arXiv:1209.0703v1 [math.ST] 4 Sep 2012

ADAPTIVE MARKOV CHAIN MONTE CARLO CONFIDENCE INTERVALS

YVES F. ATCHADÉ

(Aug. 2012)

ABSTRACT. In Adaptive Markov Chain Monte Carlo (AMCMC) simulation, classical estimators of asymptotic variances are inconsistent in general. In this work we establish that despite this inconsistency, confidence interval procedures based on these estimators remain consistent. We study two classes of confidence intervals, one based on the standard Gaussian limit theory, and the class of so-called fixed-b confidence intervals. We compare the two procedures by deriving upper bounds on their convergence rates. We establish that the rate of convergence of fixed-b variance estimators is at least $\log(n)/\sqrt{n}$, while the convergence rate of the classical procedure is typically of order $n^{-1/3}$. We use simulation examples to illustrate the results.

1. INTRODUCTION

Throughout the paper we consider the following setting that covers standard MCMC and many AMCMC algorithms. π denotes a probability measure of interest on some measure space $(\mathsf{X}, \mathcal{B})$. $\{P_\theta, \theta \in \Theta\}$ is a family of Markov transition kernels on $(\mathsf{X}, \mathcal{B})$, for some measurable space $(\Theta, \mathcal{A})$. We assume that each P_θ admits π as unique invariant distribution, and that the map $(x, \theta) \mapsto P_\theta(x, \cdot)$ is $(\mathcal{B} \times \mathcal{A})$-measurable. On some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a nondecreasing filtration $\{\mathcal{F}_n,\ n \geq 0\}$, we consider a $\mathcal{F}_n$-adapted random process $\{(X_n, \theta_n),\ n \geq 0\}$ with values in $\mathsf{X} \times \Theta$ such that for any nonnegative function $f:\ \mathsf{X} \to \mathbb{R}$, and $n \geq 1$,

$$\mathbb{E}\left(f(X_n)|\mathcal{F}_{n-1}\right) = \int P_{\theta_{n-1}}(X_{n-1}, dz) f(z), \quad \mathbb{P}-\text{a.s.} \tag{1}$$

We write $\mathbb{E}$ for the expectation operator wrt to $\mathbb{P}$. We call the marginal sequence $\{X_n,\ n \geq 0\}$ an adaptive Markov chain. Notice that when there is no adaptation (that is $\theta_n \equiv \theta$), $\{X_n,\ n \geq 0\}$ reduces to a standard Markov chain. AMCMC algorithms have recently gained popularity in Monte Carlo simulation due to their ability for producing efficient samplers with limited tuning from the user. For an introduction and literature review on AMCMC, see e.g. Andrieu and Thoms (2008).

Let $h:\ \mathsf{X} \to \mathbb{R}$ be some function of interest, and suppose that we wish to estimate $\pi(h) \stackrel{\text{def}}{=} \int_{\mathsf{X}} h(x)\pi(dx)$ (for instance, $h(x) = x$ if we wish to estimate the mean of π). Under some fairly general conditions, it is known that $\hat{\pi}_n(h) \stackrel{\text{def}}{=} n^{-1}\sum_{k=1}^n h(X_k)$ converges to $\pi(h)$ and satisfies a

2000 *Mathematics Subject Classification.* 60F15, 60G42.

Key words and phrases. Adaptive Markov chain Monte Carlo, asymptotic variance, lag-window estimation, martingale approximation, Berry-Esseen bound.

Y. F. Atchadé: University of Michigan, 1085 South University, Ann Arbor, 48109, MI, United States. *E-mail address:* yvesa@umich.edu.

central limit theorem (Andrieu and Moulines (2006); Atchade and Fort (2010, 2012); Saksman and Vihola (2010)). It is also known that in most practical cases, as $n \to \infty$, the bias of $\hat{\pi}_n(h)$ satisfies $\mathbb{E}\left(\hat{\pi}_n(h)\right) - \pi(h) = o(n^{-1/2})$ and the variance is such that $n\mathsf{Var}\left(\hat{\pi}_n(h)\right)$ converges to a limit called the asymptotic variance of h. In these cases, assessing the Monte Carlo error of $\hat{\pi}_n(h)$ boils down to estimating the asymptotic variance. For a Markov chain with transition kernel P, the asymptotic variance can be written as

$$\sigma^2_P(h) \stackrel{\text{def}}{=} \sum_{\ell=-\infty}^{+\infty} \gamma_\ell(P, h), \tag{2}$$

where for $\ell \geq 0$, $\gamma_\ell(P, h) \stackrel{\text{def}}{=} \int (h(x) - \pi(h)) P^\ell h(x) \pi(dx)$, and $\gamma_{-\ell}(P, h) = \gamma_\ell(P, h)$. A well established approach for estimating $\sigma^2_P(h)$ is by lag-window obtained by taking a weighted average of the sample auto-covariances. More precisely, for $0 \leq \ell \leq n-1$, set

$$\gamma_{n,\ell} \stackrel{\text{def}}{=} n^{-1} \sum_{j=1}^{n-\ell} \left(h(X_j) - \hat{\pi}_n(h)\right)\left(h(X_{j+\ell}) - \hat{\pi}_n(h)\right), \quad \text{and} \quad \gamma_{n,-\ell} = \gamma_{n,\ell}.$$

Let $\{c_n, ; n \geq 1\}$ be an increasing sequence of integers such that $c_n \uparrow \infty$, and $w : \mathbb{R} \to \mathbb{R}$, an even weight function ($w(-x) = w(x)$). The lag-window estimator of $\sigma^2_P(h)$ is

$$\Gamma^2_n(h) = \sum_{k=-n+1}^{n-1} w\left(\frac{k}{c_n}\right) \gamma_{n,k}. \tag{3}$$

The precision of the Monte Carlo estimate is then gauged by computing the Monte Carlo error $\sqrt{\Gamma^2_n(h)/n}$ or equivalently the effective sample size $n\gamma_{n,0}/\Gamma^2_n(h)$. Alternatively a confidence interval for $\pi(h)$ can be formed using $\hat{\pi}_n(h) \pm z_\alpha \sqrt{\Gamma^2_n(h)/n}$, where z_α is the appropriate quantile of the standard normal distribution. The width of this confidence interval can be used as a stopping rule for the simulation (Jones et al. (2006)). All this is common practice in MCMC backed by the fact that for $c_n = o(n)$, and under some regularity conditions (e.g. geometric ergodicity and existence of $(2+\epsilon)$-moment for h under π), $\Gamma^2_n(h)$ converges in probability to $\sigma^2_P(h)$ (Damerdji (1995); Flegal and Jones (2010); Atchade (2011)).

Asymptotic variance estimation for AMCMC may behave differently. With $\{(X_n, \theta_n),\ n \geq 0\}$ as defined above, if θ_n converges to a (possibly random) limit $\theta_\star$, say, the asymptotic variance for h is typically

$$\lim_{n\to\infty} n\mathsf{Var}\left(\hat{\pi}_n(h)\right) = \mathbb{E}\left[\sigma^2_{\theta_\star}(h)\right], \tag{4}$$

where $\sigma^2_\theta(h) \stackrel{\text{def}}{=} \sigma^2_{P_\theta}(h)$. The same lag-window estimate $\Gamma^2_n(h)$ given in (3) can still be computed from the adaptive chain $\{X_n,\ n \geq 0\}$. But as it turns out, if $\theta_\star$ is random, $\Gamma^2_n(h)$ is inconsistent in general in estimating the right-hand-side of (4) (Atchade (2011)). More precisely, $\Gamma^2_n(h)$ converges to the random limit $\sigma^2_{\theta_\star}(h)$, instead of the asymptotic variance $\mathbb{E}\left[\sigma^2_{\theta_\star}(h)\right]$. The following example illustrates this issue.

Example 1. The example is adapted from the PhD thesis of David Hastie (2005, University of Bristol). Consider the toy density $\pi(x) = 0.5\mathbf{1}_D(x)$, where $D = [-\beta-1, -\beta] \cup [\beta, \beta+1]$, where

$\beta = 3/4$. Consider a Random Walk Metropolis (RWM) with proposal kernel $Q_\theta(x, \cdot)$ taken as the density of the uniform $\mathcal{U}(x-\theta, x+\theta)$, assuming $\theta > 2\beta$. It is well known that if θ is too large or too small the resulting RWM kernel will mix poorly. An adaptive version of this algorithm will adaptively tune θ so as to achieve an acceptance probability in stationarity of about 23%. This is a common strategy in AMCMC. It turns out that in this case the 23% acceptance probability in stationarity can be achieved at three (3) distinct solutions $\{\vartheta_1, \vartheta_2, \vartheta_3\}$, say. For the resulting AMCMC sampler, θ_n can convergence to any one of these three solutions. For this example, $\Gamma_n^2(h)$ converges to the random limit $\sigma^2_{\theta_\star}(h)$ that takes values in $\{\sigma^2_{\vartheta_1}(h), \sigma^2_{\vartheta_2}(h), \sigma^2_{\vartheta_3}(h)\}$, whereas the asymptotic variance is $p_1\sigma^2_{\vartheta_1}(h) + p_2\sigma^2_{\vartheta_2}(h) + p_3\sigma^2_{\vartheta_3}(h)$, where $p_i = \mathbb{P}(\theta_\star = \vartheta_i)$ (which depends in general on the initialization of the algorithm). Figure 1 (c) and (d) show two sample paths of $\Gamma_n^2(h)$ with very different limits. However the adaptive chain $\{X_n,\ n \geq 0\}$ remains ergodic in the sense that $\hat{\pi}_n(h)$ converges to $\pi(h)$.

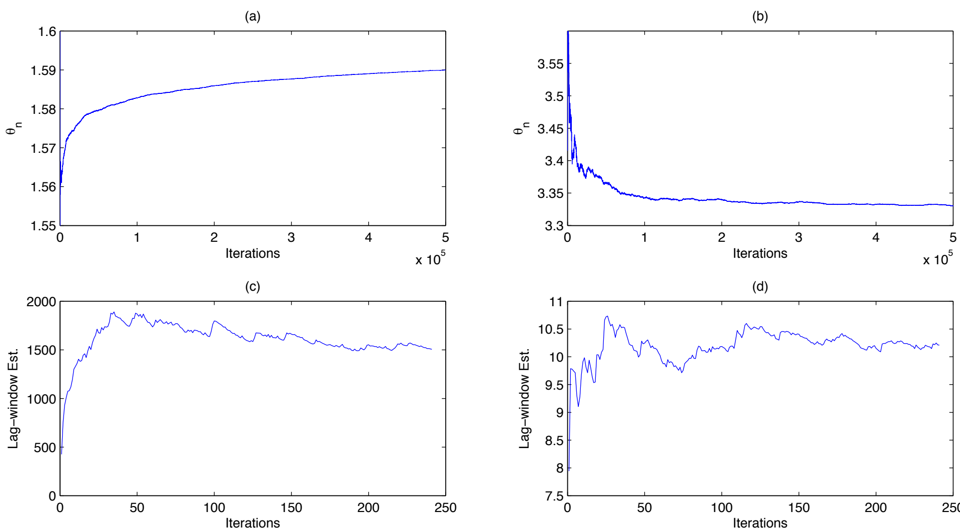

Figure 1: Two sample paths of θ_n and $\Gamma_n^2(h)$ from the toy example. Sample path 1 is (a) and (c). In both cases, $h(x) = x$.

Despite its lack of consistency in estimating the asymptotic variance, we establish in this paper that the lag-window estimators $\Gamma_n^2(h)$ can be used to derive asymptotically valid confidence interval for $\pi(h)$ in AMCMC simulation. The confidence interval is obtained by deriving the limiting distribution of the random variable

$$\mathsf{T}_n \overset{\text{def}}{=} \frac{\sqrt{n}\left(\hat{\pi}_n(h) - \pi(h)\right)}{\sqrt{\Gamma_n^2(h)}} = \frac{\frac{1}{\sqrt{n}}\sum_{j=1}^n \bar{h}(X_j)}{\sqrt{\Gamma_n^2(h)}}, \tag{5}$$

where $\bar{h} = h - \pi(h)$. The key insight of the analysis is that $n\Gamma_n^2(h)$ behaves precisely like the quadratic variation of the approximating martingale of $\sum_{k=1}^n \bar{h}(X_k)$. As a result, $\sqrt{n\Gamma_n^2(h)}$ is roughly the correct scaling statistic in the central limit theorem for $\sum_{k=1}^n \bar{h}(X_k)$, even in a general AMCMC setting.

The limiting behavior of T_n depends on the choice of c_n in computing $\Gamma_n^2(h)$. When $c_n = o(n)$, we show (Theorem 2.1) that T_n has a standard Gaussian limit. When $c_n = n$, we show (Theorem 2.2) that T_n converges in distribution to a standard Gaussian random variable scaled by an infinite sum of chi-squared. The case $c_n = n$ corresponds to the so-called fixed-b asymptotics well-known in Econometrics (Kiefer and Vogelsang (2005)). Theorems 2.1-2.2 are therefore extension to adaptive Markov chains of results that have been established for other type of stochastic models. See for instance Kiefer and Vogelsang (2005); Sun et al. (2008) for certain classes of stationary processes, and see Atchade and Cattaneo (2011) for Markov chains. These two results allow us to derive asymptotically valid confidence intervals for $\pi(h)$ in MCMC and AMCMC simulation. We compare these confidence interval procedures by simulation. We notice that the approach $c_n = o(n)$ is very sensitive to the actual choice of c_n. In contrast, the case $c_n = n$ requires no tuning (since $c_n = n$), produces slightly wider confidence intervals but with very good coverage probabilities.

The simulation results suggest that the lag-window estimator converges faster when $c_n = n$, as opposed $c_n = o(n)$. Similar conclusion has been reported elsewhere in the literature, but there are very few rigorous results on the topic. Jansson (2004) studied stationary Gaussian moving average models and established that when $c_n = n$, the rate of convergence of T_n is $n^{-1}\log(n)$. Sun et al. (2008) obtained the rate n^{-1}, under the main assumption that the underlying process is Gaussian and stationary. It is unlikely that these rates remain true without the Gaussian assumptions. We study in this paper the rate of convergence of $\Gamma_n^2(h)$ when $\{X_n,\ n \geq 0\}$ is a Markov chain. For $c_n = o(n)$, we obtain that the convergence rate of $\Gamma_n^2(h)$ toward $\sigma^2(h)$ is of order $n^{-1/3}$. But when $c_n = n$, we show that the rate of weak convergence of $\Gamma_n^2(h)$ is at least $n^{-1/2}\log(n)$.

We organize the paper as follows. Section 2 contains the main results. We illustrate the results with two simulation examples presented in Section 3. Most of the proofs are postponed to Section 5.

1.1. **Notation.** Throughout the paper, we use the notations: $\bar{f} \stackrel{\text{def}}{=} f - \pi(f)$, $\pi(f) \stackrel{\text{def}}{=} \int f(x)\pi(dx)$, $Pf(x) \stackrel{\text{def}}{=} \int f(y)P(x,dy)$, and $P^j f(x) \stackrel{\text{def}}{=} P\{P^{j-1}f\}(x)$, with $P^0 f(x) = f(x)$. For $V : \mathsf{X} \to [0,\infty)$, we define $\mathcal{L}_V$ as the space of all measurable real-valued functions $f : \mathsf{X} \to \mathbb{R}$ s.t. $\|f\|_V \stackrel{\text{def}}{=} \sup_{x\in\mathsf{X}} |f(x)|/V(x) < \infty$.

