

# Globally-centered autocovariances in MCMC

Medha Agarwal

Dept. of Mathematics and Statistics

IIT Kanpur

medhaaga@iitk.ac.in

Dootika Vats\*

Dept. of Mathematics and Statistics

IIT Kanpur

dootika@iitk.ac.in

September 3, 2020

## Abstract

Autocovariances are a fundamental quantity of interest in Markov chain Monte Carlo (MCMC) simulations with autocorrelation function (ACF) plots being an integral performance assessment visualization tool. Unfortunately, for slow mixing Markov chains, the empirical autocovariance can highly underestimate the truth. For multiple chain MCMC sampling, we propose a globally-centered estimator of the autocovariance function (G-ACvF) that exhibits significant theoretical and empirical improvements. We show that the bias of the G-ACvF estimator is smaller than the bias of the current state-of-the-art. The impact of this improved estimator is evident in three critical output analysis applications: (1) ACF plots, (2) estimates of the Monte Carlo asymptotic covariance matrix, and (3) estimates of the effective sample size. Under weak conditions, we establish strong consistency of an improved asymptotic covariance estimator, and obtain its large sample bias and variance. The performance of the new G-ACvF estimator is demonstrated through various examples.

## 1 Introduction

Advancements in modern personal computing have made it easy to run parallel Markov chain Monte Carlo (MCMC) implementations. This is particularly useful for slow mixing Markov chains where the starting points of the chains are spread over the state-space in order to more accurately capture characteristics of the target distribution. The output from  $m$  parallel chains is then summarized, visually and quantitatively, to assess the empirical mixing properties of the chains and the quality of Monte Carlo estimators.

A key quantity that drives MCMC output analysis is the autocovariance function (ACvF). Estimators of ACvF are only available for single-chain implementations, with a parallel-chain version

---

\*Corresponding author

obtained by naive averaging. As we will demonstrate, this defeats the purpose of running parallel Markov chains from dispersed starting values.

Let  $F$  be the target distribution with mean  $\mu$  defined on a space  $\mathcal{X} \subseteq \mathbb{R}^d$  equipped with a countably generated  $\sigma$ -field  $\mathcal{B}(\mathcal{X})$ . For  $s = 1, \dots, m$ , let  $\{X_{st}\}_{t \geq 1}$  be the  $s$ th Harris ergodic  $F$ -invariant Markov chain (see Meyn and Tweedie, 2009, for definitions) employed to learn characteristics about  $F$ . The process is covariance stationary so that the ACvF at lag  $k$  is defined as

$$\Gamma(k) = \text{Cov}_F(X_{s1}, X_{s1+k}) = \mathbb{E}_F \left[ (X_{s1} - \mu)(X_{s1+k} - \mu)^T \right].$$

Estimating  $\Gamma(k)$  is critical to assessing the quality of the sampler and the reliability of Monte Carlo estimators. Let  $\bar{X}_s = n^{-1} \sum_{t=1}^n X_{st}$  denote the Monte Carlo estimator of  $\mu$  from the  $s$ th chain. The standard estimator for  $\Gamma(k)$  is the sample autocovariance matrix at lag  $k \geq 0$ :

$$\hat{\Gamma}_s(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_{st} - \bar{X}_s) (X_{s,t+k} - \bar{X}_s)^T, \quad (1)$$

and for  $k < 0$ ,  $\hat{\Gamma}_s(k) = \hat{\Gamma}_s(-k)^T$ . For a single-chain MCMC run, the estimator  $\hat{\Gamma}_s(k)$  is used to construct ACF plots, to estimate the long-run variance of Monte Carlo estimators (Hannan, 1970; Damerджи, 1991), and to estimate effective sample size (ESS) (Kass et al., 1998; Gong and Flegal, 2016; Vats et al., 2019). However, there is no unified approach to constructing estimators of  $\Gamma(k)$  for parallel-chain implementations. Specifically, parallel chains are often spread across the state space so as to adequately capture high density areas. For slow mixing chains, the chains take time to traverse the space so that  $\bar{X}_s$  are all vastly different. Consequently,  $\hat{\Gamma}_s(k)$  typically underestimates  $\Gamma(k)$  leading to a false sense of security about the quality of the process.

