathbb{P} . Therefore we have

$$\widetilde{\mathbb{V}}\left(\sum_{t=1}^T L_t^T\right) = T\widetilde{\mathbb{V}}\left(L_1^T\right) = O\left(\frac{N}{T^2}T\right) = O\left(\frac{N}{T}\right).$$

This concludes the proof of the lemma.

Proof of Lemma 16. We check here that the conditions of the conditional CLT given in Lemma 27 of Section A.9 are satisfied. Consider the term M^T given by (104) which can be decomposed as

$$M^{T} = \sum_{t=1}^{T} M_{t}^{T} = \sum_{t=1}^{T} \sum_{i=1}^{N} M_{t,i}^{T}, \tag{107}$$

where

$$M_{t,i}^{T} = \frac{\sqrt{2}}{N\widehat{W}^{T}} \int_{0}^{\delta_{T}} \partial_{u} \varpi \left(Y_{t}, U_{t,i} \left(s \right); \theta \right) dB_{t,i} \left(s \right). \tag{108}$$

It is straightforward to see that

$$\widetilde{\mathbb{E}}\left(M_{t}^{T}\middle|\mathcal{F}^{T}\right) = \mathbb{E}\left(\frac{\sqrt{2}}{N}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{t},U_{t,i}\left(s\right);\theta\right)\mathrm{d}B_{t,i}\left(s\right)\middle|\mathcal{F}^{T}\right) = 0$$

and

$$s_T^2 = \widetilde{\mathbb{V}}\left(M^T \middle| \mathcal{F}^T\right) = \sum_{t=1}^T \widetilde{\mathbb{V}}\left(M_t^T \middle| \mathcal{F}^T\right). \tag{109}$$

The term $\widetilde{\mathbb{V}}\left(\left.M_{t}^{T}\right|\mathcal{F}^{T}\right)$ satisfies

$$\widetilde{\mathbb{V}}\left(M_{t}^{T}\middle|\mathcal{F}^{T}\right) = \sum_{i=1}^{N} \widetilde{\mathbb{V}}\left(M_{t,i}^{T}\middle|\mathcal{F}^{T}\right)$$

$$= \sum_{i=1}^{N} \frac{2}{N^{2}\left(\widehat{W}_{t}^{T}\right)^{2}} \int_{0}^{\delta_{T}} \widetilde{\mathbb{E}}\left(\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,i}\left(s\right);\theta\right)\right\}^{2}\middle|\mathcal{F}^{T}\right) ds.$$

Letting

$$g(y, u; \theta) = \{\partial_u \varpi(y, u; \theta)\}^2,$$

and using Itô's formula, we obtain

$$\int_{0}^{\delta_{T}} \widetilde{\mathbb{E}}\left(\left\{\partial_{u}\varpi(Y_{t}, U_{t,i}\left(s\right);\theta)\right\}^{2} \middle| \mathcal{F}^{T}\right) ds$$

$$= \int_{0}^{\delta_{T}} \left\{\partial_{u}\varpi(Y_{t}, U_{t,i}\left(0\right);\theta)\right\}^{2} ds + \int_{0}^{\delta_{T}} \int_{0}^{s} \widetilde{\mathbb{E}}\left(\mathcal{S}\left\{g(Y_{t}, U_{t,i}\left(s\right);\theta)\right\} \middle| \mathcal{F}^{T}\right) dr ds$$

$$= \delta_{T}\left\{\partial_{u}\varpi(Y_{t}, U_{t,i}\left(0\right);\theta)\right\}^{2} + \int_{0}^{\delta_{T}} \int_{0}^{s} \widetilde{\mathbb{E}}\left(\mathcal{S}\left\{g(Y_{t}, U_{t,i}\left(s\right);\theta)\right\} \middle| \mathcal{F}^{T}\right) dr ds,$$

where

$$\int_{0}^{s} \widetilde{\mathbb{E}}\left(\mathcal{S}\left\{g(Y_{t}, U_{t, i}\left(s\right); \theta\right)\right\} | \mathcal{F}^{T}\right) dr = \int_{0}^{s} \int \mathcal{S}\left\{g(Y_{t}, e^{-r}U_{t, i} + \sqrt{1 - e^{-2r}}\varepsilon; \theta)\right\} \varphi\left(\varepsilon; 0, 1\right) d\varepsilon dr.$$

Therefore

$$s_T^2 = \sum_{t=1}^T \sum_{i=1}^N \frac{2}{N^2 \left(\widehat{W}_t^T\right)^2} \int_0^{\delta_T} \widetilde{\mathbb{E}} \left(\left\{ \partial_u \varpi(Y_t, U_{t,i}(s); \theta) \right\}^2 \middle| \mathcal{F}^T \right) ds$$

$$= \sum_{t=1}^T \sum_{i=1}^N \frac{2\delta_T}{N^2 \left(\widehat{W}_t^T\right)^2} \left\{ \partial_u \varpi(Y_t, U_{t,i}; \theta) \right\}^2$$
(110)

$$+\sum_{t=1}^{T}\sum_{i=1}^{N}\frac{2}{N^{2}\left(\widehat{W}_{t}^{T}\right)^{2}}\int_{0}^{\delta_{T}}\int_{0}^{s}\widetilde{\mathbb{E}}\left(\mathcal{S}\left\{g(Y_{t},U_{t,i}\left(s\right);\theta\right)\right\}|\mathcal{F}^{T}\right)\mathrm{d}r\mathrm{d}s.\tag{111}$$

To show that the term (111) vanishes in probability, we show that it vanishes in absolute mean

$$\begin{split} &\frac{1}{N^2}\mathbb{E}\left(\left|\sum_{t=1}^T\sum_{i=1}^N\frac{2}{\left(\widehat{W}_t^T\right)^2}\int_0^{\delta_T}\int_0^s\widetilde{\mathbb{E}}\left(\mathcal{S}\left\{g(Y_t,U_{t,i}\left(r\right);\theta\right)\right\}|\mathcal{F}^T\right)\mathrm{d}r\mathrm{d}s\right|\right)\\ &\leq \frac{1}{N^2}\sum_{t=1}^T\sum_{i=1}^N\widetilde{\mathbb{E}}\left(\frac{2}{\left(\widehat{W}_t^T\right)^2}\int_0^{\delta_T}\int_0^s\left|\widetilde{\mathbb{E}}\left(\mathcal{S}\left\{g(Y_t,U_{t,i}\left(r\right);\theta\right)\right\}|\mathcal{F}^T\right)\right|\mathrm{d}r\mathrm{d}s\right)\\ &=\frac{1}{N^2}\sum_{t=1}^T\sum_{i=1}^N\widetilde{\mathbb{E}}\left(\frac{2}{\left(\widehat{W}_t^T\right)^2}\int_0^{\delta_T}\int_0^s\left|\widetilde{\mathbb{E}}\left(\mathcal{S}\left\{g(Y_t,U_{t,i}\left(r\right);\theta\right)\right\}|\mathcal{F}^T\right)\right|\mathrm{d}r\mathrm{d}s\right)\\ &=\frac{1}{N^2}\sum_{t=1}^T\sum_{i=1}^N\int_0^{\delta_T}\int_0^s\widetilde{\mathbb{E}}\left(\widetilde{\mathbb{E}}\left[\frac{2}{\left(\widehat{W}_t^T\right)^2}\left|\mathcal{S}\left\{g(Y_t,U_{t,i}\left(r\right);\theta\right)\right\}\right|\right|\mathcal{F}^T\right)\right)\mathrm{d}r\mathrm{d}s\\ &=\frac{2}{N^2}\sum_{t=1}^T\sum_{i=1}^N\int_0^{\delta_T}\int_0^s\mathbb{E}\left(\widetilde{\mathbb{E}}\left[\frac{1}{\widehat{W}_t^T}\left|\mathcal{S}\left\{g(Y_t,U_{t,i}\left(r\right);\theta\right)\right\}\right]\right)\mathrm{d}r\mathrm{d}s\\ &=\frac{2}{N^2}\sum_{t=1}^T\sum_{i=1}^N\int_0^{\delta_T}\int_0^s\mathbb{E}\left[\left(\widehat{W}_t^T\right)^{-2}\right]^{1/2}\mathbb{E}\left[\left(\mathcal{S}\left\{g(Y_t,U_{t,i}\left(r\right);\theta\right)\right\}\right)^2\right]^{1/2}\mathrm{d}r\mathrm{d}s\\ &=\delta_T^2\frac{NT}{N^2}\mathbb{E}\left[\left(\widehat{W}_t^T\right)^{-2}\right]^{1/2}\mathbb{E}\left[\left(\mathcal{S}\left\{g(Y_t,U_{t,1};\theta\right)\right\}\right)^2\right]^{1/2}\\ &=\delta_T\mathbb{E}\left[\left(\widehat{W}_t^T\right)^{-2}\right]^{1/2}\mathbb{E}\left[\left(\mathcal{S}\left\{g(Y_t,U_{t,1};\theta\right)\right\}\right)^2\right]^{1/2}=O(\delta_T), \end{split}$$

by Assumption 7 and the fact that $U_{t,i}(r)$ are stationary and i.i.d. over t, i under \mathbb{P} . Going back to our calculation of s_T^2 , we now treat the term (110)

$$\sum_{t=1}^{T} \sum_{i=1}^{N} \frac{2\delta_{T}}{N^{2} \left(\widehat{W}_{t}^{T}\right)^{2}} \left\{ \partial_{u} \varpi(Y_{t}, U_{t,i}; \theta) \right\}^{2} = \frac{2}{T} \sum_{t=1}^{T} g_{T} \left(Y_{t}, U_{t}\right),$$

where

$$g_T\left(Y_t, U_t\right) := \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{W}_t^T\right)^{-2} \left\{\partial_u \varpi(Y_t, U_{t,i}; \theta)\right\}^2.$$

In order to apply the WLLN we have to check that

$$\sum_{t=1}^{T} \widetilde{\mathbb{E}}\left(\frac{\left|g_{T}\left(Y_{t}, U_{t}\right)\right|}{T} \mathbb{I}\left\{\left|g_{T}\left(Y_{t}, U_{t}\right)\right| \geq \epsilon T\right\}\right) = \widetilde{\mathbb{E}}\left(\left|g_{T}\left(Y_{1}, U_{1}^{T}\right)\right| \mathbb{I}\left\{\left|g_{T}\left(Y_{1}, U_{1}^{T}\right)\right| \geq \epsilon T\right\}\right) \to 0,$$

or equivalently that

$$\left\{g_{T}\left(Y_{1}, U_{1}^{T}\right)\right\}_{T \geq 1} = \left\{\frac{1}{N} \sum_{i=1}^{N} \left(\widehat{W}_{1}^{T}\right)^{-2} \left\{\partial_{u} \varpi(Y_{1}, U_{1,i}^{T}; \theta)\right\}^{2}; T \geq 1\right\},\,$$

is uniformly integrable. We use de la Vallée-Poussin theorem; i.e., $\{X_n; n \geq 1\}$ is uniformly integrable if and only if there exists a non-negative increasing convex function g such that $g(x)/x \to \infty$ as $x \to \infty$ and $\sup_{n \geq 1} \widetilde{\mathbb{E}}\left[g\left(|X_n|\right)\right] < \infty$.

If g is convex then by Jensen's inequality

$$\begin{split} \widetilde{\mathbb{E}}\left[g\left(\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{W}_{1}^{T}\right)^{-2}\left\{\partial_{u}\varpi(Y_{1},U_{1,i}^{T};\theta)\right\}^{2}\right)\right] &\leq \widetilde{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^{N}g\left(\left(\widehat{W}_{1}^{T}\right)^{-2}\left\{\partial_{u}\varpi(Y_{1},U_{1,i}^{T};\theta)\right\}^{2}\right)\right] \\ &= \widetilde{\mathbb{E}}\left[g\left(\left(\widehat{W}_{1}^{T}\right)^{-2}\left\{\partial_{u}\varpi(Y_{1},U_{1,1};\theta)\right\}^{2}\right)\right], \end{split}$$

since the variables $\{U_{1,i}^T; i \in 1 : N\}$ are exchangeable under $\overline{\pi}(\cdot | \theta)$. Therefore it will suffice to assume that for some non-negative, increasing convex function g such that $g(x)/x \to \infty$

$$\limsup_{T} \widetilde{\mathbb{E}} \left[g \left(\left(\widehat{W}_{1}^{T} \right)^{-2} \left\{ \partial_{u} \varpi(Y_{1}, U_{1,1}; \theta) \right\}^{2} \right) \right] < \infty$$

or equivalently that the family

$$\left\{ \left(\widehat{W}_{1}^{T}\right)^{-2} \left\{ \partial_{u} \varpi(Y_{1}, U_{1,1}; \theta) \right\}^{2}; T \geq 1 \right\}$$

is uniformly integrable under $\widetilde{\mathbb{P}}$. However, this is satisfied as there exists $\varepsilon > 0$ such that

$$\lim \sup_{T} \widetilde{\mathbb{E}} \left[\left(\left(\widehat{W}_{1}^{T} \right)^{-2} \left\{ \partial_{u} \varpi(Y_{1}, U_{1,1}; \theta) \right\}^{2} \right)^{1+\varepsilon} \right] < \infty,$$

which can be verified by using Cauchy-Schwarz inequality and Assumptions 4 and 8. By applying now the WLLN, we have

$$\frac{1}{T} \sum_{t=1}^{T} \left(g_T \left(Y_t, U_t \right) - \widetilde{\mathbb{E}} \left[g_T \left(Y_t, U_t \right) \right] \right) \stackrel{\mathbb{P}}{\to} 0,$$

where

$$\widetilde{\mathbb{E}}\left[g_T\left(Y_1, U_1^T\right)\right] = \widetilde{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^N \left(\widehat{W}_1^T\right)^{-2} \left\{\partial_u \varpi(Y_1, U_{1,i}^T; \theta)\right\}^2\right] \\
= \widetilde{\mathbb{E}}\left[\left(\widehat{W}_1^T\right)^{-2} \left\{\partial_u \varpi(Y_1, U_{1,1}; \theta)\right\}^2\right] = \mathbb{E}\left[\left(\widehat{W}_1^T\right)^{-1} \left\{\partial_u \varpi(Y_1, U_{1,1}; \theta)\right\}^2\right].$$

By Cauchy-Schwarz, Assumptions 4 and 8 again, we have

$$\limsup_{T} \mathbb{E}\left\{ \left(\widehat{W}_{1}^{T}\right)^{-1-\varepsilon} \left\{ \partial_{u} \varpi(Y_{1}, U_{1,1}; \theta) \right\}^{2+2\varepsilon} \right\} < \infty.$$

Therefore the family $\{\left(\widehat{W}_{1}^{T}\right)^{-1} \{\partial_{u}\varpi(Y_{1},U_{1,1};\theta)\}^{2}; T \geq 1\}$ is also uniformly integrable under \mathbb{P} and, since $\widehat{W}_{t}^{T} \stackrel{\mathbb{P}}{\to} 1$, we have

$$\mathbb{E}\left(\left(\widehat{W}_{1}^{T}\right)^{-1}\left\{\partial_{u}\varpi(Y_{1},U_{1,1};\theta)\right\}^{2}\right)\to\mathbb{E}\left(\left\{\partial_{u}\varpi(Y_{1},U_{1,1};\theta)\right\}^{2}\right)=\frac{\kappa\left(\theta\right)^{2}}{2}.$$

Hence, it follows that $s_T^2 \stackrel{\widetilde{\mathbb{P}}}{\to} \kappa(\theta)^2$ and condition (159) of Lemma 27 is satisfied.

We now need to check the Lindeberg condition (160), i.e., that for any $\varepsilon > 0$

$$\sum_{t=1}^{T} \widetilde{\mathbb{E}} \left(\left| M_{t}^{T} \right|^{2} \mathbb{I} \left\{ \left| M_{t}^{T} \right| \geq \varepsilon \right\} \middle| \mathcal{F}^{T} \right) \xrightarrow{\widetilde{\mathbb{P}}} 0. \tag{112}$$

Since the l.h.s. of (112) is non-negative, it is enough to show that its unconditional expectation vanishes or equivalently that $T |M_1^T|^2$ is uniformly integrable. We have

$$\begin{split} \sum_{t=1}^{T} \widetilde{\mathbb{E}} \left(\left| M_{t}^{T} \right|^{2} \mathbb{I} \left\{ \left| M_{t}^{T} \right| \geq \varepsilon \right\} \right) &= T \widetilde{\mathbb{E}} \left(\left| M_{1}^{T} \right|^{2} \mathbb{I} \left\{ \left| M_{1}^{T} \right| \geq \varepsilon \right\} \right) \\ &= T \widetilde{\mathbb{E}} \left[\left\{ \frac{\sqrt{2}}{N \widehat{W}_{1}^{T}} \sum_{i=1}^{N} \int_{0}^{\delta_{T}} \partial_{u} \varpi \left(Y_{1}, U_{1,i}^{T}(s); \theta \right) \mathrm{d}B_{1,i}(s) \right\}^{2} \mathbb{I} \left\{ \left| M_{1}^{T} \right| \geq \varepsilon \right\} \right] \\ &= \frac{2T}{N^{2}} \widetilde{\mathbb{E}} \left[\left\{ \frac{\mathbb{I} \left\{ \left| M_{1}^{T} \right| \geq \varepsilon \right\}}{\left(\widehat{W}_{1}^{T} \right)^{3/2}} \right\} \frac{1}{\left(\widehat{W}_{1}^{T} \right)^{1/2}} \left\{ \sum_{i=1}^{N} \int_{0}^{\delta_{T}} \partial_{u} \varpi \left(Y_{1}, U_{1,i}^{T}(s); \theta \right) \mathrm{d}B_{1,i}(s) \right\}^{2} \right] \\ &\leq \frac{2T}{N^{2}} \widetilde{\mathbb{E}} \left(\frac{\mathbb{I} \left\{ \left| M_{1}^{T} \right| \geq \varepsilon \right\}}{\left(\widehat{W}_{1}^{T} \right)^{3}} \right)^{1/2} \widetilde{\mathbb{E}} \left(\frac{1}{\left(\widehat{W}_{1}^{T} \right)} \left\{ \sum_{i=1}^{N} \int_{0}^{\delta_{T}} \partial_{u} \varpi \left(Y_{1}, U_{1,i}^{T}(s); \theta \right) \mathrm{d}B_{1,i}(s) \right\}^{4} \right)^{1/2} \end{split}$$

by Cauchy-Schwartz and

$$\widetilde{\mathbb{E}}\left(\frac{1}{\left(\widehat{W}_{1}^{T}\right)}\left\{\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{1},U_{1,i}^{T}\left(s\right);\theta\right)\mathrm{d}B_{1,i}\left(s\right)\right\}^{4}\right) \\
= \mathbb{E}\left(\left\{\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{1},U_{1,i}^{T}\left(s\right);\theta\right)\mathrm{d}B_{1,i}\left(s\right)\right\}^{4}\right) \\
= \mathbb{E}\left(\mathbb{E}\left(\left\{\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{1},U_{1,i}^{T}\left(s\right);\theta\right)\mathrm{d}B_{1,i}\left(s\right)\right\}^{4}\middle|\mathcal{Y}^{T}\right)\right) \\
\leq cN^{2}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{1},U_{1,1}\left(s\right);\theta\right)\mathrm{d}B_{1,1}\left(s\right)\right)^{4}\right] \\
\leq cN^{2}\left\{3\int_{0}^{\delta_{T}}\mathbb{E}\left[\left(\partial_{u}\varpi\left(Y_{1},U_{1,1}\left(s\right);\theta\right)\right)^{4}\right]^{1/2}\mathrm{d}s\right\}^{2} \\
= c'N^{2}\delta_{T}^{2}\mathbb{E}\left[\left(\partial_{u}\varpi\left(Y_{1},U_{1,1};\theta\right)\right)^{4}\right] < \infty,$$

where the penultimate inequality follows from [54, Theorem 1] and the last one by Assumption 8. Therefore, we have

$$\sum_{t=1}^{T} \widetilde{\mathbb{E}} \left(\left| M_{t}^{T} \right|^{2} \mathbb{I} \left\{ \left| M_{t}^{T} \right| \geq \varepsilon \right\} \right) \leq \sqrt{c'} \frac{2T}{N^{2}} N \delta_{T} \mathbb{E} \left[\left(\partial_{u} \varpi \left(Y_{1}, U_{1,1}; \theta \right) \right)^{4} \right]^{1/2} \widetilde{\mathbb{E}} \left(\frac{\mathbb{I} \left\{ \left| M_{1}^{T} \right| \geq \varepsilon \right\}}{\left(\widehat{W}_{1}^{T} \right)^{3}} \right)^{1/2} \\
= 2\sqrt{c'} \mathbb{E} \left[\left(\partial_{u} \varpi \left(Y_{1}, U_{1,1}; \theta \right) \right)^{4} \right]^{1/2} \widetilde{\mathbb{E}} \left(\frac{\mathbb{I} \left\{ \left| M_{1}^{T} \right| \geq \varepsilon \right\}}{\left(\widehat{W}_{1}^{T} \right)^{3}} \right)^{1/2} . \tag{113}$$

Using Holder's inequality, Assumption 4 then Chebyshev's inequality, we have

$$\widetilde{\mathbb{E}}\left(\frac{\mathbb{I}\left\{\left|M_{1}^{T}\right| \geq \varepsilon\right\}}{\left(\widehat{W}_{t}^{T}\right)^{3}}\right) \leq \widetilde{\mathbb{E}}\left[\left(\widehat{W}_{t}^{T}\right)^{-3-3\epsilon}\right]^{1/(1+\epsilon)} \,\widetilde{\mathbb{P}}\left(\left|M_{1}^{T}\right| \geq \epsilon\right)^{\epsilon/(1+\epsilon)}$$

$$\leq c'' \,\widetilde{\mathbb{P}}\left(\left|M_{1}^{T}\right| \geq \epsilon\right)^{\epsilon/(1+\epsilon)}$$

$$\leq c'' \,\left(\frac{\widetilde{\mathbb{E}}\left[\left(M_{1}^{T}\right)^{2}\right]}{\epsilon^{2}}\right)^{\epsilon/(1+\epsilon)}.$$
(114)

To proceed, we need to control the second moment of M_1^T

$$\widetilde{\mathbb{E}}\left[\left(M_{1}^{T}\right)^{2}\right] = \widetilde{\mathbb{E}}\left[\left(\frac{\sqrt{2}}{N\widehat{W}_{1}^{T}}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{1},U_{1,i}^{T}(s);\theta\right)dB_{1,i}(s)\right)^{2}\right] \\
= \frac{2}{N^{2}}\mathbb{E}\left[\frac{1}{\widehat{W}_{1}^{T}}\left(\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{1},U_{1,i}^{T}(s);\theta\right)dB_{1,i}(s)\right)^{2}\right] \\
\leq \frac{2}{N^{2}}\mathbb{E}\left[\left(\widehat{W}_{1}^{T}\right)^{-2}\right]^{1/2}\mathbb{E}\left[\left(\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{1},U_{1,i}^{T}(s);\theta\right)dB_{1,i}(s)\right)^{4}\right]^{1/2} \\
= \frac{2}{N^{2}}\mathbb{E}\left[\left(\widehat{W}_{1}^{T}\right)^{-2}\right]^{1/2}\mathbb{E}\left[\mathbb{E}\left\{\left(\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{1},U_{1,i}^{T}(s);\theta\right)dB_{1,i}(s)\right)^{4}\middle|\mathcal{Y}^{T}\right\}\right]^{1/2} \\
\leq c\frac{N}{N^{2}}\mathbb{E}\left[\left(\widehat{W}_{1}^{T}\right)^{-2}\right]^{1/2}\mathbb{E}\left[\mathbb{E}\left\{\left(\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{1},U_{1,i}^{T}(s);\theta\right)dB_{1,i}(s)\right)^{4}\middle|\mathcal{Y}^{T}\right\}\right]^{1/2} \\
\leq c\frac{1}{N}\mathbb{E}\left[\left(\widehat{W}_{1}^{T}\right)^{-2}\right]^{1/2}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{1},U_{1,1}(s);\theta\right)dB_{1,1}(s)\right)^{4}\right]^{1/2} \\
\leq \frac{c}{N}\mathbb{E}\left[\left(\widehat{W}_{1}^{T}\right)^{-2}\right]^{1/2}\mathcal{E}\left[\left(\partial_{u}\varpi\left(Y_{1},U_{1,1};\theta\right)\right)^{4}\right]^{1/2} = O(1/T), \tag{115}$$

by Cauchy-Schwartz, [54, Theorem 1] and Assumptions 4 and 8.

By combining (113), (114) and (115), it follows that (112) holds. Therefore by the Lindeberg central limit theorem of Lemma 27 applied conditionally on \mathcal{F}^T and using $s_T^2 \xrightarrow{\widetilde{\mathbb{P}}} \kappa(\theta)^2$, we obtain

$$\sum_{t=1}^{T} M_{t}^{T} \middle| \mathcal{F}^{T} \Rightarrow \mathcal{N}\left(0, \frac{\kappa\left(\theta\right)^{2}}{2}\right).$$

Proof of Lemma 17. Notice that

$$\frac{1}{2} \sum_{t=1}^{T} (\eta_t)^2 = \frac{1}{2} \sum_{t=1}^{T} \left\{ \frac{\widehat{W}_t^T(Y_t \mid \theta'; V_t) - \widehat{W}_t^T}{\widehat{W}_t^T} \right\}^2 = \frac{1}{2} \sum_{t=1}^{T} \left\{ J_t^T + H_t^T \right\}^2 = \frac{1}{2} \sum_{t=1}^{T} \left\{ \left[J_t^T \right]^2 + \left[H_t^T \right]^2 + 2J_t^T H_t^T \right\}.$$

We know from Lemma 14 that $\sum_{t=1}^{T} (J_t^T)^2 \stackrel{\widetilde{\mathbb{P}}}{\to} 0$. The $(H_t^T)^2$ terms are given by

$$\sum_{t=1}^{T} (H_t^T)^2 = \sum_{t=1}^{T} (L_t^T + M_t^T)^2$$
$$= \sum_{t=1}^{T} (L_t^T)^2 + 2L_t^T M_t^T + (M_t^T)^2.$$

The first term vanishes in probability since by Lemma 15

$$\widetilde{\mathbb{E}}\left(\sum_{t=1}^{T} (L_t^T)^2\right) = \sum_{t=1}^{T} \widetilde{\mathbb{V}}\left(L_t^T\right) \to 0.$$

For the product term notice that by two applications of the Cauchy-Schwartz inequality

$$\begin{split} \widetilde{\mathbb{E}}\left(\left|\sum_{t=1}^{T} L_{t}^{T} M_{t}^{T}\right|\right) &\leq \widetilde{\mathbb{E}}\left(\left\{\sum_{t=1}^{T} \left(L_{t}^{T}\right)^{2}\right\}^{1/2} \left\{\sum_{t=1}^{T} \left(M_{t}^{T}\right)^{2}\right\}^{1/2}\right) \\ &\leq \widetilde{\mathbb{E}}\left(\sum_{t=1}^{T} \left(L_{t}^{T}\right)^{2}\right)^{1/2} \widetilde{\mathbb{E}}\left(\sum_{t=1}^{T} \left(M_{t}^{T}\right)^{2}\right)^{1/2} \\ &= \left(\sum_{t=1}^{T} \widetilde{\mathbb{V}}(L_{t}^{T})\right)^{1/2} \left(\sum_{t=1}^{T} \widetilde{\mathbb{V}}(M_{t}^{T})\right)^{1/2} \to 0, \end{split}$$

by Lemmas 15 and 16. Finally, we also have

$$\sum_{t=1}^{T} \widetilde{\mathbb{E}}\left(\left(M_{t}^{T}\right)^{2}\right) = O\left(1\right)$$

by Lemma 16.

For the term involving the product $J_t^T H_t^T$, we have similarly by two applications of the Cauchy-Schwartz inequality

$$\begin{split} \widetilde{\mathbb{E}}\left(\left|\sum_{t=1}^{T} J_{t}^{T} H_{t}^{T}\right|\right) &\leq \widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \left(J_{t}^{T}\right)^{2}\right)^{1/2} \left(\sum_{t=1}^{T} \left(H_{t}^{T}\right)^{2}\right)^{1/2}\right] \\ &\leq \widetilde{\mathbb{E}}\left(\sum_{t=1}^{T} \left(J_{t}^{T}\right)^{2}\right)^{1/2} \widetilde{\mathbb{E}}\left(\sum_{t=1}^{T} \left(H_{t}^{T}\right)^{2}\right)^{1/2}. \end{split}$$

By Lemmas 14, 15 and 16, the first factor vanishes, while we have just shown that the second factor is O(1).