For sequences $\{a_n, b_n\}$ of real nonnegative numbers, the notation $a_n \lesssim b_n$ means that $a_n \leq cb_n$ for all n, and for some constant c that does not depend on n. For a random sequence $\{X_n\}$, we write $X_n = O_p(a_n)$ if the sequence $|X_n|/a_n$ is bounded in probability. We say that $X_n = o_p(a_n)$ is X_n/a_n converges in probability to zero as $n \to \infty$. The notation $X_n \stackrel{\mathsf{w}}{\to} X$ means that X_n converges weakly to X. If X, Y are random variables, $X \stackrel{dist.}{=} Y$ means that X and Y have the same distribution. For a random variable X, and $q \geq 1$, we use the notation $\|X\|_q \stackrel{\text{def}}{=} \mathbb{E}^{1/q}(|X|^q)$. Throughout the paper c_n denotes the cut-off point of the lag-window estimator and we assume without further mention that $c_n \uparrow \infty$, as $n \to \infty$. Also we shall use $\varrho_n^{(1)}$, $\varrho_n^{(2)}$, etc... as a generic notation for negligible random terms.

2. Consistency: statement of the results

We rely on martingale approximation and martingale theory. Throughout this section $h:\ \mathsf{X}\to\mathbb{R}$ is a fixed measurable function. For each $\theta\in\Theta$, we assume well-defined the functions g_θ and $P_\theta g_\theta$, where

$$g_\theta(x) \stackrel{\text{def}}{=} \sum_{j\geq 0} P_\theta^j \bar{h}(x), \quad \text{and} \quad P_\theta g_\theta(x) \stackrel{\text{def}}{=} \int P_\theta(x,dz) g_\theta(z),\ x\in\mathsf{X}.$$

For each $\theta\in\Theta$, the function g_θ satisfies the so-called Poisson's equation

$$g_\theta(x) - P_\theta g_\theta(x) = \bar{h}(x). \tag{6}$$

For integer $n\geq 1$, set $D_n \stackrel{\text{def}}{=} g_{\theta_{n-1}}(X_n) - P_{\theta_{n-1}} g_{\theta_{n-1}}(X_{n-1})$. For $p>1$, and integers $n\geq k\geq 1$, let

$$a_n \stackrel{\text{def}}{=} \mathbb{E}^{\frac{1}{2p}}\left(|P_{\theta_n} g_{\theta_n}(X_n)|^{2p}\right), \quad b_n \stackrel{\text{def}}{=} \mathbb{E}^{\frac{1}{2p}}\left(|P_{\theta_n} g_{\theta_n}(X_n) - P_{\theta_{n-1}} g_{\theta_{n-1}}(X_n)|^{2p}\right),$$

$$\kappa_n \stackrel{\text{def}}{=} \mathbb{E}^{\frac{1}{2p}}\left(|D_n|^{2p}\right),\ \delta_{n,k}^{(1)} \stackrel{\text{def}}{=} a_{k-1} + \sum_{j=1\vee(k-c_n+1)}^{k} b_j + \frac{1}{c_n}\sum_{j=1\vee(k-c_n+1)}^{k} a_{j-1},$$

$$\text{and} \quad \delta_{n,k}^{(2)} \stackrel{\text{def}}{=} \sqrt{\sum_{j=1\vee(k-c_n+1)}^{k} \kappa_j^2}.$$

In keeping the notations simple, we omit the dependence on p in these terms. We shall convene that if $a>b$, $\sum_a^b \cdot = 0$. The main regularity assumption is the following.

A1 *For each $\theta\in\Theta$, g_θ and $P_\theta g_\theta$ are well defined, and there exists $p>1$ such that, as $n\to\infty$,*

$$a_n + \frac{1}{c_n}\sum_{k=1}^n a_k + \sum_{k=1}^n b_k + \sqrt{\sum_{k=1}^n \kappa_k^2} \;=\; O\left(\sqrt{n}\right). \tag{7}$$

A2 *There exists a random variable $\sigma_\star^2$, positive almost surely such that*

$$\sum_{k=1}^n \kappa_k^{2p} = o(n^p), \quad \textit{and} \quad n^{-1}\sum_{k=1}^n D_k^2 \stackrel{a.s.}{\to} \sigma_\star^2,$$

as $n\to\infty$, where p is the same as in A1.

A3 *The function $w:\mathbb{R}\to[0,1]$ has support $[-1,1]$, is even and satisfies: $w(0)=1$, $w(1)=0$.*

Remark 1. We give below in Section 3.1 some drift conditions under which A1 holds. A2 depends in general on the behavior of θ_n which depends on the specific AMCMC considered. We give an example in Section 3. Assumption A3 holds for most kernels used in practice, such as the class of Bartlett kernels $w(u) = (1-|u|^q)\mathbf{1}_{(-1,1)}(u)$, for $q\geq 1$.

When $c_n = o(n)$, T_n has a Gaussian limit. To describe this case, we introduce the sequence

$$r_n = \left(\frac{1}{n^{p\wedge 2}}\sum_{k=1}^n \left\{\kappa_k\left(1+\delta_{n,k}^{(1)}\right)\right\}^{p\wedge 2}\right)^{\frac{1}{p\wedge 2}} + \frac{1}{nc_n}\sum_{k=1}^n a_k\left(a_k+\delta_{n,k}^{(1)}+\delta_{n,k}^{(2)}\right)$$
$$+\frac{1}{n}\sum_{k=1}^n b_k\left(a_k+\delta_{n,k}^{(1)}+\delta_{n,k}^{(2)}\right)+n^{-1}a_n\left(a_{n-1}+\delta_{n,n}^{(1)}+\delta_{n,n}^{(2)}\right)+n^{-1}a_0^2. \quad (8)$$

Theorem 2.1. *Assume A1-A3 and* $\lim_n n^{-1}c_n = 0$. *If* $\lim_n r_n = 0$, *and* $\lim_n n^{-p\wedge 2}\sum_{k=1}^n\left\{\kappa_k\delta_{n,k}^{(2)}\right\}^{p\wedge 2} = 0$, *then as* $n\to\infty$, $\Gamma_n^2(h)$ *converges in probability to* $\sigma_\star^2$, *and* $\mathsf{T}_n \stackrel{w}{\to} \mathsf{N}(0,1)$.

Proof. See Section 5.2. □

When $c_n = n$, the limit of T_n is a Gaussian distribution scaled by a sum of chi-squared random variables. Define the kernel $\rho_\star:\ [0,1]\times[0,1]\to\mathbb{R}$ by

$$\rho_\star(s,t) = w(t-s) - g(t) - g(s) + \int_0^1 g(u)du,$$

where $g(t) = \int_0^1 w(t-u)du$. Notice that $\rho_\star$ is symmetric: $\rho_\star(s,t) = \rho_\star(t,s)$. The kernel $\rho_\star$ induces a compact operator $\phi \mapsto (s\mapsto \int_0^1 \rho_\star(s,t)\phi(t)dt)$ on $L^2[0,1]$ that we also denote $\rho_\star$. We will assume that the kernel $\rho_\star$ is positive definite: for all $n\geq 1$, all $a_1,\ldots,a_n\in\mathbb{R}$, and $t_1,\ldots,t_n\in[0,1]$, $\sum_{i=1}^n\sum_{j=1}^n a_ia_j\rho_\star(t_i,t_j)\geq 0$. The positive definiteness assumption of the kernel $\rho_\star$ would imply that the operator $\rho_\star$ has nonnegative eigenvalues. In which case we will denote $\{\alpha_j, j\in\mathsf{I}\}$ the (countable) set of positive eigenvalues of $\rho_\star$ (each repeated according to its multiplicity), which we assume non-empty to avoid trivialities.

Theorem 2.2. *Assume A1-A3 and suppose that* $\rho_\star$ *is positive definite. If* $c_n = n$, *and* $\lim_n r_n = 0$ *then*

$$\mathsf{T}_n \stackrel{w}{\to} \frac{Z_0}{\sqrt{\sum_{i\in\mathsf{I}}\alpha_i Z_i^2}}, \quad (9)$$

where $\{Z_0, Z_i,\ i\in\mathsf{I}\}$ *are i.i.d.* $\boldsymbol{N}(0,1)$, *and* $\{\alpha_i,\ i\in\mathsf{I}\}$ *is the set of positive eigenvalues of* $\rho_\star$.

Proof. See Section 5.3. □

It is not very convenient to work with the random variable $Z_0/\sqrt{\sum_{i\in\mathsf{I}}\alpha_i Z_i^2}$, because the eigenvalues of $\rho_\star$ are difficult to find in general. The next result gives an alternative representation of the distribution of $Z_0/\sqrt{\sum_{i\in\mathsf{I}}\alpha_i Z_i^2}$ that is more amenable to simulation. The proof is straightforward.

Proposition 2.3. *Let* $\{B(t),\ 0\leq t\leq 1\}$ *denotes the standard Brownian motion. Set*

$$\chi^2 \stackrel{\text{def}}{=} 1 - \int_0^1 g(t)dt + 2\int_0^1\left[\int_0^t \rho_\star(s,t)dB(s)\right]dB(t).$$

Then

$$\frac{B(1)}{\sqrt{\chi^2}} \overset{dist.}{=} \frac{Z_0}{\sqrt{\sum_{i\in I} \alpha_i Z_i^2}}, \tag{10}$$

where $\{Z_0, Z_i,\ i \in I\}$ *are i.i.d.* $\boldsymbol{N}(0,1)$*, and* $\{\alpha_i,\ i \in I\}$ *is the set of positive eigenvalues of* $\rho_\star$*.*

Remark 2. In fact, $\chi^2 = \sum_{i\in I} \alpha_i Z_i^2$ has many equivalent representations. It can be written as a double Ito-Wiener integral or a double Wiener integral. More precisely

$$\chi^2 = 1 - \int_0^1 g(t)dt + \underbrace{\int_0^1 \int_0^1 \rho_\star(s,t)dB(t)dB(s)}_{\text{double Ito-Wiener integral}} = \underbrace{\int_0^1 \int_0^1 \rho_\star(s,t)dB(t)dB(s)}_{\text{double Wiener integral}}.$$

The difference being that the double Ito-Wiener integral excludes the diagonal $\int_0^1 \rho_\star(t,t)dt = 1 - \int_0^1 g(t)dt$. We prefer the representation given in Proposition 2.3 as a standard (iterated) stochastic integral. In fact, it can be easily shown that χ^2 also has a representation as a double Wiener integral of w wrt to the Brownian bridge $\{\bar{B}(t),\ 0 \le t \le 1\}$:

$$\chi^2 = \underbrace{\int_0^1 \int_0^1 w(s-t)d\bar{B}(t)d\bar{B}(t)}_{\text{double Wiener integral}}.$$

Remark 3. Theorem 2.2 requires the positivity of $\rho_\star$, whereas Theorem 2.1 does not require any positivity assumption. When w turns out to be positive (in the sense that the kernel $k(s,t) = w(t-s)$ is positive definite), then $\rho_\star$ is also positive. This is easy to show and we omit details. This result applies for example to the Bartlett kernel given by $w(u) = (1-|u|)\mathbf{1}_{(-1,1)}(u)$. This function is the characteristic function of the distribution with density $(1-\cos(x))/\pi x^2$, $x \in \mathbb{R}$, and by Bochner's theorem $(s,t) \mapsto w(t-s)$ is positive definite on $[0,1]$.

Theorems 2.1-2.2 yield two asymptotically valid confidence procedures for $\pi(h)$. From Theorem 2.1, we can form the classical confidence interval

$$\hat{\pi}_n(h) \pm z_{1-\alpha/2}\sqrt{\frac{\Gamma_n^2(h)}{n}}, \tag{11}$$

where $\Gamma_n^2(h)$ is computed using $c_n = o(n)$, and $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. We can also use Theorem 2.2 to propose the fixed-b confidence interval

$$\hat{\pi}_n(h) \pm t_{1-\alpha/2}\sqrt{\frac{\Gamma_n^2(h)}{n}}, \tag{12}$$

where $\Gamma_n^2(h)$ is computed using $c_n = n$, and $t_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the distribution of $Z_0/\sqrt{\sum_{i\in I} \alpha_i Z_i^2}$. The quantiles t_α are intractable in general. But using Proposition 2.3, we have $Z_0/\sqrt{\sum_{i\in I} \alpha_i Z_i^2} \overset{dist.}{=} B(1)/\sqrt{\chi^2}$, so that these quantiles can be obtained by simulating the Brownian motion and approximating the iterated Ito integral $\int_0^1 \int_0^t \rho_\star(s,t)dB(s)dB(t)$. For illustration, we consider the following two kernels

(1) The Bartlett kernel given by $w(u) = (1 - |u|)\mathbf{1}_{(0,1)}(u)$. As pointed out in Remark (3), $\rho_\star$ in this case is known to be positive definite and is given by

$$\rho_\star(s,t) = \frac{2}{3} - s(1-s) - t(1-t) - |s-t|.$$

(2) We also consider the kernel $w(u) = (1-u^2)\mathbf{1}_{(0,1)}(u)$, for which

$$\rho_\star(s,t) = 2(s-0.5)(t-0.5).$$

Thus obviously, $\rho_\star$ is positive definite. In fact, in this case $\chi^2 = Z^2/6$, where $Z \sim \mathbf{N}(0,1)$. In Figure 2, we plot the cdfs of $Z_0/\sqrt{\sum_{i\in\mathsf{I}} \alpha_i Z_i^2}$ for these two kernels in comparison with the standard Gaussian cdf. We also give some quantiles in Table 1.

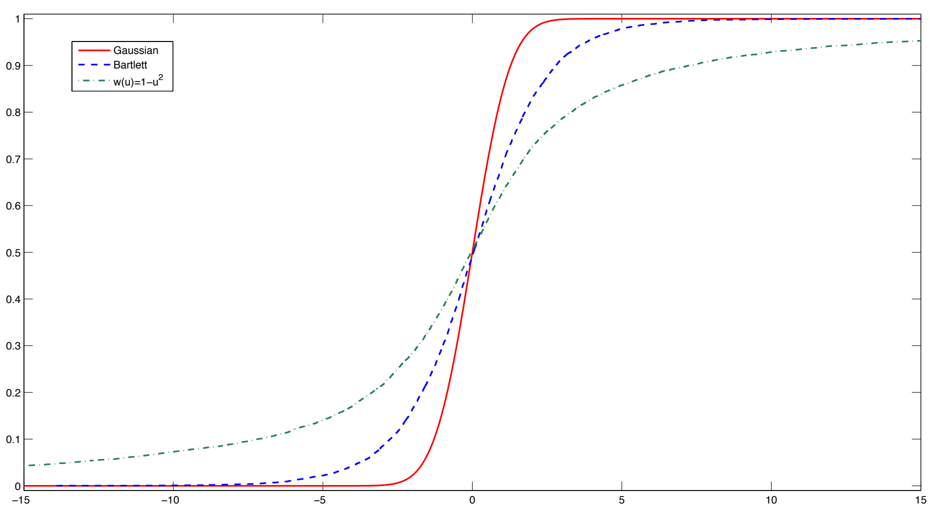

Figure 2: CDF of $Z_0/\sqrt{\sum_{i\in\mathsf{I}} \alpha_i Z_i^2}$. The standard normal CDF is given as a reference.

	$\alpha = 10\%$	$\alpha = 5\%$
$w(u) = 1 - \lvert u\rvert$	3.796	4.784
$w(u) = 1 - u^2$	15.590	31.520

TABLE 1. The table reports t such that $\mathbb{P}(T > t) = \alpha/2$, where $T = Z_0/\sqrt{\sum_{i\in\mathsf{I}} \alpha_i Z_i^2}$.

3. EXAMPLES

3.1. **Application to adaptive Markov chains with geometric drift conditions.** We will now illustrate how the assumptions stated above can be checked using drift conditions. We consider the following assumptions.

B1 *For each $\theta \in \Theta$, P_θ has invariant distribution π. Uniformly for $\theta \in \Theta$, there exist a measurable function $V : \mathsf{X} \to [1, \infty)$, $\mathcal{C} \in \mathcal{B}$, ν a probability measure on $(\mathsf{X}, \mathcal{B})$, $b, \epsilon > 0$ and $\lambda \in (0,1)$ such that $\nu(\mathcal{C}) > 0$, $P_\theta(x, \cdot) \geq \epsilon\nu(\cdot)\mathbb{1}_{\mathcal{C}}(x)$ and*

$$P_\theta V \leq \lambda V + b\mathbb{1}_{\mathcal{C}}. \tag{13}$$

B2 *There exists $\eta \in [0, 1/2)$, positive $\gamma_n \downarrow 0$, with $\gamma_n = O\left(n^{-\alpha}\right)$, $\alpha \geq 1/2$, and a finite constant c such that for all $n \geq 1$, all $\beta \in (0, 1-\eta)$, and all $f \in \mathcal{L}_{V^\beta}$, with $|f|_{V^\beta} \leq 1$,*

$$\left|P_{\theta_n} f(X_n) - P_{\theta_{n-1}} f(X_n)\right| \leq c\gamma_n V^{\beta+\eta}(X_n), \quad \mathbb{P} - a.s. \tag{14}$$

Remark 4. B1 is the well known geometric drift condition. In general these drift conditions are difficult to check on specific examples, but there are known to hold for a number of target probability distributions and algorithms. On the other hand, B2 is the so-called diminishing adaptation condition. This condition is in general easier to check and is known to hold for the Random Walk Metropolis (RMW) (Andrieu and Moulines (2006)) and the Metropolis adjusted Langevin algorithm (Atchade and Fort (2012)). Finally, we point out that in the case of a standard Markov chain, B2 trivially holds.

In the present context Theorem 2.1-2.2 can be transposed as follows.

Theorem 3.1. *Assume A3, B1-B2, and take $h \in \mathcal{L}_{V^\delta}$, for $\delta \in [0, 1/2 - \eta)$. Suppose that there exists a random variable $\theta_\star$ with $\sigma^2_{P_{\theta_\star}}(h)$ positive almost surely, such that $\theta_n \overset{a.s.}{\to} \theta_\star$, as $n \to \infty$. Set $\sigma^2_\star \overset{\text{def}}{=} \sigma^2_{P_{\theta_\star}}(h)$, and assume that $\sqrt{n} \lesssim c_n$.*

(1) If $\sqrt{c_n} = o\left(n^{1-\frac{1}{p\wedge 2}}\right)$, as $n \to \infty$, then $\Gamma^2_n(h)$ converges in probability to $\sigma^2_\star$, and $\mathsf{T}_n \overset{w}{\to} \boldsymbol{N}(0,1)$.

(2) If $c_n = n$, $\rho_\star$ is positive definite, and $\alpha > \frac{1}{p\wedge 2}$, then (9) holds.

Remark 5. The assumption that θ_n converges almost surely to a limit depends on the specific adaptive algorithm under consideration. Many adaptive algorithms rely on stochastic approximation. In this case, conditions under which θ_n converges can be found for instance in Andrieu and Moulines (2006); Atchade and Fort (2012) and the references therein. In practice, a simple plot of the sample path of θ_n (or some function of it) can give a good indication whether the assumption hold.

Proof of Theorem 3.1. The real number p in A1 can be taken as $p = \frac{1}{2(\delta+\eta)}$, and we notice that $p > 1$, and $2p(\delta + \eta) = 1$. It is also well known that under B1 g_θ and $P_\theta g_\theta$ are well-defined, $\sup_{\theta\in\Theta} |g_\theta|_{V^\beta} + |P_\theta g_\theta|_{V^\beta} < \infty$, and

$$\sup_{n\geq 0} \mathbb{E}\left(V(X_n)\right) < \infty. \tag{15}$$

These results can be found for instance in Andrieu and Moulines (2006). This implies that

$$\kappa_n \lesssim 1, \quad \text{and} \quad a_n \lesssim 1.$$

(14) and (15) imply that $b_n \lesssim \gamma_n$. We deduce that

$$a_n + c_n^{-1}\sum_{k=1}^n a_k + \sum_{k=1}^n b_k + \sqrt{\sum_{k=1}^n \kappa_k^2} = O\left(nc_n^{-1} + \sum_{k=1}^n \gamma_k + \sqrt{n}\right) = O\left(nc_n^{-1} + \sqrt{n}\right) = O(\sqrt{n}),$$

by assumption. Thus A1 holds. Recall that $D_n = g_{\theta_{n-1}}(X_n) - P_{\theta_{n-1}} g_{\theta_{n-1}}(X_{n-1})$, and $\sigma_\theta^2(h) = \int \pi(dx)\int P_\theta(x,dz)\left(g_\theta(z) - P_\theta g_\theta(x)\right)^2$. Therefore, by the law of large numbers, we have the almost sure convergence to zero of the sequence $n^{-1}\sum_{k=1}^n \left(D_k^2 - \sigma^2_{\theta_{k-1}}(h)\right)$ (see e.g. Atchade (2011) Proposition 3.3 for a proof). It follows that $n^{-1}\sum_{k=1}^n D_k^2$ converges almost surely to $\sigma_\star^2(h)$. Thus clearly A2 holds.

We also check that $\delta_{n,k}^{(2)} \lesssim \sqrt{c_n}$, and $\delta_{n,k}^{(1)} \lesssim 1 + \sum_{k=1}^{c_n} b_j \lesssim \sum_{k=1}^{c_n} j^{-\alpha} \lesssim 1 + c_n^{1-\alpha}$. It follows that

$$r_n = \frac{1 + c_n^{1-\alpha}}{n^\alpha} + \frac{1}{\sqrt{c_n}} + \frac{1 + c_n^{1-\alpha}}{n^{1-\frac{1}{p\wedge 2}}}. \tag{16}$$

(1) The conditions $c_n = o\left(n^{2(1-\frac{1}{p\wedge 2})}\right)$, and $\alpha \geq 1/2$ imply that $\lim_n n^{-p\wedge 2}\sum_{k=1}^n \left\{\kappa_k \delta_{n,k}^{(2)}\right\}^{p\wedge 2} = 0$ and $\lim_n r_n = 0$. Therefore, the conclusions of Theorem 2.1 hold.
(2) If $c_n = n$, the condition $\alpha > \frac{1}{p\wedge 2}$ implies that $\lim_n r_n = 0$, and the conclusions of Theorem 2.2 hold.

□

3.2. **A logistic regression example.** We assume that

$$y_i \sim \mathcal{B}\left(p(x_i'\beta)\right), \quad i = 1, \ldots, n,$$

where $y_i \in \{0,1\}$, $x_i \in \mathbb{R}^d$ is a vector of covariate, and $\beta \in \mathbb{R}^d$ is the vector of parameter. $\mathcal{B}(p)$ denotes the Bernoulli distribution with parameter $p \in (0,1)$, and $p(x) = \frac{e^x}{1+e^x}$ is the cdf of the logistic distribution. Assume a Gaussian prior $\mathbf{N}(0, s^2 I_d)$ for β, with $s = 20$. The posterior distribution of β then becomes

$$\pi\left(\beta|X\right) \propto e^{\ell(\beta|X)} e^{-\frac{1}{2s^2}|\beta|^2}.$$

To illustrate the ideas above, we will consider two commonly used algorithms to sample from π: a plain Random Walk Metropolis (RWM) with Gaussian proposal and the adaptive version of the same algorithm presented in Atchade and Fort (2010) (Algorithm 3.1). This algorithm adaptively and simultaneously estimates the covariance matrix of the target distribution and implements the 0.23 acceptance rule. It is known that for this problem, both algorithms satisfy B1-2 (see e.g. Atchade (2011) Section 5.2).

As a simulation example I test the model with the *Heart dataset* which has $n = 217$ cases and $d = 14$ covariates. I first run the adaptive chain for 10^6 iterations and takes the sample posterior mean of β as the true posterior mean. I repeat the confidence intervals (95% confidence intervals) $K = 200$ times to estimate coverage probability and half-length. Each sampler is run for $30,000$ iterations. The result is summarized in Figure 3. For the case $c_n = o(n)$, I use $c_n = n^\delta$ for different values of $\delta \in (0,1)$.

We see from the results that using $c_n = n$ gives very good coverage, but slightly wider intervals. The interval width is significantly wider for the quadratic kernel $w(u) = 1 - u^2$, which is somewhat expected given the very fat tail of the limiting distribution (Figure 2). In contrast, the result show that in the setting $c_n = o(n)$ careful tuning of c_n is necessary to obtain good coverage. As expected, the results in this case are similar for both kernels.

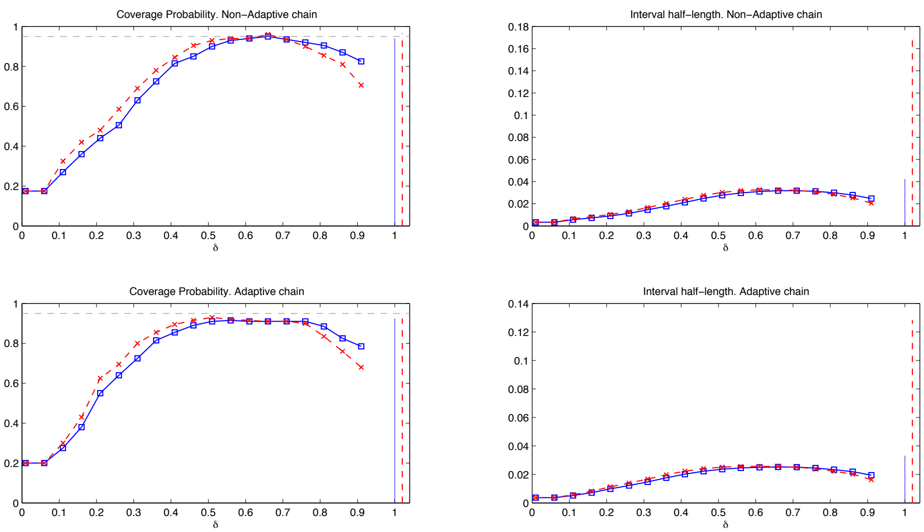

Figure 3: Coverage probability and confidence interval half-length for parameter β_3 and for different values of δ using $c_n = n^\delta$. The vertical lines correspond to $c_n = n$. The dashed lines correspond to the kernel $w(u) = 1 - u^2$.

3.3. **A random effect Poisson regression example.** We now consider a random effect Poisson regression example taken from Gelman et al. (2004). For $e = 1, \ldots, N_e$ and $p = 1, \ldots, N_p$, the variables y_{ep} are conditionally independent given $(\{\beta_p\}, \{\varepsilon_{ep}\}) \in \mathbb{R}^{N_p} \times \mathbb{R}^{N_e \times N_p}$, with conditional distribution

$$y_{ep} \sim \mathcal{P}\left(n_{ep} e^{\mu + \alpha_e + \beta_p + \varepsilon_{ep}}\right), \quad e = 1, \ldots, N_e, \quad p = 1, \ldots, N_p,$$

where $\mathcal{P}(\lambda)$ is the Poisson distribution with parameter λ. In the above display, $\{n_{ep}\}$ is a deterministic baseline covariate, and $\mu \in \mathbb{R}$, $\{\alpha_e\} \in \mathbb{R}^{N_e}$ are parameters. We assume that the random effects $\{\beta_p\}$ and $\{\varepsilon_{ep}\}$ are independent with prior distributions

$$\beta_p \overset{iid}{\sim} N(0, \sigma_\beta^2), \quad \varepsilon_{ep} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2), \quad e = 1, \ldots, N_e, \quad p = 1, \ldots, N_p,$$

for some parameters $\sigma_\beta^2 > 0$, $\sigma_\varepsilon^2 > 0$. We assume a diffuse prior for $(\mu, \alpha, \sigma_\beta^2, \sigma_\varepsilon^2)$ $(\sigma_\epsilon^2 > 0, \sigma_\beta^2 > 0)$. For identifiability, we assume that $\sum_{k=1}^{N_e} \alpha_k = 0$. Let $\theta = (\mu, \alpha, \beta, \epsilon, \sigma_\epsilon^2, \sigma_\beta^2) \in \mathbb{R}^{3 + N_e - 1 + (N_p+1)N_e}$.

The posterior distribution of θ given $\mathcal{D} = (y_{ep}, n_{ep})$ takes the form

$$\pi(\theta|\mathcal{D}) \propto \exp\left\{\sum_{e,p} y_{e,p}(\mu + \alpha_e + \beta_p + \epsilon_{e,p}) - n_{ep}e^{\mu+\alpha_e+\beta_p+\epsilon_{ep}} \right.$$
$$\left. -\frac{N_e N_p}{2}\log\sigma_\epsilon^2 - \frac{N_p}{2}\log\sigma_\beta^2 - \frac{1}{2\sigma_\epsilon^2}\sum_{e,p}\epsilon_{e,p}^2 - \frac{1}{2\sigma_\beta^2}\sum_{p=1}^{N_p}\beta_p^2\right\}.$$

This posterior distribution is typical of probability distributions for which AMCMC are useful. A possible MCMC strategy to sample from this posterior is a Metropolis-within-Gibbs. One can update σ_ε^2 and σ_b^2 exactly as inverse-Gamma $\mathsf{IG}(0.5(3N_p - 2), 0.5\sum_{e,p}\epsilon_{e,p}^2)$ and $\mathsf{IG}(0.5(N_p - 2), 0.5\sum_{p=1}^{N_p}\beta_p^2)$ respectively. The parameter μ can be updated exactly as the log of the Gamma distribution $\mathsf{G}(\sum_{e,p} y_{e,p}, \sum_{e,p} n_{e,p}e^{\alpha_e+\beta_p+\epsilon_{e,p}})$. The rest of the parameter α_1, α_2, β_p and $\epsilon_{e,p}$ can be updated one at the time, using one step of a RWM with a Gaussian proposal $\mathcal{N}(x, \sigma^2)$ with $\sigma = e^{-2}$. One can compare this Metropolis-within-Gibbs sampler with its adaptive version where the scaling parameters σ of the RWM steps are adaptively tuned using the 23% acceptance rule. It is unknown whether these algorithms satisfies the assumptions above.

I set $N_e = 3$ and $N_p = 27$ (thus $\theta \in \mathbb{R}^{89}$). I generate an artificial dataset with $(\alpha_1, \alpha_2, \mu, \sigma_\varepsilon^2, \sigma_\beta^2) = (0.35, 0.15, -1.0, 0.1, 0.3)$, and run a preliminary MCMC sampler for 2 millions (2×10^6) iterations and compute its sample mean. This gives $\bar{\alpha}_1 = 0.3948$ that I take as $\int \alpha_1 \pi(\theta|\mathcal{D})d\theta$. We wish to construct 95% confidence intervals for α_1. I run each algorithm for $60,000$ iterations and discard the first $10,000$ iterations as burn-in. This is repeated $K = 200$ times to estimate the properties of the confidence intervals. The asymptotic variance are estimated using only the Bartlett kernel. The results are reported in Figure 4 and yield similar conclusions as the previous example.