Finally, conditionally on \mathcal{F}^T , the terms $(M_t^T)^2$ are independent. We want to apply the conditional WLLN to show that

$$\sum_{t=1}^{T} \left(M_{t}^{T} \right)^{2} - \widetilde{\mathbb{E}} \left(\left(M_{t}^{T} \right)^{2} \middle| \mathcal{F}^{T} \right) \stackrel{\widetilde{\mathbb{P}}}{\to} 0.$$

As we have already shown that Lemma 16 holds, we only need to check that for any $\epsilon > 0$

$$\sum_{t=1}^{T} \widetilde{\mathbb{E}} \left(\left(M_{t}^{T} \right)^{2} \mathbb{I} \left\{ \left| M_{t}^{T} \right| \geq \varepsilon \right\} \middle| \mathcal{F}^{T} \right) \stackrel{\widetilde{\mathbb{P}}}{\to} 0.$$

This has already been established in the proof of Lemma 16.

A.6 Sufficient conditions to ensure Assumption 3

We will provide here sufficient conditions to ensure convergence happens almost surely, hence in probability. In the notation of Section A.5, let μ^T denote the conditional law of

$$R^T := \sum_{t=1}^T \log \left(\frac{\widehat{W}_t^T(Y_t \mid \theta'; U_t')}{\widehat{W}_t^T} \right) = \sum_{t=1}^T \log(1 + \eta_t^T) = \sum_{t=1}^T \eta_t^T - \frac{1}{2} \sum_{t=1}^T [\eta_t^T]^2 + \sum_{t=1}^T h(\eta_t^T) [\eta_t^T]^2,$$

given \mathcal{F}^T where θ, ξ are fixed, $\theta' = \theta + \xi/\sqrt{T}$, $\xi \sim v(\cdot)$, $U \sim \overline{\pi}_T(\cdot|\theta)$, $U' \sim K_{\rho_T}(U,\cdot)$ with ρ_T given by (25) and $N_T \to \infty$ as $T \to \infty$ with $N_T/T \to 0$. We want to control the term

$$\begin{split} \sup_{\theta \in N(\bar{\theta})} & \widetilde{\mathbb{E}} \left[\left. d_{BL}(\mu^T, \varphi\left(\cdot; -\frac{\kappa^2(\theta)}{2}, \kappa^2(\theta)\right) \right| \mathcal{Y}^T \right] \\ &= \sup_{\theta \in N(\bar{\theta})} \iint \left\{ \overline{\pi}_T \left(\mathrm{d}u_0 | \, \theta \right) \upsilon \left(\mathrm{d}\xi \right) \right. \\ & \times \sup_{f: \|f\|_{BL} \le 1} \left| \int K_{\rho_T} \left(u_0, \mathrm{d}u_1 \right) f \left\{ \log \left(\frac{\hat{p}(Y_{1:T} \mid \theta_0 + \xi/\sqrt{T}, u_1)/p(Y_{1:T} \mid \theta + \xi/\sqrt{T})}{\hat{p}(Y_{1:T} \mid \theta, u_0)//p(Y_{1:T} \mid \theta)} \right) \right\} \\ & - \int \varphi(\mathrm{d}w; -\frac{\kappa^2(\theta)}{2}, \kappa^2(\theta)) f\left(w \right) \left| \right\}. \end{split}$$

We state sufficient conditions under which this result holds in the setting where d=1, p=1 and $\psi=1$. The extension to the multivariate scenario is straightforward albeit tedious. As in Theorem 3, we define

$$\kappa(\theta)^2 = 2\mathbb{E}\left(\left\{\partial_u \varpi(Y_1, U_{1,1}; \theta)\right\}^2\right).$$

We will also write

$$\kappa(y,\theta)^2 = 2\mathbb{E}\left(\left\{\partial_u \varpi(Y_1, U_{1,1}; \theta)\right\}^2 \middle| Y_1 = y\right).$$

Assumption 10. Let $B: \mathcal{Y} \to \mathbb{R}^+$ be a measurable function such that $\mathbb{E}B(Y_1)^{10} < \infty$, and let $\epsilon_T \to 0$ as $T \to \infty$. Assume that $\int \xi^{10} v(\mathrm{d}\xi) < \infty$, that $\kappa^2(\cdot, \theta)$ is measurable for all θ and $\kappa^2(y, \cdot)$ is continuous in θ for all y, that $\kappa(\theta)$ is locally Lispschitz around $\bar{\theta}$ and that the following inequalities hold:

$$\kappa^2(y,\theta) \le B(y),\tag{116}$$

$$\sup_{\theta \in N(\bar{\theta})} \mathbb{E}\left[\left(\widehat{W}_1^T\right)^{-6} \middle| Y_1 = y\right] \le B(y),\tag{117}$$

$$\sup_{\theta \in N(\bar{\theta})} \mathbb{E}^{1/2} \left[\left| \left(\widehat{W}_t^T \right)^2 \right| \middle| Y_t = y \right] \le B(y), \tag{118}$$

$$\sup_{\theta \in N(\bar{\theta})} \mathbb{E}^{1/2} \left[\left\{ 2\partial_{\theta} \varpi(y, U_{t,1}; \theta)^2 \right\}^2 \middle| \right] \le B(y), \tag{119}$$

$$\sup_{\theta \in N(\bar{\theta})} \mathbb{E}\left[\left|\frac{1}{\widehat{W}_t^T} - 1\right|^2 \middle| Y_t = y\right], \sup_{\theta \in N(\bar{\theta})} \mathbb{E}\left[\left|\frac{1}{\left(\widehat{W}_t^T\right)^2} - 1\right|^2 \middle| Y_t = y\right] \le \epsilon_T B(y), \tag{120}$$

$$\sup_{\theta \in N(\bar{\theta})} \mathbb{E}\left[\left|\widehat{W}_t^T - 1\right|^2 \middle| Y_t = y\right] \le \epsilon_T B(y), \tag{121}$$

$$\sup_{\theta \in N(\bar{\theta})} \mathbb{E}\left[\left. \frac{\eta_t^T}{1 + \eta_t^T} \right| Y_t = y\right] \le B(y), \tag{122}$$

$$\sup_{\theta \in N(\bar{\theta})} \mathbb{E}\left[\left.\partial_{\theta}\varpi(Y_{t}, U_{t, 1}; \theta)^{10}\right| Y_{t} = y\right] \leq B(y),\tag{123}$$

$$\sup_{\theta \in N(\bar{\theta})} \mathbb{E}\left[\left(\partial_u \varpi\left(Y_t, U_{t,1}; \theta\right)\right)^{10} \middle| Y_t = y\right] \le B(y), \tag{124}$$

$$\sup_{\theta \in N(\bar{\theta})} \mathbb{E}\left[\left[\partial_{uuu}^{3} \varpi\left(Y_{t}, U_{t, 1}\left(0\right); \theta\right)\right]^{4} \middle| Y_{t} = y\right] \leq B(y), \tag{125}$$

$$\sup_{\theta \in N(\bar{\theta})} \mathbb{E}\left[\mathcal{S}\left\{-\partial_u \varpi\left(Y_t, U_{t,1}; \theta\right) U_{t,1} + \partial_{u,u}^2 \varpi\left(Y_t, U_{t,1}; \theta\right)\right\}^{10} \middle| Y_t = y\right] \le B(y).$$
(126)

Under Assumption 10 then Assumption 3 is satisfied as established in the following theorem.

Theorem 18. Under Assumption 10, we have as $T \to \infty$

$$\sup_{\theta \in N(\bar{\theta})} \widetilde{\mathbb{E}} \left[\left. d_{BL}(\mu^T, \varphi\left(\cdot; -\frac{\kappa^2(\theta)}{2}, \kappa^2(\theta)\right) \right| \mathcal{Y}^T \right] \to 0 \ \mathbb{P}^Y - \text{a.s.}$$

The proof of this result will require establish a few preliminary lemmas. Let us first recall the decomposition

$$\eta_t^T = J_t^T + L_t^T + M_t^T, (127)$$

where J_t^T, L_t^T and M_t^T are defined in (102)-(104). We rearrange the above expression as

$$R^{T} = M^{T} - \frac{1}{2} \sum (\eta_{t}^{T})^{2} + \mathcal{R}_{1}^{T}, \tag{128}$$

where $M^T := \sum_{t=1}^T M_t^T = \sum_{t=1}^T \sum_{i=1}^N M_{t,i}^T$ where $M_{t,i}^T$ is defined in (108).

We can further decompose M_t^T as

$$\begin{split} \sum_{t=1}^{T} M_{t}^{T} &= \sum_{t=1}^{T} \int_{0}^{\delta_{T}} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \partial_{u} \varpi \left(Y_{t}, U_{t,i}\left(s\right); \theta \right) \mathrm{d}B_{t,i}\left(s\right) \\ &= \sum_{t=1}^{T} \int_{0}^{\delta_{T}} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \partial_{u} \varpi \left(Y_{t}, U_{t,i}\left(0\right); \theta \right) \mathrm{d}B_{t,i}\left(s\right) \\ &+ \sum_{t=1}^{T} \int_{0}^{\delta_{T}} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \left[-\int_{0}^{s} \partial_{uu}^{2} \varpi \left(Y_{t}, U_{t,i}\left(r\right); \theta \right) U_{t,i}\left(r\right) \mathrm{d}r + \int_{0}^{s} \partial_{uuu}^{3} \left(Y_{t}, U_{t,i}\left(r\right); \theta \right) \mathrm{d}B_{t,i}\left(s\right) \\ &+ \sum_{t=1}^{T} \int_{0}^{\delta_{T}} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \sqrt{2} \int_{0}^{s} \partial_{uu}^{2} \varpi \left(Y_{t}, U_{t,i}\left(r\right); \theta \right) \mathrm{d}B_{t,i}\left(r\right) \mathrm{d}B_{t,i}\left(s\right) \\ &= \sum_{t=1}^{T} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \partial_{u} \varpi \left(Y_{t}, U_{t,i}\left(0\right); \theta \right) B_{t,i}\left(\delta_{T}\right) + \mathcal{R}_{2}^{T}, \end{split}$$

where

$$\mathcal{R}_{2}^{T} := -\sum_{t=1}^{T} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} \mathcal{S} \left\{ \partial_{u} \varpi \left(Y_{t}, U_{t,i} \left(r \right) ; \theta \right) \right\} dr dB_{t,i} \left(s \right)$$

$$+ \sum_{t=1}^{T} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} \sqrt{2} \int_{0}^{s} \widetilde{\mathbb{E}} \left[d_{BL} (\mu^{T}, \varphi \left(\cdot ; -\frac{\kappa^{2}(\theta)}{2}, \kappa^{2}(\theta) \right) \middle| \mathcal{Y}^{T} \right]$$

$$\partial_{uuu}^{3} \varpi \left(Y_{t}, U_{t,i} \left(r \right) ; \theta \right) dB_{t,i} \left(r \right) dB_{t,i} \left(s \right).$$

Thus we can write

$$M^{T} = \left\{ \sum_{i,t} \frac{2}{NT \left(\widehat{W}_{t}^{T}\right)^{2}} \left[\partial_{u} \varpi \left(Y_{t}, U_{t,i} \left(0 \right) ; \theta \right) \right]^{2} \right\}^{1/2} Z + \mathcal{R}_{2}^{T}$$
$$= \hat{s}_{T}(\theta) Z + \mathcal{R}_{2}^{T},$$

where $Z \sim \mathcal{N}(0, 1)$. Finally let

$$\mathcal{R}_{3}^{T} := \frac{1}{2} \widetilde{\mathbb{E}} \left[\left| \sum_{t=1}^{T} \left(\eta_{t}^{T} \right)^{2} - \kappa^{2}(\theta) \right| \right| \mathcal{F}^{T} \right].$$

Lemma 19. We have

$$\widetilde{\mathbb{E}}\left[d_{BL}\left(\mu^{T}, \varphi\left(\cdot; -\frac{\kappa^{2}(\theta)}{2}, \kappa^{2}(\theta)\right)\right) \middle| \mathcal{Y}^{T}\right] \leq \sqrt{\frac{2}{\pi}} \widetilde{\mathbb{E}}\left[\left|\hat{s}_{T}(\theta) - \kappa(\theta)\right|\right| \mathcal{Y}^{T}\right] \\
+ \widetilde{\mathbb{E}}\left[\left|\mathcal{R}_{1}^{T}\right|\right| \mathcal{Y}^{T}\right] + \widetilde{\mathbb{E}}\left[\left|\mathcal{R}_{2}^{T}\right|\right| \mathcal{Y}^{T}\right] + \widetilde{\mathbb{E}}\left[\left|\mathcal{R}_{3}^{T}\right|\right| \mathcal{Y}^{T}\right]. \tag{129}$$

Proof. We first notice that if $f \in BL(1)$ we have

$$\begin{split} &\left| \widetilde{\mathbb{E}} \left[f \left(R^T \right) \middle| \mathcal{F}^T \right] - \int f(z) \varphi \left(z; -\frac{\kappa^2(\theta)}{2}, \kappa^2(\theta) \right) \mathrm{d}z \right| \\ &= \left| \widetilde{\mathbb{E}} \left[f \left(M^T - \frac{1}{2} \sum \left(\eta_t^T \right)^2 + \mathcal{R}_1^T \right) \middle| \mathcal{F}^T \right] - \int f(z) \varphi \left(z; -\frac{\kappa^2(\theta)}{2}, \kappa^2(\theta) \right) \mathrm{d}z \right| \\ &\leq \left| \widetilde{\mathbb{E}} \left[f \left(M^T - \frac{1}{2} \sum \left(\eta_t^T \right)^2 + \mathcal{R}_1^T \right) \middle| \mathcal{F}^T \right] - \widetilde{\mathbb{E}} \left[f \left(M^T - \frac{1}{2} \sum \left(\eta_t^T \right)^2 \right) \middle| \mathcal{F}^T \right] \right| \\ &+ \left| \widetilde{\mathbb{E}} \left[f \left(M^T - \frac{1}{2} \sum \left(\eta_t^T \right)^2 \right) \middle| \mathcal{F}^T \right] - \int f(z) \varphi \left(z; -\frac{\kappa^2(\theta)}{2}, \kappa^2(\theta) \right) \mathrm{d}z \right| \\ &\leq \left| \widetilde{\mathbb{E}} \left[f \left(M^T - \frac{1}{2} \sum \left(\eta_t^T \right)^2 \right) \middle| \mathcal{F}^T \right] - \int f(z) \varphi \left(z; -\frac{\kappa^2(\theta)}{2}, \kappa^2(\theta) \right) \mathrm{d}z \right| + \left| \mathbb{E} \left[\mathcal{R}_1^T \middle| \mathcal{F}^T \right] \middle|, \end{split}$$

since $|f(x) - f(y)| \le |x - y|$. Notice that for all θ the function $f_{\theta}(z) := f(z - \kappa^2(\theta)/2)$ also belongs to BL(1). Continuing with our estimate we therefore have

$$\begin{split} &\left| \widetilde{\mathbb{E}} \left[f \left(M^T - \frac{1}{2} \sum_{t=1}^{T} (\eta_t^T)^2 \right) \right| \mathcal{F}^T \right] - \int_{t=1}^{T} f(z) \varphi \left(z; -\frac{\kappa^2(\theta)}{2}, \kappa^2(\theta) \right) dz \right| \\ &= \left| \widetilde{\mathbb{E}} \left[f_{\theta} \left(M^T - \frac{1}{2} \sum_{t=1}^{T} (\eta_t^T)^2 + \frac{\kappa^2(\theta)}{2} \right) \right| \mathcal{F}^T \right] - \int_{t=1}^{T} f_{\theta}(z) \varphi \left(z; 0, \kappa^2(\theta) \right) dz \right| \\ &\leq \left| \widetilde{\mathbb{E}} \left[f_{\theta} \left(M^T - \frac{1}{2} \sum_{t=1}^{T} (\eta_t^T)^2 + \frac{\kappa^2(\theta)}{2} \right) \right| \mathcal{F}^T \right] - \widetilde{\mathbb{E}} \left[f_{\theta} \left(M^T \right) \right| \mathcal{F}^T \right] \right| \\ &+ \left| \widetilde{\mathbb{E}} \left[f_{\theta} \left(M^T \right) \right| \mathcal{F}^T \right] - \int_{t=1}^{T} f_{\theta}(z) \varphi \left(z; 0, \kappa^2(\theta) \right) dz \right| \\ &\leq \frac{1}{2} \widetilde{\mathbb{E}} \left[\left| \sum_{t=1}^{T} \left(\eta_t^T \right)^2 - \kappa^2(\theta) \right| \right| \mathcal{F}^T \right] + \left| \widetilde{\mathbb{E}} \left[f_{\theta} \left(M^T \right) \right| \mathcal{Y}^T \right] - \int_{t=1}^{T} f_{\theta}(z) \varphi \left(z; 0, \kappa^2(\theta) \right) dz \right| \\ &\leq \left| \widetilde{\mathbb{E}} \left[\mathcal{R}_3^T \right| \mathcal{F}^T \right] \right| + \left| \widetilde{\mathbb{E}} \left[\mathcal{R}_2^T \right| \mathcal{F}^T \right] + \left| \int_{t=1}^{T} f_{\theta}(z) \varphi \left(z; 0, \hat{\kappa}^2(\theta) \right) dz - \int_{t=1}^{T} f_{\theta}(z) \varphi \left(z; 0, \kappa^2(\theta) \right) dz \right| \end{aligned}$$

Collecting terms and taking supremum over BL(1), we have

$$d_{BL}\left(\mu^{T},\varphi\left(\cdot;-\frac{\kappa^{2}(\theta)}{2},\kappa^{2}(\theta)\right)\right) := \sup_{f \in BL(1)} \left|\widetilde{\mathbb{E}}\left[f(R^{T})\big|\mathcal{F}^{T}\right] - \int f(z)\varphi\left(z;-\frac{\kappa^{2}(\theta)}{2},\kappa^{2}(\theta)\right) dz\right|$$

$$\leq \sup_{f \in BL(1)} \left|\int f_{\theta}(z)\varphi\left(z;0,\hat{s}_{T}^{2}(\theta)\right) dz - \int f_{\theta}(z)\varphi\left(z;0,\kappa^{2}(\theta)\right) dz\right| +$$

$$+ \left|\mathbb{E}\left[\mathcal{R}_{1}^{T}\big|\mathcal{F}^{T}\right]\right| + \left|\mathbb{E}\left[\mathcal{R}_{2}^{T}\big|\mathcal{F}^{T}\right]\right| + \left|\mathbb{E}\left[\mathcal{R}_{3}^{T}\big|\mathcal{F}^{T}\right]\right| + \left|\mathbb{E}\left[\mathcal{R}_{3}^{T}\big|\mathcal{F}^{T}\right]\right| + \left|\mathbb{E}\left[\mathcal{R}_{3}^{T}\big|\mathcal{F}^{T}\right]\right| + \left|\mathbb{E}\left[\mathcal{R}_{3}^{T}\big|\mathcal{F}^{T}\right]\right| + \left|\mathbb{E}\left[\mathcal{R}_{3}^{T}\big|\mathcal{F}^{T}\right]\right|$$

$$\leq \sqrt{\frac{2}{\pi}}\left|\hat{s}_{T}(\theta) - \kappa(\theta)\right| + \left|\mathbb{E}\left[\mathcal{R}_{1}^{T}\big|\mathcal{F}^{T}\right]\right| + \left|\mathbb{E}\left[\mathcal{R}_{2}^{T}\big|\mathcal{F}^{T}\right]\right| + \left|\mathbb{E}\left[\mathcal{R}_{3}^{T}\big|\mathcal{F}^{T}\right]\right|$$

since $\{f_{\theta}: f \in BL(1)\} = BL(1)$. The result follows by taking the conditional expectation w.r.t. \mathcal{Y}^T and elementary manipulations.

We now need to control the four terms appearing on the r.h.s. of (129). This is done in the following four subsections.

A.6.1 Control of $|\hat{s}_T(\theta) - \kappa(\theta)|$

Lemma 20. As $T \to \infty$ we have that

$$\sup_{\theta \in N(\bar{\theta})} \widetilde{\mathbb{E}} \left[|\hat{s}_T(\theta) - \kappa(\theta)| |\mathcal{Y}^T \right] \to 0 \ \mathbb{P}^Y - \text{a.s.}$$

Proof. This result is established as follows. For any real numbers α, β , we have

$$\sqrt{\alpha^2} - \sqrt{\beta^2} \le \sqrt{|\alpha^2 - \beta^2|},$$

thus

$$\widetilde{\mathbb{E}}\left[\left|\hat{s}_T(\theta) - \kappa(\theta)\right| \middle| \mathcal{Y}^T\right] \leq \widetilde{\mathbb{E}}\left[\left. \sqrt{\left|\hat{s}_T^2(\theta) - \kappa^2(\theta)\right|} \middle| \mathcal{Y}^T\right] \leq \widetilde{\mathbb{E}}^{1/2}\left[\left. \left|\hat{s}_T^2(\theta) - \kappa^2(\theta)\right| \middle| \mathcal{Y}^T\right]\right]$$

We will control the last term in the above expression. Let us write $g(y, u, \theta) := [\partial_u \varpi(y; u, \theta)]^2$ and define

$$\kappa^2(y,\theta) := 2\mathbb{E}g(y, U_{1,1}, \theta).$$

Therefore $\kappa^2(\theta) = \mathbb{E}\kappa^2(Y_1, \theta)$. We next compute

$$\hat{s}_{T}^{2}(\theta) - \kappa^{2}(\theta) = \sum_{t=1}^{T} \frac{1}{T} \sum_{i=1}^{N} \frac{1}{N\left(\widehat{W}_{t}^{T}\right)^{2}} \left[2g\left(Y_{t}, U_{t,i}, \theta\right) - \kappa^{2}(Y_{t}, \theta) \right] + \sum_{t=1}^{T} \frac{1}{T} \left[\frac{\kappa^{2}(Y_{t}, \theta)}{\left(\widehat{W}_{t}^{T}\right)^{2}} - \kappa^{2}(\theta) \right].$$

First we notice that

$$\begin{split} &\widetilde{\mathbb{E}}\left[\left|\frac{1}{N\left(\widehat{W}_{t}^{T}\right)^{2}}\sum_{i=1}^{N}\left[2g\left(Y_{t},U_{t,i},\theta\right)-\kappa^{2}(Y_{t},\theta)\right]\right|\left|\mathcal{Y}^{T}\right] \\ &=\mathbb{E}\left[\left|\frac{1}{N\left(\widehat{W}_{t}^{T}\right)}\sum_{i=1}^{N}\left[2g\left(Y_{t},U_{t,i},\theta\right)-\kappa^{2}(Y_{t},\theta)\right]\right|\left|\mathcal{Y}^{T}\right] \\ &\leq \mathbb{E}^{1/2}\left[\left(\widehat{W}_{t}^{T}\right)^{-2}\left|\mathcal{Y}^{T}\right]\mathbb{E}^{1/2}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\left[2g\left(Y_{t},U_{t,i},\theta\right)-\kappa^{2}(Y_{t},\theta)\right]\right)^{2}\right|\mathcal{Y}^{T}\right] \\ &=\frac{1}{\sqrt{N}}\mathbb{E}^{1/2}\left[\left(\widehat{W}_{t}^{T}\right)^{-2}\left|Y_{t}\right]\mathbb{V}^{1/2}\left[2g(Y_{t},U_{t,1},\theta)|Y_{t}\right], \end{split}$$

since the terms are mean zero and independent over i. Therefore by Assumptions 116 and 119

$$\widetilde{\mathbb{E}} \left[\left| \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N\left(\widehat{W}_{t}^{T}\right)^{2}} \sum_{i=1}^{N} \left[2g\left(Y_{t}, U_{t,i}, \theta\right) - \kappa^{2}(Y_{t}, \theta) \right] \right| \mathcal{Y}^{T} \right] \\
\leq \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} \left[\left| \frac{1}{N\left(\widehat{W}_{t}^{T}\right)} \sum_{i=1}^{N} \left[2g\left(Y_{t}, U_{t,i}, \theta\right) - \kappa^{2}(Y_{t}, \theta) \right] \right| \mathcal{Y}^{T} \right] \\
\leq \frac{1}{\sqrt{N}} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}^{1/2} \left[\left(\widehat{W}_{t}^{T}\right)^{-2} \middle| Y_{t} \right] \mathbb{V}^{1/2} \left[2g(Y_{t}, U_{t,1}, \theta) \middle| Y_{t} \right]. \\
\leq \frac{1}{\sqrt{N}} \frac{1}{T} \sum_{t=1}^{T} B(Y_{t})^{1/3+1}.$$

Continuing we have to control the remainder term

$$\begin{split} &\widetilde{\mathbb{E}}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\kappa^{2}(Y_{t},\theta)}{\left(\widehat{W}_{t}^{T}\right)^{2}}-\kappa^{2}(\theta)\right]\right|\right|\mathcal{Y}^{T}\right] \\ &\leq \widetilde{\mathbb{E}}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\kappa^{2}(Y_{t},\theta)}{\left(\widehat{W}_{t}^{T}\right)^{2}}-\kappa^{2}(Y_{t},\theta)\right]\right|\right|\mathcal{Y}^{T}\right]+\left|\sum_{t=1}^{T}\frac{1}{T}\left[\kappa^{2}(Y_{t},\theta)-\kappa^{2}(\theta)\right]\right| \\ &\leq \sum_{t=1}^{T}\frac{1}{T}\kappa^{2}(Y_{t},\theta)\mathbb{E}\left[\left|\frac{1}{\widehat{W}_{t}^{T}}-\widehat{W}_{t}^{T}\right|\right|Y_{t}\right]+\left|\sum_{t=1}^{T}\frac{1}{T}\left[\kappa^{2}(Y_{t},\theta)-\kappa^{2}(\theta)\right]\right| \\ &\leq \sum_{t=1}^{T}\frac{1}{T}\kappa^{2}(Y_{t},\theta)\left\{\mathbb{E}\left[\left|\frac{1}{\widehat{W}_{t}^{T}}-1\right|\right|Y_{t}\right]+\mathbb{E}\left[\left|\widehat{W}_{t}^{T}-1\right|\right|Y_{t}\right]\right\}+\left|\sum_{t=1}^{T}\frac{1}{T}\left[\kappa^{2}(Y_{t},\theta)-\kappa^{2}(\theta)\right]\right| \\ &\leq \sum_{t=1}^{T}\frac{2}{T}\kappa^{2}(Y_{t},\theta)\epsilon_{T}B(Y_{t})+\left|\sum_{t=1}^{T}\frac{1}{T}\left[\kappa^{2}(Y_{t},\theta)-\kappa^{2}(\theta)\right]\right|, \end{split}$$

by (120) and (121). Finally, by assumption $\kappa^2(y,\theta)$, defined for $\theta \in N(\bar{\theta})$, is continuous in θ for all y, and a measurable function of y for each θ and Assumption 116 ensures that $\kappa^2(y,\theta) \leq B(y)$ for all $y \in \mathcal{Y}$ and $\theta \in N(\bar{\theta})$ and we also have $\mathbb{E}B(Y_1) < \infty$ by assumption. Thus, by [Theorem 2, 30] it follows that as $T \to \infty$

$$\sup_{\theta \in N(\bar{\theta})} \left| \sum_{t=1}^{T} \frac{1}{T} \left[\kappa^2(Y_t, \theta) - \kappa^2(\theta) \right] \right| \to 0 \ \mathbb{P}^Y - \text{a.s.}, \tag{130}$$

In addition we have that as $T \to \infty$

$$\sum_{t=1}^{T} \frac{2}{T} \kappa^2(Y_t, \theta) \epsilon_T B(Y_t) \le \frac{2\epsilon_T}{T} \sum_{t=1}^{T} B^2(Y_t) \to 0 \ \mathbb{P}^Y - \text{a.s.}$$

A.6.2 Control of \mathcal{R}_1^T

Lemma 21. As $T \to \infty$ we have that

$$\sup_{\theta \in N(\bar{\theta})} \widetilde{\mathbb{E}} \left[\left| \mathcal{R}_1^T \right| \right| \mathcal{Y}^T \right] \to 0 \ \mathbb{P}^Y - \text{a.s.}$$

Proof. It follows from (128) that

$$\mathcal{R}_{1}^{T} := \sum_{t=1}^{T} J_{t}^{T} + \sum_{t=1}^{T} L_{t}^{T} + \sum_{t=1}^{T} h(\eta_{t}^{T})[\eta_{t}^{T}]^{2}$$

with h(x) = o(x). Recall that h was defined through the Taylor expansion

$$\log(1+x) = x - \frac{x^2}{2} + h(x)x^2$$
$$= x - \frac{x^2}{2} + \int_0^x \left[\frac{y^2}{1+y} \right] dy.$$

Therefore we can write

$$\sum_{t=1}^{T} \widetilde{\mathbb{E}} \left[h(\eta_{t}^{T}) \left[\eta_{t}^{T} \right]^{2} \middle| Y_{t} \right] = \sum_{t=1}^{T} \widetilde{\mathbb{E}} \left[\int_{0}^{\eta_{t}^{T}} \left[\frac{y^{2}}{1+y} \right] dy \middle| Y_{t} \right] \\
\leq \sum_{t=1}^{T} \widetilde{\mathbb{E}} \left[\left[\int_{0}^{\eta_{t}^{T}} y^{4} dy \right]^{1/2} \left[\int_{0}^{\eta_{t}^{T}} \frac{dy}{(1+y)^{2}} \right]^{1/2} \middle| Y_{t} \right] \\
\leq C \sum_{t=1}^{T} \mathbb{E} \left[\widehat{W}_{t}^{T} \left[\eta_{t}^{T} \right]^{5/2} \frac{\left(\eta_{t}^{T} \right)^{1/2}}{\left(1 + \eta_{t}^{T} \right)^{1/2}} \middle| Y_{t} \right] \\
\leq C \sum_{t=1}^{T} \mathbb{E}^{1/2} \left[\left[\widehat{W}_{t}^{T} \right]^{2} \left[\eta_{t}^{T} \right]^{5} \middle| Y_{t} \right] \mathbb{E}^{1/2} \left[\frac{\eta_{t}^{T}}{(1+\eta_{t}^{T})} \middle| Y_{t} \right] \\
\leq C \sum_{t=1}^{T} \mathbb{E}^{1/2} \left[\left[\widehat{W}_{t}^{T} \right]^{2} \left[\eta_{t}^{T} \right]^{5} \middle| Y_{t} \right] B(Y_{t})^{1/2} \tag{131}$$

by (122). Letting $\widetilde{\eta}_t^T := \widehat{W}_t^T \eta_t^T$ we have

$$\mathbb{E}^{1/2} \left[\left[\widehat{W}_t^T \right]^2 \left[\eta_t^T \right]^5 \middle| Y_t \right]$$

$$= \mathbb{E}^{1/2} \left[\left[\widehat{W}_t^T \right]^{2-5} \left[\widetilde{\eta}_t^T \right]^5 \middle| Y_t \right]$$

$$\leq \mathbb{E}^{1/4} \left[\left[\widehat{W}_t^T \right]^{-6} \middle| Y_t \right] \mathbb{E}^{1/4} \left[\left[\widetilde{\eta}_t^T \right]^{10} \middle| Y_t \right],$$
(132)

and

$$\mathbb{E}\left[\left.\left[\widetilde{\eta}_{t}^{T}\right]^{10}\right|Y_{t}\right] \leq C\left\{\mathbb{E}\left[\left.\left[\widetilde{J}_{t}^{T}\right]^{10}\right|Y_{t}\right] + \mathbb{E}\left[\left.\left[\widetilde{L}_{t}^{T}\right]^{10}\right|Y_{t}\right] + \mathbb{E}\left[\left.\left[\widetilde{M}_{t}^{T}\right]^{10}\right|Y_{t}\right]\right\},\tag{133}$$

where $\widetilde{J}_t^T := \widehat{W}_t^T J_t^T$, $\widetilde{L}_t^T := \widehat{W}_t^T L_t^T$ and $\widetilde{M}_t^T := \widehat{W}_t^T M_t^T$. We now control the terms $\widetilde{\mathbb{E}}\left[\left|\sum_{t=1}^T J_t^T\right|\left|Y_t\right|^2\right]$ and $\widetilde{\mathbb{E}}\left[\left|\sum_{t=1}^T L_t^T\right|\left|Y_t\right|^2\right]$ and the terms on the r.h.s. of (133).