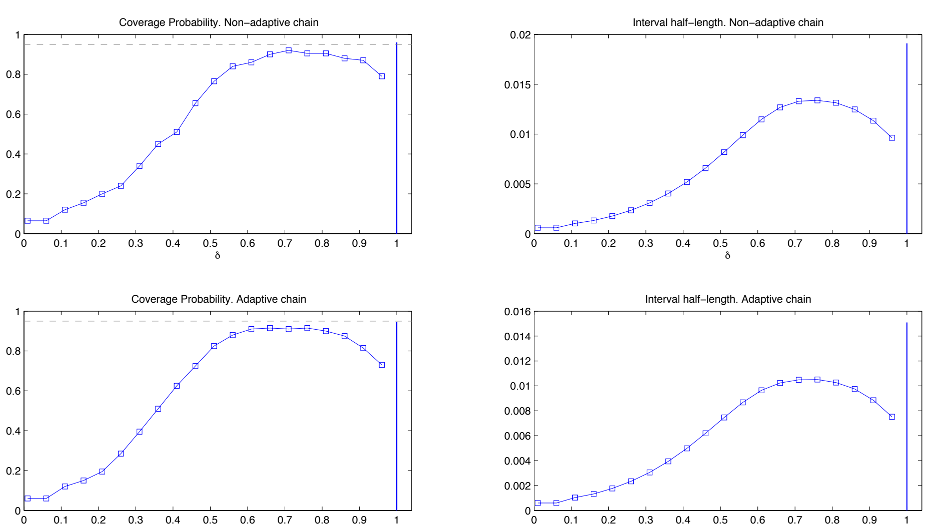

Figure 4: Coverage probability and confidence interval half-length for parameter α_1 and for different values of δ using $c_n = n^\delta$. The vertical lines correspond to $c_n = n$.

4. Rate of convergence

The simulation results presented above suggest that $\Gamma_n^2(h)$ has better convergence properties when $c_n = n$. We consider this issue here in the case where $\{X_n,\ n \geq 0\}$ is a Markov chain with transition kernel P and invariant distribution π, so that the asymptotic variance $\sigma^2(h)$, is non-random, and given by (2). The initial distribution of the chain is arbitrary. We assume that $\sigma_P^2(h) > 0$ and without any loss of generality, we take $\pi(h) = 0$ and $\sigma_P^2(h) = 1$; otherwise, simply replace h by $(h - \pi(h))/\sqrt{\sigma_P^2(h)}$. We further simplify the analysis by assuming that P satisfies the following geometric ergodicity assumption

GE *There exists a measurable function* $V : \mathsf{X} \to [1, \infty)$ *such that*

$$\sup_{n\geq 0} \mathbb{E}\left(V(X_n)\right) < \infty, \tag{17}$$

and for all $\beta \in (0, 1]$,

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{V^\beta} \leq C\rho^n\ V^\beta(x), \quad n \geq 0,\ x \in \mathsf{X}. \tag{18}$$

When $c_n = o(n)$, we have the following.

Theorem 4.1. *Suppose that A3 and (GE) hold. Let* $\delta \in [0, 1/4)$, *and* $h \in \mathcal{L}_{V^\delta}$. *As* $n \to \infty$, *if* $c_n = o(n)$, *then*

$$\mathbb{E}^{1/2}\left[\left(\Gamma_n^2(h) - 1\right)^2\right] \lesssim \frac{1}{c_n} + \sqrt{\frac{c_n}{n}}. \tag{19}$$

Proof. See Section 5.5. □

Remark 6. Theorem 4.1 implies that the rate of convergence of $\Gamma_n^2(h)$ is of order $n^{-1/3}$, using $c_n = n^{1/3}$. This rate is known to be tight for kernels with characteristic exponent 1. The characteristic exponent of w is the largest number $r \geq 1$ such that $\lim_{x\to 0} |x|^{-r}(1 - w(x)) \in (0, \infty)$. Our analysis does not make use of this concept. If w has characteristic exponent r, it is known (see e.g. Parzen (1957) Theorem 5A-5B) that the rate of convergence of $\Gamma_n^2(h)$ is $\frac{1}{c_n^r} + \sqrt{\frac{c_n}{n}}$, using $c_n = n^{(1+2r)^{-1}}$, for certain classes of stationary processes. The Bartlett kernel has characteristic exponent 1, and the kernel $w(u) = 1 - u^2$ has characteristic exponent 2.

We also consider the rate of weak convergence of $\Gamma_n^2(h)$ towards the limiting distribution of Theorem 2.2, when $c_n = n$. Denote $\mathsf{Lip}_1(\mathbb{R})$ the set of all bounded Lipschitz functions $f : \mathbb{R} \to \mathbb{R}$ such that

$$|f|_{\mathrm{Lip}} \stackrel{\mathrm{def}}{=} \sup_{x\neq y} \frac{|f(x) - f(y)|}{|x - y|} \leq 1.$$

For P, Q two probability measures on $\mathbb{R}$, we define

$$\mathsf{d}_1(P, Q) \stackrel{\mathrm{def}}{=} \sup_{f\in \mathsf{Lip}_1(\mathbb{R})} \left| \int f dP - \int f dQ \right|.$$

$\mathsf{d}_1(P,Q)$ is the Wasserstein metric between P,Q. It is well known (see e.g. Dudley (2002) Section 11.8, Problem 1) that in the case of $\mathbb{R}$, d_1 can be written as

$$\mathsf{d}_1(P,Q)=\int_{-\infty}^{\infty}|P(A_u)-Q(A_u)|\,du,\quad A_u=(-\infty,u].$$

Thus an upper bound on $\mathsf{d}_1(P_n,P)$ gives a Berry-Esseen-type bound on the rate of weak convergence of P_n to P. In a slight abuse of notation, if X,Y are random variables, and $X\sim P$ and $Y\sim Q$, we shall also write $\mathsf{d}_1(X,Y)$ to mean $\mathsf{d}_1(P,Q)$.

Theorem 4.2. *Suppose that A3 and (GE) hold. Suppose also that I is finite. Let $\delta\in[0,1/4)$, and $h\in\mathcal{L}_{V^\delta}$. If $c_n=n$, then*

$$\mathsf{d}_1\left(\Gamma_n^2(h),\chi^2\right)\lesssim\frac{\log(n)}{\sqrt{n}},\quad \text{as } n\to\infty,\tag{20}$$

where $\chi^2=\sum_{i\in\mathsf{I}}\alpha_i Z_i^2$, $\{Z_i,\ i\in\mathsf{I}\}$ are i.i.d. $\boldsymbol{N}(0,1)$, and $\{\alpha_i,\ i\in\mathsf{I}\}$ is the set of positive eigenvalues of $\rho_\star$.

Proof. See Section 5.5. □

For the proof we use the Bergstrom method, well known in studying convergence rates in the CLT for partial sums (see e.g. Dedecker and Rio (2008)). The $\log(n)$ term in (20) is an artefact of the method. We conjecture that in general the convergence rate of $\Gamma_n^2(h)$ is $n^{-1/2}$. If we further assume that $\{h(X_n),\ n\geq 0\}$ is a Gaussian process with a martingale structure, then the convergence rate in (20) can actually be improved to $\log(n)/n$; we omit the details. Quadratic forms has also been studied elsewhere in the literature. In a series of papers, F. Gotze and co-authors have studied the convergence rate of quadratic forms and obtained the optimal rate of n^{-1} (see e.g. Götze and Tikhomirov (2005) and references therein). But their setting is different as they assume i.i.d. sequence and consider quadratic forms for which the weights do not depend with n.

Remark 7. The assumption that I is finite is mostly technical and it seems plausible that this result continues to hold without this assumption. There are known kernels for which I is finite. For example I is finite for the kernel $w(u)=(1-u^2)\mathbf{1}_{(-1,1)}(u)$. This is because in this case, $\rho_\star(s,t)=2\left(s-\frac{1}{2}\right)\left(t-\frac{1}{2}\right)$. Thus $\rho_\star$ admits a unique positive eigenvalue $\alpha_1=1/6$ with eigenfunction $\phi_1(t)=t-\frac{1}{2}$.

5. Proofs

5.1. **Martingale approximation.** Much of the analysis relies on the ability to approximate a partial sum of the form $\sum_{k=1}^n\alpha_k\bar{h}(X_k)$ by the martingale $\sum_{k=1}^n\alpha_k D_k$. This is well known and we skip some of the details and refer the reader for instance to Andrieu and Moulines (2006). It is easy to see from property (1) of the adaptive chain that $\mathbb{E}\left(D_k|\mathcal{F}_{k-1}\right)=0$. Therefore, under A1, $\{(D_k,\mathcal{F}_k),\ k\geq 1\}$ is a $2p$-integrable martingale-difference. Such martingale satisfy Burkeholder's inequality (Hall and Heyde (1980) Theorem 2.10) that we will use repetitively: for any sequence of real numbers $\{\alpha_k,\ 1 \leq k \leq n\}$, and for any $q > 1$,

$$\left\|\sum_{k=1}^{n} \alpha_k D_k\right\|_q \lesssim \left[\sum_{k=1}^{n} \|\alpha_k D_k\|_q^{q\wedge 2}\right]^{\frac{1}{q\wedge 2}}. \tag{21}$$

The following martingale approximation for partial sums plays an important role in the sequel.

Lemma 5.1. *Under A1, and for any sequence of real numbers $\{\alpha_k,\ 0 \leq k \leq n\}$,*

$$\sum_{k=1}^{n} \alpha_k \bar{h}(X_k) = \sum_{k=1}^{n} \alpha_k D_k + \epsilon_n^{(0)}, \tag{22}$$

where the remainder is given by

$$\epsilon_n^{(0)} \stackrel{\text{def}}{=} \alpha_0 P_{\theta_0} g_{\theta_0}(X_0) - \alpha_n P_{\theta_n} g_{\theta_n}(X_n) + \sum_{k=1}^{n} (\alpha_k - \alpha_{k-1}) P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1}) \\ + \sum_{k=1}^{n} \alpha_k \left(P_{\theta_k} g_{\theta_k}(X_k) - P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k)\right),$$

and satisfies

$$\left\|\epsilon_n^{(0)}\right\|_{2p} \leq |\alpha_0| a_0 + |\alpha_n| a_n + \sum_{k=1}^{n} |\alpha_k - \alpha_{k-1}| a_{k-1} + \sum_{k=1}^{n} |\alpha_k| b_k.$$

Combined with (21), this lemma implies that

$$\left\|\sum_{k=1}^{n} \alpha_k \bar{h}(X_k)\right\|_{2p} \lesssim 1 + |\alpha_n| a_n + \sum_{k=1}^{n} |\alpha_k - \alpha_{k-1}| a_{k-1} + \sum_{k=1}^{n} |\alpha_k| b_k + \sqrt{\sum_{k=1}^{n} \alpha_k^2 \kappa_k^2}. \tag{23}$$

We now show that a similar martingale approximation holds for quadratic forms. This extends Lemma 2.1 of Atchade and Cattaneo (2011) which considered the case where $\{X_n,\ n \geq 0\}$ is a Markov chain. The proof is postponed to the Appendix.

Lemma 5.2. *Assume A1 and A3. Consider the quadratic form*

$$Q_n \stackrel{\text{def}}{=} \frac{1}{n} \sum_{k=1}^{n} \sum_{j=1}^{n} w\left(\frac{k-j}{c_n}\right) \bar{h}(X_k) \bar{h}(X_j).$$

Then we have

$$Q_n = \frac{1}{n} \sum_{k=1}^{n} \sum_{j=1}^{n} w\left(\frac{k-j}{c_n}\right) D_k D_j + \epsilon_n^{(1)} + \epsilon_n^{(2)},$$

for remainders $\epsilon_n^{(1)}, \epsilon_n^{(2)}$ for which $\mathbb{E}^{1/p}(|\epsilon_n^{(i)}|^p) \lesssim r_n^{(i)}$, where p is as in A1, and

$$r_n^{(1)} = \frac{1}{nc_n} \sum_{k=1}^{n} a_k \left(a_k + \delta_{n,k}^{(1)} + \delta_{n,k}^{(2)}\right), \quad \textit{and}$$

$$r_n^{(2)} = \left(\frac{1}{n^{p\wedge 2}} \sum_{k=1}^{n} \left\{\kappa_k \left(1 + \delta_{n,k}^{(1)}\right)\right\}^{p\wedge 2}\right)^{\frac{1}{p\wedge 2}} + \frac{1}{n} \sum_{k=1}^{n} b_k \left(a_k + a_{k-1}\right) + \frac{1}{nc_n} \sum_{k=1}^{n} a_k \kappa_k + n^{-1} a_n \kappa_n.$$

Proof. See Section 6.3 in the Appendix. □

5.2. **Proof of Theorem 2.1.** The idea is to show that $\Gamma_n^2(h)$ behaves asymptotically like $n^{-1}\sum_{k=1}^n D_k^2$. And since the partial sum $n^{-1/2}\sum_{k=1}^n \bar{h}(X_k)$ behaves like $n^{-1/2}\sum_{k=1}^n D_k$ as shown in Lemma 5.1, it would follow that T_n behaves asymptotically like $\sum_{k=1}^n D_k/\sqrt{\sum_{k=1}^n D_k^2}$ which satisfies a CLT as recalled in Theorem 6.1 of the Appendix.

As a matter of re-arranging the summations, we can rewrite $\Gamma_n^2(h)$ as follows

$$\Gamma_n^2(h) = \frac{1}{n}\sum_{k=1}^n\sum_{j=1}^n w\left(\frac{k-j}{c_n}\right)\bar{h}(X_j)\bar{h}(X_k) - \frac{2}{n}\left(\sum_{k=1}^n \bar{h}(X_k)\right)\left(\sum_{k=1}^n v_n(k)\bar{h}(X_k)\right) + \frac{u_n}{n}\left(\sum_{k=1}^n \bar{h}(X_k)\right)^2 \quad (24)$$

where $v_n(k) \stackrel{\text{def}}{=} n^{-1}\sum_{\ell=1}^n w\left(\frac{k-\ell}{c_n}\right)$, and $u_n \stackrel{\text{def}}{=} n^{-2}\sum_{\ell=1}^n\sum_{k=1}^n w((\ell-k)/c_n)$. Under A3, it is easy to check that $v_n(k)$ and u_n satisfy

$$v_n(k) = O\left(\frac{c_n}{n}\right), \quad v_n(k) - v_n(k-1) = O\left(\frac{1}{n}\right), \quad \text{and} \quad u_n = O\left(\frac{c_n}{n}\right), \quad \text{(uniformly in } k\text{)}.$$

Therefore, using (23) and (7), we can write (24) as

$$\Gamma_n^2(h) = \frac{1}{n}\sum_{k=1}^n\sum_{j=1}^n w\left(\frac{k-j}{c_n}\right)\bar{h}(X_j)\bar{h}(X_k) + \varrho_n^{(1)}, \quad (25)$$

where $\mathbb{E}^{1/p}(|\varrho_n^{(1)}|^p) \lesssim c_n/n$. Combined with Lemma 5.2, it follows that

$$\Gamma_n^2(h) = \frac{1}{n}\sum_{k=1}^n D_k^2 + \frac{2}{n}\sum_{k=2}^n D_k\sum_{j=1}^{k-1} w\left(\frac{k-j}{c_n}\right)D_j + \varrho_n^{(2)}, \quad (26)$$

where $\mathbb{E}^{1/p}(|\varrho_n^{(2)}|^p) \lesssim r_n^{(1)} + r_n^{(2)} + c_n/n$. The term $\sum_{k=2}^n D_k\sum_{j=1}^{k-1} w((k-j)/c_n)D_j$ is a martingale array and Burkeholder inequality (21) (applied twice) yields

$$\left\|\frac{1}{n}\sum_{k=2}^n\sum_{j=1}^{k-1}\left(\frac{k-j}{c_n}\right)D_kD_j\right\|_p \lesssim \left\{n^{-p\wedge 2}\sum_{k=2}^n\left[\kappa_k\delta_{n,k}^{(2)}\right]^{p\wedge 2}\right\}^{\frac{1}{p\wedge 2}}. \quad (27)$$

We then obtain that

$$\Gamma_n^2(h) = \frac{1}{n}\sum_{k=1}^n D_k^2 + \varrho_n^{(3)}, \quad (28)$$

where

$$\mathbb{E}^{1/p}\left(|\varrho_n^{(3)}|^p\right) \lesssim r_n^{(1)} + r_n^{(2)} + \frac{c_n}{n} + \left\{n^{-p\wedge 2}\sum_{k=2}^n\left[\kappa_k\delta_{n,k}^{(2)}\right]^{p\wedge 2}\right\}^{\frac{1}{p\wedge 2}} = o(1),$$

by assumption. A2 implies that $n^{-1}\sum_{j=1}^n D_j^2$ converges almost surely to $\sigma_\star^2$, and we conclude that $\Gamma_n^2(h)$ converges in probability to $\sigma_\star^2$. It remains to deal with T_n. We have

$$\mathsf{T}_n = \frac{\frac{1}{\sqrt{n}}\sum_{k=1}^n \bar{h}(X_k)}{\sqrt{\Gamma_n^2(h)}} = \frac{\frac{\sum_{k=1}^n D_k}{\sqrt{\sum_{k=1}^n D_k^2}} + \frac{o_p(1)}{\sqrt{\frac{1}{n}\sum_{k=1}^n D_k^2}}}{\sqrt{1+\frac{o_p(1)}{\frac{1}{n}\sum_{k=1}^n D_k^2}}}.$$

Under A2, $n^{-1}\sum_{k=1}^n D_k^2$ converges in probability to $\sigma_\star^2$ that is positive almost surely, and by the martingale weak invariance principle (see Theorem 6.1 and the following remark), we have that $\sum_{k=1}^n D_k/\sqrt{\sum_{k=1}^n D_k^2} \stackrel{\mathsf{w}}{\to} \mathbf{N}(0,1)$. It follows that $\mathsf{T}_n \stackrel{\mathsf{w}}{\to} \mathbf{N}(0,1)$, and the theorem is proved.

5.3. **Proof of Theorem 2.2.** The idea of the proof is that for large n, and for $c_n = n$, $\Gamma_n^2(h)$ behaves like $\sum_{i\in\mathsf{I}}\alpha_i\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n \phi_j\left(\frac{k}{n}\right)D_k\right)^2$, and that $\frac{1}{\sqrt{n}}\sum_{k=1}^n \phi_j\left(\frac{k}{n}\right)D_k$ behaves like $\int_0^1 \phi_j(t)dB(t) \sim \mathbf{N}(0,1)$. To carry the details, we start again from (24) which gives

$$\Gamma_n^2(h) = \frac{1}{n}\sum_{k=1}^n\sum_{j=1}^n w\left(\frac{k-j}{n}\right)\bar{h}(X_j)\bar{h}(X_k) - \frac{2}{n}\left(\sum_{k=1}^n \bar{h}(X_k)\right)\left(\sum_{k=1}^n v_n(k)\bar{h}(X_k)\right) + \frac{u_n}{n}\left(\sum_{k=1}^n \bar{h}(X_k)\right)^2.$$

With $c_n = n$, $v_n(k) = n^{-1}\sum_{\ell=1}^n w\left(\frac{k-\ell}{n}\right)$ is the right-Riemann sum approximation of $g(kn^{-1})$, where $g(t) = \int_0^1 w(t-u)du$. Thus with the smoothness of w,

$$\left|v_n(k) - g(kn^{-1})\right| = O\left(\frac{1}{n}\right), \quad \left|v_n(k) - g(kn^{-1}) - v_n(k-1) + g((k-1)n^{-1})\right| = O\left(\frac{1}{n^2}\right),$$
$$\text{and} \quad \left|u_n - \int_0^1 g(t)dt\right| = O\left(\frac{1}{n}\right), \quad \text{(uniformly in } k\text{)}.$$

By combining this with the linear and quadratic martingale approximation (Lemma 5.1, Lemma 5.2), we obtain that

$$\Gamma_n^2(h) = \frac{1}{n}\sum_{k=1}^n\sum_{j=1}^n\left\{w\left(\frac{k-j}{n}\right) - g\left(\frac{k}{n}\right) - g\left(\frac{j}{n}\right) + \int_0^1 g(t)dt\right\}D_jD_k + \varrho_n^{(1)}$$
$$= \frac{1}{n}\sum_{k=1}^n\sum_{j=1}^n \rho_\star\left(\frac{k}{n},\frac{j}{n}\right)D_jD_k + \varrho_n^{(1)}, \quad (29)$$

where

$$\mathbb{E}^{1/p}\left(|\varrho_n^{(1)}|^p\right) \lesssim r_n^{(1)} + r_n^{(2)} + \frac{1}{n}.$$

It is assumed that the kernel $\rho_\star : [0,1]\times[0,1] \to \mathbb{R}$ with $\rho_\star(s,t) = w(s-t) - g(t) - g(s) + \int_0^1 g(u)du$ is twice continuously differentiable and positive definite. By Mercer's theorem, there exist positive eigenvalues $\{\alpha_j,\ j\in\mathsf{I}\}$ and orthonormal eigenfunctions $\{\phi_j,\ j\in\mathsf{I}\}$ $\phi_j \in L^2([0,1])$ such that

$$\rho_\star(s,t) = \sum_{j\in\mathsf{I}}\alpha_j\phi_j(s)\phi_j(t). \quad (30)$$

Notice that $\int_0^1 \rho_\star(s,t)dt = 0$ which means that 0 is also eigenvalue of $\rho_\star$ with associated eigenfunction ϕ_0 ($\phi_0(t) = 1$). Using the expansion (30), we conclude that

$$\Gamma_n^2(h) = \sum_{j\in\mathsf{I}} \alpha_j \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n \phi_j\left(\frac{k}{n}\right) D_k\right)^2 + \varrho_n^{(1)},$$

where $\varrho_n^{(1)}$ is as in (29). It follows that

$$\mathsf{T}_n = \frac{\frac{1}{\sqrt{n}}\sum_{k=1}^n \bar{h}(X_k)}{\sqrt{\Gamma_n^2(h)}} = \frac{\sum_{k=1}^n \bar{h}(X_k)}{\sqrt{n\Gamma_n^2(h)}} = \frac{\frac{\sum_{k=1}^n \phi_0\left(\frac{k}{n}\right) D_k}{\sqrt{\sum_{k=1}^n D_k^2}} + \frac{o_p(1)}{\sqrt{\frac{1}{n}\sum_{k=1}^n D_k^2}}}{\sqrt{\sum_{j\in\mathsf{I}} \alpha_j \left(\frac{\sum_{k=1}^n \phi_j\left(\frac{k}{n}\right) D_k}{\sqrt{\sum_{k=1}^n D_k^2}}\right)^2 + \frac{o_p(1)}{\frac{1}{n}\sum_{k=1}^n D_k^2}}}.$$

We shall write $\bar{\mathsf{I}} \stackrel{\text{def}}{=} \{0\} \cup \mathsf{I}$. Consider the Hilbert space $\ell^2(\alpha) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^{\bar{\mathsf{I}}} :\ \sum_{k\in\bar{\mathsf{I}}} \alpha_k x_k^2 < \infty\}$, where $\alpha_0 = 1$, equipped with the norm $\|x\|_2 = \sqrt{\sum_{j\in\bar{\mathsf{I}}} \alpha_j x_j^2}$ and the inner product $\langle x, y\rangle \stackrel{\text{def}}{=} \sum_{j\in\bar{\mathsf{I}}} \alpha_j x_j y_j$. The random variable $\Psi_n \stackrel{\text{def}}{=} \left\{\frac{\sum_{k=1}^n \phi_j\left(\frac{k}{n}\right) D_k}{\sqrt{\sum_{k=1}^n D_k^2}},\ j \in \bar{\mathsf{I}}\right\}$ is a $\ell^2(\alpha)$-valued random process. We prove in Lemma 5.3 below that $\Psi_n \stackrel{\mathsf{w}}{\to} \Psi$, where

$$\Psi \stackrel{\text{def}}{=} \left\{\int_0^1 \phi_j(t) dB(t),\ j \in \bar{\mathsf{I}}\right\},$$

and $\{B(t),\ 0 \le t \le 1\}$ is the standard Brownian motion. The theorem follows from the continuous mapping theorem, since $\{\int_0^1 \phi_j(t)dB(t),\ j \in \bar{\mathsf{I}}\}$ are i.i.d. $\mathbf{N}(0,1)$.

Lemma 5.3. *Under the assumptions of Theorem 2.2, $\Psi_n \stackrel{\mathsf{w}}{\to} \Psi$ as $n \to \infty$.*

Proof. We set the notation $S_k = \sum_{j=1}^k D_j$ ($S_0 = 0$), $U_k^2 = \sum_{j=1}^k D_j^2$, ($U_0 = 0$), and $s_k = \sum_{j=1}^k \mathbb{E}(D_j^2)$ ($s_0 = 0$). Consider $\zeta_n :\ [0,1] \to \mathbb{R}$, the $C[0,1]$-valued process obtained by interpolating the points $\left(\frac{s_0}{s_n}, \frac{S_0}{U_n}\right), \ldots, \left(\frac{s_k}{s_n}, \frac{S_k}{U_n}\right), \ldots, \left(1, \frac{S_n}{U_n}\right)$. That is

$$\zeta_n(t) = \frac{D_k}{U_n}\frac{(s_n t - s_{k-1})}{s_k - s_{k-1}} + \frac{S_{k-1}}{U_n}, \quad \text{for} \quad \frac{s_{k-1}}{s_n} \le t < \frac{s_k}{s_n}.$$

Then we have

$$\int_0^1 \phi_j(t) d\zeta_n(t) = \sum_{k=1}^n \underbrace{\left(\frac{s_n}{s_k - s_{k-1}} \int_{\frac{s_{k-1}}{s_n}}^{\frac{s_k}{s_n}} \phi_j(t)dt\right)}_{\delta_{n,k}(j)} \frac{D_k}{\sqrt{\sum_{k=1}^n D_k^2}} = \sum_{k=1}^n \delta_{n,k}(j) \frac{D_k}{\sqrt{\sum_{k=1}^n D_k^2}}.$$

Now, conveniently, we introduce the integration map $\mathcal{M} :\ C[0,1] \to \ell^2(\alpha)$ defined as $\mathcal{M}(h) = \{\int_0^1 \phi_j(t)dh(t),\ j \in \bar{\mathsf{I}}\}$. These integrals are well-defined since the functions ϕ_j are continuously differentiable (Theorem 6.2 (ii)). Furthermore, the inegration by part formula gives $\int_0^1 \phi_j(t)dh(t) =$

$\phi_j(1)h(1) - \phi_j(0)h(0) - \int_0^1 h(t)\phi_j'(t)dt$. Therefore, for $h, h_0 \in C[0,1]$,

$$\|\mathcal{M}(h) - \mathcal{M}(h_0)\|_2^2 = \sum_{j\in\bar{\mathrm{I}}} \alpha_j \left|\int_0^1 \phi_j dh - \int_0^1 \phi_j dh_0\right|^2$$
$$\leq 4\|h-h_0\|_\infty^2 \sup_{0\leq t\leq 1} \sum_{j\in\bar{\mathrm{I}}} \alpha_j|\phi_j(t)|^2 + 2\|h-h_0\|_\infty^2 \sum_{j\in\bar{\mathrm{I}}} \alpha_j \int_0^1 \{\phi_j'\}^2(t)dt \leq c_0\|h-h_0\|_\infty^2,$$

where the last inequality uses (36-37). This establishes that $\mathcal{M}$ in fact takes values in $\ell^2(\alpha)$ and is Lipschitz. Now, it is clear that we can write

$$\Psi_n = \mathcal{M}(\zeta_n) + \epsilon_n.$$

where the j-th component of ϵ_n is

$$\epsilon_n(j) = \sum_{k=1}^n \left(\phi_j\left(\frac{k}{n}\right) - \delta_{n,k}(j)\right)\frac{D_k}{U_n}.$$

With A1-A2, we have the weak convergence of ζ_n towards the standard Brownian motion (see Remark 8), and by the continuous mapping theorem, $\mathcal{M}(\zeta_n) \to \Psi$. The lemma is proved by showing that ϵ_n converges in probability to zero in $\ell^2(\alpha)$.

`Negligibility of` ϵ_n.

$$\|\epsilon_n\|_2 = \sum_{j\in\bar{\mathrm{I}}} \alpha_j\epsilon_n^2(j) = \frac{\sum_{j\in\bar{\mathrm{I}}} \alpha_j \left\{\sum_{k=1}^n \left(\delta_{n,k}(j) - \phi_j\left(\frac{k}{n}\right)\right)\frac{D_k}{\sqrt{n}}\right\}^2}{\frac{1}{n}\sum_{k=1}^n D_k^2}.$$

Since $n^{-1}\sum_{k=1}^n D_k^2$ has a positive limit almost surely, it is enough to show that the numerator converges to zero in probability. Towards that end, we have

$$\mathbb{E}\left(\sum_{j\in\bar{\mathrm{I}}} \alpha_j \left\{\sum_{k=1}^n \left(\delta_{n,k}(j) - \phi_j\left(\frac{k}{n}\right)\right)\frac{D_k}{\sqrt{n}}\right\}^2\right) = \frac{1}{n}\sum_{k=1}^n \mathbb{E}(D_k^2)\sum_{j\in\bar{\mathrm{I}}} \alpha_j\left(\delta_{n,k}(j) - \phi_j\left(\frac{k}{n}\right)\right)^2.$$

For any arbitrary continuously differentiable function $f:\ [0,1]\to\mathbb{R}$, $w\in[0,1]$, and $0\leq a<b\leq 1$, the following bound holds true:

$$\left(\frac{1}{b-a}\int_a^b f(t)dt - f(w)\right)^2 \leq (b-a)\int_0^1 \{f'(u)\}^2 du.$$

With $\delta_{n,k}(j) = \frac{s_n}{s_k - s_{k-1}}\int_{\frac{s_{k-1}}{s_n}}^{\frac{s_k}{s_n}} \phi_j(t)dt$, we use the above inequality to write

$$\sum_{j\in\bar{\mathrm{I}}} \alpha_j\left(\delta_{n,k}(j) - \phi_j\left(\frac{k}{n}\right)\right)^2 \leq \frac{s_k - s_{k-1}}{s_n}\sum_{j\in\bar{\mathrm{I}}} \alpha_j\int_0^1 \{\phi_j'(t)\}^2 dt \lesssim \frac{s_k - s_{k-1}}{s_n}.$$

We conclude that

$$\mathbb{E}\left(\sum_{j\in\bar{\mathrm{I}}} \alpha_j \left\{\sum_{k=1}^n \left(\delta_{n,k}(j) - \phi_j\left(\frac{k}{n}\right)\right)\frac{D_k}{\sqrt{n}}\right\}^2\right) \lesssim \frac{\sup_{1\leq k\leq n}\mathbb{E}(D_k^2)}{n} = o(1),$$

as $n \to \infty$, since $\sum_{k=1}^n \mathbb{E}(|D_k|^{2p}) = o(n^p)$. This completes the proof.