Terms \widetilde{J}_t^T . Since $\widetilde{\mathbb{E}}\left[\left.J_t^T\right|Y_t\right]=0$, and the terms are independent over t we have

$$\begin{split} \widetilde{\mathbb{E}} \left[\left| \sum_{t=1}^{T} J_{t}^{T} \right| \left| Y_{t} \right|^{2} &\leq \widetilde{\mathbb{E}} \left[\left(\sum_{t=1}^{T} J_{t}^{T} \right)^{2} \middle| Y_{t} \right] = \sum_{t=1}^{T} \widetilde{\mathbb{E}} \left[\left(J_{t}^{T} \right)^{2} \middle| Y_{t} \right] \\ &= \sum_{t=1}^{T} \mathbb{E} \left[\left[\widehat{W}_{t}^{T} \right]^{-1} \left(\widetilde{J}_{t}^{T} \right)^{2} \middle| Y_{t} \right] \\ &\leq \sum_{t=1}^{T} \mathbb{E}^{1/2} \left[\left[\widehat{W}_{t}^{T} \right]^{-2} \middle| Y_{t} \right] \mathbb{E}^{1/2} \left[\left(\widetilde{J}_{t}^{T} \right)^{4} \middle| Y_{t} \right] \\ &\leq \sum_{t=1}^{T} B(Y_{t})^{1/2} \mathbb{E}^{1/5} \left[\left(\widetilde{J}_{t}^{T} \right)^{10} \middle| Y_{t} \right], \end{split}$$

by Holder's inequality. We thus have to control

$$\mathbb{E}\left[\left|\widetilde{J}_{t}^{T}\right|^{10} \middle| Y_{t}\right] \leq \mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N} \left\{\varpi(Y_{t}, U_{t,i}\left(\delta_{T}\right); \theta + \xi/\sqrt{T}) - \varpi(Y_{t}, U_{t,i}\left(\delta_{T}\right); \theta)\right\}\right]^{10} \middle| Y_{t}\right]$$

$$= \mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N} \left\{\varpi(Y_{t}, U_{t,i}; \theta + \xi/\sqrt{T}) - \varpi(Y_{t}, U_{t,i}; \theta)\right\}\right]^{10} \middle| Y_{t}\right].$$

Since the terms are i.i.d. over i and have zero mean we will use the following fact: let X_1, \ldots, X_N be i.i.d. and zero mean, then

$$\mathbb{E}\left[\left(\sum X_i\right)^{10}\right] = \sum_{k=1}^{5} \sum_{i_1 \neq \dots \neq i_k}^{N} \sum_{\alpha \in A(k)} \prod_{j=1}^{k} \mathbb{E}\left[X_{i_j}^{2+\alpha_j}\right],$$

where

$$A(k) = \{(\alpha_1, \dots, \alpha_k) : \alpha_1 + \dots + \alpha_k = 10 - 2k\}.$$

Using Holder's inequality notice that since the factors are i.i.d. we have

$$\prod_{j=1}^k \mathbb{E}\left[X_{i_j}^{2+\alpha_j}\right] \leq \prod_{j=1}^k \mathbb{E}\left[X_{i_j}^{10}\right]^{(2+\alpha_j)/10} = \mathbb{E}\left[X_1^{10}\right].$$

Therefore overall we have

$$\mathbb{E}\left[\left(\sum X_i\right)^{10}\right] \leq \mathbb{E}\left[X_1^{10}\right] \sum_{k=1}^5 \binom{N}{k} C(k) \leq C \mathbb{E}\left[X_1^{10}\right] \sum_{k=1}^5 \binom{N}{k} \leq C N^5 \mathbb{E}\left[X_1^{10}\right],$$

since $C(k) := \sharp A(k)$ are combinatorial factors not depending on N. Thus

$$\mathbb{E}\left[\left[\widetilde{J}_{t}^{T}\right]^{10}\middle|Y_{t},\xi\right] \leq \mathbb{E}\left[\left[\frac{1}{N}\sum_{i=1}^{N}\{\varpi(Y_{t},U_{t,i};\theta+\xi/\sqrt{T})-\varpi(Y_{t},U_{t,i};\theta)\}\right]^{10}\middle|Y_{t},\xi\right]$$

$$\leq \frac{C}{N^{10}}N^{5}\mathbb{E}\left[\left(\varpi(Y_{t},U_{t,1};\theta+\xi/\sqrt{T})-\varpi(Y_{t},U_{t,1};\theta)\right)^{10}\middle|Y_{t},\xi\right]$$

$$=\frac{C}{N^{10}}N^{5}\mathbb{E}\left[\left(\int_{0}^{\xi/\sqrt{T}}\partial_{\theta}\varpi(Y_{t},U_{t,1};\theta+s)\mathrm{d}s\right)^{10}\middle|Y_{t},\xi\right]$$

$$=\frac{C}{N^{5}}\left(\frac{\xi}{\sqrt{T}}\right)^{10}\mathbb{E}\left[\left(\int_{0}^{\xi/\sqrt{T}}\partial_{\theta}\varpi(Y_{t},U_{t,1};\theta+s)\frac{\mathrm{d}s}{\xi/\sqrt{T}}\right)^{10}\middle|Y_{t},\xi\right]$$

$$\leq \frac{C}{N^{5}}\left(\frac{\xi}{\sqrt{T}}\right)^{9}\mathbb{E}\left[\int_{0}^{\xi/\sqrt{T}}\partial_{\theta}\varpi(Y_{t},U_{t,1};\theta+s)^{10}\mathrm{d}s\middle|Y_{t},\xi\right]$$

$$=\frac{C}{N^{5}}\left(\frac{\xi}{\sqrt{T}}\right)^{9}\int_{0}^{\xi/\sqrt{T}}\mathbb{E}\left[\partial_{\theta}\varpi(Y_{t},U_{t,1};\theta+s)^{10}\middle|Y_{t}\right]\mathrm{d}s$$

$$\leq \frac{C}{N^{5}}\left(\frac{\xi}{\sqrt{T}}\right)^{10}B(Y_{t}),$$

by (123).

Since $\mathbb{E}[\xi^{10}] < \infty$, we conclude that

$$\mathbb{E}\left[\left|\widetilde{J}_{t}^{T}\right|^{10} \middle| Y_{t}\right] = \mathbb{E}\left[\mathbb{E}\left[\left|\widetilde{J}_{t}^{T}\right|^{10} \middle| Y_{t}, \xi\right] \middle| Y_{t}\right]$$

$$\leq \frac{C}{N^{5}T^{5}}B(Y_{t}), \tag{134}$$

Therefore we have

$$\widetilde{\mathbb{E}} \left[\left| \sum_{t=1}^{T} J_{t}^{T} \right| \left| Y_{t} \right|^{2} \leq \sum_{t=1}^{T} B(Y_{t})^{1/2} \mathbb{E}^{1/5} \left[\left(\widetilde{J}_{t}^{T} \right)^{10} \middle| Y_{t} \right] \right] \\
\leq \frac{1}{NT} \sum_{t=1}^{T} B(Y_{t})^{1/2} B(Y_{t})^{1/5}.$$
(135)

Terms \widetilde{L}_t^T . Since $\widetilde{\mathbb{E}}\left[\left.L_t^T\right|Y_t\right]=0$, and the terms are independent over t we have

$$\begin{split} \widetilde{\mathbb{E}} \left[\left| \sum_{t=1}^{T} L_{t}^{T} \right| \left| Y_{t} \right|^{2} &\leq \widetilde{\mathbb{E}} \left[\left(\sum_{t=1}^{T} L_{t}^{T} \right)^{2} \middle| Y_{t} \right] = \sum_{t=1}^{T} \widetilde{\mathbb{E}} \left[\left(L_{t}^{T} \right)^{2} \middle| Y_{t} \right] \\ &= \sum_{t=1}^{T} \mathbb{E} \left[\left[\widehat{W}_{t}^{T} \right]^{-1} \left(\widetilde{L}_{t}^{T} \right)^{2} \middle| Y_{t} \right] \\ &\leq \sum_{t=1}^{T} \mathbb{E}^{1/2} \left[\left[\widehat{W}_{t}^{T} \right]^{-2} \middle| Y_{t} \right] \mathbb{E}^{1/2} \left[\left(\widetilde{L}_{t}^{T} \right)^{4} \middle| Y_{t} \right] \\ &\leq \sum_{t=1}^{T} B(Y_{t})^{1/2} \mathbb{E}^{1/5} \left[\left(\widetilde{L}_{t}^{T} \right)^{10} \middle| Y_{t} \right]. \end{split}$$

To proceed we estimate

$$\mathbb{E}\left[\left[\widetilde{L}_{t}^{T}\right]^{10}\middle|Y_{t}\right] = \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,i}\left(s\right);\theta\right)U_{t,i}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,i}\left(s\right);\theta\right)\right\}\mathrm{d}s\right)^{10}\middle|Y_{t}\right] \\
\leq \frac{C}{N^{5}}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,1}\left(s\right);\theta\right)U_{t,1}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,1}\left(s\right);\theta\right)\right\}\mathrm{d}s\right)^{10}\middle|Y_{t}\right] \\
\leq \frac{C}{N^{5}}\delta_{T}^{10}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,1}\left(s\right);\theta\right)U_{t,1}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,1}\left(s\right);\theta\right)\right\}\frac{\mathrm{d}s}{\delta_{T}}\right)^{10}\middle|Y_{t}\right] \\
\leq \frac{C}{N^{5}}\delta_{T}^{10}\mathbb{E}\left[\int_{0}^{\delta_{T}}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,1}\left(s\right);\theta\right)U_{t,1}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,1}\left(s\right);\theta\right)\right\}^{10}\frac{\mathrm{d}s}{\delta_{T}}\middle|Y_{t}\right] \\
= C\frac{N^{4}}{T^{9}}\mathbb{E}\left[\int_{0}^{\delta_{T}}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,1}\left(s\right);\theta\right)U_{t,1}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,1}\left(s\right);\theta\right)\right\}^{10}\mathrm{d}s\middle|Y_{t}\right] \\
= C\frac{N^{4}}{T^{9}}\int_{0}^{\delta_{T}}\mathbb{E}\left[\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,1}\left(s\right);\theta\right)U_{t,1}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,1}\left(s\right);\theta\right)\right\}^{10}\right]Y_{t}\right]\mathrm{d}s \\
\leq C\frac{N^{5}}{T^{10}}B(Y_{t}), \tag{136}$$

by (126). Therefore we conclude that

$$\widetilde{\mathbb{E}} \left[\left| \sum_{t=1}^{T} L_{t}^{T} \right| \left| Y_{t} \right|^{2} \leq \sum_{t=1}^{T} B(Y_{t})^{1/2} \mathbb{E}^{1/5} \left[\left(\widetilde{L}_{t}^{T} \right)^{10} \middle| Y_{t} \right] \right] \\
\leq \left(\frac{N^{5}}{T^{10}} \right)^{1/5} \sum_{t=1}^{T} B(Y_{t})^{1/2} B(Y_{t})^{1/5} \\
\leq \frac{N}{T} \frac{1}{T} \sum_{t=1}^{T} B(Y_{t})^{7/10}. \tag{137}$$

Terms \widetilde{M}_t^T . Finally we have, using [Corollary 1, 54], that

$$\mathbb{E}\left[\left[\widetilde{M}_{t}^{T}\right]^{10}\middle|Y_{t}\right] = \mathbb{E}\left[\left(\int_{0}^{\delta_{T}} \frac{\sqrt{2}}{N} \sum_{i=1}^{N} \partial_{u}\varpi\left(Y_{t}, U_{t,i}\left(s\right);\theta\right) dB_{t,i}\left(s\right)\right)^{10}\middle|Y_{t}\right] \\
\leq \frac{C2^{5}}{N^{5}} \mathbb{E}\left[\left(\int_{0}^{\delta_{T}} \partial_{u}\varpi\left(Y_{t}, U_{t,1}\left(s\right);\theta\right) dB_{t,1}\left(s\right)\right)^{10}\middle|Y_{t}\right] \\
\leq \frac{C}{N^{5}} \delta_{T}^{4} \int_{0}^{\delta_{T}} \mathbb{E}\left[\left(\partial_{u}\varpi\left(Y_{t}, U_{t,1}\left(s\right);\theta\right)\right)^{10}\middle|Y_{t}\right] ds \\
= \frac{C}{N^{5}} \delta_{T}^{5} \mathbb{E}\left[\left(\partial_{u}\varpi\left(Y_{t}, U_{t,1};\theta\right)\right)^{10}\middle|Y_{t}\right] \leq \frac{C}{T^{5}} B(Y_{t}), \tag{138}$$

by (124).

Overall control. Bringing all terms together we thus have using (131), (132), (133) and (138), (134) and

(136) that

$$\sum_{t=1}^{T} \widetilde{\mathbb{E}} \left[h(\eta_{t}^{T}) \left[\eta_{t}^{T} \right]^{2} \middle| Y_{t} \right] \leq C \sum_{t=1}^{T} B(Y_{t})^{1/2} \mathbb{E}^{1/2} \left[\left[\widehat{W}_{t}^{T} \right]^{2} \left[\eta_{t}^{T} \right]^{5} \middle| Y_{t} \right] \\
\leq C \sum_{t=1}^{T} B(Y_{t})^{1/2} \mathbb{E}^{1/4} \left[\left[\widehat{W}_{t}^{T} \right]^{-6} \middle| Y_{t} \right] \mathbb{E}^{1/4} \left[\left[\widetilde{\eta}_{t}^{T} \right]^{10} \middle| Y_{t} \right] \\
\leq C \sum_{t=1}^{T} B(Y_{t})^{3/2} \mathbb{E}^{1/4} \left[\left[\widetilde{\eta}_{t}^{T} \right]^{10} \middle| Y_{t} \right] \\
\leq C \sum_{t=1}^{T} B(Y_{t})^{3/2} \left(\mathbb{E} \left[\left[\widetilde{J}_{t}^{T} \right]^{10} \middle| Y_{t} \right] + \mathbb{E} \left[\left[\widetilde{L}_{t}^{T} \right]^{10} \middle| Y_{t} \right] + \mathbb{E} \left[\left[\widetilde{M}_{t}^{T} \right]^{10} \middle| Y_{t} \right] \right)^{1/4} \\
\leq C \sum_{t=1}^{T} B(Y_{t}) \left(\frac{1}{(NT)^{5}} + \frac{N^{5}}{T^{10}} + \frac{1}{T^{5}} \right)^{1/4} \tag{139}$$

$$\leq \frac{C}{T^{1/4}} \frac{1}{T} \sum_{t=1}^{T} B(Y_{t})$$

for T large enough as $N_T/T \to 0$. Hence by combining the bounds (139), (135) and (137), Lemma 21 follows.

A.6.3 Control of \mathcal{R}_2^T

Lemma 22. As $T \to \infty$ we have that

$$\sup_{\theta \in N(\bar{\theta})} \widetilde{\mathbb{E}} \left[\left| \mathcal{R}_2^T \right| \middle| \mathcal{Y}^T \right] \to 0 \, \mathbb{P}^Y - \text{a.s.}$$

Proof. We have

$$\mathcal{R}_{2}^{T} := -\sum_{t=1}^{T} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} \mathcal{S} \left\{ \partial_{u} \varpi \left(Y_{t}, U_{t,i} \left(r \right) ; \theta \right) \right\} dr dB_{t,i} \left(s \right)$$

$$+ \sum_{t=1}^{T} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} \sqrt{2} \int_{0}^{s} \partial_{uuu}^{3} \varpi \left(Y_{t}, U_{t,i} \left(r \right) ; \theta \right) dB_{t,i} \left(r \right) dB_{t,i} \left(s \right)$$

$$=: \mathcal{R}_{21}^{T} + \mathcal{R}_{22}^{T}.$$

We control these two terms separately. We have

$$\widetilde{\mathbb{E}}\left[\left|\mathcal{R}_{21}^{T}\right|\left|\mathcal{Y}^{T}\right] \leq \widetilde{\mathbb{E}}^{1/2}\left[\left|\sum_{t=1}^{T} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} \mathcal{S}\left\{\partial_{u} \varpi\left(Y_{t}, U_{t, i}\left(r\right); \theta\right)\right\} dr ds\right|^{2} \middle| \mathcal{Y}^{T}\right].$$

Since conditionally on \mathcal{Y}^T the vectors $\{U_{t,i}:i\}$ are independent across t and

$$\widetilde{\mathbb{E}}\left[\frac{\sqrt{2}}{N\widehat{W}_{t}^{T}}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\int_{0}^{s}\mathcal{S}\left\{\partial_{u}\varpi\left(Y_{t},U_{t,i}\left(r\right);\theta\right)\right\}\mathrm{d}r\mathrm{d}s\middle|\mathcal{Y}^{T}\right]$$

$$=\mathbb{E}\left[\frac{\sqrt{2}}{N}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\int_{0}^{s}\mathcal{S}\left\{\partial_{u}\varpi\left(Y_{t},U_{t,i}\left(r\right);\theta\right)\right\}\mathrm{d}r\mathrm{d}s\middle|\mathcal{Y}^{T}\right]=0,$$

it follows that

$$\begin{split} &\widetilde{\mathbb{E}}\left[\left|\sum_{t=1}^{T} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} S\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,i}\left(r\right);\theta\right)\right\} \mathrm{d}r\mathrm{d}s\right|^{2} \middle|\mathcal{Y}^{T}\right] \\ &= \sum_{t=1}^{T} \widetilde{\mathbb{E}}\left[\frac{2}{N^{2}\left[\widehat{W}_{t}^{T}\right]^{2}}\left(\sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} S\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,i}\left(r\right);\theta\right)\right\} \mathrm{d}r\mathrm{d}s\right)^{2} \middle|\mathcal{Y}^{T}\right] \\ &= \sum_{t=1}^{T} \mathbb{E}\left[\frac{2}{\widehat{W}_{t}^{T}}\left(\frac{1}{N}\sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} S\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,i}\left(r\right);\theta\right)\right\} \mathrm{d}r\mathrm{d}s\right)^{2} \middle|\mathcal{Y}^{T}\right] \right] \\ &\leq 2\sum_{t=1}^{T} \mathbb{E}^{1/2}\left[\left(\widehat{W}_{t}^{T}\right)^{-2} \middle|\mathcal{Y}^{T}\right] \mathbb{E}^{1/2}\left[\left(\frac{1}{N}\sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} S\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,i}\left(r\right);\theta\right)\right\} \mathrm{d}r\mathrm{d}s\right)^{4} \middle|\mathcal{Y}^{T}\right] \\ &\leq 2\sum_{t=1}^{T} B(Y_{t}) \times \mathbb{E}^{1/2}\left[\left(\frac{1}{N}\sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} S\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,i}\left(r\right);\theta\right)\right\} \mathrm{d}r\mathrm{d}s\right)^{4} \middle|\mathcal{Y}^{T}\right] \\ &\leq 2\sum_{t=1}^{T} B(Y_{t}) \times \left(\frac{C}{N^{2}} \mathbb{E}\left[\left(\int_{0}^{\delta_{T}} \int_{0}^{s} S\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,1}\left(r\right);\theta\right)\right\} \mathrm{d}r\mathrm{d}s\right)^{4} \middle|\mathcal{Y}^{T}\right] \right] \\ &\leq C\sum_{t=1}^{T} B(Y_{t}) \frac{1}{N} \frac{\delta_{T}^{4}}{4} \mathbb{E}^{1/2} \left[\left(\int_{0}^{\delta_{T}} \int_{0}^{s} S\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,1}\left(r\right);\theta\right)\right\} \frac{\mathrm{d}r\mathrm{d}s}{\delta_{T}^{2}/2}\right)^{4} \middle|\mathcal{Y}^{T}\right] \\ &\leq C\sum_{t=1}^{T} B(Y_{t}) \frac{1}{N} \frac{\delta_{T}^{4}}{4} \left[\int_{0}^{\delta_{T}} \int_{0}^{s} \mathbb{E}\left[S\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,1}\left(r\right);\theta\right)\right\}^{4} \middle|\mathcal{Y}^{T}\right] \frac{\mathrm{d}r\mathrm{d}s}{\delta_{T}^{2}/2}\right]^{1/2} \\ &\leq C\sum_{t=1}^{T} B(Y_{t}) \frac{1}{N} \frac{\delta_{T}^{4}}{4} \mathbb{E}^{1/2} \left[S\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,1}\left(0\right);\theta\right)\right\}^{4} \middle|\mathcal{Y}^{T}\right] \mathrm{d}r\mathrm{d}s\right]^{1/2} \\ &= C\sum_{t=1}^{T} B(Y_{t}) \frac{1}{N} \frac{\delta_{T}^{4}}{4} \mathbb{E}^{1/2} \left[S\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,1}\left(0\right);\theta\right)\right\}^{4} \middle|Y_{t}\right] \\ &= C\sum_{t=1}^{T} B(Y_{t}) \frac{1}{N} \frac{\delta_{T}^{4}}{4} = C\sum_{t=1}^{T} B(Y_{t})^{2} \frac{1}{N} \frac{N^{4}}{4} = C\sum_{t=1}^{T} B(Y_{t})^{2} \frac{1}{N} \frac{1$$

by (126). Thus

$$\widetilde{\mathbb{E}}\left[\left|\mathcal{R}_{21}^T\right|\right|\mathcal{Y}^T\right]^2 \leq C\delta_T^3 \frac{1}{T}\sum_{t=1}^T B(Y_t)^2.$$

On the other hand using [Corollary 1, 54] twice, we have

$$\begin{split} &\tilde{\mathbb{E}}\left[|\mathcal{R}_{22}^{T}||\mathcal{Y}^{T}|^{2}\right] \\ &\leq \tilde{\mathbb{E}}\left[|\mathcal{R}_{22}^{T}|^{2}|\mathcal{Y}^{T}\right] \\ &= \tilde{\mathbb{E}}\left[\left|\sum_{l=1}^{T} \frac{\sqrt{2}}{NW_{l}^{T}}\sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} \partial_{uuu}^{3}\varpi\left(Y_{t}, U_{t,i}\left(r\right);\theta\right) \mathrm{d}B_{t,i}\left(r\right) \mathrm{d}B_{t,i}\left(s\right)\right|^{2} |\mathcal{Y}^{T}\right] \\ &= \sum_{t=1}^{T} \tilde{\mathbb{E}}\left[\left(\frac{\sqrt{2}}{NW_{l}^{T}}\sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} \partial_{uuu}^{3}\varpi\left(Y_{t}, U_{t,i}\left(r\right);\theta\right) \mathrm{d}B_{t,i}\left(r\right) \mathrm{d}B_{t,i}\left(s\right)\right)^{2} |\mathcal{Y}^{T}\right] \\ &\leq \sum_{t=1}^{T} \mathbb{E}^{1/2}\left[\left(\widehat{W}_{l}^{T}\right)^{-2} |\mathcal{Y}^{T}\right] \mathbb{E}^{1/2}\left[\left(\frac{\sqrt{2}}{N}\sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} \partial_{uuu}^{3}\varpi\left(Y_{t}, U_{t,i}\left(r\right);\theta\right) \mathrm{d}B_{t,i}\left(r\right) \mathrm{d}B_{t,i}\left(s\right)\right)^{4} |\mathcal{Y}^{T}\right] \\ &\leq \sum_{t=1}^{T} B(Y_{t})^{1/2} \mathbb{E}^{1/2}\left[\left(\frac{\sqrt{2}}{N}\sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} \partial_{uuu}^{3}\varpi\left(Y_{t}, U_{t,1}\left(r\right);\theta\right) \mathrm{d}B_{t,1}\left(r\right) \mathrm{d}B_{t,1}\left(s\right)\right)^{4} |\mathcal{Y}^{T}\right] \\ &\leq \sum_{t=1}^{T} B(Y_{t})^{1/2} \left\{\frac{C}{N^{2}} \mathbb{E}\left[\left(\int_{0}^{\delta_{T}} \int_{0}^{s} \sqrt{2} \int_{0}^{s} \partial_{uuu}^{3}\varpi\left(Y_{t}, U_{t,1}\left(r\right);\theta\right) \mathrm{d}B_{t,1}\left(r\right) \mathrm{d}B_{t,1}\left(s\right)\right)^{4} |\mathcal{Y}^{T}\right] \right\} \\ &\leq \sum_{t=1}^{T} B(Y_{t})^{1/2} \left\{C\frac{4}{N^{2}} \delta_{T} \int_{0}^{\delta_{T}} S \int_{0}^{s} \mathbb{E}\left[\left[\partial_{uuu}^{3}\varpi\left(Y_{t}, U_{t,1}\left(r\right);\theta\right) \mathrm{d}B_{t,1}\left(r\right)\right]^{4} |Y_{t}\right] \mathrm{d}r \mathrm{d}s\right\}^{1/2} \\ &\leq \sum_{t=1}^{T} B(Y_{t})^{1/2} \left\{C\frac{4}{N^{2}} \delta_{T} \int_{0}^{\delta_{T}} S \int_{0}^{s} \mathbb{E}\left[\left[\partial_{uuu}^{3}\varpi\left(Y_{t}, U_{t,1}\left(r\right);\theta\right)\right]^{4} |Y_{t}\right] \delta_{T} \int_{0}^{\delta_{T}} s^{2} \mathrm{d}s\right\}^{1/2} \\ &\leq \sum_{t=1}^{T} B(Y_{t})^{1/2} \left\{C\frac{4}{N^{2}} \mathbb{E}\left[\left[\partial_{uuu}^{3}\varpi\left(Y_{t}, U_{t,1}\left(0\right);\theta\right)\right]^{4} |Y_{t}\right] \delta_{T} \right\}^{1/2} \\ &\leq \sum_{t=1}^{T} B(Y_{t})^{1/2} \left\{C\frac{4}{N^{2}} \mathbb{E}\left[\left[\partial_{uuu}^{3}\varpi\left(Y_{t}, U_{t,1}\left(0\right);\theta\right)\right]^{4} |Y_{t}\right] \delta_{T}^{4} \right\}^{1/2} \\ &\leq \sum_{t=1}^{T} B(Y_{t})^{1/2} \left\{C\frac{4}{N^{2}} \mathbb{E}\left[\left[\partial_{uuu}^{3}\varpi\left(Y_{t}, U_{t,1}\left(0\right);\theta\right)\right]^{4} |Y_{t}\right] \delta_{T}^{4} \right\}^{1/2} \\ &\leq \frac{N}{T} \frac{T}{T} \sum_{t=1}^{T} B(Y_{t}), \end{cases}$$

by (125). Since all bounds obtained are independent of θ we have the following result.