□

5.4. **Proof theorem 4.1.** The idea is to use the decomposition (25), together with Lemma 5.2 and a more careful bound on the term $\epsilon_n^{(1)}$ in Lemma 5.2. The main ingredient of the proof is again martingale approximation. We recall that $\sigma^2(h) = 1$ and $\pi(h) = 0$. Since P no longer depend on θ, we write g instead g_θ, Pg instead $P_\theta g_\theta$ etc... We gather from (25) in the proof of Theorem 2.1, and Lemma 5.2, that

$$\Gamma_n^2(h) - 1 = \frac{1}{n}\sum_{k=1}^n \left(D_k^2 - 1\right) + \frac{2}{n}\sum_{k=2}^n D_k \sum_{j=1}^{k-1} w\left(\frac{k-j}{c_n}\right) D_j + \epsilon_n^{(1)} + \varrho_n^{(1)},$$

where

$$\epsilon_n^{(1)} = \sum_{k=3}^n Pg(X_{k-1}) \sum_{j=1}^{k-2} \left(\tilde{w}_n(k-j) - \tilde{w}_n(k-1-j)\right) h(X_j),$$

and

$$\mathbb{E}^{1/2}(|\varrho_n^{(1)}|^2) \lesssim r_n^{(2)} + \frac{c_n}{n} \lesssim \frac{1}{\sqrt{n}} + \frac{c_n}{n}.$$

Using (27), if follows that $\Gamma_n^2(h) - 1 = \frac{1}{n}\sum_{k=1}^n \left(D_k^2 - 1\right) + \epsilon_n^{(1)} + \varrho_n^{(2)}$, where $\mathbb{E}^{1/2}(|\varrho_n^{(2)}|^2) \lesssim \frac{1}{\sqrt{n}} + \frac{c_n}{n} + \sqrt{\frac{c_n}{n}} \lesssim \sqrt{\frac{c_n}{n}}$.

We know from Lemma 5.2 that $\mathbb{E}^{1/2}(|\epsilon_n^{(1)}|^2) \lesssim r_n^{(1)} \lesssim \frac{1}{\sqrt{c_n}}$. But under the current assumptions, it is possible to obtain a better bound. We further use the Poisson equation approach to write $Pg(x) = U(x) - PU(x)$, where $U(x) \stackrel{\text{def}}{=} \sum_{j\geq 0} P^{j+1}g(x)$. By the assumption (GE), and since $g \in \mathcal{L}_{V^\delta}$, U is well defined. Set $B_{n,k-2} \stackrel{\text{def}}{=} \sum_{j=1}^{k-2}\left(\tilde{w}_n(k-j) - \tilde{w}_n(k-1-j)\right) h(X_j)$, and $\delta_{n,j} \stackrel{\text{def}}{=} \tilde{w}_n(k-j) - \tilde{w}_n(k-1-j)$. By A3, and using the mean value theorem, we get $|\delta_{n,j}| \lesssim \frac{1}{nc_n}$, and $|\delta_{n,j} - \delta_{n,j-1}| \leq \frac{1}{nc_n^2}$, uniformly in j. Then using (23), it comes that

$$\mathbb{E}^{1/4}\left(|B_{n,k-2}|^4\right) \lesssim \frac{1}{n\sqrt{c_n}}, \quad \text{and} \quad \mathbb{E}^{1/4}\left(|B_{n,k-2} - B_{n,k-3}|^4\right) \lesssim \frac{1}{nc_n}.$$

The second inequality follows from the bounds on $|\delta_{n,j}|$ and $|\delta_{n,j} - \delta_{n,j-1}|$, and the fact that

$$B_{n,k-2} - B_{n,k-3} = \left(\tilde{w}_n(2) - \tilde{w}_n(1)\right) h(X_{k-2}) + \sum_{j=1}^{k-3}\left(\delta_{n,j} - \delta_{n,j-1}\right) h(X_j).$$

Now, from $Pg(x) = U(x) - PU(x)$, we get

$$\begin{aligned}\epsilon_n^{(1)} = \sum_{k=3}^n \left(U(X_{k-1}) - PU(X_{k-1})\right) B_{n,k-2} &= \sum_{k=3}^n \left(U(X_{k-1}) - PU(X_{k-2})\right) B_{n,k-2} \\ &+ \sum_{k=3}^n PU(X_{k-2})\left(B_{n,k-2} - B_{n,k-3}\right) - PU(X_{n-1})B_{n,n-2}.\end{aligned}$$

Noticing that $\sum_{k=3}^n \left(U(X_{k-1}) - PU(X_{k-2})\right) B_{n,k-2}$ is a martingale array, we deduce easily from the above that

$$\mathbb{E}^{1/2}\left(|\epsilon_n^{(1)}|^2\right) \lesssim \frac{1}{c_n}.$$

Under assumption (GE), and since $\delta \in (0, 1/4)$, we have $\mathbb{E}^{1/2}\left[\left(\sum_{k=1}^n \left(D_k^2 - 1\right)\right)^2\right] \lesssim \sqrt{n}$. We conclude that

$$\mathbb{E}^{1/2}\left[\left(\Gamma_n^2(h) - 1\right)^2\right] \lesssim \frac{1}{\sqrt{n}} + \frac{1}{c_n} + \sqrt{\frac{c_n}{n}} \lesssim \frac{1}{c_n} + \sqrt{\frac{c_n}{n}},$$

which ends the proof.

5.5. **Proof theorem 4.2.** We define

$$\bar{\Gamma}_n^2 = \frac{1}{n}\sum_{k=1}^n \sum_{j=1}^n \rho_\star\left(\frac{k}{n}, \frac{j}{n}\right) D_k D_j.$$

We recall from (29) that

$$\Gamma_n^2(h) = \bar{\Gamma}_n^2 + \varrho_n^{(1)},$$

where $\mathbb{E}^{1/2}\left(|\varrho_n^{(1)}|^2\right) \lesssim r_n^{(1)} + r_n^{(2)} + n^{-1} \lesssim n^{-1/2}$ (see the proof of Theorem 3.1 for the bound on $r_n^{(1)} + r_n^{(2)}$). This implies that

$$\mathsf{d}_1\left(\Gamma_n^2(h), \chi^2\right) \lesssim \mathsf{d}_1\left(\bar{\Gamma}_n^2, \chi^2\right) + \frac{1}{\sqrt{n}}. \tag{31}$$

Therefore we only need to focus on the term $\mathsf{d}_1\left(\bar{\Gamma}_n^2, \chi^2\right)$.

On the Euclidean space $\mathbb{R}^{\mathsf{I}}$, we shall use the norms $\|x\|_\alpha^2 = \sum_{i\in\mathsf{I}} \alpha_i x_i^2$, $\|x\|^2 = \sum_{i\in\mathsf{I}} x_i^2$ and the inner-products $\langle x, y\rangle_\alpha = \sum_{i\in\mathsf{I}} \alpha_i x_i y_i$, and $\langle x, y\rangle = \sum_{i\in\mathsf{I}} x_i y_i$. For a sequence $(a_1, a_2, \ldots)$, we use the notation $a_{i:k} = (a_i, \ldots, a_k)$ (and $a_{i:k}$ is the empty set if $i > k$). We introduce new random variables $\{Z_{i,j},\ i \in \mathsf{I},\ 1 \le j \le n\}$ which are i.i.d. $\mathbf{N}(0, 1)$, and set $\mathbf{S}_{\ell:k} \stackrel{\text{def}}{=} \left(\sum_{j=\ell}^k Z_{1j}, \ldots, \sum_{j=\ell}^k Z_{\mathsf{I}j}\right)^\mathsf{T} \in \mathbb{R}^{\mathsf{I}}$, so that

$$\chi^2 \stackrel{dist.}{=} \sum_{i\in\mathsf{I}} \alpha_i \left(\frac{1}{\sqrt{n}}\sum_{j=1}^n Z_{i,j}\right)^2 = \|\frac{1}{\sqrt{n}}\mathbf{S}_{1:n}\|_\alpha^2.$$

For $1 \le \ell \le k \le n$, and omitting the dependence on n, we set $\mathbf{B}_{\ell:k}$ as the $\mathbb{R}^{\mathsf{I}\times(k-\ell+1)}$ matrix

$$\mathbf{B}_{\ell:k}(i, j) = \phi_i\left(\frac{j}{n}\right),\ i \in \mathsf{I},\ \ell \le j \le k.$$

By the Mercer's expansion for $\rho_\star$, we have

$$\bar{\Gamma}_n^2 = \sum_{i\in\mathsf{I}} \alpha_i \left(\frac{1}{\sqrt{n}}\sum_{k=1}^n \phi_i\left(\frac{k}{n}\right) D_k\right)^2 = \|\frac{1}{\sqrt{n}}\mathbf{B}_{1:n} D_{1:n}\|_\alpha^2.$$

For $f \in \mathsf{Lip}_1(\mathbb{R})$, we introduce the function $f_\alpha :\ \mathbb{R}^{|\mathsf{I}|} \to \mathbb{R}$, defined as $f_\alpha(x) = f\left(\|x\|_\alpha^2\right)$. As a matter of telescoping the sums, we have

$$\begin{aligned}
\mathbb{E}\left[f(\bar{\Gamma}_n^2) - f(\chi^2)\right] &= \mathbb{E}\left[f_\alpha\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:n} D_{1:n}\right) - f_\alpha\left(\frac{1}{\sqrt{n}}\mathbf{S}_{1:n}\right)\right] \\
&= \sum_{\ell=1}^n \mathbb{E}\left[f_\alpha\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell} D_{1:\ell} + \frac{1}{\sqrt{n}}\mathbf{S}_{\ell+1:n}\right) - f_\alpha\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1} D_{1:\ell-1} + \frac{1}{\sqrt{n}}\mathbf{S}_{\ell:n}\right)\right] \\
&= \sum_{\ell=1}^n \mathbb{E}\left[f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1} D_{1:\ell-1} + \frac{1}{\sqrt{n}}\mathbf{B}_{\ell:\ell} D_\ell\right) - f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1} D_{1:\ell-1} + \frac{1}{\sqrt{n}}\mathbf{S}_{\ell:\ell}\right)\right],
\end{aligned}$$

where we define

$$f_{\alpha,n,\ell}(x) \stackrel{\text{def}}{=} \mathbb{E}\left[f_\alpha\left(x+\frac{1}{\sqrt{n}}\mathbf{S}_{\ell:n}\right)\right], \quad \text{and set} \quad f_{\alpha,n,n+1}(x) = f_\alpha(x).$$

We deal with the case $\ell = n$ separately. Indeed, it is easy to check using the Lipschitz property of f that

$$\left|\mathbb{E}\left[f_\alpha\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:n-1}D_{1:n-1}+\frac{1}{\sqrt{n}}\mathbf{B}_{n:n}D_n\right)-f_\alpha\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:n-1}D_{1:n-1}+\frac{1}{\sqrt{n}}\mathbf{S}_{n:n}\right)\right]\right|$$
$$\leq \mathbb{E}\left[\left|\|\frac{1}{\sqrt{n}}\mathbf{B}_{1:n-1}D_{1:n-1}+\frac{1}{\sqrt{n}}\mathbf{B}_{n:n}D_n\|_\alpha^2-\|\frac{1}{\sqrt{n}}\mathbf{B}_{1:n-1}D_{1:n-1}+\frac{1}{\sqrt{n}}\mathbf{S}_{n:n}\|_\alpha^2\right|\right] \lesssim \frac{1}{\sqrt{n}}.$$

For the rest of the proof, we assume $1 \leq \ell \leq n-1$. First, we claim that $f_{\alpha,n,\ell}$ is differentiable everywhere on $\mathbb{R}^{\mathsf{I}}$. To prove this, it suffices to obtain the almost everywhere differentiability of $z \in \mathbb{R}^{\mathsf{I}} \mapsto f_\alpha(x+z)$ for any $x \in \mathbb{R}^{\mathsf{I}}$. By Rademacher's theorem, f as a Lipschitz function is differentiable almost everywhere on $\mathbb{R}$. If E is the set of points where f is not differentiable, we conclude that f_α is differentiable at all points $z \notin \{z \in \mathbb{R}^{\mathsf{I}} :\ \|x+z\|_\alpha^2 \in E\}$. Now by Ponomarëv (1987) Theorem 2, the Lebesgue measure of the set $\{z \in \mathbb{R}^{\mathsf{I}} :\ \|x+z\|_\alpha^2 \in E\}$ is zero, which proves the claim.

As a result, the function $x \mapsto f_{\alpha,n,\ell}(x)$ is differentiable with derivative

$$\nabla f_{\alpha,n,\ell}(x)\cdot h = 2\mathbb{E}\left[f_\alpha'\left(x+\frac{1}{\sqrt{n}}\mathbf{S}_{\ell:n}\right)\left\langle x+\frac{1}{\sqrt{n}}\mathbf{S}_{\ell:n}, h\right\rangle_\alpha\right].$$

By writing his expectation wrt the distribution of $x+\frac{1}{\sqrt{n}}\mathbf{S}_{\ell:n}$, we get

$$\nabla f_{\alpha,n,\ell}(x)\cdot h = 2\int f_\alpha'(z)\left\langle z,h\right\rangle_\alpha \exp\left(-\frac{n}{2(n-\ell+1)}\left(\|x\|^2-2\left\langle x,z\right\rangle\right)\right)\mu_{n,\ell}(dz),$$

where $\mu_{n,\ell}$ is the distribution of $\frac{1}{\sqrt{n}}\mathbf{S}_{\ell:n}$. This implies that $f_{\alpha,n,\ell}$ is infinitely differentiable with second and third derivatives given by

$$\begin{aligned}
&\nabla^{(2)} f_{\alpha,n,\ell}(x)\cdot(h_1,h_2)\\
&= -2\left(\frac{n}{n-\ell+1}\right)\int f_\alpha'(z)\left\langle z,h_1\right\rangle_\alpha\left\langle x-z,h_2\right\rangle \exp\left(-\frac{n}{2(n-\ell+1)}\left(\|x\|^2-2\left\langle x,z\right\rangle\right)\right)\mu_{n,\ell}(dz)\\
&\qquad = 2\mathbb{E}\left[f_\alpha'\left(x+\frac{1}{\sqrt{n}}\mathbf{S}_{\ell:n}\right)\left\langle x\sqrt{\frac{n}{n-\ell+1}}+\frac{\mathbf{S}_{\ell:n}}{\sqrt{n-\ell+1}}, h_1\right\rangle_\alpha\left\langle\frac{\mathbf{S}_{\ell:n}}{\sqrt{n-\ell+1}}, h_2\right\rangle\right],
\end{aligned}$$

which implies after soe easy calculations that for $h \in \mathbb{R}^{\mathsf{I}}$,

$$\left|\nabla^{(2)} f_{\alpha,n,\ell}(x)\cdot(h,h)\right| \lesssim \|h\|^2\left(1+\sqrt{\frac{n}{n-\ell+1}}\|x\|_\alpha\right). \tag{32}$$

Similarly,

$$\begin{aligned}
\nabla^{(3)} f_{\alpha,n,\ell}(x)\cdot(h_1,h_2,h_3) = 2\sqrt{\frac{n}{n-\ell+1}}\mathbb{E}\Bigg[f_\alpha'\left(x+\frac{1}{\sqrt{n}}\mathbf{S}_{\ell:n}\right)\left\langle x\sqrt{\frac{n}{n-\ell+1}}+\frac{\mathbf{S}_{\ell:n}}{\sqrt{n-\ell+1}}, h_1\right\rangle_\alpha\\
\times\left(\left\langle h_2,h_3\right\rangle-\left\langle\frac{\mathbf{S}_{\ell:n}}{\sqrt{n-\ell+1}}, h_2\right\rangle\left\langle\frac{\mathbf{S}_{\ell:n}}{\sqrt{n-\ell+1}}, h_3\right\rangle\right)\Bigg].
\end{aligned}$$

and for $h \in \mathbb{R}^{\mathsf{I}}$,

$$\left|\nabla^{(3)} f_{\alpha,n,\ell}(x)\cdot(h,h,h)\right| \lesssim \sqrt{\frac{n}{n-\ell+1}}\|h\|^3\left(1+\sqrt{\frac{n}{n-\ell+1}}\|x\|_\alpha\right). \tag{33}$$