A.6.4 Control of \mathcal{R}_3^T

Lemma 23. As $T \to \infty$ we have that

$$\sup_{\theta \in N(\bar{\theta})} \widetilde{\mathbb{E}} \left[\left| \mathcal{R}_3^T \right| \right| \mathcal{Y}^T \right] \to 0 \ \mathbb{P}^Y - \text{a.s.}$$

Proof. This result is established as follows. Let us write $K_t^T = J_t^T + L_t^T$, then we have

$$\begin{split} \widetilde{\mathbb{E}}\left[\left|\sum_{t=1}^{T}\left(\eta_{t}^{T}\right)^{2}-\kappa^{2}(\theta)\right|\left|\mathcal{Y}^{T}\right] &\leq \widetilde{\mathbb{E}}\left[\left|\sum_{t=1}^{T}\left(\eta_{t}^{T}\right)^{2}-\kappa^{2}(\theta)\right|\left|\mathcal{Y}^{T}\right]\right] \\ &\leq \widetilde{\mathbb{E}}\left[\left|\sum_{t=1}^{T}\left(K_{t}^{T}\right)^{2}\right|+\left|\sum_{t=1}^{T}K_{t}^{T}M_{t}^{T}\right|\left|\mathcal{Y}^{T}\right]+\widetilde{\mathbb{E}}\left[\left|\sum_{t=1}^{T}\left(M_{t}^{T}\right)^{2}-\kappa^{2}(\theta)\right|\left|\mathcal{Y}^{T}\right]\right] \\ &\leq \widetilde{\mathbb{E}}\left[\left|\sum_{t=1}^{T}\left(K_{t}^{T}\right)^{2}\right|\left|\mathcal{Y}^{T}\right]+\sum_{t=1}^{T}\widetilde{\mathbb{E}}^{1/2}\left[\left(K_{t}^{T}\right)^{2}\right|\mathcal{F}^{T}\right]\widetilde{\mathbb{E}}^{1/2}\left[\left(M_{t}^{T}\right)^{2}\right|\mathcal{Y}^{T}\right] \\ &+\widetilde{\mathbb{E}}\left[\left|\sum_{t=1}^{T}\left(M_{t}^{T}\right)^{2}-\kappa^{2}(\theta)\right|\left|\mathcal{Y}^{T}\right]. \end{split}$$

Now from (135) and (137) it easily follows that

$$\widetilde{\mathbb{E}}\left[\left|\sum_{t=1}^{T} \left(K_{t}^{T}\right)^{2} \right\| \mathcal{Y}^{T}\right] \leq C\left(\frac{1}{NT} + \frac{N}{T^{2}}\right) \sum_{t=1}^{T} B(Y_{t})^{7/10}.$$

In addition, from (138) and (116)

$$\begin{split} \widetilde{\mathbb{E}}\left[\left.\left(M_{t}^{T}\right)^{2}\middle|\mathcal{Y}^{T}\right] &= \mathbb{E}\left[\left.\left(\widehat{W}_{t}^{T}\right)^{-1}\left(\widetilde{M}_{t}^{T}\right)^{2}\middle|\mathcal{Y}^{T}\right] \right. \\ &\leq \mathbb{E}^{1/2}\left[\left.\left(\widehat{W}_{t}^{T}\right)^{-2}\middle|\mathcal{Y}^{T}\right]\mathbb{E}^{1/2}\left[\left.\left(\widetilde{M}_{t}^{T}\right)^{4}\middle|\mathcal{Y}^{T}\right] \right. \\ &\leq B(Y_{t})^{1/2}\mathbb{E}^{1/5}\left[\left.\left(\widetilde{M}_{t}^{T}\right)^{10}\middle|\mathcal{Y}^{T}\right] \right. \\ &\leq \frac{1}{T}B(Y_{t})^{1/2}B(Y_{t})^{1/5}, \end{split}$$

and thus we find that

$$\sum_{t=1}^{T} \widetilde{\mathbb{E}}^{1/2} \left[\left(K_t^T \right)^2 \middle| \mathcal{Y}^T \right] \widetilde{\mathbb{E}}^{1/2} \left[\left(M_t^T \right)^2 \middle| \mathcal{Y}^T \right] \leq \sum_{t=1}^{T} \left[C \left(\frac{1}{NT} + \frac{N}{T^2} \right) \frac{1}{T} \right]^{1/2} B(Y_t)^{7/10}.$$

Finally we have that

$$M_{t}^{T} = \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \partial_{u} \varpi \left(Y_{t}, U_{t,i}\left(0\right); \theta\right) \sqrt{\delta_{T}} \xi_{t,i}$$

$$- \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \int_{0}^{\delta_{T}} \int_{0}^{s} \mathcal{S} \left\{ \partial_{u} \varpi \left(Y_{t}, U_{t,i}\left(r\right); \theta\right) \right\} dr dB_{t,i}\left(s\right)$$

$$+ \int_{0}^{\delta_{T}} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}} \sum_{i=1}^{N} \sqrt{2} \int_{0}^{s} \partial_{uu}^{2} \varpi \left(Y_{t}, U_{t,i}\left(r\right); \theta\right) dB_{t,i}\left(r\right) dB_{t,i}\left(s\right)$$

$$= \frac{\sqrt{2}}{\widehat{W}_{t}^{T}} \frac{1}{\sqrt{T}} \left[\frac{1}{N} \sum_{i=1}^{N} \left(\partial_{u} \varpi \left(Y_{t}, U_{t,i}\left(0\right); \theta\right) \right)^{2} \right]^{1/2} \xi_{t} + \mathcal{R}_{2,t}^{T}.$$

Therefore

$$\sum_{t=1}^{T} (M_{t}^{T})^{2} = \sum_{t=1}^{T} \frac{2}{\left[\widehat{W}_{t}^{T}\right]^{2}} \frac{1}{TN} \sum_{i=1}^{N} (\partial_{u} \varpi (Y_{t}, U_{t,i}(0); \theta))^{2} \xi_{t}^{2} + \mathcal{R}_{2,t}^{T}.$$

From Section A.6.3 it follows that

$$\sup_{\theta \in B(\bar{\theta}, \epsilon)} \widetilde{\mathbb{E}} \left[\sum_{t=1}^T \mathcal{R}_{2,t}^T \middle| \mathcal{Y}^T \right] \to 0 \ \mathbb{P}^Y - \text{a.s.}$$

Thus we can focus on the remaining term. We have

$$\begin{split} &\widetilde{\mathbb{E}}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\frac{2}{\left[\widehat{W}_{t}^{T}\right]^{2}}\frac{1}{N}\sum_{i=1}^{N}\left(\partial_{u}\varpi\left(Y_{t},U_{t,i}\left(0\right);\theta\right)\right)^{2}\xi_{t}^{2}-\kappa^{2}(\theta)\right|\left|\mathcal{Y}^{T}\right]\right] \\ &\leq\widetilde{\mathbb{E}}\left[\left|\frac{2}{T}\sum_{t=1}^{T}\xi_{t}^{2}\left(\frac{1}{\left[\widehat{W}_{t}^{T}\right]^{2}}-1\right)\frac{1}{N}\sum_{i=1}^{N}\left(\partial_{u}\varpi\left(Y_{t},U_{t,i}\left(0\right);\theta\right)\right)^{2}\right|\left|\mathcal{Y}^{T}\right]\right] \\ &+\widetilde{\mathbb{E}}\left[\left|\frac{2}{T}\sum_{t=1}^{T}\xi_{t}^{2}\frac{1}{N}\sum_{i=1}^{N}\left(\partial_{u}\varpi\left(Y_{t},U_{t,i}\left(0\right);\theta\right)\right)^{2}-\kappa^{2}(\theta)\right|\left|\mathcal{Y}^{T}\right]\right] \\ &\leq\frac{2}{T}\sum_{t=1}^{T}\widetilde{\mathbb{E}}\left[\left|\xi_{t}^{2}\left(\frac{1}{\left[\widehat{W}_{t}^{T}\right]^{2}}-1\right)\frac{1}{N}\sum_{i=1}^{N}\left(\partial_{u}\varpi\left(Y_{t},U_{t,i}\left(0\right);\theta\right)\right)^{2}\right|\left|\mathcal{Y}^{T}\right]\right] \\ &+\widetilde{\mathbb{E}}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\xi_{t}^{2}\frac{1}{N}\sum_{i=1}^{N}2\left(\partial_{u}\varpi\left(Y_{t},U_{t,i}\left(0\right);\theta\right)\right)^{2}-\kappa^{2}(\theta)\right|\left|\mathcal{Y}^{T}\right]\right] \\ &\leq\frac{2}{T}\sum_{t=1}^{T}\widetilde{\mathbb{E}}^{1/2}\left[\left(\frac{1}{\left[\widehat{W}_{t}^{T}\right]^{2}}-1\right)^{2}\right|Y_{t}\right]\widetilde{\mathbb{E}}^{1/2}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\left(\partial_{u}\varpi\left(Y_{t},U_{t,i}\left(0\right);\theta\right)\right)^{2}\right|^{2}\right|Y_{t}\right] \\ &+\widetilde{\mathbb{E}}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\xi_{t}^{2}\frac{1}{N}\sum_{i=1}^{N}2\left(\partial_{u}\varpi\left(Y_{t},U_{t,i}\left(0\right);\theta\right)\right)^{2}-\kappa^{2}(\theta)\right|\left|\mathcal{Y}^{T}\right] \\ &=:J_{1}+J_{2}. \end{split}$$

since ξ_t is independent of the remaining terms and $\mathbb{E}\xi_t^2 = 1$. We control the first term using the Cauchy-Schwarz inequality, (118), (120) and the triangle inequality

$$J_{1} \leq \frac{2}{T} \sum_{t=1}^{T} \mathbb{E}^{1/2} \left[\widehat{W}_{t}^{T} \middle| \left(\frac{1}{\left[\widehat{W}_{t}^{T}\right]^{2}} - 1 \right)^{2} \middle| Y_{t} \right] \widetilde{\mathbb{E}}^{1/2} \left[\left| \frac{1}{N} \sum_{i=1}^{N} \left(\partial_{u} \varpi \left(Y_{t}, U_{t,i} \left(0 \right) ; \theta \right) \right)^{2} \middle|^{2} \middle| Y_{t} \right]$$

$$\leq \frac{2}{T} \sum_{t=1}^{T} \sqrt{\epsilon_{T}} B(Y_{t}) \widetilde{\mathbb{E}}^{1/2} \left[\left| \left(\partial_{u} \varpi \left(Y_{t}, U_{t,i} \left(0 \right) ; \theta \right) \right) \right|^{4} \middle| Y_{t} \right]$$

$$\leq \frac{2}{T} \sum_{t=1}^{T} \sqrt{\epsilon_{T}} B(Y_{t}) \mathbb{E}^{1/4} \left[\left| \left(\partial_{u} \varpi \left(Y_{t}, U_{t,i} \left(0 \right) ; \theta \right) \right) \right|^{8} \middle| Y_{t} \right] \mathbb{E}^{1/2} \left[\left| \left(\widehat{W}_{t}^{T} \right)^{2} \middle| Y_{t} \right]$$

$$\leq \frac{2}{T} \sum_{t=1}^{T} \sqrt{\epsilon_{T}} B(Y_{t})^{k}.$$

Next we control

$$\begin{split} J_2 &= \widetilde{\mathbb{E}} \left[\left| \frac{1}{T} \sum_{t=1}^T \xi_t^2 \frac{1}{N} \sum_{i=1}^N 2 \left(\partial_u \varpi \left(Y_t, U_{t,i} \left(0 \right) ; \theta \right) \right)^2 - \kappa^2(\theta) \right| \left| \mathcal{Y}^T \right] \right. \\ &= \widetilde{\mathbb{E}} \left[\left| \frac{1}{T} \sum_{t=1}^T \xi_t^2 \left[\frac{1}{N} \sum_{i=1}^N 2 g(Y_t, U_{t,i}; \theta) - \kappa^2(Y_t, \theta) \right] \right| \left| \mathcal{Y}^T \right] + \widetilde{\mathbb{E}} \left[\left| \frac{1}{T} \sum_{t=1}^T \xi_t^2 \kappa^2(Y_t, \theta) - \kappa^2(\theta) \right| \right| \mathcal{Y}^T \right] \\ &\leq \frac{1}{T} \sum_{t=1}^T \widetilde{\mathbb{E}} \left[\left| \xi_t^2 \left[\frac{1}{N} \sum_{i=1}^N 2 g(Y_t, U_{t,i}; \theta) - \kappa^2(Y_t, \theta) \right] \right| \left| \mathcal{Y}^T \right] + \widetilde{\mathbb{E}} \left[\left| \frac{1}{T} \sum_{t=1}^T \xi_t^2 \kappa^2(Y_t, \theta) - \kappa^2(\theta) \right| \right| \mathcal{Y}^T \right] \\ &\leq \frac{1}{T} \sum_{t=1}^T \mathbb{E}^{1/2} \left[\left| \left(\widehat{W}_t^T \right)^2 \right| \left| \mathcal{Y}^T \right| \mathbb{E}^{1/2} \left[\left| \left[\frac{2}{N} \sum_{i=1}^N g(Y_t, U_{t,i}; \theta) - \kappa^2(Y_t, \theta) \right]^2 \right| Y_t \right] \\ &+ \widetilde{\mathbb{E}} \left[\left| \frac{1}{T} \sum_{t=1}^T \xi_t^2 \kappa^2(Y_t, \theta) - \kappa^2(\theta) \right| \right| \mathcal{Y}^T \right] \\ &\leq \frac{1}{T} \sum_{t=1}^T B(Y_t)^{1/2} \frac{1}{\sqrt{N}} \mathbb{V}^{1/2} \left[\left| g(Y_t, U_{t,i}; \theta) \right| \left| Y_t \right| + \widetilde{\mathbb{E}} \left[\left| \frac{1}{T} \sum_{t=1}^T \left[\xi_t^2 - 1 \right] \kappa^2(Y_t, \theta) \right| \right| Y_t \right] \\ &+ \left| \frac{1}{T} \sum_{t=1}^T \kappa(Y_t, \theta) - \kappa^2(\theta) \right| \\ &\leq \frac{1}{\sqrt{N}} \frac{1}{T} \sum_{t=1}^T B(Y_t) + \widetilde{\mathbb{E}}^{1/2} \left[\left| \frac{1}{T} \sum_{t=1}^T \left[\xi_t^2 - 1 \right] \kappa^2(Y_t, \theta) \right|^2 \right| Y_t \right] + \left| \frac{1}{T} \sum_{t=1}^T \kappa(Y_t, \theta) - \kappa^2(\theta) \right| \\ &\leq \frac{1}{\sqrt{N}} \frac{1}{T} \sum_{t=1}^T B(Y_t) + \frac{1}{T^2} \sum_{t=1}^T \kappa^4(Y_t, \theta) \mathbb{V}(\xi_t^2) + \left| \frac{1}{T} \sum_{t=1}^T \kappa(Y_t, \theta) - \kappa^2(\theta) \right|. \end{split}$$

Except from the very last term, everything else is independent of θ . Therefore from (130) and the strong law of large numbers we have the following result.

A.6.5 Proof of Theorem 18

The result follows now directly from Lemmas 19, 20, 21, 22 and 23.

A.7 Proof of Theorem 4

Let $\{\widetilde{\vartheta}_n^T; n \geq 0\}$ be the projection on the first component of the stationary Markov chain $\{(\widetilde{\vartheta}_n^T, \mathsf{U}_n^T); n \geq 0\}$ of invariant distribution $\widetilde{\pi}_T$ defined in (28) and transition kernel Q_T defined in (31) and $\{\widetilde{\vartheta}_n; n \geq 0\}$ the stationary Markov chain of invariant distribution $\varphi(\mathrm{d}\widetilde{\theta}; 0, \overline{\Sigma})$ and transition kernel P defined in (32). The proof of Theorem 4 relies on a set of preliminary propositions. All the expectations in this section have to be understood as conditional expectations w.r.t. \mathcal{Y}^T .

Proposition 24. Under the assumptions of Theorem 4, for any subsequence $\{T_k; k \geq 0\}$ we can extract a further subsequence $\{T'_k; k \geq 0\}$ such that almost surely along this subsequence we have

$$\mathbb{E}\left(\left|Q_T f(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T) - P f(\widetilde{\vartheta}_0^T)\right|\right) \to 0 \tag{140}$$

for all $f \in B(\mathbb{R}^d)$.

Remark. We emphasize that the subset on which this almost sure convergence occurs is independent of f.

Proof of Proposition 24. We define

$$\widetilde{r}_T\left(\widetilde{\theta}_0,\widetilde{\theta}_1\right) = \frac{\widetilde{\pi}_T(\widetilde{\theta}_1)\widetilde{q}(\widetilde{\theta}_1,\widetilde{\theta}_0)}{\widetilde{\pi}_T(\widetilde{\theta}_0)\widetilde{q}(\widetilde{\theta}_0,\widetilde{\theta}_1)}, \quad \ \widetilde{r}\left(\widetilde{\theta}_0,\widetilde{\theta}_1\right) = \frac{\varphi(\widetilde{\theta}_1;0,\overline{\Sigma})\widetilde{q}(\widetilde{\theta}_1,\widetilde{\theta}_0)}{\varphi(\widetilde{\theta}_0;0,\overline{\Sigma})\widetilde{q}(\widetilde{\theta}_0,\widetilde{\theta}_1)}$$

and write

$$\overline{p}(Y_{1:T} \mid \theta_i, u_i) = \frac{\widehat{p}(Y_{1:T} \mid \theta_i, u_i)}{p(Y_{1:T} \mid \theta_i)},$$

where $\theta_i = \hat{\theta}_T + \tilde{\theta}_i / \sqrt{T}$ for i = 0, 1. As Assumption 2 holds, we have

$$Q_{T}f(\widetilde{\theta}_{0}, u_{0}) = \iint \widetilde{q}(\widetilde{\theta}_{0}, d\widetilde{\theta}_{1}) f(\widetilde{\theta}_{1}) K_{\rho_{T}}(u_{0}, du_{1}) \min \left(1, \widetilde{r}_{T}\left(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}\right) \frac{\overline{p}(Y_{1:T} \mid \theta_{1}, u_{1})}{\overline{p}(Y_{1:T} \mid \theta_{0}, u_{0})}\right) + f(\widetilde{\theta}_{0}) \left[1 - \iint \widetilde{q}(\widetilde{\theta}_{0}, d\widetilde{\theta}_{1}) K_{\rho_{T}}(u_{0}, du_{1}) \min \left(1, \widetilde{r}_{T}\left(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}\right) \frac{\overline{p}(Y_{1:T} \mid \theta_{1}, u_{1})}{\overline{p}(Y_{1:T} \mid \theta_{0}, u_{0})}\right)\right]$$

$$(141)$$

and

$$Pf(\widetilde{\theta}_{0}) = \iint \widetilde{q}(\widetilde{\theta}_{0}, d\widetilde{\theta}_{1}) f(\widetilde{\theta}_{1}) \varphi\left(dw; -\frac{\kappa^{2}}{2}, \kappa^{2}\right) \min\left(1, \widetilde{r}\left(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}\right) \exp\left(w\right)\right)$$

$$+ f(\widetilde{\theta}_{0}) \left[1 - \iint \widetilde{q}(\widetilde{\theta}_{0}, d\widetilde{\theta}_{1}) \varphi\left(dw; -\frac{\kappa^{2}}{2}, \kappa^{2}\right) \min\left(1, \widetilde{r}\left(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}\right) \exp\left(w\right)\right)\right].$$

$$(142)$$

It follows that

$$\begin{split} &\frac{1}{2}\mathbb{E}\left(\left|Q_{T}f(\widetilde{\boldsymbol{\vartheta}}_{0}^{T},\mathsf{U}_{0}^{T})-Pf(\widetilde{\boldsymbol{\vartheta}}_{0}^{T})\right|\right) \\ &\leq \frac{1}{2}\iiint\widetilde{\pi}_{T}(\mathrm{d}\widetilde{\boldsymbol{\theta}}_{0},\mathrm{d}u_{0})\widetilde{q}(\widetilde{\boldsymbol{\theta}}_{0},\mathrm{d}\widetilde{\boldsymbol{\theta}}_{1})\left|f(\widetilde{\boldsymbol{\theta}}_{1})\right|\left|\int K_{\rho_{T}}\left(u_{0},\mathrm{d}u_{1}\right)\min\left(1,\widetilde{r}_{T}\left(\widetilde{\boldsymbol{\theta}}_{0},\widetilde{\boldsymbol{\theta}}_{1}\right)\frac{\overline{p}(Y_{1:T}\mid\boldsymbol{\theta}_{1},u_{1})}{\overline{p}(Y_{1:T}\mid\boldsymbol{\theta}_{0},u_{0})}\right) \\ &-\int\varphi\left(\mathrm{d}w;-\frac{\kappa^{2}}{2},\kappa^{2}\right)\min\left(1,\widetilde{r}\left(\widetilde{\boldsymbol{\theta}}_{0},\widetilde{\boldsymbol{\theta}}_{1}\right)\exp\left(w\right)\right)\right| \\ &+\frac{1}{2}\iiint\widetilde{\pi}_{T}(\mathrm{d}\widetilde{\boldsymbol{\theta}}_{0},\mathrm{d}u_{0})\widetilde{q}(\widetilde{\boldsymbol{\theta}}_{0},\mathrm{d}\widetilde{\boldsymbol{\theta}}_{1})\left|f\left(\widetilde{\boldsymbol{\theta}}_{0}\right)\right|\left|\int K_{\rho_{T}}\left(u_{0},\mathrm{d}u_{1}\right)\min\left(1,\widetilde{r}_{T}\left(\widetilde{\boldsymbol{\theta}}_{0},\widetilde{\boldsymbol{\theta}}_{1}\right)\frac{\overline{p}(Y_{1:T}\mid\boldsymbol{\theta}_{1},u_{1})}{\overline{p}(Y_{1:T}\mid\boldsymbol{\theta}_{0},u_{0})}\right) \\ &-\int\varphi\left(\mathrm{d}w;-\frac{\kappa^{2}}{2},\kappa^{2}\right)\min\left(1,\widetilde{r}\left(\widetilde{\boldsymbol{\theta}}_{0},\widetilde{\boldsymbol{\theta}}_{1}\right)\exp\left(w\right)\right)\right| \\ &\leq \iiint\widetilde{\pi}_{T}(\mathrm{d}\widetilde{\boldsymbol{\theta}}_{0},\mathrm{d}u_{0})\widetilde{q}(\widetilde{\boldsymbol{\theta}}_{0},\mathrm{d}\widetilde{\boldsymbol{\theta}}_{1})\left|\int K_{\rho_{T}}\left(u_{0},\mathrm{d}u_{1}\right)\min\left(1,\widetilde{r}_{T}\left(\widetilde{\boldsymbol{\theta}}_{0},\widetilde{\boldsymbol{\theta}}_{1}\right)\frac{\overline{p}(Y_{1:T}\mid\boldsymbol{\theta}_{1},u_{1})}{\overline{p}(Y_{1:T}\mid\boldsymbol{\theta}_{0},u_{0})}\right) \\ &-\int\varphi\left(\mathrm{d}w;-\frac{\kappa^{2}}{2},\kappa^{2}\right)\min\left(1,\widetilde{r}\left(\widetilde{\boldsymbol{\theta}}_{0},\widetilde{\boldsymbol{\theta}}_{1}\right)\exp\left(w\right)\right)\right|. \end{split}$$

Hence, we have

$$\frac{1}{2}\mathbb{E}\left(\left|Q_{T}f(\widetilde{\vartheta}_{0}^{T}, \mathsf{U}_{0}^{T}) - Pf(\widetilde{\vartheta}_{0}^{T})\right|\right) \\
= \iiint \widetilde{\pi}_{T}(du_{0}|\widetilde{\theta}_{0})\widetilde{q}(\widetilde{\theta}_{0}, d\widetilde{\theta}_{1}) \left|\int K_{\rho_{T}}\left(u_{0}, du_{1}\right) \min\left(\widetilde{\pi}_{T}(\widetilde{\theta}_{0}), \widetilde{\pi}_{T}(\widetilde{\theta}_{1}) \frac{\widetilde{q}(\widetilde{\theta}_{1}, \widetilde{\theta}_{0})}{\widetilde{q}(\widetilde{\theta}_{0}, \widetilde{\theta}_{1})} \frac{\overline{p}(Y_{1:T} \mid \theta_{1}, u_{1})}{\overline{p}(Y_{1:T} \mid \theta_{0}, u_{0})}\right) \\
- \widetilde{\pi}_{T}(\widetilde{\theta}_{0}) \int \varphi\left(dw; -\frac{\kappa^{2}}{2}, \kappa^{2}\right) \min\left(1, \widetilde{r}\left(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}\right) \exp\left(w\right)\right) \left|d\widetilde{\theta}_{0}\right| \\
\leq \iiint \widetilde{\pi}_{T}(du_{0}|\widetilde{\theta}_{0}) \int K_{\rho_{T}}\left(u_{0}, du_{1}\right) \left|\min\left(\widetilde{\pi}_{T}(\widetilde{\theta}_{0})\widetilde{q}(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}), \widetilde{\pi}_{T}(\widetilde{\theta}_{1})\widetilde{q}(\widetilde{\theta}_{1}, \widetilde{\theta}_{0}) \frac{\overline{p}(Y_{1:T} \mid \theta_{1}, u_{1})}{\overline{p}(Y_{1:T} \mid \theta_{0}, u_{0})}\right) \\
- \min\left(\varphi(\widetilde{\theta}_{0}; 0, \overline{\Sigma})\widetilde{q}(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}), \varphi(\widetilde{\theta}_{1}; 0, \overline{\Sigma})\widetilde{q}(\widetilde{\theta}_{1}, \widetilde{\theta}_{0}) \frac{\overline{p}(Y_{1:T} \mid \theta_{1}, u_{1})}{\overline{p}(Y_{1:T} \mid \theta_{0}, u_{0})}\right) \left|d\widetilde{\theta}_{0}d\widetilde{\theta}_{1}\right. \\
+ \iiint \widetilde{\pi}_{T}(du_{0}|\widetilde{\theta}_{0})\widetilde{q}(\widetilde{\theta}_{0}, d\widetilde{\theta}_{1}) \left|\varphi(\widetilde{\theta}_{0}; 0, \overline{\Sigma})\int K_{\rho_{T}}\left(u_{0}, du_{1}\right) \min\left(1, \widetilde{r}\left(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}\right) \frac{\overline{p}(Y_{1:T} \mid \theta_{1}, u_{1})}{\overline{p}(Y_{1:T} \mid \theta_{0}, u_{0})}\right)\right) (144) \\
- \widetilde{\pi}_{T}\left(\widetilde{\theta}_{0}\right)\int \varphi\left(dw; -\frac{\kappa^{2}}{2}, \kappa^{2}\right) \min\left(1, \widetilde{r}\left(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}\right) \exp\left(w\right)\right) \left|d\widetilde{\theta}_{0}\right.\right]$$

For the first term given in (143), using the inequality $|\min(x,y) - \min(w,z)| \leq |x-w| + |y-z|$, we

obtain the bound

$$\iiint \widetilde{\pi}_{T}(du_{0}|\widetilde{\theta}_{0}) \int K_{\rho_{T}}(u_{0}, du_{1}) \left| \min \left(\widetilde{\pi}_{T}(\widetilde{\theta}_{0}) \widetilde{q}(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}), \widetilde{\pi}_{T}(\widetilde{\theta}_{1}) \widetilde{q}(\widetilde{\theta}_{1}, \widetilde{\theta}_{0}) \frac{\overline{p}(Y_{1:T} \mid \theta_{1}, u_{1})}{\overline{p}(Y_{1:T} \mid \theta_{0}, u_{0})} \right) \right. \\
\left. - \min \left(\varphi(\widetilde{\theta}_{0}; 0, \overline{\Sigma}) \widetilde{q}(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}), \varphi(\widetilde{\theta}_{1}; 0, \overline{\Sigma}) \widetilde{q}(\widetilde{\theta}_{1}, \widetilde{\theta}_{0}) \frac{\overline{p}(Y_{1:T} \mid \theta_{1}, u_{1})}{\overline{p}(Y_{1:T} \mid \theta_{0}, u_{0})} \right) \right| d\widetilde{\theta}_{0} d\widetilde{\theta}_{1} \\
\leq \iiint \widetilde{\pi}_{T}(du_{0}|\widetilde{\theta}_{0}) \widetilde{q}(\widetilde{\theta}_{0}, d\widetilde{\theta}_{1}) K_{\rho_{T}}(u_{0}, du_{1}) \left| \widetilde{\pi}_{T}(\widetilde{\theta}_{0}) - \varphi(\widetilde{\theta}_{0}; 0, \overline{\Sigma}) \right| d\widetilde{\theta}_{0} \\
+ \iiint \widetilde{\pi}_{T}(du_{0}|\widetilde{\theta}_{0}) K_{\rho_{T}}(u_{0}, du_{1}) \left| \widetilde{\pi}_{T}(\widetilde{\theta}_{1}) - \varphi(\widetilde{\theta}_{1}; 0, \overline{\Sigma}) \right| \widetilde{q}(\widetilde{\theta}_{1}, d\widetilde{\theta}_{0}) \frac{\overline{p}(Y_{1:T} \mid \theta_{1}, u_{1})}{\overline{p}(Y_{1:T} \mid \theta_{0}, u_{0})} d\widetilde{\theta}_{1} \tag{146}$$

The term (145) satisfies

$$\iiint \widetilde{\pi}_{T}(du_{0}|\widetilde{\theta}_{0})\widetilde{q}(\widetilde{\theta}_{0},d\widetilde{\theta}_{1})K_{\rho_{T}}(u_{0},du_{1})\left|\widetilde{\pi}_{T}(\widetilde{\theta}_{0})-\varphi(\widetilde{\theta}_{0};0,\overline{\Sigma})\right|d\widetilde{\theta}_{0}$$

$$=\int \left|\widetilde{\pi}_{T}(\widetilde{\theta}_{0})-\varphi(\widetilde{\theta}_{0};0,\overline{\Sigma})\right|d\widetilde{\theta}_{0}\overset{\mathbb{P}^{Y}}{\to}0$$

by Assumption 1. Therefore, for any subsequence $\{T_k; k \geq 0\}$ we can extract a further subsequence $\{T_k^1; k \geq 0\}$ such that along this subsequence

$$\int \left| \widetilde{\pi}_T(\widetilde{\theta}_0) - \varphi(\widetilde{\theta}_0; 0, \overline{\Sigma}) \right| d\widetilde{\theta}_0 \to 0$$

almost surely. Since

$$\widetilde{\pi}_T(du_0|\widetilde{\theta}_0) = m_T(du_0)\,\overline{p}(Y_{1:T}|\theta_0,u_0)$$

then the term (146) satisfies along $\{T_k^1; k \geq 0\}$

$$\iiint \widetilde{\pi}_{T}(du_{0}|\widetilde{\theta}_{0})\widetilde{q}(\widetilde{\theta}_{1},d\widetilde{\theta}_{0})K_{\rho_{T}}(u_{0},du_{1}) \left| \widetilde{\pi}_{T}(\widetilde{\theta}_{1}) - \varphi(\widetilde{\theta}_{1};0,\overline{\Sigma}) \right| \frac{\overline{p}(Y_{1:T}|\theta_{1},u_{1})}{\overline{p}(Y_{1:T}|\theta_{0},u_{0})} d\widetilde{\theta}_{1}$$

$$= \iiint m_{T}(du_{0})\widetilde{q}(\widetilde{\theta}_{1},d\widetilde{\theta}_{0})K_{\rho_{T}}(u_{0},du_{1}) \left| \widetilde{\pi}_{T}(\widetilde{\theta}_{1}) - \varphi(\widetilde{\theta}_{1};0,\overline{\Sigma}) \right| \overline{p}(Y_{1:T}|\theta_{1},u_{1}) d\widetilde{\theta}_{1}$$

$$= \iiint \widetilde{q}(\widetilde{\theta}_{1},d\widetilde{\theta}_{0})m_{T}(du_{1}) \left| \widetilde{\pi}_{T}(\widetilde{\theta}_{1}) - \varphi(\widetilde{\theta}_{1};0,\overline{\Sigma}) \right| \overline{p}(Y_{1:T}|\theta_{1},u_{1}) d\widetilde{\theta}_{1} \quad (K_{\rho_{T}} \text{ m-invariant})$$

$$= \iiint \widetilde{q}(\widetilde{\theta}_{1},d\widetilde{\theta}_{0})\widetilde{\pi}_{T}(du_{1}|\widetilde{\theta}_{1}) \left| \widetilde{\pi}_{T}(\widetilde{\theta}_{1}) - \varphi(\widetilde{\theta}_{1};0,\overline{\Sigma}) \right| d\widetilde{\theta}_{1}$$

$$= \int \left| \widetilde{\pi}_{T}(\widetilde{\theta}_{1}) - \varphi(\widetilde{\theta}_{1};0,\overline{\Sigma}) \right| d\widetilde{\theta}_{1} \to 0$$

almost surely.

Going back to the term given by (144), we note that

$$\iiint \widetilde{\pi}_{T}(du_{0}|\widetilde{\theta}_{0})\widetilde{q}(\widetilde{\theta}_{0},d\widetilde{\theta}_{1}) \left| \varphi(\widetilde{\theta}_{0};0,\overline{\Sigma}) \int K_{\rho_{T}}(u_{0},du_{1}) \min \left(1,\widetilde{r}\left(\widetilde{\theta}_{0},\widetilde{\theta}_{1}\right) \frac{\overline{p}(Y_{1:T}|\theta_{1},u_{1})}{\overline{p}(Y_{1:T}|\theta_{0},u_{0})}\right) - \widetilde{\pi}_{T}(\widetilde{\theta}_{0}) \int \varphi\left(dw;-\frac{\kappa^{2}}{2},\kappa^{2}\right) \min\left(1,\widetilde{r}\left(\widetilde{\theta}_{0},\widetilde{\theta}_{1}\right) \exp\left(w\right)\right) \left| d\widetilde{\theta}_{0} \right| \\
\leq \iiint \widetilde{\pi}_{T}(du_{0}|\widetilde{\theta}_{0})\widetilde{q}(\widetilde{\theta}_{0},d\widetilde{\theta}_{1}) \left| \varphi(\widetilde{\theta}_{0};0,\overline{\Sigma}) \int K_{\rho_{T}}(u_{0},du_{1}) \min\left(1,\widetilde{r}\left(\widetilde{\theta}_{0},\widetilde{\theta}_{1}\right) \frac{\overline{p}(Y_{1:T}|\theta_{1},u_{1})}{\overline{p}(Y_{1:T}|\theta_{0},u_{0})}\right) - \varphi\left(\widetilde{\theta}_{0};0,\overline{\Sigma}\right) \int \varphi\left(dw;-\frac{\kappa^{2}}{2},\kappa^{2}\right) \min\left(1,\widetilde{r}\left(\widetilde{\theta}_{0},\widetilde{\theta}_{1}\right) \exp\left(w\right)\right) \left| d\widetilde{\theta}_{0} \right| \\
+ \iiint \left| \widetilde{\pi}_{T}(\widetilde{\theta}_{0}) - \varphi(\widetilde{\theta}_{0};0,\overline{\Sigma}) \right| \widetilde{q}(\widetilde{\theta}_{0},d\widetilde{\theta}_{1})\varphi\left(dw;-\frac{\kappa^{2}}{2},\kappa^{2}\right) \min\left(1,\widetilde{r}\left(\widetilde{\theta}_{0},\widetilde{\theta}_{1}\right) \exp\left(w\right)\right) d\widetilde{\theta}_{0} \tag{148}$$

where (148) satisfies

$$\iiint \left| \widetilde{\pi}_{T}(\widetilde{\theta}_{0}) - \varphi(\widetilde{\theta}_{0}; 0, \overline{\Sigma}) \right| \widetilde{q}(\widetilde{\theta}_{0}, d\widetilde{\theta}_{1}) \varphi\left(dw; -\frac{\kappa^{2}}{2}, \kappa^{2}\right) \min\left(1, \widetilde{r}\left(\widetilde{\theta}_{0}, \widetilde{\theta}_{1}\right) \exp\left(w\right)\right) d\widetilde{\theta}_{0}$$

$$\leq \iiint \left| \widetilde{\pi}_{T}(\widetilde{\theta}_{0}) - \varphi(\widetilde{\theta}_{0}; 0, \overline{\Sigma}) \right| \widetilde{q}(\widetilde{\theta}_{0}, d\widetilde{\theta}_{1}) \varphi\left(dw; -\frac{\kappa^{2}}{2}, \kappa^{2}\right) d\widetilde{\theta}_{0}$$

$$= \int \left| \widetilde{\pi}_{T}(\widetilde{\theta}_{0}) - \varphi(\widetilde{\theta}_{0}; 0, \overline{\Sigma}) \right| d\widetilde{\theta}_{0} \to 0$$

almost surely along $\{T_k^1; k \geq 0\}$. We can rewrite (147) as

$$\iiint \varphi\left(\mathrm{d}\theta_{0};\widehat{\theta}_{T}, \frac{\overline{\Sigma}}{T}\right) \overline{\pi}_{T} \left(\mathrm{d}u_{0}|\theta_{0}\right) q\left(\theta_{0}, \mathrm{d}\theta_{1}\right) \left| \int K_{\rho_{T}} \left(u_{0}, \mathrm{d}u_{1}\right) \min\left(1, \frac{\varphi(\sqrt{T}(\theta_{1} - \widehat{\theta}_{T}); 0, \overline{\Sigma})}{\varphi(\sqrt{T}(\theta_{0} - \widehat{\theta}_{T}); 0, \overline{\Sigma})} \frac{\overline{p}(Y_{1:T} | \theta_{1}, u_{1})}{\overline{p}(Y_{1:T} | \theta_{0}, u_{0})}\right) - \int \varphi\left(\mathrm{d}w; -\frac{\kappa^{2}}{2}, \kappa^{2}\right) \min\left(1, \frac{\varphi(\sqrt{T}(\theta_{1} - \widehat{\theta}_{T}); 0, \overline{\Sigma})}{\varphi(\sqrt{T}(\theta_{0} - \widehat{\theta}_{T}); 0, \overline{\Sigma})} \exp\left(w\right)\right) \right| \\
= \iiint \varphi\left(\mathrm{d}\theta_{0}; \widehat{\theta}_{T}, \frac{\overline{\Sigma}}{T}\right) \overline{\pi}_{T} \left(\mathrm{d}u_{0}|\theta_{0}\right) \upsilon\left(\mathrm{d}\xi\right) \left| \int K_{\rho_{T}} \left(u_{0}, \mathrm{d}u_{1}\right) \min\left(1, \frac{\varphi(\sqrt{T}(\theta_{0} + \xi/\sqrt{T} - \widehat{\theta}_{T}); 0, \overline{\Sigma})}{\varphi(\sqrt{T}(\theta_{0} - \widehat{\theta}_{T}); 0, \overline{\Sigma})} \right) \right. \\
\times \left. \frac{\overline{p}(Y_{1:T} | \theta_{0} + \xi/\sqrt{T}, u_{1})}{\overline{p}(Y_{1:T} | \theta_{0}, u_{0})} \right) - \int \varphi\left(\mathrm{d}w; -\frac{\kappa^{2}}{2}, \kappa^{2}\right) \min\left(1, \frac{\varphi(\sqrt{T}(\theta_{0} + \xi/\sqrt{T} - \widehat{\theta}_{T}); 0, \overline{\Sigma})}{\varphi(\sqrt{T}(\theta_{0} - \widehat{\theta}_{T}); 0, \overline{\Sigma})} \exp\left(w\right)\right) \right|.$$

As $\widehat{\theta}_T \overset{\mathbb{P}^Y}{\to} \overline{\theta}$, we extract a further subsequence $\left\{T_k^2; k \geq 0\right\}$ of $\left\{T_k^1; k \geq 0\right\}$ such that along this subsequence $\widehat{\theta}_T \to \overline{\theta}$ almost surely. Hence if we let $A^T(\varepsilon) = \left\{Y_{1:T} : \left\|\widehat{\theta}_T - \overline{\theta}\right\| < \varepsilon/2\right\}$ which satisfies $\mathbb{P}^Y\left(\left(A^T(\varepsilon)\right)^{\mathsf{C}}\right) = o\left(1\right)$ then along this subsequence $\mathbb{E}\left(A^T(\varepsilon)\right) \to 1$ almost surely and therefore (147) is equal to

$$\mathbb{I}\left(A^{T}\left(\varepsilon\right)\right)\iiint\varphi\left(\mathrm{d}\theta_{0};\widehat{\theta}_{T},\frac{\overline{\Sigma}}{T}\right)\overline{\pi}_{T}\left(\mathrm{d}u_{0}|\theta_{0}\right)\upsilon\left(\mathrm{d}\xi\right)\left|\int K_{\rho_{T}}\left(u_{0},\mathrm{d}u_{1}\right)\min\left(1,\frac{\varphi(\sqrt{T}(\theta_{0}+\xi/\sqrt{T}-\widehat{\theta}_{T});0,\overline{\Sigma})}{\varphi(\sqrt{T}(\theta_{0}-\widehat{\theta}_{T});0,\overline{\Sigma})}\right)\right| \times \frac{\overline{p}(Y_{1:T}|\theta_{0}+\xi/\sqrt{T},u_{1})}{\overline{p}(Y_{1:T}|\theta_{0},u_{0})}\right) - \int\varphi(\mathrm{d}w;-\kappa^{2}/2,\kappa^{2})\min\left(1,\frac{\varphi(\sqrt{T}(\theta_{0}+\xi/\sqrt{T}-\widehat{\theta}_{T});0,\overline{\Sigma})}{\varphi(\sqrt{T}(\theta_{0}-\widehat{\theta}_{T});0,\overline{\Sigma})}\exp\left(w\right)\right)\right| + o\left(1\right)$$

almost surely. Along $\{T_k^2; k \geq 0\}$, we can rewrite the integral in the above expression as

$$\begin{split} & \mathbb{I}\left(A^{T}\left(\varepsilon\right)\right) \iiint \varphi\left(\mathrm{d}\theta_{0};\widehat{\theta}_{T},\frac{\overline{\Sigma}}{T}\right) \mathbb{I}\left(\left\|\widehat{\theta}_{T}-\theta_{0}\right\|<\varepsilon/2\right) \overline{\pi}_{T}\left(\mathrm{d}u_{0}|\,\theta_{0}\right) \upsilon\left(\mathrm{d}\xi\right) \\ & \left|\int K_{\rho_{T}}\left(u_{0},\mathrm{d}u_{1}\right) \min\left(1,\frac{\varphi(\sqrt{T}(\theta_{0}+\xi/\sqrt{T}-\widehat{\theta}_{T});0,\overline{\Sigma})}{\varphi(\sqrt{T}(\theta_{0}-\widehat{\theta}_{T});0,\overline{\Sigma})} \frac{\overline{p}(Y_{1:T}|\,\theta_{0}+\xi/\sqrt{T},u_{1})}{\overline{p}(Y_{1:T}|\,\theta_{0},u_{0})}\right) - \int \varphi(\mathrm{d}w;-\kappa^{2}/2,\kappa^{2}) \min\left(1,\frac{\varphi(\sqrt{T}(\theta_{0}+\xi/\sqrt{T}-\widehat{\theta}_{T});0,\overline{\Sigma})}{\varphi(\sqrt{T}(\theta_{0}-\widehat{\theta}_{T});0,\overline{\Sigma})} \exp\left(w\right)\right)\right| + o\left(1\right). \end{split}$$

Notice that the functions

$$x \mapsto \min \left(1, \frac{\varphi(\sqrt{T}(\theta_0 + \xi/\sqrt{T} - \widehat{\theta}_T); 0, \overline{\Sigma})}{\varphi(\sqrt{T}(\theta_0 - \widehat{\theta}_T); 0, \overline{\Sigma})} \exp(x) \right)$$

are bounded above by 1 and Lipschitz, with Lipschitz constants bounded by 1 uniformly in all parameters. Therefore (147) is bounded almost surely along $\{T_k^2; k \geq 0\}$ by

$$\mathbb{I}\left(A^{T}\left(\varepsilon\right)\right) \iiint \varphi\left(\mathrm{d}\theta_{0}; \widehat{\theta}_{T}, \frac{\overline{\Sigma}}{T}\right) \mathbb{I}\left(\left\|\widehat{\theta}_{T} - \theta_{0}\right\| < \varepsilon/2\right) \overline{\pi}_{T}\left(\mathrm{d}u_{0}|\theta_{0}\right) \upsilon\left(\mathrm{d}\xi\right) \\
\times \sup_{f:\left\|f\right\|_{BL} \leq 2} \left|\int K_{\rho_{T}}\left(u_{0}, \mathrm{d}u_{1}\right) f\left\{\log\left(\frac{\overline{p}(Y_{1:T}|\theta_{0} + \xi/\sqrt{T}, u_{1})}{\overline{p}(Y_{1:T}|\theta_{0}, u_{0})}\right)\right\} - \int \varphi(\mathrm{d}w; -\kappa^{2}/2, \kappa^{2}) f\left(w\right)\right| + o\left(1\right), \tag{149}$$

where $||f||_{BL}$ is defined in (158).

We further decompose (149) as

$$\mathbb{I}\left(A^{T}\left(\varepsilon\right)\right) \iiint \varphi\left(\mathrm{d}\theta_{0};\widehat{\theta}_{T}, \frac{\overline{\Sigma}}{T}\right) \mathbb{I}\left(\left\|\widehat{\theta}_{T} - \theta_{0}\right\| < \varepsilon/2\right) \overline{\pi}_{T}\left(\mathrm{d}u_{0}|\theta_{0}\right) \upsilon\left(\mathrm{d}\xi\right) \\
\times \sup_{f:\left\|f\right\|_{BL} \leq 2} \left| \int K_{\rho_{T}}\left(u_{0}, \mathrm{d}u_{1}\right) f\left\{\log\left(\frac{\overline{p}(Y_{1:T}|\theta_{0} + \xi/\sqrt{T}, u_{1})}{\overline{p}(Y_{1:T}|\theta_{0}, u_{0})}\right)\right\} - \int \varphi\left(\mathrm{d}w; -\kappa^{2}(\theta_{0})/2, \kappa^{2}(\theta_{0})\right) f\left(w\right) \right| \\
+ \mathbb{I}\left(A^{T}\left(\varepsilon\right)\right) \iiint \left\{\varphi\left(\mathrm{d}\theta_{0}; \widehat{\theta}_{T}, \frac{\overline{\Sigma}}{T}\right) \mathbb{I}\left(\left\|\widehat{\theta}_{T} - \theta_{0}\right\| < \varepsilon/2\right) \overline{\pi}_{T}\left(\mathrm{d}u_{0}|\theta_{0}\right) \upsilon\left(\mathrm{d}\xi\right) \\
\times d_{BL}\left(\mathcal{N}\left(-\kappa^{2}(\theta_{0})/2, \kappa^{2}(\theta_{0})\right), \mathcal{N}\left(-\kappa^{2}/2, \kappa^{2}\right)\right)\right\} + o(1).$$
(150)

The second term can be easily bounded above by

$$\mathbb{I}\left(A^{T}\left(\varepsilon\right)\right) \iiint \varphi\left(\mathrm{d}\theta_{0};\widehat{\theta}_{T}, \frac{\overline{\Sigma}}{T}\right) \mathbb{I}\left(\left\|\widehat{\theta}_{T} - \theta_{0}\right\| < \varepsilon/2\right) \left[\frac{1}{2}\left|\kappa^{2}(\theta_{0}) - \kappa^{2}(\bar{\theta})\right| + \left|\kappa(\theta_{0}) - \kappa(\bar{\theta})\right| \sqrt{\frac{2}{\pi}}\right] \\
\leq \mathbb{I}\left(A^{T}\left(\varepsilon\right)\right) \iiint \varphi\left(\mathrm{d}\theta_{0};\widehat{\theta}_{T}, \frac{\overline{\Sigma}}{T}\right) \mathbb{I}\left(\left\|\widehat{\theta}_{T} - \theta_{0}\right\| < \varepsilon/2\right) \left|\kappa(\theta_{0}) - \kappa(\bar{\theta})\right| \left[\frac{1}{2}\left(\kappa(\theta_{0}) + \kappa(\bar{\theta})\right) + \sqrt{\frac{2}{\pi}}\right] \\
\leq C\mathbb{I}\left(A^{T}\left(\varepsilon\right)\right) \iiint \varphi\left(\mathrm{d}\theta_{0};\widehat{\theta}_{T}, \frac{\overline{\Sigma}}{T}\right) \mathbb{I}\left(\left\|\widehat{\theta}_{T} - \theta_{0}\right\| < \varepsilon/2\right) \left(\left|\kappa(\theta_{0}) - \kappa(\widehat{\theta}_{T})\right| + \left|\kappa(\bar{\theta}) - \kappa(\widehat{\theta}_{T})\right|\right) \\
\leq C\mathbb{I}\left(A^{T}\left(\varepsilon\right)\right) \int \varphi\left(\mathrm{d}\theta_{0};\widehat{\theta}_{T}, \frac{\overline{\Sigma}}{T}\right) \mathbb{I}\left(\left\|\widehat{\theta}_{T} - \theta_{0}\right\| < \varepsilon/2\right) \left[\left\|\theta_{0} - \widehat{\theta}_{T}\right\| + \left\|\widehat{\theta}_{T} - \bar{\theta}\right\|\right] \\
\leq C\mathbb{I}\left(A^{T}\left(\varepsilon\right)\right) \left[T^{-1/2} \int \varphi\left(\mathrm{d}\zeta; 0, I_{d}\right) \left\|\left(\overline{\Sigma}\right)^{1/2} \zeta\right\| + \left\|\widehat{\theta}_{T} - \bar{\theta}\right\|\right],$$

where we have used the fact that κ is locally Lipschitz around $\bar{\theta}$. As we are in a subsequence along which $\hat{\theta}_T \to \bar{\theta}$ almost surely, then this quantity converges to zero almost surely along this subsequence. Finally the first term of (150) can be controlled for ε small enough by

$$\begin{split} & \mathbb{I}\left(A^{T}\left(\varepsilon\right)\right) \int \varphi\left(\mathrm{d}\theta_{0};\widehat{\theta}_{T},\frac{\overline{\Sigma}}{T}\right) \mathbb{I}\left(\left\|\widehat{\theta}_{T}-\theta_{0}\right\|<\varepsilon/2\right) \\ & \times \sup_{\theta_{0} \in N(\bar{\theta})} \iint \left\{\overline{\pi}_{T}\left(\mathrm{d}u_{0}|\theta_{0}\right) \upsilon\left(\mathrm{d}\xi\right) \right. \\ & \times \sup_{f:\|f\|_{BL} \leq 2} \left| \int K_{\rho_{T}}\left(u_{0},\mathrm{d}u_{1}\right) f\left\{\log\left(\frac{\overline{p}(Y_{1:T}\mid\theta_{0}+\xi/\sqrt{T},u_{1})}{\overline{p}(Y_{1:T}\mid\theta_{0},u_{0})}\right)\right\} - \int \varphi(\mathrm{d}w;-\kappa^{2}(\theta_{0})/2,\kappa^{2}(\theta_{0})) f\left(w\right) \right| \right\} \\ & = \mathbb{I}\left(A^{T}\left(\varepsilon\right)\right) \int \varphi\left(\mathrm{d}\theta_{0};\widehat{\theta}_{T},\frac{\overline{\Sigma}}{T}\right) \mathbb{I}\left(\left\|\widehat{\theta}_{T}-\theta_{0}\right\|<\varepsilon/2\right) \\ & \times \sup_{\theta_{0} \in N(\bar{\theta})} \iint \left\{\overline{\pi}_{T}\left(\mathrm{d}u_{0}|\theta_{0}\right) \upsilon\left(\mathrm{d}\xi\right),\right. \\ & \times \sup_{f:\|f\|_{BL} \leq 2} \left| \int K_{\rho_{T}}\left(u_{0},\mathrm{d}u_{1}\right) f\left\{\log\left(\frac{\overline{p}(Y_{1:T}\mid\theta_{0}+\xi/\sqrt{T},u_{1})}{\overline{p}(Y_{1:T}\mid\theta_{0},u_{0})}\right)\right\} - \int \varphi(\mathrm{d}w;-\kappa^{2}(\theta_{0})/2,\kappa^{2}(\theta_{0})) f\left(w\right) \right| \right\} \end{split}$$

which vanishes in probability by Assumption 3. Hence we can extract a further subsequence along which this convergence happens almost surely. The result follows. \Box

Lemma 25. If along a subsequence $\{T_k; k \geq 0\}$ we have almost surely

$$\mathbb{E}\left(\left|Q_T f(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T) - P f(\widetilde{\vartheta}_0^T)\right|\right) \to 0$$

for all $f \in B(\mathbb{R}^d)$, then along $\{T_k; k \geq 0\}$ we have almost surely

$$\mathbb{E}\left(\left|Q_T^k f(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T) - P^k f(\widetilde{\vartheta}_0^T)\right|\right) \to 0$$

for all $f \in B(\mathbb{R}^d)$ and all $k \geq 1$.

Remark. We emphasize again that the subset on which this almost sure convergence occurs is independent of f and k.

Proof of Lemma 25. We prove the result by induction. For k=1, this follows from the assumption. Now we have

$$\begin{split} &Q_T^{k+1}f(\widetilde{\theta}_0,u_0)-P^{k+1}f(\widetilde{\theta}_0)\\ &=Q_T^{k+1}f(\widetilde{\theta}_0,u_0)-Q_T(P^kf)(\widetilde{\theta}_0,u_0)+Q_T(P^kf)(\widetilde{\theta}_0,u_0)-P^{k+1}f(\widetilde{\theta}_0). \end{split}$$

and therefore

$$\begin{split} & \mathbb{E}\left(\left|Q_T^{k+1}f(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T) - P^{k+1}f(\widetilde{\vartheta}_0^T)\right|\right) \\ & \leq \mathbb{E}\left(\left|Q_T^{k+1}f(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T) - Q_T(P^kf)(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T)\right|\right) + \mathbb{E}\left(\left|Q_T(P^kf)(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T) - P^{k+1}f(\widetilde{\vartheta}_0^T)\right|\right) \\ & \leq \mathbb{E}\left(\left|Q_T^kf(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T) - P^kf(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T)\right|\right) + \mathbb{E}\left(\left|Q_T(P^kf)(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T) - P(P^kf)(\widetilde{\vartheta}_0^T)\right|\right), \end{split}$$

since Q_T is $\widetilde{\pi}_T$ -invariant. We can now apply the induction hypothesis to the functions f and $P^k f$ as $P^k f \in B(\mathbb{R}^d)$.

Proposition 26. Under the assumptions of Theorem 4, for any subsequence $\{T_k; k \geq 0\}$ we can extract a further subsequence $\{T'_k; k \geq 0\}$ such that almost surely along this subsequence we have

$$\mathbb{E}\left[\prod_{i=0}^{n} f_{i}\left(\widetilde{\vartheta}_{k_{i}}^{T}\right)\right] \to \mathbb{E}\left[\prod_{i=0}^{n} f_{i}\left(\widetilde{\vartheta}_{k_{i}}\right)\right]$$

for any $n \ge 0$, any $0 \le k_0 < k_1 < k_2 < \dots < k_n \in \mathbb{N} \text{ and } f_0, \dots, f_n \in B(\mathbb{R}^d)$.