Then by Taylor expansion we have

$$\begin{aligned} f_{\alpha,n,\ell+1}&\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+\frac{1}{\sqrt{n}}\mathbf{B}_{\ell:\ell}D_\ell\right) - f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+\frac{1}{\sqrt{n}}\mathbf{S}_{\ell:\ell}\right)\\ &= \frac{1}{\sqrt{n}}\nabla f_{\alpha,n,\ell+1}(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1})\cdot(\mathbf{B}_{\ell:\ell}D_\ell-\mathbf{S}_{\ell:\ell})\\ &+\frac{1}{2n}\nabla^{(2)} f_{\alpha,n,\ell+1}(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1})\cdot[(\mathbf{B}_{\ell:\ell}D_\ell,\mathbf{B}_{\ell:\ell}D_\ell)-(\mathbf{S}_{\ell:\ell},\mathbf{S}_{\ell:\ell})]+\varrho^{(3)}_{n,\ell}, \end{aligned}$$

where, using (33),

$$\left|\varrho^{(3)}_{n,\ell}\right| \lesssim \sqrt{\frac{n}{n-\ell+1}}n^{-3/2}\left(1+\sqrt{\frac{\ell-1}{n-\ell+1}}\left\|\frac{\mathbf{B}_{1:\ell-1}D_{1:\ell-1}}{\sqrt{\ell-1}}\right\|_\alpha\right)\left(\|\mathbf{B}_{\ell:\ell}D_\ell\|_\alpha^3+\|\mathbf{S}_{\ell:\ell}\|_\alpha^3\right).$$

It follows that

$$\sum_{\ell=1}^{n-1}\mathbb{E}\left(\left|\varrho^{(3)}_{n,\ell}\right|\right) \lesssim n^{-1}\sum_{\ell=1}^{n}\frac{1}{\sqrt{\ell}}+n^{-1/2}\sum_{\ell=1}^{n}\frac{1}{\ell}\lesssim n^{-1/2}\log(n). \tag{34}$$

By first conditioning on $\mathcal{F}_{\ell-1}$, we have

$$\mathbb{E}\left[\nabla f_{\alpha,n,\ell+1}(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1})\cdot(\mathbf{B}_{\ell:\ell}D_\ell-\mathbf{S}_{\ell:\ell})\right]=0.$$

Writing $K_{n,\ell}\stackrel{\text{def}}{=}\frac{1}{2}\nabla^{(2)} f_{\alpha,n,\ell}(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1})$, we have

$$\begin{aligned} \nabla^{(2)} &f_{\alpha,n,\ell+1}(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1})\cdot[(\mathbf{B}_{\ell:\ell}D_\ell,\mathbf{B}_{\ell:\ell}D_\ell)-(\mathbf{S}_{\ell:\ell},\mathbf{S}_{\ell:\ell})]\\ &= D_\ell^2\sum_{i,j}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)K_{n,\ell}(i,j)-\sum_{i,j}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)K_{n,\ell}(i,j)Z_{i,\ell}Z_{j\ell} \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E}&\left(\nabla^{(2)} f_{\alpha,n,\ell+1}(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1})\cdot[(\mathbf{B}_{\ell:\ell}D_\ell,\mathbf{B}_{\ell:\ell}D_\ell)-(\mathbf{S}_{\ell:\ell},\mathbf{S}_{\ell:\ell})]\,|\mathcal{F}_{\ell-1}\right)=\\ &\sum_{i,j}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)K_{n,\ell+1}(i,j)\left[\mathbb{E}\left(D_\ell^2|\mathcal{F}_{\ell-1}\right)-\delta_{ij}\right]\\ &=\sum_{i,j}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)K_{n,\ell+1}(i,j)\left[\mathbb{E}\left(D_\ell^2|\mathcal{F}_{\ell-1}\right)-1\right]+\sum_{i\neq j}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)K_{n,\ell+1}(i,j), \end{aligned}$$

where $\delta_{ij}=1$ if $i=j$ and zero otherwise. We claim that the proof will be finished if we show that for all $i,j\in\mathsf{I}$, and $1\le\ell\le n$,

$$\mathbb{E}^{1/2}\left[(K_{n,\ell}(i,j)-K_{n,\ell+1}(i,j))^2\right]\lesssim\frac{\sqrt{n}}{n-\ell+1}. \tag{35}$$

To prove this claim, it suffice to use (35) to show that $\left|n^{-1}\sum_{\ell=1}^n \phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)\mathbb{E}\left(K_{n,\ell+1}(i,j)\right)\right| \lesssim n^{-1/2}\log(n)$ for $i\neq j$, and $\left|n^{-1}\sum_{\ell=1}^n \phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)K_{n,\ell+1}(i,j)\left[\mathbb{E}\left(D_\ell^2|\mathcal{F}_{\ell-1}\right)-1\right]\right| \lesssim n^{-1/2}\log(n)$ for all $i,j\in\mathsf{I}$. To show this, write

$$\frac{1}{n}\sum_{\ell=1}^{n-1}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)\mathbb{E}\left(K_{n,\ell+1}(i,j)\right) = \left\{\frac{1}{n}\sum_{\ell=1}^{n-1}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)\right\}\mathbb{E}\left(K_{n,n}(i,j)\right)$$
$$+\frac{1}{n}\sum_{\ell=1}^{n-1}\left[\frac{1}{n}\sum_{k=1}^{\ell-1}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)\right]\left[\mathbb{E}\left(K_{n,\ell}(i,j)-K_{n,\ell+1}(i,j)\right)\right]$$

By the convergence of Riemann sums, $\left|\frac{1}{n}\sum_{\ell=1}^{n-1}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)\right| \lesssim n^{-1}$. Combined with (32) and (35), this implies that

$$\left|\frac{1}{n}\sum_{\ell=1}^{n}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)\mathbb{E}\left(K_{n,\ell+1}(i,j)\right)\right| \le \frac{1}{n}\left(\sqrt{n}+\sqrt{n}\sum_{k=1}^n\frac{1}{k}\right)\lesssim\frac{\log(n)}{\sqrt{n}}.$$

For the second term, let U denotes the Poisson equation solution associated to $\mathbb{E}\left(D_\ell^2|\mathcal{F}_{\ell-1}\right)-1$, so that we have almost surely

$$U(X_{\ell-1})-PU(X_{\ell-1}) = \mathbb{E}\left(D_\ell^2|\mathcal{F}_{\ell-1}\right)-1.$$

Therefore, by the usual martingale approximation trick

$$\frac{1}{n}\sum_{\ell=1}^{n-1}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)\mathbb{E}\left(K_{n,\ell+1}(i,j)\left[\mathbb{E}\left(D_\ell^2|\mathcal{F}_{\ell-1}\right)-1\right]\right) = \frac{1}{n}\phi_i\left(\frac{1}{n}\right)\phi_j\left(\frac{1}{n}\right)\mathbb{E}\left(K_{n,2}(i,j)U(X_0)\right)$$
$$-\frac{1}{n}\phi_i\left(1-\frac{1}{n}\right)\phi_j\left(1-\frac{1}{n}\right)\mathbb{E}\left(K_{n,n}(i,j)U(X_{n-1})\right)$$
$$+\frac{1}{n}\sum_{\ell=1}^{n-1}\mathbb{E}\left[\left\{\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)K_{n,\ell+1}(i,j)-\phi_i\left(\frac{\ell-1}{n}\right)\phi_j\left(\frac{\ell-1}{n}\right)K_{n,\ell}(i,j)\right\}U(X_{\ell-1})\right].$$

We now use the fact that $\phi_i\phi_j$ is of class C^1 (see Theorem 6.2 (ii)), (32), and (35) to conclude that

$$\left|\frac{1}{n}\sum_{\ell=1}^{n-1}\phi_i\left(\frac{\ell}{n}\right)\phi_j\left(\frac{\ell}{n}\right)\mathbb{E}\left(K_{n,\ell+1}(i,j)\left[\mathbb{E}\left(D_\ell^2|\mathcal{F}_{\ell-1}\right)-1\right]\right)\right|$$
$$\lesssim\frac{1}{\sqrt{n}}+\frac{1}{n}\sum_{\ell=1}^{n-1}\mathbb{E}^{1/2}\left(|K_{n,\ell+1}(i,j)-K_{n,\ell+2}(i,j)|^2\right)\lesssim\frac{\log(n)}{\sqrt{n}}.$$

This proves the claim. It remains to establish (35). Write $\mathbb{E}_\ell$ to denote the expectation operator wrt $n^{-1/2}\mathbf{S}_{\ell:n}$. We then have for any $h_1,h_2\in\mathbb{R}^{\mathsf{I}}$,

$$2K_{n,\ell}\cdot(h_1,h_2) = \nabla^{(2)}f_{\alpha,n,\ell}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}\right)\cdot(h_1,h_2)$$
$$=2\left(\frac{n}{n-\ell+1}\right)\mathbb{E}_\ell\left[f'_\alpha\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+\frac{\mathbf{S}_{\ell:n}}{\sqrt{n}}\right)\left\langle\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+\frac{\mathbf{S}_{\ell:n}}{\sqrt{n}},h_1\right\rangle_\alpha\left\langle\frac{\mathbf{S}_{\ell:n}}{\sqrt{n}},h_2\right\rangle\right]$$
$$=\left(\frac{n-\ell}{n-\ell+1}\right)\nabla^{(2)}f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+\frac{\mathbf{S}_\ell}{\sqrt{n}}\right)\cdot(h_1,h_2)+\left(\frac{n}{n-\ell+1}\right)O\left(\frac{1}{\sqrt{n}}\right).$$

Therefore

$$
\begin{aligned}
&2\left(K_{n,\ell}-K_{n,\ell+1}\right)\cdot(h_1,h_2)= \\
&\nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+\frac{\mathbf{S}_\ell}{\sqrt{n}}\right)\cdot(h_1,h_2)-\nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+\frac{\mathbf{B}_\ell D_\ell}{\sqrt{n}}\right)\cdot(h_1,h_2) \\
&\qquad -\frac{1}{n-\ell+1}\nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+\frac{\mathbf{S}_\ell}{\sqrt{n}}\right)\cdot(h_1,h_2)+\left(\frac{n}{n-\ell+1}\right)O\left(\frac{1}{\sqrt{n}}\right) \\
&\qquad =\nabla^{(3)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+t\frac{\mathbf{S}_\ell}{\sqrt{n}}+(1-t)\frac{\mathbf{B}_\ell D_\ell}{\sqrt{n}}\right)\cdot\left(h_1,h_2,\frac{\mathbf{S}_\ell}{\sqrt{n}}-\frac{\mathbf{B}_\ell D_\ell}{\sqrt{n}}\right) \\
&\qquad\quad -\frac{1}{n-\ell+1}\nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+\frac{\mathbf{S}_\ell}{\sqrt{n}}\right)\cdot(h_1,h_2)+\left(\frac{n}{n-\ell+1}\right)O\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

for some $t\in(0,1)$. Using (32) and (33), (35) follows from the above.

6. Appendix

6.1. **A weak invariance principle for martingales.**

We recall a martingale weak invariance principle from Hall and Heyde (1980) Theorem 4.2. Let $\{D_n,\mathcal{F}_n,\ n\geq 0\}$ be a martingale difference sequence. Set $S_k=\sum_{j=1}^k D_j$, $U_k^2=\sum_{j=1}^k D_j^2$, and $s_k^2=\sum_{j=1}^k\mathbb{E}(D_j^2)$ ($S_0=0$, $U_0^2=0$, $s_0^2=0$). Consider $\zeta_n:\ [0,1]\to\mathbb{R}$, the $C[0,1]$-valued process obtained by interpolating the points $\left(0,\frac{S_0}{U_n}\right),\ldots,\left(\frac{s_k^2}{s_n^2},\frac{S_k}{U_n}\right),\ldots,\left(1,\frac{S_n}{U_n}\right)$, where $U_k=\sqrt{U_k^2}$.

Theorem 6.1. *As $n\to\infty$, suppose that*

(1) $s_n^2\to\infty$, and $\frac{U_n^2}{s_n^2}$ converges almost surely to a random variable that is positive almost surely.
(2) For all $\epsilon>0$, $s_n^{-2}\sum_{j=1}^n\mathbb{E}\left(D_j^2\mathbf{1}_{\{|D_j|>\epsilon s_n\}}\right)\to 0$.

Then ζ_n, as a random process in $C[0,1]$ (equipped with the uniform norm), converges weakly to the standard Brownian motion.

Remark 8. Assumptions A1-A2 imply that the conclusion of Theorem 6.1 holds for the martingale difference $\{D_k,\ 1\leq k\leq n\}$ of Section 2. To see this, notice that for $p>1$ as in A1 and for any $M>0$,

$$\mathbb{E}\left(\frac{U_n^2}{n}\mathbf{1}_{\{U_n>nM\}}\right)\lesssim\left(\frac{1}{M}\right)^{p-1}\left(\frac{1}{n}\sum_{k=1}^n\kappa_k^2\right)^p.$$

Therefore, under A1, $n^{-1}U_n^2$ is uniformly integrable. Now, by A2, $n^{-1}U_n^2$ converges almost surely to $\sigma_\star^2$. We conclude that $n^{-1}s_n=n^{-1}\mathbb{E}(U_n^2)$ converges to $\mathbb{E}(\sigma_\star^2)>0$, so that $s_n\to\infty$, and U_n^2/s_n converges to $\sigma_\star^2/\mathbb{E}(\sigma_\star^2)$. Thus part (1) of the theorem holds. For all $\epsilon>0$,

$$s_n^{-2}\sum_{j=1}^n\mathbb{E}\left(D_j^2\mathbf{1}_{\{|D_j|>\epsilon s_n\}}\right)\lesssim\frac{1}{s_n^{2p}}\sum_{j=1}^n\kappa_j^{2p}\lesssim\frac{1}{n^p}\sum_{j=1}^n\kappa_j^{2p},$$

thus part (2) also holds under A2.

6.2. **Mercer's Theorem.** We recall Mercer's theorem below. Part (i) is the standard Mercer's theorem, and part (ii) is a special case of a result due to T. Kadota (Kadota (1967)).

Theorem 6.2 (Mercer's Theorem). **(i):** *Let $k:\ [0,1]\times[0,1]\to\mathbb{R}$ be a continuous positive semi-definite kernel. Then there exist nonnegative numbers $\{\lambda_j,\ j\geq 0\}$, and orthonormal functions $\{\phi_j,\ j\geq 0\}$, $\phi_j\in L^2([0,1])$, such that $\int_0^1 k(x,y)\phi_j(y)dy=\lambda_j\phi_j(x)$ for all $x\in[0,1]$, $j\geq 0$, and*

$$\lim_{n\to\infty}\sup_{x,y\in[0,1]}\left|k(x,y)-\sum_{j=0}^{n}\lambda_j\phi_j(x)\phi_j(y)\right|=0.$$

Furthermore, if $\lambda_j\neq 0$, ϕ_j is continuous.

(ii): *Let k as above. If in addition k is of class $\mathcal{C}^2$ on $[0,1]\times[0,1]$, then for $\lambda_j\neq 0$, ϕ_j is of class $\mathcal{C}^1$ on $[0,1]$ and*

$$\lim_{n\to\infty}\sup_{x,y\in[0,1]}\left|\frac{\partial^2}{\partial x\partial y}k(x,y)-\sum_{j=0}^{n}\lambda_j\phi_j'(x)\phi_j'(y)\right|=0.$$

By setting $x=y$, in both expansions, it follows that

$$\sup_{0\leq x\leq 1}\sum_{j\geq 0}\lambda_j|\phi_j(x)|^2\leq\sup_{0\leq x\leq 1}k(x,x)<\infty. \tag{36}$$

and

$$\sup_{0\leq x\leq 1}\sum_{j\geq 0}\lambda_j|\phi_j'(x)|^2\leq\sup_{0\leq x\leq 1}\left|\frac{\partial^2}{\partial u\partial v}k(u,v)|_{u=x,v=x}\right|<\infty. \tag{37}$$

6.3. **Proof of Lemma 5.2.**

Proof. Set $\tilde{w}_n(0)=1/n$, and for $k>0$ integer, $\tilde{w}_n(k)=2n^{-1}w(k/c_n)$. Then we can rewrite Q_n as

$$Q_n=\sum_{k=1}^{n}\sum_{j=1}^{k}\tilde{w}_n(k-j)\bar{h}(X_k)\bar{h}(X_j).$$

Using the Poisson's equation $\bar{h}(x)=g_\theta(x)-P_\theta g_\theta(x)$, it holds almost surely that $\bar{h}(X_k)=g_{\theta_{k-1}}(X_k)-P_{\theta_{k-1}}g_{\theta_{k-1}}(X_k)$. Therefore

$$\begin{aligned}\bar{h}(X_k)\bar{h}(X_j)-D_kD_j&=\left(\bar{h}(X_k)-D_k\right)\bar{h}(X_j)+D_k\left(\bar{h}(X_j)-D_j\right)\\&=\left(P_{\theta_{k-1}}g_{\theta_{k-1}}(X_{k-1})-P_{\theta_{k-1}}g_{\theta_{k-1}}(X_k)\right)\bar{h}(X_j)\\&\quad+D_k\left(P_{\theta_{j-1}}g_{\theta_{j-1}}(X_{j-1})-P_{\theta_{j-1}}g_{\theta_{j-1}}(X_j)\right).\end{aligned}$$

Then setting $\epsilon_n=Q_n-\sum_{k=1}^n\sum_{j=1}^k\tilde{w}_n(k-j)D_kD_j$, we obtain

$$\begin{aligned}\epsilon_n&=\sum_{k=1}^{n}\sum_{j=1}^{k}\tilde{w}_n(k-j)\left(P_{\theta_{k-1}}g_{\theta_{k-1}}(X_{k-1})-P_{\theta_{k-1}}g_{\theta_{k-1}}(X_k)\right)\bar{h}(X_j)\\&\quad+\sum_{k=1}^{n}\sum_{j=1}^{k}\tilde{w}_n(k-j)D_k\left(P_{\theta_{j-1}}g_{\theta_{j-1}}(X_{j-1})-P_{\theta_{j-1}}g_{\theta_{j-1}}(X_j)\right)=\varrho_n^{(1)}+\varrho_n^{(2)}.\end{aligned} \tag{38}$$

We start with the second term on the right hand side of (38) that yields after some rearrangements

$$
\begin{aligned}
\varrho_n^{(2)} = \sum_{k=1}^n D_k \sum_{j=1}^k \tilde{w}_n(k-j)\left(P_{\theta_{j-1}} g_{\theta_{j-1}}(X_{j-1}) - P_{\theta_{j-1}} g_{\theta_{j-1}}(X_j)\right) \\
= \sum_{k=2}^n D_k \left(\sum_{j=1}^{k-1} \left(\tilde{w}_n(k-j) - \tilde{w}_n(k-j+1)\right) P_{\theta_{j-1}} g_{\theta_{j-1}}(X_{j-1})\right) \\
+ \sum_{k=2}^n D_k \left(\sum_{j=1}^{k-1} \tilde{w}_n(k-j)\left(P_{\theta_j} g_{\theta_j}(X_j) - P_{\theta_{j-1}} g_{\theta_{j-1}}(X_j)\right)\right) \\
+ P_{\theta_0} g_{\theta_0}(X_0) \sum_{k=1}^n \tilde{w}_n(k) D_k + \left(\tilde{w}_n(0) - \tilde{w}_n(1)\right) \sum_{k=1}^n D_k P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1}) \\
- \tilde{w}_n(0) \sum_{k=1}^n D_k P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k) = T_n^{(1)} + T_n^{(2)},
\end{aligned}
$$

where $T_n^{(2)} = -\tilde{w}_n(0) \sum_{k=1}^n D_k P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k)$. We deal with $T_n^{(2)}$ below. Notice that $T_n^{(1)}$ has the general form $\sum_{k=1}^n B_{n,k} D_k$, where $B_{n,k}$ is $\mathcal{F}_{k-1}$-measurable. Thus by Burkeholder's inequality applied to the martingale $\sum_{k=1}^n B_{n,k} D_k$, we derive after some calculations that

$$
\left\|T_n^{(1)}\right\|_p^{p\wedge 2} \lesssim \sum_{k=1}^n \|B_{n,k} D_k\|_p^{p\wedge 2} \le \sum_{k=1}^n \|B_{n,k}\|_{2p}^{p\wedge 2} \kappa_k^{p\wedge 2} \lesssim \frac{1}{n^{p\wedge 2}} \sum_{k=1}^n \left\{\kappa_k \left(1 + \delta_{n,k}^{(1)}\right)\right\}^{p\wedge 2}.
$$

The first term on the right hand side of (38) gives

$$
\begin{aligned}
\varrho_n^{(1)} = \sum_{k=1}^n \left(P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1}) - P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k)\right) \sum_{j=1}^k \tilde{w}_n(k-j)\bar{h}(X_j) \\
= \sum_{k=1}^n \left(P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1}) - P_{\theta_k} g_{\theta_k}(X_k)\right) \sum_{j=1}^k \tilde{w}_n(k-j)\bar{h}(X_j) \\
+ \sum_{k=1}^n \left(P_{\theta_k} g_{\theta_k}(X_k) - P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k)\right) \sum_{j=1}^k \tilde{w}_n(k-j)\bar{h}(X_j).
\end{aligned}
$$

Then we write

$$
\begin{aligned}
\sum_{k=1}^n \left(P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1}) - P_{\theta_k} g_{\theta_k}(X_k)\right) \sum_{j=1}^k \tilde{w}_n(k-j)\bar{h}(X_j) \\
= \sum_{k=1}^n P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1}) \left\{\sum_{j=1}^k \tilde{w}_n(k-j)\bar{h}(X_j) - \sum_{j=1}^{k-1} \tilde{w}_n(k-1-j)\bar{h}(X_j)\right\} \\
- P_{\theta_n} g_{\theta_n}(X_n) \sum_{j=1}^n \tilde{w}_n(n-j)\bar{h}(X_j).
\end{aligned}
$$

This implies that

$$
\varrho_n^{(1)} = \epsilon_n^{(1)} + R_n^{(1)} + R_n^{(2)},
$$

where

$$\epsilon_n^{(1)} = \sum_{k=3}^{n} P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1}) \sum_{j=1}^{k-2} \left(\tilde{w}_n(k-j) - \tilde{w}_n(k-1-j)\right) \bar{h}(X_j),$$

$$R_n^{(1)} = \sum_{k=1}^{n} \left(P_{\theta_k} g_{\theta_k}(X_k) - P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k)\right) \sum_{j=1}^{k} \tilde{w}_n(k-j)\bar{h}(X_j) - P_{\theta_n} g_{\theta_n}(X_n) \sum_{j=1}^{n} \tilde{w}_n(n-j)\bar{h}(X_j),$$

and

$$R_n^{(2)} = \tilde{w}_n(0) \sum_{k=1}^{n} P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1})\bar{h}(X_k) + (\tilde{w}_n(1) - \tilde{w}_n(0)) \sum_{k=2}^{n} P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1})\bar{h}(X_{k-1}).$$

We gather these terms together and rewrite (38) as

$$\epsilon_n = \epsilon_n^{(1)} + T_n^{(1)} + R_n^{(1)} + T_n^{(2)} + R_n^{(2)}. \tag{39}$$

Using (23) we get:

$$\begin{aligned} \left\| R_n^{(1)} \right\|_p &\leq \sum_{k=1}^{n} b_k \left\| \sum_{j=1}^{k} \tilde{w}_n(k-j)\bar{h}(X_j) \right\|_{2p} + a_n \left\| \sum_{j=1}^{n} \tilde{w}_n(n-j)\bar{h}(X_j) \right\|_{2p} \\ &\lesssim n^{-1} \sum_{k=1}^{n} b_k \left(a_{k-1} + \delta_{n,k}^{(1)} + \delta_{n,k}^{(2)}\right) + n^{-1} a_n \left(a_{n-1} + \delta_{n,n}^{(1)} + \delta_{n,n}^{(2)}\right). \end{aligned}$$

With the same technique we get

$$\|\epsilon_n^{(1)}\|_p \lesssim \frac{1}{nc_n} \sum_{k=2}^{n} a_k \left(\delta_{n,k}^{(1)} + \delta_{n,k}^{(2)}\right).$$

Remark 9. In fact, as we shall show later, with additional assumptions, the term $\epsilon_n^{(1)}$ has better convergence rate than shown above. See the proof of Theorem 4.1 in Section 5.4.

The last two terms in (39) are

$$\begin{aligned} T_n^{(2)} + R_n^{(2)} = -\tilde{w}_n(0) \sum_{k=1}^{n} D_k P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k) + \tilde{w}_n(0) \sum_{k=1}^{n} P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1})\bar{h}(X_k) \\ + (\tilde{w}_n(1) - \tilde{w}_n(0)) \sum_{k=2}^{n} P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1})\bar{h}(X_{k-1}). \end{aligned} \tag{40}$$

Replacing $\bar{h}(X_k)$ by $g_{\theta_{k-1}}(X_k) - P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k)$, the first and third terms on the right hand side of (40) gives after some easy re-arrangements

$$\begin{aligned} &- \tilde{w}_n(0) \sum_{k=1}^{n} P_{\theta_{k-1}} g_{\theta_{k-1}}(X_k) D_k + (\tilde{w}_n(1) - \tilde{w}_n(0)) \sum_{k=1}^{n-1} P_{\theta_k} g_{\theta_k}(X_k)\bar{h}(X_k) \\ &= (\tilde{w}_n(1) - \tilde{w}_n(0)) \sum_{k=1}^{n-1} P_{\theta_k} g_{\theta_k}(X_k) \left(P_{\theta_{k-1}} g_{\theta_{k-1}}(X_{k-1}) - P_{\theta_k} g_{\theta_k}(X_k)\right) + \varrho_n^{(3)}, \end{aligned}$$

where $\mathbb{E}^{1/p}(|\varrho_n^{(3)}|^p) \lesssim n^{-1}a_n\kappa_n + n^{-1}c_n^{-1}\sum_{k=1}^n a_k\kappa_k + n^{-1}\sum_{k=1}^n b_k(a_k+\kappa_k)$. The second term on the right-hand side of (40) gives

$$\tilde{w}_n(0)\sum_{k=1}^n P_{\theta_{k-1}}g_{\theta_{k-1}}(X_{k-1})\bar{h}(X_k) = \tilde{w}_n(0)\sum_{k=1}^n P_{\theta_{k-1}}g_{\theta_{k-1}}(X_{k-1})\left(P_{\theta_{k-1}}g_{\theta_{k-1}}(X_{k-1}) - P_{\theta_k}g_{\theta_k}(X_k)\right) + \varrho_n^{(4)}$$

where $\mathbb{E}^{1/p}\left(|\varrho_n^{(4)}|\right) \lesssim n^{-1}\left(\sum_{k=1}^n (a_{k-1}\kappa_k)^{p\wedge 2}\right)^{\frac{1}{p\wedge 2}} + n^{-1}\sum_{k=1}^n a_{k-1}b_k$. Therefore

$$T_n^{(2)} + R_n^{(2)} = \varrho_n^{(3)} + \varrho_n^{(4)} + \tilde{w}_n(0)(P_{\theta_0}g_{\theta_0}(X_0))^2 - \tilde{w}_n(0)P_{\theta_n}g_{\theta_n}(X_n)P_{\theta_{n-1}}g_{\theta_{n-1}}(X_{n-1})$$
$$+ (\tilde{w}_n(1) - 2\tilde{w}_n(0))\sum_{k=1}^{n-1} P_{\theta_k}g_{\theta_k}(X_k)P_{\theta_{k-1}}g_{\theta_{k-1}}(X_{k-1}) + (2\tilde{w}_n(0) - \tilde{w}_n(1))\sum_{k=1}^{n-1}\left(P_{\theta_k}g_{\theta_k}(X_k)\right)^2,$$

and

$$\left\|T_n^{(2)} + R_n^{(2)}\right\|_p \lesssim \left\|\varrho_n^{(3)}\right\|_p + \left\|\varrho_n^{(4)}\right\|_p + \frac{1}{n}\left(a_0^2 + a_na_{n-1}\right) + \frac{1}{nc_n}\sum_{k=1}^n a_k(a_k + a_{k-1}).$$

We obtain the lemma by putting all the remainders together. □

Acknowledgments: I'm grateful to Matias Cattaneo for very helpful discussions. This work is partly supported by NSF grant DMS 0906631.

References

Andrieu, C. and Moulines, E. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.* **16** 1462–1505.

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing* **18** 343–373.

Atchade, Y. (2011). Kernel estimators of asymptotic variance for adaptive Markov Chain Monte Carlo. *Annals of Statistics* **39** 990–1011.

Atchade, Y. and Fort, G. (2010). Limit theorems for some adaptive mcmc algorithms with sub-geometric kernels. *Bernoulli* **16** 116–154.

Atchade, Y. and Fort, G. (2012). Limit theorems for some adaptive MCMC algorithms with subgeometric kernels: Part II. *Bernoulli* **18** 975–1001.

Atchade, Y. F. and Cattaneo, M. D. (2011). Limit theorems for quadratic forms of Markov chains. *Technical Report, Dept of Statistics, University of Michigan* .

Damerdji, H. (1995). Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Oper. Res.* **43** 282–291.

Dedecker, J. and Rio, E. (2008). On mean central limit theorems for stationary sequences. *Annales de l'Institut Henri Poincare - Probabilités et Statistiques* **44** 693–726.

Dudley, R. (2002). *Real Analysis and Probability.* Cambridge Series in advanced mathematics, Cambridge University Press, NY.

Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov Chain Monte Carlo. *Ann. Statist.* **38** 1034–1070.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian data analysis.* 2nd ed. Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, FL.

Götze, F. and Tikhomirov, A. N. (2005). Asymptotic expansions in non-central limit theorems for quadratic forms. *J. Theoret. Probab.* **18** 757–811.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit theory and its application.* Academic Press, New York.

Jansson, M. (2004). The error in rejection probability of simple autocorrelation robust tests. *Econometrica* **72** 937–946.

Jones, G. L., Haran, M., Caffo, B. S. and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547.

Kadota, T. T. (1967). Term-by-term differentiability of mercer's expansion. *Proceedings of the American Mathematical Society* **18** 69–72.

Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* **21** 1130–1164.

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* **28** 329–348.

Ponomarëv, S. P. (1987). Submersions and pre-images of sets of measure zero. *Sibirsk. Mat. Zh.* **28** 199–210.

Saksman, E. and Vihola, M. (2010). On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Probab.* **20** 2178–2203.

Sun, Y., Phillips, P. C. B. and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* **76** 175–194.