Proof of Proposition 26. In Proposition 24, we have extracted a subsequence $\{T'_k; k \geq 0\}$ of $\{T_k; k \geq 0\}$ such that along this subsequence

$$\int \left| \widetilde{\pi}_T(\widetilde{\theta}_0) - \varphi(\widetilde{\theta}_0; 0, \overline{\Sigma}) \right| d\widetilde{\theta}_0 \to 0$$

almost surely. Hence, along this subsequence, the result holds for n = 0. For n = 1, we have

$$\begin{split} & \left| \mathbb{E} \left[f_0(\widetilde{\vartheta}_{k_0}^T) f_1(\widetilde{\vartheta}_{k_1}^T) \right] - \mathbb{E} \left[f_0(\widetilde{\vartheta}_{k_0}) f_1(\widetilde{\vartheta}_{k_1}) \right] \right| \\ & = \left| \int f_0(\widetilde{\theta}_0) \widetilde{\pi}_T(\widetilde{\theta}_0, u_0) Q_T^{k_1 - k_0} f_1(\widetilde{\theta}_0, u_0) \mathrm{d}\widetilde{\theta}_0 \mathrm{d}u_0 - \int f_0(\widetilde{\theta}_0) \varphi(\widetilde{\theta}_0; 0, \overline{\Sigma}) P^{k_1 - k_0} f_1(\widetilde{\theta}_0) \mathrm{d}\widetilde{\theta}_0 \right| \\ & \leq \left| \int f_0(\widetilde{\theta}_0) \widetilde{\pi}_T(\widetilde{\theta}_0, u_0) \{ Q_T^{k_1 - k_0} f_1(\widetilde{\theta}_0, u_0) - P^{k_1 - k_0} f_1(\widetilde{\theta}_0) \} \mathrm{d}\widetilde{\theta}_0 \mathrm{d}u_0 \right| \\ & + \left| \int f_0(\widetilde{\theta}_0) \{ \widetilde{\pi}_T(\widetilde{\theta}_0) - \varphi(\widetilde{\theta}_0; 0, \overline{\Sigma}) \} P^{k_1 - k_0} f_1(\widetilde{\theta}_0) \mathrm{d}\widetilde{\theta}_0 \right| \\ & \leq \mathbb{E} \left[\left| Q_T^{k_1 - k_0} f_1(\widetilde{\vartheta}_0^T, \mathsf{U}_0^T) - P^{k_1 - k_0} f_1(\widetilde{\vartheta}_0^T) \right| \right] + \int \left| \widetilde{\pi}_T(\widetilde{\theta}_0) - \varphi(\widetilde{\theta}_0; 0, \overline{\Sigma}) \right| \mathrm{d}\widetilde{\theta}_0. \end{split}$$

Hence from Lemma 25, the result also follows for n=1. Now for any $n\geq 1$, we have

$$\mathbb{E}\left[\prod_{j=0}^{n+1} f_{j}(\widetilde{\vartheta}_{k_{j}}^{T})\right] = \mathbb{E}\left[\prod_{j=0}^{n} f_{j}(\widetilde{\vartheta}_{k_{j}}^{T}) Q_{T}^{k_{n+1}-k_{n}} f_{n+1}(\widetilde{\vartheta}_{k_{n}}^{T}, U_{k_{n}}^{T})\right] \\
= \mathbb{E}\left[\prod_{j=0}^{n} f_{j}(\widetilde{\vartheta}_{k_{j}}^{T}) P^{k_{n+1}-k_{n}} f_{n+1}(\widetilde{\vartheta}_{k_{n}}^{T})\right] \\
+ \mathbb{E}\left[\prod_{j=0}^{n} f_{j}(\widetilde{\vartheta}_{k_{j}}^{T}) \left\{Q_{T}^{k_{n+1}-k_{n}} f_{n+1}(\widetilde{\vartheta}_{k_{n}}^{T}, U_{k_{n}}^{T}) - P^{k_{n+1}-k_{n}} f_{n+1}(\widetilde{\vartheta}_{k_{n}}^{T})\right\}\right]. \tag{152}$$

By the induction hypothesis, the first term (151) converges to

$$\mathbb{E}\left[\prod_{j=0}^n f_j(\widetilde{\vartheta}_{k_j}) P^{k_{n+1}-k_n} f_{n+1}(\widetilde{\vartheta}_{k_n})\right] = \mathbb{E}\left[\prod_{j=0}^{n+1} f_j(\widetilde{\vartheta}_{k_j})\right].$$

So it remains to show that the remainder (152) vanishes. We have

$$\begin{split} & \left| \mathbb{E} \left[\prod_{j=0}^n f_j(\widetilde{\vartheta}_{k_j}^T) \{ Q_T^{k_{n+1}-k_n} f_{n+1}(\widetilde{\vartheta}_{k_n}^T, U_{k_n}^T) - P^{k_{n+1}-k_n} f_{n+1}(\widetilde{\vartheta}_{k_n}^T) \} \right] \right| \\ & \leq \mathbb{E} \left[\left| Q_T^{k_{n+1}-k_n} f_{n+1}(\widetilde{\vartheta}_{k_n}^T, U_{k_n}^T) - P^{k_{n+1}-k_n} f_{n+1}(\widetilde{\vartheta}_{k_n}^T) \right| \right]. \end{split}$$

So using Lemma 25, this term vanishes and the result follows.

Proof of Theorem 4. We have shown that for any subsequence $\{T_k; k \geq 0\}$ there exists a further subsequence $\{T'_k; k \geq 0\}$ such that almost surely we have

$$\mathbb{E}\left(\prod_{j=0}^{n} f_{j}(\widetilde{\vartheta}_{k_{j}}^{T})\right) \to \mathbb{E}\left(\prod_{j=0}^{n} f_{j}(\widetilde{\vartheta}_{k_{j}})\right),\tag{153}$$

for any $n \geq 0$, any $0 \leq k_0 < k_1 < k_2 < \dots < k_n \in \mathbb{N}$ and any bounded functions f_0, \dots, f_n . Therefore, we have by [20, Proposition 3.4.6] that on this subsequence the probability measures on $(\mathbb{R}^d)^{\infty}$ given by the laws of $\{\Theta_T; T \geq 1\}$ converge weakly towards the probability measure induced by the law of $\{\widetilde{\vartheta}_n; n \geq 0\}$ almost surely. From this, the result follows from a standard argument; see, e.g., [19, Theorem 2.3.2]. \square

A.8 Proofs for the bounding chain 4

Proof of Proposition 5. It is straightforward to check that Q^* is π -reversible as it follows from (34) that

$$\pi (d\theta) Q^* (\theta, d\theta') = \varrho_{U} (\kappa) \pi (d\theta) Q_{EX} (\theta, d\theta') + \{1 - \varrho_{U} (\kappa)\} \pi (d\theta) \delta_{\theta} (d\theta')$$

$$= \varrho_{U} (\kappa) \pi (d\theta') Q_{EX} (\theta', d\theta) + \{1 - \varrho_{U} (\kappa)\} \pi (d\theta') \delta_{\theta'} (d\theta)$$

$$= \pi (d\theta') Q^* (\theta', d\theta)$$

given Q_{EX} is π -reversible. Now, we can also rewrite Q^* as

$$Q^{*}(\theta, d\theta') = \varrho_{U}(\kappa) \alpha_{EX}(\theta, \theta') q(\theta, d\theta') + \{1 - \varrho_{U}(\kappa) \varrho_{EX}(\theta)\} \delta_{\theta}(d\theta')$$

so the acceptance probability of a proposal is given by $\varrho_{\mathrm{U}}(\kappa) \alpha_{\mathrm{EX}}(\theta, \theta') \leq \alpha_{\widehat{Q}}(\theta, \theta')$ for any (θ, θ') as $\min{(1,a)}\min{(1,b)} \leq \min{(1,ab)}$ for $a,b \geq 0$. The inequality (37) follows directly from this result. Moreover, it also follows that IF $(h,Q^*) \leq \mathrm{IF}(h,\widehat{Q})$ from [50, Theorem 4], which is a general state-space version of the main result in [43]. To establish the expression of IF (h,Q^*) , we first note that there exists a probability measure $e(h,Q_{\mathrm{EX}})$ on [-1,1] such that

$$\phi_n(h, Q_{\mathrm{EX}}) = \int_{-1}^1 \lambda^n e(h, Q_{\mathrm{EX}})(\mathrm{d}\lambda), \quad \mathrm{IF}(h, Q_{\mathrm{EX}}) = \int_{-1}^1 \frac{1+\lambda}{1-\lambda} e(h, Q_{\mathrm{EX}})(\mathrm{d}\lambda).$$

This follows from the spectral representation of reversible Markov chains; see e.g. [32]. From the expression (34) of Q^* , we have

$$\left(Q^*\right)^n = \sum_{k=0}^n \binom{n}{k} \varrho_{\mathrm{U}}^k \left(\kappa\right) \left\{1 - \varrho_{\mathrm{U}}\left(\kappa\right)\right\}^{n-k} \varrho_{\mathrm{Ex}}^k.$$

Therefore, if we denote by $X \sim \text{Bin}(n; \varrho_{\text{U}}(\kappa))$ the number of acceptances from 0 to n, we have

$$\phi_n(h, Q^*) = \sum_{k=0}^n \int \lambda^k \Pr(X = k) e(h, Q_{\text{EX}})(\mathrm{d}\lambda) = \int \{(1 - \varrho_{\text{U}}(\kappa)) + \varrho_{\text{U}}(\kappa) \lambda\}^n e(h, Q_{\text{EX}})(\mathrm{d}\lambda), \quad (154)$$

Hence, it follows that

$$\begin{aligned} \text{IF}\left(h,Q^*\right) &= 1 + 2\sum_{n=1}^{\infty} \phi_n(h,Q^*) \\ &= \int \frac{1 + (1 - \varrho_{\text{U}}\left(\kappa\right)) + \varrho_{\text{U}}\left(\kappa\right)\lambda}{\varrho_{\text{U}}\left(\kappa\right) - \varrho_{\text{U}}\left(\kappa\right)\lambda} e(h,Q_{\text{EX}}) (\text{d}\lambda) = \frac{1}{\varrho_{\text{U}}\left(\kappa\right)} (\text{IF}(h,Q_{\text{EX}}) + 1) - 1. \end{aligned}$$

Assuming Q_{EX} is geometrically ergodic, then $\phi_n(h,Q_{\text{EX}}) = \int_{-1+\epsilon}^{1-\epsilon} \lambda^n e(h,Q_{\text{EX}}) (\mathrm{d}\lambda)$, where $0 < \epsilon < 1$ is the spectral gap. From (154), a simple change of variables yields

$$\phi_{n}(h, Q^{*}) = \int_{-1+\epsilon}^{1-\epsilon} \left[(1 - \varrho_{\mathrm{U}}(\kappa)) + \varrho_{\mathrm{U}}(\kappa) \lambda \right]^{n} e(h, Q_{\mathrm{EX}}) (\mathrm{d}\lambda) = \int_{1-2\varrho_{\mathrm{U}}(\kappa) + \epsilon\varrho_{\mathrm{U}}(\kappa)}^{1-\epsilon\varrho_{\mathrm{U}}(\kappa)} \widetilde{\lambda}^{n} \widetilde{e}(h, Q^{*}) (\mathrm{d}\widetilde{\lambda}).$$

Thus Q^* is also geometrically ergodic.

Proof of Proposition 6. Parts (i) and (ii) are immediate. To simplify notation, we write here ARCT = ARCT (h, Q^*) , IF = IF (h, Q_{EX}) , $\varrho_{U}(\kappa) = \varrho(\kappa)$, $\varphi(x) = \varphi(x; 0, 1)$. We note that

$$\log\left(\mathrm{ARCT}\right) = \frac{1}{2}\log\left\{\mathrm{IF} + 1 - \varrho\left(\kappa\right)\right\} - \log\left(\kappa\right) - \log\left(\varrho\left(\kappa\right)\right) - \frac{1}{2}\log\left\{\mathrm{IF}\right\},$$

so we obtain

$$\frac{\partial \log \left(\mathrm{ARCT} \right)}{\partial \mathrm{IF}} = \frac{1}{2} \left\{ \frac{1}{\mathrm{IF} + 1 - \varrho \left(\kappa \right)} - \frac{1}{\mathrm{IF}} \right\} < 0,$$

which shows that ARCT decreases with IF. We also have

$$\frac{\partial \log \left(\text{ARCT} \right)}{\partial \kappa} = \frac{1}{2} \varphi \left(\frac{\kappa}{2} \right) G \left(\kappa \right) - \frac{1}{\kappa}, \quad G \left(\kappa \right) = \frac{1}{2 - \varrho \left(\kappa \right)} + \frac{2}{\varrho \left(\kappa \right)}.$$

The minimizing argument $\hat{\kappa}$ satisfies

$$J(\hat{\kappa}, \text{IF}) = \left. \frac{\partial \log (\text{ARCT})}{\partial \kappa} \right|_{\hat{\kappa}} = 0.$$

By implicit differentiation

$$\frac{\mathrm{d}\hat{\kappa}}{\mathrm{dIF}} = -\frac{\partial J\left(\hat{\kappa}, \mathrm{IF}\right)}{\partial \mathrm{IF}} / \frac{\partial J\left(\hat{\kappa}, \mathrm{IF}\right)}{\partial \hat{\kappa}},\tag{155}$$

where

$$\frac{\partial J\left(\widehat{\kappa}, \text{IF}\right)}{\partial \text{IF}} = -\frac{1}{2}\varphi\left(\frac{\widehat{\kappa}}{2}\right) \frac{1}{\left(IF + 1 - \rho\left(\widehat{\kappa}\right)\right)^{2}} < 0 \tag{156}$$

and

$$\begin{split} \frac{\partial J\left(\widehat{\kappa},\text{IF}\right)}{\partial \widehat{\kappa}} &= \frac{1}{2} \varphi\left(\frac{\widehat{\kappa}}{2}\right) \left\{ \frac{\partial G\left(\widehat{\kappa},IF\right)}{\partial \widehat{\kappa}} - \frac{\widehat{\kappa}}{4} G\left(\widehat{\kappa},IF\right) \right\} + \frac{1}{\widehat{\kappa}^2} \\ &= \frac{1}{2} \left\{ \varphi\left(\frac{\widehat{\kappa}}{2}\right) \frac{\partial G\left(\widehat{\kappa},IF\right)}{\partial \widehat{\kappa}} - \frac{1}{2} \right\} + \frac{1}{\widehat{\kappa}^2} \\ &\geq \frac{1}{2} \left\{ \varphi\left(\frac{\widehat{\kappa}}{2}\right)^2 \left[\frac{2}{\varrho\left(\widehat{\kappa}\right)^2} - \frac{1}{\left(2 - \varrho\left(\widehat{\kappa}\right)\right)^2} \right] - \frac{1}{2} \right\} + \frac{1}{\widehat{\kappa}^2} > 0 \end{split} \tag{157}$$

where we have used $J(\hat{\kappa}, \text{IF}) = 0$ to simplify the expression of this derivative. It follows from (155), (156) and (157) that the minimizing argument of ARCT increases with IF.

A.9 Conditional weak convergence

Let C_b the space of bounded continuous functions and BL(1) the space of bounded Lipschitz functions f with $||f||_{BL} \leq 1$ where

$$||f||_{BL} = ||f||_{\infty} + \sup_{x,y:x \neq y} \left| \frac{f(y) - f(x)}{y - x} \right|.$$
 (158)

For sake of completeness, we present a version of the conditional CLT for triangular arrays which allows us to conclude that the expectations of any function $f \in C_b$ converge in probability. We have not been able to find this precise statement in the literature so we present a proof mimicking the steps of the proof of [8, Theorem 7.2] without any claim of originality.

Lemma 27. Let $\{X_{n,i}\}_{1\leq i\leq k_n}$ be a triangular array of real-valued random variables on a common probability space (Ω, \mathcal{F}, P) and $\{\mathcal{F}_n\}_{n\geq 0}$ a sequence of sub- σ -algebras of \mathcal{F} such that $\{X_{n,i}\}_{1\leq i\leq k_n}$ are conditionally independent given \mathcal{F}_n , $\mathbb{E}(X_{n,i} \mid \mathcal{F}_n) = 0$ and $\sigma_{n,i}^2 := \mathbb{E}(X_{n,i}^2 \mid \mathcal{F}_n) < \infty$. Suppose also that for some $\sigma^2 > 0$, as $n \to \infty$

$$s_n^2 := \sum_{i=1}^{k_n} \sigma_{n,i}^2 \xrightarrow{P} \sigma^2,$$
 (159)

and that for all $\epsilon > 0$,

$$\sum_{i=1}^{k_n} \mathbb{E}\left(X_{n,i}^2 \mathbf{1}\left\{|X_{n,i}| \ge \epsilon\right\} \middle| \mathcal{F}_n\right) \xrightarrow{P} 0.$$
 (160)

Then we have

$$\sum_{i=1}^{k_n} X_{n,i} | \mathcal{F}_n \Rightarrow \mu,$$

where $\mu(dx) = \varphi(dx; 0, \sigma^2)$ in the sense that for all $f \in C_b$

$$\mathbb{E}\left[f\left(\sum\nolimits_{i=1}^{k_{n}}X_{n,i}\right)\middle|\mathcal{F}_{n}\right]\overset{P}{\rightarrow}\mu\left(f\right).$$

In particular, the random measures μ_n defined by a regular version of the conditional probability distributions

$$\mu_n(A) = \mathbb{P}\left(\left.\sum_{i=1}^{k_n} X_{n,i} \in A\right| \mathcal{F}_n\right) \text{ for } A \in \mathcal{B}(\mathbb{R})$$
(161)

converge weakly to μ in probability in the sense that

$$d_{\rm BL}(\mu_n, \mu) := \sup_{f \in BL(1)} |\mu_n(f) - \mu(f)| \stackrel{P}{\to} 0.$$
 (162)

Remark 28. A random probability measure μ on a metric space \mathcal{X} equipped with its Borel σ -algebra $\mathcal{B}(\mathcal{X})$ is usually defined as a map μ from some probability space (Ω, \mathcal{F}, P) to the space $\mathcal{P}(\mathcal{X})$ of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that for all $\omega \in \Omega$, $\mu(\omega, \cdot) \in \mathcal{P}(\mathcal{X})$ and for all $A \in \mathcal{B}(\mathcal{X})$, the map $\omega \longmapsto \mu(\omega, A)$ is measurable. As explained in [13, Remark 3.20], such a random measure is a measurable map from Ω to $\mathcal{P}(\mathcal{X})$ w.r.t. the Borel σ -algebra on $\mathcal{P}(\mathcal{X})$ induced by the topology of weak convergence. Indeed, from the above definition of random measures, it follows that for any function $g \in C_b(\mathcal{X})$, the map $\omega \longmapsto \mu(\omega)(g)$ is measurable. Since the map $\omega \longmapsto \mu(\omega)(g)$ can be written as a composition of

$$\omega \longmapsto \mu(\omega, \cdot) \stackrel{\mathcal{I}_g}{\longmapsto} \mu(\omega)(g),$$

measurability of this map implies that for any $B \in \mathcal{B}(\mathbb{R})$ we have $(\mathcal{I}_g \circ \mu)^{-1}(B) \in \mathcal{F}$ or equivalently that $\mu^{-1}(\mathcal{I}_g^{-1}(B)) \in \mathcal{F}$. Since the collection of sets $\{\mathcal{I}_g^{-1}(B); B \in \mathcal{B}(\mathbb{R}), g \in C_b(\mathcal{X})\}$ generates $\mathcal{B}(\mathcal{P}(\mathcal{X}))$, the mapping $\omega \longmapsto \mu(\omega, \cdot)$ is measurable w.r.t. $\mathcal{P}(\mathcal{X})$. In particular if \mathcal{X} is Polish, the topology of weak convergence is metrized by the bounded Lipschitz metric which is then continuous and therefore measurable. One can easily check that the random probability measures specified in Lemma 27 falls within this context. Therefore the quantity on the l.h.s. of (162) is measurable.

Proof of Lemma 27. We first prove the result for f bounded and infinitely differentiable, with bounded derivatives of all orders. Without loss of generality, we can assume that the probability space also supports a triangular array of independent standard normal random variables $\{\xi_{n,i}\}_{1\leq i\leq k_n}$, independent of $\{X_{n,i}\}_{n,i}$ and of \mathcal{F}_n for all n. For all n and $1\leq i\leq k_n$ define $\eta_{n,i}:=\sigma_{n,i}\xi_{n,i}$.

Then using the standard Lindeberg approach, as employed in the proof of [8, Theorem 7.2], we use the following telescoping identity

$$f\left(\sum_{i=1}^{k_n} X_{n,i}\right) = f\left(\sum_{i=1}^{k_n} X_{n,i}\right) - f\left(\sum_{i=1}^{k_n-1} X_{n,i} + \eta_{n,k_n}\right)$$

$$+ f\left(\sum_{i=1}^{k_n-1} X_{n,i} + \eta_{n,k_n}\right) - f\left(\sum_{i=1}^{k_n-2} X_{n,i} + \eta_{n,k_n-1} + \eta_{n,k_n}\right)$$

$$+ \cdots$$

$$+ f\left(X_{n,1} + \sum_{j=2}^{k_n} \eta_{n,j}\right) - f\left(\sum_{i=1}^{k_n} \eta_{n,i}\right)$$

$$+ f\left(\sum_{i=1}^{k_n} \eta_{n,i}\right).$$

Writing Z for a standard normal, independent of all other variables and \mathcal{F}_n and $\{X_{n,i}\}_i$, notice first that

$$\mathbb{E}\left[f\left(\sum\nolimits_{i=1}^{k_n}\eta_{n,i}\right)\bigg|\mathcal{F}_n\right]=\mathbb{E}\left[f(s_nZ)\bigg|\mathcal{F}_n\right]\overset{P}{\to}\mathbb{E}\left[f(\sigma Z)\right].$$

Therefore

$$\mathbb{E}\left[f\left(\sum\nolimits_{i=1}^{k_n}X_{n,i}\right)\Big|\mathcal{F}_n\right] = o_P(1) + \mathbb{E}\left[f(\sigma Z)\right] + \sum\nolimits_{i=1}^{k_n}\mathbb{E}[\mathcal{E}_{n,i}|\mathcal{F}_n],$$

where

$$\mathbb{E}[\mathcal{E}_{n,i}|\mathcal{F}_{n}] := \mathbb{E}\left[f\left(\sum_{j=1}^{i} X_{n,j} + \sum_{j=i+1}^{k_{n}} \eta_{n,j}\right) \middle| \mathcal{F}_{n}\right] - \mathbb{E}\left[f\left(\sum_{j=1}^{i-1} X_{n,j} + \sum_{j=i}^{k_{n}} \eta_{n,j}\right) \middle| \mathcal{F}_{n}\right]$$

$$= \mathbb{E}\left[f\left(\sum_{j=1}^{i-1} X_{n,j} + \sum_{j=i+1}^{k_{n}} \eta_{n,j} + X_{n,i}\right) \middle| \mathcal{F}_{n}\right]$$

$$- \mathbb{E}\left[f\left(\sum_{j=1}^{i-1} X_{n,j} + \sum_{j=i+1}^{k_{n}} \eta_{n,j} + \eta_{n,i}\right) \middle| \mathcal{F}_{n}\right].$$

Letting

$$g(h) := \sup_{x} |f(x+h) - f(x) - f'(x)h - \frac{1}{2}f''(x)h^{2}|,$$

we have by the mean value theorem, and the fact that f has bounded derivative of order two that

$$f(x+h) - f(x) = \int_{x}^{x+h} f'(s)ds = f'(x)h + \int_{x}^{x+h} \int_{x}^{s} f''(t)dtds$$
$$= f'(x)h + \frac{1}{2}f''(x)h^{2} + \int_{x}^{x+h} \int_{x}^{s} f''(t) - f''(x)dtds,$$

and the last term can be bounded above by

$$\left| \int_{s=x}^{x+h} \int_{t=x}^{s} f''(t) - f''(x) dt ds \right| \le \int_{s=x}^{x+h} \int_{t=x}^{s} |f''(t) - f''(x)| dt ds \le h^2 ||f''||_{\infty},$$

and by

$$\int_{s=x}^{x+h} \int_{t=x}^{s} |f''(t) - f''(x)| \, \mathrm{d}t \, \mathrm{d}s \le ch^3 \|f'''\|_{\infty}$$

Therefore there exists K such that

$$g(h) \le K \min\{h^2, |h|^3\}.$$

Let us look at one of these remainder terms. Write

$$\mathcal{E}_{n,i} = f\left(\sum_{j=1}^{i-1} X_{n,j} + \sum_{j=i+1}^{k_n} \eta_{n,j} + X_{n,i}\right) - f\left(\sum_{j=1}^{i-1} X_{n,j} + \sum_{j=i+1}^{k_n} \eta_{n,j} + \eta_{n,i}\right)$$

$$= f'\left(\sum_{j=1}^{i-1} X_{n,j} + \sum_{j=i+1}^{k_n} \eta_{n,j}\right) (X_{n,i} - \eta_{n,i})$$

$$+ \frac{1}{2} f''\left(\sum_{j=1}^{i-1} X_{n,j} + \sum_{j=i+1}^{k_n} \eta_{n,j}\right) (X_{n,i}^2 - \eta_{n,i}^2) + R_{n,i},$$

where

$$|R_{n,i}| \le g(X_{n,i}) + g(\eta_{n,i}).$$

Taking conditional expectations we observe that

$$\mathbb{E}\left[f'\left(\sum_{j=1}^{i-1} X_{n,j} + \sum_{j=i+1}^{k_n} \eta_{n,j}\right) (X_{n,i} - \eta_{n,i}) \middle| \mathcal{F}_n\right]$$

$$= \mathbb{E}\left[f'\left(\sum_{j=1}^{i-1} X_{n,j} + \sum_{j=i+1}^{k_n} \eta_{n,j}\right) \middle| \mathcal{F}_n\right] \times \mathbb{E}\left[(X_{n,i} - \eta_{n,i}) \middle| \mathcal{F}_n\right] = 0.$$

by independence, conditional independence and the fact that $X_{n,i}$ are conditionally centred. Similarly

$$\mathbb{E}\left[f''\left(\sum_{j=1}^{i-1} X_{n,j} + \sum_{j=i+1}^{k_n} \eta_{n,j}\right) (X_{n,i}^2 - \eta_{n,i}^2) \middle| \mathcal{F}_n\right]$$

$$= \mathbb{E}\left[f'\left(\sum_{j=1}^{i-1} X_{n,j} + \sum_{j=i+1}^{k_n} \eta_{n,j}\right) \middle| \mathcal{F}_n\right] \times \mathbb{E}\left[(X_{n,i}^2 - \eta_{n,i}^2) \middle| \mathcal{F}_n\right] = 0,$$

since

$$\mathbb{E}[\eta_{n,i}^2|\mathcal{F}_n] = \sigma_{n,i}^2 \mathbb{E}[\xi_{n,i}^2|\mathcal{F}_n] = \sigma_{n,i}^2.$$

It remains to control the expression

$$\sum_{i=1}^{k_n} \mathbb{E}[g(X_{n,i})|\mathcal{F}_n] + \mathbb{E}[g(\eta_{n,i})|\mathcal{F}_n].$$

For the first term, letting $\epsilon > 0$ we have

$$\begin{split} \sum_{i=1}^{k_{n}} \mathbb{E}[g(X_{n,i})|\mathcal{F}_{n}] &= \sum_{i=1}^{k_{n}} \mathbb{E}\left[g(X_{n,i})\mathbf{1}\{|X_{n,i}| < \epsilon\} \middle| \mathcal{F}_{n}\right] + \sum_{i=1}^{k_{n}} \mathbb{E}\left[g(X_{n,i})\mathbf{1}\{|X_{n,i}| \ge \epsilon\} \middle| \mathcal{F}_{n}\right] \\ &\leq K \sum_{i=1}^{k_{n}} \mathbb{E}\left[|X_{n,i}|^{3}\mathbf{1}\{|X_{n,i}| < \epsilon\} \middle| \mathcal{F}_{n}\right] + K \sum_{i=1}^{k_{n}} \mathbb{E}\left[|X_{n,i}|^{2}\mathbf{1}\{|X_{n,i}| \ge \epsilon\} \middle| \mathcal{F}_{n}\right] \\ &\leq K\epsilon \sum_{i=1}^{k_{n}} \mathbb{E}\left[|X_{n,i}|^{2}\mathbf{1}\{|X_{n,i}| < \epsilon\} \middle| \mathcal{F}_{n}\right] + K \sum_{i=1}^{k_{n}} \mathbb{E}\left[|X_{n,i}|^{2}\mathbf{1}\{|X_{n,i}| \ge \epsilon\} \middle| \mathcal{F}_{n}\right] \\ &\leq K\epsilon \sum_{i=1}^{k_{n}} \sigma_{n,i}^{2} + K \sum_{i=1}^{k_{n}} \mathbb{E}\left[|X_{n,i}|^{2}\mathbf{1}\{|X_{n,i}| \ge \epsilon\} \middle| \mathcal{F}_{n}\right] \xrightarrow{P} 0, \end{split}$$

because $\epsilon > 0$ is arbitrary, and the second term vanishes in probability by hypothesis.

For the second term, we obtain similarly

$$\sum_{i=1}^{k_n} \mathbb{E}[g(\eta_{n,i})|\mathcal{F}_n] \leq K\epsilon \sum_{i=1}^{k_n} \mathbb{E}\left[|\eta_{n,i}|^2 \mathbf{1}\{|\eta_{n,i}| < \epsilon\} \middle| \mathcal{F}_n\right] + K \sum_{i=1}^{k_n} \mathbb{E}\left[|\eta_{n,i}|^2 \mathbf{1}\{|\eta_{n,i}| \ge \epsilon\} \middle| \mathcal{F}_n\right]$$

$$\leq KC\epsilon \sum_{i=1}^{k_n} \sigma_{n,i}^2 + K \sum_{i=1}^{k_n} \mathbb{E}\left[\sigma_{n,i}^2 |Z|^2 \mathbf{1}\{\sigma_{n,i}|Z| \ge \epsilon\} \middle| \mathcal{F}_n\right],$$

where the second term on the r.h.s. of this inequality satisfies

$$K\sum\nolimits_{i=1}^{k_n} \epsilon^2 \mathbb{E}\left[\epsilon^{-2} \sigma_{n,i}^2 |Z|^2 \mathbf{1}\{\sigma_{n,i}|Z| \geq \epsilon\} \middle| \mathcal{F}_n\right] \leq \frac{K}{\epsilon} \sum\nolimits_{i=1}^{k_n} \mathbb{E}\left[\sigma_{n,i}^3 |Z|^3 \middle| \mathcal{F}_n\right] = \frac{K}{\epsilon} \sum\nolimits_{i=1}^{k_n} \sigma_{n,i}^3 \mathbb{E}[|Z|^3].$$

Since

$$\begin{split} \sigma_{n,i}^2 &= \mathbb{E}[X_{n,i}^2 | \mathcal{F}_n] \\ &= \mathbb{E}\left[X_{n,i}^2 \mathbf{1} \left\{ |X_{n,i}| \leq \epsilon \right\} \middle| \mathcal{F}_n \right] + \mathbb{E}\left[X_{n,i}^2 \mathbf{1} \left\{ |X_{n,i}| > \epsilon \right\} \middle| \mathcal{F}_n \right] \\ &= \epsilon^2 + \mathbb{E}\left[X_{n,i}^2 \mathbf{1} \left\{ |X_{n,i}| > \epsilon \right\} \middle| \mathcal{F}_n \right], \end{split}$$

we have that

$$\max_{i \le k_n} \sigma_{n,i}^2 \le \epsilon^2 + \sum_{i=1}^{k_n} \mathbb{E}\left[X_{n,i}^2 \mathbf{1}\left\{|X_{n,i}| > \epsilon\right\} \middle| \mathcal{F}_n\right].$$

Since $\epsilon > 0$ is arbitrary, $\max_{i \leq k_n} \sigma_{n,i}^2 \stackrel{P}{\to} 0$, and therefore

$$\sum_{i=1}^{k_n} \sigma_{n,i}^3 \le \max_{i \le k_n} \sigma_{n,i}^2 \sum_{i=1}^{k_n} \sigma_{n,i}^2 \stackrel{P}{\to} 0.$$

To complete the proof, let $f \in C_b$ and $Z_n := \sum X_{n,i}$. Let K > 0 be arbitrary and notice that

$$\mathbb{E}\left[f(Z_n)|\mathcal{F}_n\right] = \mathbb{E}\left[f_K(Z_n)|\mathcal{F}_n\right] + E_1,$$

$$|E_1| = |\mathbb{E}\left[f_K(Z_n)|\mathcal{F}_n\right] - \mathbb{E}\left[f(Z_n)|\mathcal{F}_n\right]|$$

$$\leq \mathbb{E}\left[|f_K(Z_n) - f(Z_n)||\mathcal{F}_n\right]$$

$$\leq (2||f||_{\infty} + 1)\mathbb{P}\left(|Z_n| \geq K|\mathcal{F}_n\right).$$

Since f_K is continuous and compactly supported, for any $\epsilon > 0$ we can find $g_{K,\epsilon} \in C_b^{\infty}$, the space of continuous functions with continuous bounded derivatives of all orders, such that $\sup_x |g_{K,\epsilon}(x) - f_K(x)| < \epsilon$. Therefore we also have

$$\mathbb{E}\left[f_K(Z_n)|\mathcal{F}_n\right] = \mathbb{E}\left[g_{K,\epsilon}(Z_n)|\mathcal{F}_n\right] + E_2,$$

where

$$|E_2| = |\mathbb{E}\left[f_K(Z_n)|\mathcal{F}_n\right] - \mathbb{E}\left[g_{K,\epsilon}(Z_n)|\mathcal{F}_n\right]| < \epsilon.$$

Since $g_{K,\epsilon} \in C_b^{\infty}$ we know by the first result that

$$\mathbb{E}\left[\left.q_{K,\epsilon}(Z_n)\right|\mathcal{F}_n\right] = \mathbb{E}\left[\left.q_{K,\epsilon}(\sigma Z)\right] + E_3(n),\right.$$

where $E_3(n) \stackrel{P}{\to} 0$.

Moreover, we also have that

$$\mathbb{E}\left[f(\sigma Z)\right] = \mathbb{E}\left[g_{K,\epsilon}(\sigma Z)\right] + D_1 + D_2,$$
$$|D_1| \le \mathbb{P}\left(|\sigma Z| \ge K\right),$$
$$|D_2| \le \epsilon.$$

Thus, overall we get that, for any K > 0 and $\epsilon > 0$

$$\begin{aligned} &|\mathbb{E}\left[f(Z_n)|\mathcal{F}_n\right] - \mathbb{E}\left[f(\sigma Z)\right]|\\ &\leq 2\epsilon + E_3(n) + (2\|f\|_{\infty} + 1)\mathbb{P}\left(|Z_n| \geq K|\mathcal{F}_n\right) + \mathbb{P}\left(|\sigma Z| \geq K\right).\end{aligned}$$

We know that for any $K, \epsilon > 0$, $E_3(n) \stackrel{P}{\to} 0$. It is clear that as $K \to \infty$ the last term vanishes, while we also have that

$$\mathbb{P}(|Z_n| \ge K|\mathcal{F}_n) \le \frac{\mathbb{E}(Z_n^2|\mathcal{F}_n)}{K^2}$$

$$= \frac{\sum_{i=1}^{k_n} \sigma_{n,i}^2}{K^2} \xrightarrow{P} \frac{\sigma^2}{K^2},$$

as $n \to \infty$ by assumption. Letting $K \to \infty$ we obtain the result.

Result (161) follows from Corollary 2.4 in [7] while (162) follows from the discussion after Eq. (3) in this paper since μ_n and μ are measures on \mathbb{R} .

Lemma 29. Suppose that $Z_n := \sum_{i=1}^{k_n} X_{n,i}$ and \mathcal{F}_n are as in Lemma 27. If $T_n \stackrel{P}{\to} c$, then

$$Z_n + T_n | \mathcal{F}_n \Rightarrow \mathcal{N}(c, \sigma^2).$$

Proof of Lemma 29. Let $f \in C_b$. Let K > 0 be arbitrary, and let f_K be continuous so that $f_K(x) = f(x)$ for $|x| \le K$, $f_K(x) = 0$ for |x| > K + 1 and $||f_K||_{\infty} \le ||f||_{\infty}$. Then f_K is continuous, and compactly supported, so also bounded and uniformly continuous. Then

$$\mathbb{E}\left[f(Z_n + T_n)|\mathcal{F}_n\right] = \mathbb{E}\left[f_K(Z_n + T_n)|\mathcal{F}_n\right] + E_1(n),$$
$$|E_1(n)| \le (2\|f\|_{\infty} + 1)\mathbb{P}\left(|Z_n + T_n| \ge K|\mathcal{F}_n\right).$$

Then

$$\begin{aligned} &|\mathbb{E}\left[f_K(Z_n+T_n)|\mathcal{F}_n\right] - \mathbb{E}\left[f_K(\sigma Z+c)\right]|\\ &\leq &|\mathbb{E}\left[f_K(Z_n+T_n)|\mathcal{F}_n\right] - \mathbb{E}\left[f_K(Z_n+c)|\mathcal{F}_n\right]| + &|\mathbb{E}\left[f_K(Z_n+c)|\mathcal{F}_n\right] - \mathbb{E}\left[f_K(\sigma Z+c)\right]|.\end{aligned}$$

For the first term notice that since f_K is uniformly continuous, for any $\epsilon > 0$, we can find $\epsilon' > 0$, so that $|x - y| < \epsilon'$ implies that $|f_K(x) - f_K(y)| < \epsilon$. Therefore

$$\begin{aligned} &|\mathbb{E}\left[f_K(Z_n+T_n)|\mathcal{F}_n\right] - \mathbb{E}\left[f_K(Z_n+c)|\mathcal{F}_n\right]| \\ &\leq 2\|f\|_{\infty}\mathbb{P}\left(|T_n-c| \geq \epsilon'|\mathcal{F}_n\right) + \mathbb{E}\left[|f_K(Z_n+T_n) - f_K(Z_n+c)|\mathbf{1}\left\{|T_n-c| \leq \epsilon'\right\}|\mathcal{F}_n\right] \\ &\leq 2\|f\|_{\infty}\mathbb{P}\left(|T_n-c| \geq \epsilon'|\mathcal{F}_n\right) + \epsilon. \end{aligned}$$

We know that

$$\mathbb{E}\left[\mathbb{P}\left(\left|T_{n}-c\right|\geq\epsilon'|\,\mathcal{F}_{n}\right)\right]\to0,$$

and thus

$$\mathbb{P}\left(\left|T_{n}-c\right| \geq \epsilon' | \mathcal{F}_{n}\right) \stackrel{P}{\to} 0.$$

This proves that the first term vanishes in probability. For the second term notice that $z \mapsto f_K(\cdot + c)$ is continuous and bounded, and therefore the second term also vanishes in probability.

A.10 Proof of Proposition 7

We want to study IF (Ψ, Q_T) where $\Psi(u) = \nabla_{\vartheta} \log W(\widehat{\theta}, u)$ is only a function of the auxiliary variables. To be precise, we should write $\Psi^T(u^T) = \nabla_{\vartheta} \log W^T(\widehat{\theta}_T, u^T)$. However, for presentation brevity, we drop the index T also in the following proof whenever there is no possible confusion. The kernel Q_T has been designed as a pseudo-marginal-like algorithm targetting $\pi(d\theta)$ while U are auxiliary variables. However, we can also think of Q_T as a pseudo-marginal algorithm targeting

$$\overline{\pi}(du) = \int \overline{\pi}(d\theta, du) = m(du) \int \frac{\widehat{p}(y \mid \theta, u)}{p(y \mid \theta)} \pi(d\theta),$$

while θ is an auxiliary variable. In particular, the acceptance probability of the CPM kernel (7) can be rewritten as

$$\alpha_{Q}\left\{\left(\theta,u\right),\left(\theta',u'\right)\right\} = \min\left\{1,r\left(u,u'\right)\frac{\overline{\pi}\left(\theta'|u'\right)q\left(\theta',\theta\right)}{\overline{\pi}\left(\theta|u\right)q\left(\theta,\theta'\right)}\right\},\,$$

with

$$r(u, u') = \frac{\overline{\pi}(u')m(u)}{\overline{\pi}(u)m(u')}.$$

Let us consider the following MH algorithm

$$\overline{Q}\left\{\left(\theta,u\right),\left(\mathrm{d}\theta',\mathrm{d}u'\right)\right\} = K\left(u,\mathrm{d}u'\right)\overline{\pi}\left(\mathrm{d}\theta'|u'\right)\alpha_{\overline{Q}}\left(u,u'\right) + \left\{1 - \varrho_{\overline{Q}}\left(u\right)\right\}\delta_{\left(\theta,u\right)}\left(\mathrm{d}\theta',\mathrm{d}u'\right),$$

where

$$\alpha_{\overline{O}}(u, u') = \min\{1, r(u, u')\}\$$

and $1 - \varrho_{\overline{Q}}(u)$ is the corresponding rejection probability. This kernel admits the same invariant distribution as Q and we have

$$\int_{\Theta} \overline{Q} \{ (\theta, u), (d\theta', du') \} = \overline{Q}(u, du')$$

where

$$\overline{Q}\left(u,\mathrm{d}u'\right)=K\left(u,\mathrm{d}u'\right)\alpha_{\overline{Q}}\left(u,u'\right)+\left\{ 1-\varrho_{\overline{Q}}\left(u\right)\right\} \delta_{u}\left(\mathrm{d}u'\right)$$

is the 'ideal' marginal MH algorithm. The following lemma is an adaptation from [4, Proposition 2].

Lemma 30. Let $g: \mathcal{U}^2 \to \mathbb{R}^+$ be a measurable function. Define

$$\begin{split} &\Delta_{\overline{Q}}\left(g\right) = \iint \overline{\pi}\left(\mathrm{d}\theta,\mathrm{d}u\right) \iint K\left(u,\mathrm{d}u'\right) \overline{\pi}\left(\mathrm{d}\theta'|\,u'\right) \alpha_{\overline{Q}}\left(u,u'\right) g\left(u,u'\right), \\ &\Delta_{Q}\left(g\right) = \iint \overline{\pi}\left(\mathrm{d}\theta,\mathrm{d}u\right) \iint K\left(u,\mathrm{d}u'\right) q\left(\theta,\mathrm{d}\theta'\right) \alpha_{Q}\left\{\left(\theta,u\right),\left(\theta',u'\right)\right\} g\left(u,u'\right). \end{split}$$

Then we have $\Delta_{\overline{Q}}(g) \geq \Delta_{Q}(g)$.

Proof of Lemma 30. We can write for a bounded function g

$$\Delta_{\overline{Q}}(g) - \Delta_{Q}(g) = \iint \overline{\pi}(du) K(u, du') g(u, u') \iint \overline{\pi}(d\theta | u) q(\theta, d\theta') \left[\alpha_{\overline{Q}}(u, u') - \alpha_{Q} \{(\theta, u), (\theta', u')\}\right].$$

Now we have by Jensen's inequality

$$\iint \overline{\pi} (d\theta | u) q(\theta, d\theta') \alpha_Q \{(\theta, u), (\theta', u')\} = \iint \overline{\pi} (d\theta | u) q(\theta, d\theta') \min \left\{ 1, r(u, u') \frac{\overline{\pi} (\theta' | u') q(\theta', \theta)}{\overline{\pi} (\theta | u) q(\theta, \theta')} \right\} \\
\leq \min \left\{ 1, r(u, u') \iint \overline{\pi} (d\theta | u) q(\theta, d\theta') \frac{\overline{\pi} (\theta' | u') q(\theta', \theta)}{\overline{\pi} (\theta | u) q(\theta, \theta')} \right\} \\
= \alpha_{\overline{\Omega}} (u, u').$$

Hence $\Delta_{\overline{Q}}(g) \geq \Delta_{Q}(g)$ for bounded g. Monotone convergence and a truncation argument shows this is true for general g.

The following Proposition follows now directly from Lemma 30 and by checking that the arguments of the proof of Theorem 7 in [4, Proposition 2] are still valid in our scenario.

Proposition 31. Let $h: \mathcal{U} \to \mathbb{R}$ satisfying $\overline{\pi}(h^2) < \infty$ then $\mathrm{IF}(h,Q) \geq \mathrm{IF}(h,\overline{Q})$.

Armed with Proposition 31, we will show that $\mathrm{IF}(\Psi,\overline{Q}) \geq C\mathbb{V}_{\overline{\pi}}(\Psi)$ which implies that $\mathrm{IF}(\Psi,Q) \geq C\mathbb{V}_{\overline{\pi}}(\Psi)$ almost surely. Let $e\left(\Psi,\overline{Q}\right)(\mathrm{d}\lambda)$ denote the spectral measure of Ψ w.r.t Q; see e.g. [24], [32]. This measure $e\left(\Psi,\overline{Q}\right)$ is supported on [-1,1] as \overline{Q} is reversible and $\int_{-1}^{1} e\left(\Psi,\overline{Q}\right)(\mathrm{d}\lambda) = \mathbb{V}_{\overline{\pi}}(\Psi)$. We will show that

 $\int (1 - \lambda) e\left(\Psi, \overline{Q}\right) (\mathrm{d}\lambda) \le C,\tag{163}$

almost surely where the l.h.s. of (163) is the Expected Square Jump Distance (ESJD) of Ψ . By applying Jensen's inequality w.r.t. the probability measure $e\left(\Psi,\overline{Q}\right)\left(\mathrm{d}\lambda\right)/\mathbb{V}_{\overline{\pi}}\left(\Psi\right)$, the above inequality will imply that

$$IF(\Psi, \overline{Q}) = 2 \int \frac{1}{1 - \lambda} \frac{e(\Psi, \overline{Q})(d\lambda)}{\mathbb{V}_{\overline{\pi}}(\Psi)} - 1$$

$$\geq \frac{2}{\int (1 - \lambda) \frac{e(\Psi, \overline{Q})(d\lambda)}{\mathbb{V}_{\overline{\pi}}(\Psi)}} - 1 = 2C\mathbb{V}_{\overline{\pi}}(\Psi) - 1$$

almost surely. We now show that (163) holds, at least under severe regularity conditions listed in the proof which however make the calculations tractables. We postulate that this result holds under much weaker assumptions.

Proposition 32. Under Assumptions 1 and 11-17, the inequality (163) holds almost surely.

The lengthy proof of this proposition is deferred to the next section.

A.11 Proof of Proposition 32

For presentation brevity, we will only prove the result for d=1 and p=1. Using the notation of Section A.5 and a similar continuous-time embedding approach, we have for $U^{T}(\delta_{T}) \sim K_{\rho_{T}}(U^{T}(0), \cdot)$

$$\nabla \log W^{T}(\widehat{\theta}_{T}, U(\delta_{T})) - \nabla \log W^{T}(\widehat{\theta}_{T}, U(0))$$

$$= \sum_{t=1}^{T} \nabla \log \widehat{W}_{t}^{T}(Y_{t} \mid \widehat{\theta}_{T}; U(\delta_{T})) - \nabla \log \widehat{W}_{t}^{T}(Y_{t} \mid \widehat{\theta}_{T}; U_{t}(0))$$

$$= \sum_{t=1}^{T} \nabla \log (1 + \eta_{t}^{T}) = \sum_{t=1}^{T} \frac{\nabla \eta_{t}^{T}}{1 + \eta_{t}^{T}}$$

where

$$\eta_{t}^{T} = \frac{\widehat{W}_{t}^{T}(Y_{t} \mid \widehat{\theta}_{T}; U\left(\delta_{T}\right)}{\widehat{W}_{t}^{T}(Y_{t} \mid \widehat{\theta}_{T}; U_{t}\left(0\right)} - 1.$$

To simplify notation, we have written ∇ to denote the derivative w.r.t. ϑ evaluated at $\widehat{\theta}_T$.

We will make here the following assumptions. Here $B(\overline{\theta})$ denotes a neighbourhood of $\overline{\theta}$.

Assumption 11. There exists $\epsilon > 0$ such that for T large enough we have $\eta_t^T > -1 + \epsilon$ for all t in probability.

Assumption 12. The function $u \mapsto \pi_T(u)/\pi_T(u|\widehat{\theta}_T)$ is bounded w.r.t u for T large enough in probability.

Assumption 13. We have

$$\limsup_{T} \sup_{\theta \in B(\overline{\theta})} \mathbb{E}\left[\left(\widehat{W}_{1}^{T}\left(\theta\right)\right)^{-14} \middle| Y_{1}\right] < B\left(Y_{1}\right).$$

Assumption 14. We have

$$\limsup_{T} \sup_{\theta \in B(\overline{\theta})} \mathbb{E}\left[\left(\nabla \widehat{W}_{1}^{T}\left(\theta\right)\right)^{16} \middle| Y_{1}\right] < B\left(Y_{1}\right).$$

Assumption 15. We have

$$\limsup_{T} \sup_{\theta \in B(\overline{\theta})} \left\{ \mathbb{E} \left[\left| \partial_{u} \varpi \left(Y_{1}, U_{1,1} \left(0 \right) ; \theta \right) \right|^{8} + \left| \partial_{u} \varpi \left(Y_{1}, U_{1,1} \left(0 \right) ; \theta \right) U_{1,1} \left(0 \right) \right|^{8} \middle| Y_{1} \right] + \mathbb{E} \left[\left| \partial_{u,u} \varpi \left(Y_{1}, U_{1,1} \left(0 \right) ; \theta \right) \right|^{8} \middle| Y_{1} \right] \right\} < B \left(Y_{1} \right).$$

Assumption 16. We have

$$\limsup_{T} \sup_{\theta \in B(\overline{\theta})} \mathbb{E} \left[\left| \partial_{u,\vartheta} \varpi \left(Y_{1}, U_{1,1} \left(0 \right) ; \theta \right) \right|^{16} + \left| \partial_{u,\vartheta} \varpi \left(Y_{1}, U_{1,1} \left(0 \right) ; \theta \right) U_{1,1} \left(0 \right) \right|^{8} + \left| \partial_{u,u,\vartheta} \varpi \left(Y_{1}, U_{1,1} \left(0 \right) ; \theta \right) \right|^{8} \left| Y_{1} \right| < B \left(Y_{1} \right).$$

Assumption 17. We have

$$\mathbb{E}_{Y_1 \sim \mu} \left[B \left(Y_1 \right)^4 \right] < \infty.$$

To establish the result of the proposition, it is enough to show that the ESJD is O(1) almost surely. All the expectations in this section have to be understood conditional expectations w.r.t. \mathcal{Y}^T . Under Assumption 11, we have

$$\nabla \log W^{T}(\widehat{\theta}, U(\delta_{T})) - \nabla \log W^{T}(\widehat{\theta}, U(0)) = \sum_{t=1}^{T} \nabla \eta_{t}^{T} + \nabla \eta_{t}^{T} \cdot f(\eta_{t}^{T}),$$

with $|f(x)| \lesssim x$. In the sequel, the generic notation c is used to denote a constant that is independent of T. To alleviate notations, we do not use distinct indices each time such a constant appears, and keep using the notation c even though the corresponding constant may vary from one statement to the other. However, to avoid confusion, we sometimes make a distinction between such constants by using c, c', c'' inside an argument. We also further drop the dependence of W^T , $\widehat{\theta}_T$ and δ_T on T when no confusion is possible.

Using Assumption 12, the ESJD satisfies

$$\mathbb{E}_{U(0) \sim \overline{\pi}} \left[\left(\frac{\nabla W \left(\widehat{\theta}, U \left(0 \right) \right)}{W \left(\widehat{\theta}, U \left(0 \right) \right)} - \frac{\nabla W \left(\widehat{\theta}, U \left(\delta \right) \right)}{W \left(\widehat{\theta}, U \left(\delta \right) \right)} \right)^{2} \cdot 1 \wedge \left(\frac{\overline{\pi} \left(U \left(\delta \right) \right)}{m \left(U \left(\delta \right) \right)} \right) / \left(\frac{\overline{\pi} \left(U \left(0 \right) \right)}{m \left(U \left(0 \right) \right)} \right) \right]$$

$$= \widetilde{\mathbb{E}} \left[\frac{\overline{\pi} \left(U \left(0 \right) \right)}{\overline{\pi} \left(U \left(0 \right) \right) \widehat{\theta}} \left(\frac{\nabla W \left(\widehat{\theta}, U \left(0 \right) \right)}{W \left(\widehat{\theta}, U \left(0 \right) \right)} - \frac{\nabla W \left(\widehat{\theta}, U \left(\delta \right) \right)}{W \left(\widehat{\theta}, U \left(\delta \right) \right)} \right)^{2} \cdot 1 \wedge \left(\frac{\overline{\pi} \left(U \left(\delta \right) \right)}{m \left(U \left(\delta \right) \right)} \right) / \left(\frac{\overline{\pi} \left(U \left(0 \right) \right)}{m \left(U \left(0 \right) \right)} \right) \right]$$

$$\leq \widetilde{\mathbb{E}} \left[\frac{\overline{\pi} \left(U \left(0 \right) \right)}{\overline{\pi} \left(U \left(0 \right) \right) \widehat{\theta}} \left(\frac{\nabla W \left(\widehat{\theta}, U \left(0 \right) \right)}{W \left(\widehat{\theta}, U \left(0 \right) \right)} - \frac{\nabla W \left(\widehat{\theta}, U \left(\delta \right) \right)}{W \left(\widehat{\theta}, U \left(\delta \right) \right)} \right)^{2} \right].$$

$$\leq c \widetilde{\mathbb{E}} \left[\left(\sum_{t=1}^{T} \nabla \eta_{t}^{T} + \nabla \eta_{t}^{T} \cdot f \left(\eta_{t}^{T} \right) \right)^{2} \right]$$

$$\leq c' \left(\widetilde{\mathbb{E}} \left[\left(\sum_{t=1}^{T} \nabla \eta_{t}^{T} \right)^{2} + \widetilde{\mathbb{E}} \left[\left(\sum_{t=1}^{T} \nabla \eta_{t}^{T} \cdot f \left(\eta_{t}^{T} \right) \right)^{2} \right] \right)$$

where $\widetilde{\mathbb{E}}$ is to be understood in the rest of this section as having $U\left(0\right) \sim \overline{\pi}\left(\cdot | \widehat{\theta}\right)$.

A.11.1 Decomposition of η_t^T

We have

$$\eta_t^T = L_t^T + M_t^T,$$

where

$$L_{t}^{T} = \int_{0}^{\delta_{T}} \frac{1}{N\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)} \sum_{i=1}^{N} \left\{ -\partial_{u}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right) U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right) \right\} ds,$$

$$M_{t}^{T} = \int_{0}^{\delta_{T}} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)} \sum_{i=1}^{N} \partial_{u}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right) dB_{t,s}^{i}.$$

Here we write

$$\nabla L_t^T = \nabla L_{t,1}^T + \nabla L_{t,2}^T, \ \nabla M_t^T = \nabla M_{t,1}^T + \nabla M_{t,2}^T,$$

where

$$\nabla L_{t,1}^{T} = \int_{0}^{\delta_{T}} \frac{1}{N\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)} \sum_{i=1}^{N} \left\{ -\partial_{u,\vartheta}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right) U_{t,i}^{T}\left(s\right) + \partial_{u,u,\vartheta}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right) \right\} ds,$$

$$\nabla L_{t,2}^{T} = -\int_{0}^{\delta_{T}} \frac{\sum_{i=1}^{N} \left\{ -\partial_{u}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right) U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right) \right\} \nabla \log \widehat{W}_{t}^{T}\left(\widehat{\theta}\right)}{N\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)} ds,$$

$$\nabla M_{t,1}^{T} = \int_{0}^{\delta_{T}} \frac{\sqrt{2}}{N\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)} \sum_{i=1}^{N} \partial_{u,\vartheta}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right) dB_{t,s}^{i},$$

$$\nabla M_{t,2}^{T} = -\int_{0}^{\delta_{T}} \frac{\sqrt{2}\nabla \widehat{W}_{t}^{T}\left(\widehat{\theta}\right)}{N\left(\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)\right)^{2}} \left(\sum_{i=1}^{N} \partial_{u}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right)\right) dB_{t,s}^{i}.$$

A.11.2 Control of the term $\left(\sum_{t=1}^{T} \nabla \eta_t^T\right)^2$

By the C_p inequality, we have

$$\widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \nabla L_{t,1}^{T} + \nabla L_{t,2}^{T} + \nabla M_{t}^{T}\right)^{2}\right] \leq c \left(\widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \nabla L_{t,1}^{T}\right)^{2}\right] + \widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \nabla L_{t,2}^{T}\right)^{2}\right] + \widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \nabla M_{t}^{T}\right)^{2}\right]\right).$$
(164)

We now need to control the three terms appearing on the r.h.s. of (164).

Term $\nabla L_{t,1}^T$. We have

$$\widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \nabla L_{t,1}^{T}\right)^{2}\right] = \sum_{t=1}^{T} \widetilde{\mathbb{E}}\left[\left(\nabla L_{t,1}^{T}\right)^{2}\right] + \sum_{t,s:t\neq s}^{T} \widetilde{\mathbb{E}}\left[\nabla L_{t,1}^{T}.\nabla L_{s,1}^{T}\right].$$
(165)

We have for $s \neq t$

$$\widetilde{\mathbb{E}}\left[\nabla L_{t,1}^T \ \nabla L_{s,1}^T\right] = \widetilde{\mathbb{E}}\left[\nabla L_{t,1}^T \ \nabla L_{s,1}^T\right] = \widetilde{\mathbb{E}}\left[\nabla L_{t,1}^T \right] \ \widetilde{\mathbb{E}}\left[\nabla L_{s,1}^T\right] = 0.$$

Now we have

$$\begin{split} &\widetilde{\mathbb{E}}\left[\left(\nabla L_{t,1}^{T}\right)^{2}\right] \\ &= \widetilde{\mathbb{E}}\left[\left(\int_{0}^{\delta_{T}} \frac{1}{N\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)} \sum_{i=1}^{N} \left\{-\partial_{u,\vartheta}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right) U_{t,i}^{T}\left(s\right) + \partial_{u,u,\vartheta}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right)\right\} \mathrm{d}s\right)^{2}\right] \\ &= \mathbb{E}\left[\left(\prod_{r \neq t} \widehat{W}_{r}^{T}\left(\widehat{\theta}\right)\right) \left(\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)\right)^{-1} \\ &\quad \times \left(\int_{0}^{\delta_{T}} \frac{1}{N} \sum_{i=1}^{N} \left\{-\partial_{u,\vartheta}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right) U_{t,i}\left(s\right) + \partial_{u,u,\vartheta}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \widehat{\theta}\right)\right\} \mathrm{d}s\right)^{2}\right] \\ &\leq \sup_{\theta \in B\left(\overline{\theta}\right)} \mathbb{E}\left[\left(\int_{0}^{\delta_{T}} \frac{1}{N} \sum_{i=1}^{N} \left\{-\partial_{u,\vartheta}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \theta\right) U_{t,i}^{T}\left(s\right) + \partial_{u,u,\vartheta}\varpi\left(Y_{t}, U_{t,i}\left(s\right); \theta\right)\right\} \mathrm{d}s\right)^{4}\right]^{1/2} \\ &\times \sup_{\theta \in B\left(\overline{\theta}\right)} \mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-2}\right]^{1/2} \\ &\leq c \frac{\left(\delta_{T}\right)^{2}}{N} B\left(Y_{t}\right)^{1/14+1/4} \leq c' \frac{N}{T^{2}} B\left(Y_{t}\right)^{1/14+1/4}, \end{split}$$

where we have used Assumptions 13 and 16 and Cauchy-Schwarz inequality. To establish the last inequality, we have also used the fact that

$$\mathbb{E}\left[\left(\int_{0}^{\delta_{T}} f\left(U_{t}\left(s\right)\right) \mathrm{d}s\right)^{4}\right] = \delta_{T}^{4} \mathbb{E}\left[\left(\int_{0}^{\delta_{T}} f\left(U_{t}\left(s\right)\right) \frac{\mathrm{d}s}{\delta_{T}}\right)^{4}\right]$$

$$\leq \delta_{T}^{4} \mathbb{E}\left[\int_{0}^{\delta_{T}} f^{4}\left(U_{t}\left(s\right)\right) \frac{\mathrm{d}s}{\delta_{T}}\right]$$

$$= \delta_{T}^{4} \mathbb{E}\left[f^{4}\left(U_{t}\left(0\right)\right)\right] \text{ (by stationarity)}.$$

Further on, we will not emphasize that the constant appearing in our upper bounds are a power of $B(Y_t)$, which is assumed to have a finite expectation under the distribution μ of the observations under Assumption 17. Overall (165) thus contributes O(N/T) almost surely by the SLLN. To shorten presentation, we keep the details to a minimum from now on.

Term $\nabla L_{t,2}^T$. We have

$$\widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \nabla L_{t,2}^{T}\right)^{2}\right] = \sum_{t=1}^{T} \widetilde{\mathbb{E}}\left[\left(\nabla L_{t,2}^{T}\right)^{2}\right] + \sum_{t,s:t\neq s}^{T} \widetilde{\mathbb{E}}\left[\nabla L_{t,2}^{T}.\nabla L_{s,2}^{T}\right].$$
(166)

We have for $s \neq t$

$$\widetilde{\mathbb{E}}\left[\nabla L_{t,2}^T.\nabla L_{s,2}^T\right] = \widetilde{\mathbb{E}}\left[\nabla L_{t,2}^T\right]\widetilde{\mathbb{E}}\left[\nabla L_{s,2}^T\right]$$

and

$$\begin{split} &\left|\widetilde{\mathbb{E}}\left[\nabla L_{t,2}^{T}\right]\right| = \frac{1}{N}\left|\mathbb{E}\left[-\nabla\log\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)\int_{0}^{\delta_{T}}\sum_{i=1}^{N}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\right\}\mathrm{d}s\right]\right|^{2} \\ &\leq \frac{1}{N}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\nabla\log\widehat{W}_{t}^{T}\left(\theta\right)\right)^{2}\right]^{1/2} \\ &\times\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\sum_{i=1}^{N}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)\right\}\mathrm{d}s\right)^{2}\right]^{1/2} \\ &= \frac{1}{N}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\sum_{i=1}^{N}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\right\}\mathrm{d}s\right)^{2}\right]^{1/2} \\ &\times\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\sum_{i=1}^{N}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\right\}\mathrm{d}s\right)^{2}\right]^{1/2} \\ &= \frac{1}{N}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\nabla\widehat{W}_{t}^{T}\left(\theta\right)\right)^{4}\right]^{1/2}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-4}\right]^{1/2} \\ &\times\sqrt{N}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,1}^{T}\left(s\right);\widehat{\theta}\right)U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,1}^{T}\left(s\right);\widehat{\theta}\right)\right\}\mathrm{d}s\right)^{2}\right]^{1/2} \\ &\leq c\frac{1}{N\sqrt{N}}\sqrt{N}\delta_{T} \leq c'\frac{1}{T}. \end{split}$$

We have

$$\begin{split} &\widetilde{\mathbb{E}}\left[\left(\nabla L_{t,2}^{T}\right)^{2}\right] \\ &= \mathbb{E}\left[W\left(\widehat{\theta}, U\left(0\right)\right) \cdot \left(-\int_{0}^{\delta_{T}} \frac{\sum_{i=1}^{N} \left\{-\partial_{u} \varpi\left(Y_{t}, U_{t,i}^{T}\left(s\right); \widehat{\theta}\right) U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2} \varpi\left(Y_{t}, U_{t,i}^{T}\left(s\right); \widehat{\theta}\right)\right\} \nabla \widehat{W}_{t}^{T}}{N\left(\widehat{W}_{t}^{T}\right)^{2}} \mathrm{d}s\right)^{2}\right] \\ &= \mathbb{E}\left[\frac{\left(\nabla \log \widehat{W}_{t}^{T}\left(\widehat{\theta}\right)\right)^{2}}{\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)} \cdot \left(-\int_{0}^{\delta_{T}} \frac{1}{N} \sum_{i=1}^{N} \left\{-\partial_{u} \varpi\left(Y_{t}, U_{t,i}^{T}\left(s\right); \widehat{\theta}\right) U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2} \varpi\left(Y_{t}, U_{t,i}^{T}\left(s\right); \widehat{\theta}\right)\right\} \mathrm{d}s\right)^{2}\right] \\ &= \sup_{\theta \in B(\widehat{\theta})} \mathbb{E}\left[\frac{\left(\nabla \log \widehat{W}_{t}^{T}\left(\theta\right)\right)^{4}}{\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{2}}\right]^{1/2} \\ &\times \sup_{\theta \in B(\widehat{\theta})} \mathbb{E}\left[\left(-\int_{0}^{\delta_{T}} \frac{1}{N} \sum_{i=1}^{N} \left\{-\partial_{u} \varpi\left(Y_{t}, U_{t,i}^{T}\left(s\right); \theta\right) U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2} \varpi\left(Y_{t}, U_{t,i}^{T}\left(s\right); \theta\right)\right\} \mathrm{d}s\right)^{4}\right]^{1/2} \\ &\leq c \frac{N}{T^{2}} \sup_{\theta \in B(\widehat{\theta})} \mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-12}\right]^{1/4} \sup_{\theta \in B(\widehat{\theta})} \mathbb{E}\left[\left(\nabla \widehat{W}_{t}^{T}\left(\theta\right)\right)^{8}\right]^{1/4} \\ &\leq c' \frac{N}{T^{2}}. \end{split}$$

Thus the term (166) is O(1) almost surely.

Term $\nabla M_{t,1}^T$. We have

$$\widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \nabla M_{t,1}^{T}\right)^{2}\right] = \sum_{t=1}^{T} \widetilde{\mathbb{E}}\left[\left(\nabla M_{t,1}^{T}\right)^{2}\right] + \sum_{t,s:t\neq s}^{T} \widetilde{\mathbb{E}}\left[\nabla M_{t,1}^{T}.\nabla M_{s,1}^{T}\right].$$
(167)

We have for $s \neq t$

$$\widetilde{\mathbb{E}}\left[\nabla M_{t\,1}^T.\nabla M_{s\,1}^T\right] = \widetilde{\mathbb{E}}\left[\nabla M_{t\,1}^T\right]\widetilde{\mathbb{E}}\left[\nabla M_{s\,1}^T\right] = 0.$$

Now we have

$$\begin{split} \widetilde{\mathbb{E}}\left[\left(\nabla M_{t,1}^{T}\right)^{2}\right] &= \mathbb{E}\left[W\left(\widehat{\theta},U\left(0\right)\right).\left(\int_{0}^{\delta_{T}}\frac{\sqrt{2}}{N\widehat{W}_{t}^{T}}\sum_{i=1}^{N}\partial_{u,\vartheta}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\mathrm{d}B_{t,s}^{i}\right)^{2}\right] \\ &= \mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\frac{\sqrt{2}}{N\sqrt{\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)}}\sum_{i=1}^{N}\partial_{u,\vartheta}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\mathrm{d}B_{t,s}^{i}\right)^{2}\right] \\ &= \frac{2}{N^{2}}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\mathbb{E}\left[\frac{1}{\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)}\left\{\partial_{u,\vartheta}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\right\}^{2}\right]\mathrm{d}s \\ &\leq c\frac{\sup_{\theta\in B\left(\widehat{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-2}\right]^{1/2}}{N^{2}}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\sup_{\theta\in B\left(\widehat{\theta}\right)}\mathbb{E}\left[\left\{\partial_{u,\vartheta}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)\right\}^{4}\right]^{1/2}\mathrm{d}s \\ &\leq \frac{c'}{T}. \end{split}$$

Hence the term (167) is overall O(1) almost surely.

Term $\nabla M_{t,2}^T$. We have

$$\widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \nabla M_{t,2}^{T}\right)^{2}\right] = \sum_{t=1}^{T} \widetilde{\mathbb{E}}\left[\left(\nabla M_{t,2}^{T}\right)^{2}\right] + \sum_{t,s:t\neq s}^{T} \widetilde{\mathbb{E}}\left[\nabla M_{t,2}^{T} \ \nabla M_{s,2}^{T}\right]. \tag{168}$$

We have for $s \neq t$

$$\widetilde{\mathbb{E}}\left[\nabla M_{t,2}^T \; \nabla M_{s,2}^T\right] = \widetilde{\mathbb{E}}\left[\nabla M_{t,2}^T\right] \; \widetilde{\mathbb{E}}\left[\nabla M_{s,2}^T\right]$$

where

$$\begin{split} \widetilde{\mathbb{E}} \left[-\int_{0}^{\delta_{T}} \frac{\sqrt{2} \nabla \widehat{W}_{t}^{T} \left(\widehat{\boldsymbol{\theta}} \right)}{N \left(\widehat{W}_{t}^{T} \left(\widehat{\boldsymbol{\theta}} \right) \right)^{2}} \left(\sum_{i=1}^{N} \partial_{u} \boldsymbol{\varpi} \left(Y_{t}, U_{t,i}^{T} \left(\boldsymbol{s} \right) ; \widehat{\boldsymbol{\theta}} \right) \right) \mathrm{d} B_{t,s}^{i} \right] \\ = \mathbb{E} \left[-\int_{0}^{\delta_{T}} \frac{\sqrt{2} \nabla \widehat{W}_{t}^{T} \left(\widehat{\boldsymbol{\theta}} \right)}{N \widehat{W}_{t}^{T} \left(\widehat{\boldsymbol{\theta}} \right)} \left(\sum_{i=1}^{N} \partial_{u} \boldsymbol{\varpi} \left(Y_{t}, U_{t,i}^{T} \left(\boldsymbol{s} \right) ; \widehat{\boldsymbol{\theta}} \right) \right) \mathrm{d} B_{t,s}^{i} \right] = 0 \end{split}$$

We have

$$\begin{split} \widetilde{\mathbb{E}}\left[\left(\nabla M_{t,2}^{T}\right)^{2}\right] &= \widetilde{\mathbb{E}}\left[\left(\sum_{i=1}^{N} \frac{\sqrt{2}\nabla \widehat{W}_{t}^{T}\left(\widehat{\boldsymbol{\theta}}\right)}{N\left(\widehat{W}_{t}^{T}\left(\widehat{\boldsymbol{\theta}}\right)\right)^{2}} \int_{0}^{\delta_{T}} \partial_{u}\varpi\left(Y_{t}, U_{t,i}^{T}\left(s\right); \widehat{\boldsymbol{\theta}}\right) \mathrm{d}B_{t,s}^{i}\right)^{2}\right] \\ &= \mathbb{E}\left[\left(\sum_{i=1}^{N} \frac{\sqrt{2}\nabla \widehat{W}_{t}^{T}}{N\left(\widehat{W}_{t}^{T}\right)^{3/2}} \int_{0}^{\delta_{T}} \partial_{u}\varpi\left(Y_{t}, U_{t,i}^{T}\left(s\right); \widehat{\boldsymbol{\theta}}\right) \mathrm{d}B_{t,s}^{i}\right)^{2}\right] \\ &= \frac{1}{N}\mathbb{E}\left[\frac{2\left(\nabla \widehat{W}_{t}^{T}\left(\widehat{\boldsymbol{\theta}}\right)\right)^{2}}{\left(\widehat{W}_{t}^{T}\left(\widehat{\boldsymbol{\theta}}\right)\right)^{3}} \int_{0}^{\delta_{T}} \left\{\partial_{u}\varpi\left(Y_{t}, U_{t,i}^{T}\left(s\right); \widehat{\boldsymbol{\theta}}\right)\right\}^{2} \mathrm{d}s\right] \\ &= \frac{1}{N} \int_{0}^{\delta_{T}} \mathbb{E}\left[\frac{2\left(\nabla \widehat{W}_{t}^{T}\left(\widehat{\boldsymbol{\theta}}\right)\right)^{2}}{\left(\widehat{W}_{t}^{T}\left(\widehat{\boldsymbol{\theta}}\right)\right)^{3}} \left\{\partial_{u}\varpi\left(Y_{t}, U_{t,i}^{T}\left(s\right); \widehat{\boldsymbol{\theta}}\right)\right\}^{2}\right] \mathrm{d}s \\ &\leq \frac{\delta_{T}}{N} \sup_{\theta \in B(\overline{\boldsymbol{\theta}})} \mathbb{E}\left[\left(\frac{\sqrt{2}\left(\nabla \widehat{W}_{t}^{T}\left(\boldsymbol{\theta}\right)\right)^{2}}{\left(\widehat{W}_{t}^{T}\left(\boldsymbol{\theta}\right)\right)^{3}}\right)^{2}\right]^{1/2} \sup_{\theta \in B(\overline{\boldsymbol{\theta}})} \mathbb{E}\left[\left\{\partial_{u}\varpi\left(Y_{t}, U_{t,1}^{T}\left(0\right); \boldsymbol{\theta}\right)\right\}^{4}\right]^{1/2} \\ &\leq \frac{c}{T}. \end{split}$$

Hence the term (168) is overall O(1) almost surely.

A.11.3 Control of $\sum_{t=1}^{T} \nabla \eta_t^T . f(\eta_t^T)$

We have

$$\widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \nabla \eta_{t}^{T} \cdot f(\eta_{t}^{T})\right)^{2}\right] \leq \widetilde{\mathbb{E}}\left[\sum_{t=1}^{T} \left(\nabla \eta_{t}^{T}\right)^{2} \cdot \sum_{t=1}^{T} f(\eta_{t}^{T})^{2}\right] \\
\leq c \,\widetilde{\mathbb{E}}\left[\sum_{t=1}^{T} \left(\nabla \eta_{t}^{T}\right)^{2} \cdot \sum_{t=1}^{T} \eta_{t}^{T2}\right] \\
\leq c' \,\widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \left(\nabla \eta_{t}^{T}\right)^{2}\right)^{2}\right]^{1/2} \,\widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T} \left(\eta_{t}^{T}\right)^{2}\right)^{2}\right]^{1/2}.$$

Control of $\left(\sum_{t=1}^{T} \left(\nabla \eta_{t}^{T}\right)^{2}\right)^{2}$. We have

$$\begin{split} \widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T}\left(\nabla\eta_{t}^{T}\right)^{2}\right)^{2}\right] &= \sum_{t=1}^{T}\widetilde{\mathbb{E}}\left[\left(\nabla\eta_{t}^{T}\right)^{4}\right] + \sum_{t,s:t\neq s}^{T}\widetilde{\mathbb{E}}\left[\left(\nabla\eta_{t}^{T}\right)^{2}\left(\nabla\eta_{s}^{T}\right)^{2}\right] \\ &= \sum_{t=1}^{T}\widetilde{\mathbb{E}}\left[\left(\nabla\eta_{t}^{T}\right)^{4}\right] + \sum_{t,s:t\neq s}^{T}\widetilde{\mathbb{E}}\left[\left(\nabla\eta_{t}^{T}\right)^{2}\right]\widetilde{\mathbb{E}}\left[\left(\nabla\eta_{s}^{T}\right)^{2}\right]. \end{split}$$

We have

$$\begin{split} \widetilde{\mathbb{E}}\left[\left(\nabla \eta_{t}^{T}\right)^{4}\right] &= \widetilde{\mathbb{E}}\left[\left(\nabla L_{t,1}^{T} + \nabla L_{t,2}^{T} + \nabla M_{t,1}^{T} + \nabla M_{t,2}^{T}\right)^{4}\right] \\ &\leq c \; \left(\widetilde{\mathbb{E}}\left[\left(\nabla L_{t,1}^{T}\right)^{4}\right] + \widetilde{\mathbb{E}}\left[\left(\nabla L_{t,2}^{T}\right)^{4}\right] + \widetilde{\mathbb{E}}\left[\left(\nabla M_{t,1}^{T}\right)^{4}\right] + \widetilde{\mathbb{E}}\left[\left(\nabla M_{t,2}^{T}\right)^{4}\right]\right). \end{split}$$

And now

$$\begin{split} &\widetilde{\mathbb{E}}\left[\left(\nabla L_{t,1}^{T}\right)^{4}\right] \\ &= \mathbb{E}\left[W\left(\widehat{\theta},U\left(0\right)\right).\left(\nabla L_{t,1}^{T}\right)^{4}\right] \\ &= \mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)\right)^{-3}.\left(\int_{0}^{\delta_{T}}\frac{1}{N}\sum_{i=1}^{N}\left\{-\partial_{u,\vartheta}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)U_{t,i}^{T}\left(s\right)+\partial_{u,u,\vartheta}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\right\}\mathrm{d}s\right)^{4}\right] \\ &\leq \sup_{\theta \in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-6}\right]^{1/2} \\ &\times \sup_{\theta \in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\frac{1}{N}\sum_{i=1}^{N}\left\{-\partial_{u,\vartheta}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)U_{t,i}^{T}\left(s\right)+\partial_{u,u,\vartheta}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)\right\}\mathrm{d}s\right)^{8}\right]^{1/2} \\ &\leq c\frac{N^{2}}{T^{4}}. \end{split}$$

We also have

$$\begin{split} &\widetilde{\mathbb{E}}\left[\left(\nabla L_{t,2}^{T}\right)^{4}\right] \\ &= \mathbb{E}\left[W\left(\widehat{\theta},U\left(0\right)\right).\left(\nabla L_{t,2}^{T}\right)^{4}\right] \\ &= \mathbb{E}\left[W\left(\widehat{\theta},U\left(0\right)\right).\left(\int_{0}^{\delta_{T}} \frac{\sum_{i=1}^{N}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)U_{t,i}^{T}\left(s\right)+\partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\right\}\nabla\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)}{N\left(\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)\right)^{2}}\mathrm{d}s\right)^{4}\right] \\ &= \mathbb{E}\left[\nabla\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)\left(\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)\right)^{-3}.\left(\int_{0}^{\delta_{T}} \frac{\sum_{i=1}^{N}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)U_{t,i}^{T}\left(s\right)+\partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\right\}}{N}\mathrm{d}s\right)^{4}\right] \\ &\leq \sup_{\theta \in B(\widehat{\theta})} \mathbb{E}\left[\left(\nabla\widehat{W}_{t}^{T}\left(\theta\right)\right)^{2}\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-6}\right]^{1/2} \\ &\times \sup_{\theta \in B(\widehat{\theta})} \mathbb{E}\left[\left(\nabla\widehat{W}_{t}^{T}\left(\theta\right)\right)^{4}\right]^{1/4} \sup_{\theta \in B(\widehat{\theta})} \mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-12}\right]^{1/4} \\ &\leq \sup_{\theta \in B(\widehat{\theta})} \mathbb{E}\left[\left(\int_{0}^{\delta_{T}} \frac{\sum_{i=1}^{N}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)U_{t,i}^{T}\left(s\right)+\partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)\right\}}{N}\mathrm{d}s\right)^{8}\right]^{1/2} \\ &\leq c \frac{N^{2}}{T^{4}}. \end{split}$$

We have

$$\begin{split} \widetilde{\mathbb{E}}\left[\left(\nabla M_{t,1}^{T}\right)^{4}\right] &= \mathbb{E}\left[W\left(\widehat{\theta},U\left(0\right)\right).\left(\int_{0}^{\delta_{T}}\frac{\sqrt{2}}{N\widehat{W}_{t}^{T}}\sum_{i=1}^{N}\partial_{u,\vartheta}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\mathrm{d}B_{t,s}^{i}\right)^{4}\right] \\ &\leq \sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-6}\right]^{1/2}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\frac{\sqrt{2}}{N}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\partial_{u,\vartheta}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)\mathrm{d}B_{t,s}^{i}\right)^{8}\right]^{1/2} \\ &\leq c\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-6}\right]^{1/2}\frac{1}{N^{2}}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\partial_{u,\vartheta}\varpi\left(Y_{t},U_{t,1}^{T}\left(s\right);\theta\right)\mathrm{d}B_{t,s}^{i}\right)^{8}\right]^{1/2} \\ &\leq c'\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-6}\right]^{1/2}\frac{1}{N^{2}}\left[\int_{0}^{\delta_{T}}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left(\left(\partial_{u,\vartheta}\varpi\left(Y_{t},U_{t,1}^{T}\left(0\right);\theta\right)\right)^{8}\right)^{1/4}\mathrm{d}s\right]^{2} \\ &\leq \frac{c''}{T^{2}} \end{split}$$

and

$$\begin{split} \widetilde{\mathbb{E}}\left[\left(\nabla M_{t,2}^{T}\right)^{4}\right] &= \mathbb{E}\left[W\left(\widehat{\theta},U\left(0\right)\right).\left(\int_{0}^{\delta_{T}}\frac{\sqrt{2}\nabla\widehat{W}_{t}^{T}}{N\left(\widehat{W}_{t}^{T}\right)^{2}}\left(\sum_{i=1}^{N}\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\right)\mathrm{d}B_{t,s}^{i}\right)^{4}\right] \\ &\leq \sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-14}\right]^{1/2}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\frac{\sqrt{2}\nabla\widehat{W}_{t}^{T}\left(\theta\right)}{N}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\left(\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)\right)\mathrm{d}B_{t,s}^{i}\right)^{8}\right]^{1/2} \\ &\leq c\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-14}\right]^{1/2}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\nabla\widehat{W}_{t}^{T}\left(\theta\right)\right)^{16}\right]^{1/4} \\ &\times\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\left(\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)\right)\mathrm{d}B_{t,s}^{i}\right)^{16}\right]^{1/4} \\ &\leq \frac{c}{N^{2}}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\partial_{u,\vartheta}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)\mathrm{d}B_{t,s}^{i}\right)^{16}\right]^{1/4} \\ &\leq c'\frac{N^{2}}{T^{4}}. \end{split}$$

Control of $\left(\sum_{t=1}^{T} \left(\eta_t^T\right)^2\right)^2$. We have

$$\begin{split} \widetilde{\mathbb{E}}\left[\left(\sum_{t=1}^{T}\left(\eta_{t}^{T}\right)^{2}\right)^{2}\right] &= \sum_{t=1}^{T}\widetilde{\mathbb{E}}\left[\left(\eta_{t}^{T}\right)^{4}\right] + \sum_{t,s:t\neq s}^{T}\widetilde{\mathbb{E}}\left[\left(\eta_{t}^{T}\right)^{2}\left(\eta_{s}^{T}\right)^{2}\right] \\ &= \sum_{t=1}^{T}\widetilde{\mathbb{E}}\left[\left(\eta_{t}^{T}\right)^{4}\right] + \sum_{t,s:t\neq s}^{T}\widetilde{\mathbb{E}}\left[\left(\eta_{t}^{T}\right)^{2}\right]\widetilde{\mathbb{E}}\left[\left(\eta_{s}^{T}\right)^{2}\right] \end{split}$$

where

$$\widetilde{\mathbb{E}}\left[\left(\eta_t^T\right)^2\right]\widetilde{\mathbb{E}}\left[\left(\eta_s^T\right)^2\right] \leq \widetilde{\mathbb{E}}\left[\left(\eta_t^T\right)^4\right]^{1/2}\widetilde{\mathbb{E}}\left[\left(\eta_t^T\right)^4\right]^{1/2},$$

with

$$\eta_t^T = L_t^T + M_t^T.$$

Now we have

$$\begin{split} \widetilde{\mathbb{E}}\left[\left(L_{t}^{T}\right)^{4}\right] &= \mathbb{E}\left[W\left(\widehat{\theta},U\left(0\right)\right)\left(L_{t}^{T}\right)^{4}\right] \\ &= \mathbb{E}\left[W\left(\widehat{\theta},U\left(0\right)\right)\left(\int_{0}^{\delta_{T}}\frac{1}{N\widehat{W}_{t}^{T}}\sum_{i=1}^{N}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\right\}\mathrm{d}s\right)^{4}\right] \\ &= \mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)\right)^{-3}\left(\frac{1}{N}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\right\}\mathrm{d}s\right)^{4}\right] \\ &\leq c\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-6}\right]^{1/2} \\ &\times\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\left\{-\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)U_{t,i}^{T}\left(s\right) + \partial_{u,u}^{2}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)\right\}\mathrm{d}s\right)^{8}\right]^{1/2} \\ &\leq c'\frac{N^{2}}{T^{4}} \end{split}$$

and

$$\begin{split} \widetilde{\mathbb{E}}\left[\left(M_{t}^{T}\right)^{4}\right] &= \mathbb{E}\left[W\left(\widehat{\theta},U\left(0\right)\right)\left(M_{t}^{T}\right)^{4}\right] \\ &= \mathbb{E}\left[W\left(\widehat{\theta},U\left(0\right)\right)\left(\int_{0}^{\delta_{T}}\frac{\sqrt{2}}{N\widehat{W}_{t}^{T}}\sum_{i=1}^{N}\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\mathrm{d}B_{t,s}^{i}\right)^{4}\right] \\ &= \mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\widehat{\theta}\right)\right)^{-3}\left(\frac{\sqrt{2}}{N}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\widehat{\theta}\right)\mathrm{d}B_{t,s}^{i}\right)^{4}\right] \\ &\leq \sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-6}\right]^{1/2}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\frac{\sqrt{2}}{N}\sum_{i=1}^{N}\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{t},U_{t,i}^{T}\left(s\right);\theta\right)\mathrm{d}B_{t,s}^{i}\right)^{8}\right]^{1/2} \\ &\leq \sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-6}\right]^{1/2}\frac{1}{N^{2}}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\int_{0}^{\delta_{T}}\partial_{u}\varpi\left(Y_{t},U_{t,1}^{T}\left(s\right);\theta\right)\mathrm{d}B_{t,s}^{i}\right)^{8}\right]^{1/2} \\ &\leq \sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\widehat{W}_{t}^{T}\left(\theta\right)\right)^{-6}\right]^{1/2}\frac{1}{N^{2}}\left(\int_{0}^{\delta_{T}}\sup_{\theta\in B\left(\overline{\theta}\right)}\mathbb{E}\left[\left(\partial_{u}\varpi\left(Y_{t},U_{t,1}^{T}\left(0\right);\theta\right)^{8}\right)\right]^{1/2}\mathrm{d}s\right)^{2} \\ &\leq c\frac{\delta_{T}^{2}}{N^{2}}\leq \frac{c'}{T^{2}}. \end{split}$$

Combining all the terms, we have shown that the ESJD is O(1) almost surely.