We propose a globally-centered estimator of ACvF (G-ACvF) that centers the Markov chains around the global mean from all  $m$  chains. We show that the bias for G-ACvF is lower than  $\hat{\Gamma}_s(k)$ , and through various examples, demonstrate improved estimation. We employ the G-ACvF estimators to construct ACF plots and a demonstrative example is at the end of this Section.

Estimators of ACvFs are used to estimate the long-run variance of Monte Carlo averages. Specifically, spectral variance (SV) estimators are used to estimate  $\Sigma = \sum_{k=-\infty}^{\infty} \Gamma(k)$  (Andrews, 1991; Damerджи, 1991; Flegal and Jones, 2010). We replace  $\hat{\Gamma}_s(k)$  with G-ACvF in SV estimators to obtain a globally-centered SV (G-SV) estimator and demonstrate strong consistency under weak conditions. In the spirit of Andrews (1991), we also obtain large-sample bias and variance of the resulting estimator. SV estimators can be prohibitively slow to compute (Liu and Flegal, 2018). To relieve the computational burden, we adapt the fast algorithm of Heberle and Sattarhoff (2017) for G-SV estimator that dramatically reduces computation time. The G-SV estimator is employed in the computation of ESS. We will show that using G-SV estimator for estimating  $\Sigma$  safeguards users against early termination of the MCMC process.

## 1.1 Demonstrative example

We demonstrate the striking difference in estimation of  $\Gamma(k)$  via ACF plots. Consider a random-walk Metropolis-Hastings sampler for a univariate mixture of Gaussians target density. Let

$$f(x) = 0.7 f(x; -5, 1) + 0.3 f(x; 5, 0.5),$$

be the target density where  $f(x; a, b)$  is the density of a normal distribution with mean  $a$  and variance  $b$ . We set  $m = 2$  with starting values distributed to the two modes. The trace plots in Figure 1 indicate that in the first  $10^4$  steps, the chains do not jump modes so that both Markov chains yield significantly different estimates of the population mean,  $\mu$ . At  $10^5$  sample size, both Markov chains have traversed the state space reasonably and yield similar estimates of  $\mu$ .

In Figure 1 for  $n = 10^4$ , we present the ACF plots using  $\hat{\Gamma}_s(k)$  and our proposed G-ACvF estimator. The blue curves are the respective estimates at  $n = 10^5$ . At  $n = 10^5$  when the chains have similar means, the G-ACF and locally-centered ACF are equivalent. However, for  $n = 10^4$ ,  $\hat{\Gamma}_s(k)$  critically underestimates the correlation giving users a misleading visual of the quality of the Markov chain. The G-ACF estimator accounts for the discrepancy in sample means between two chains leading to a far improved quality of estimation.

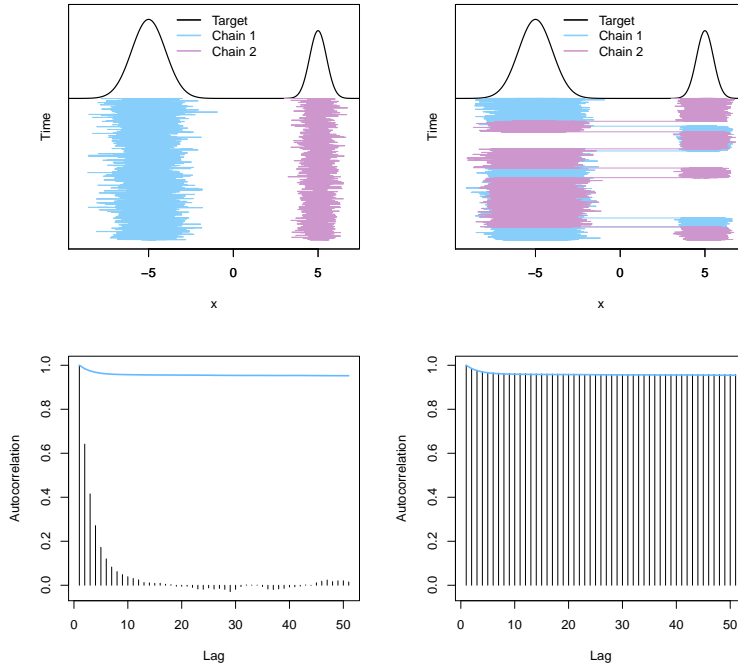


Figure 1: Top: Target density and trace plots for two chains at  $n = 10^4$  (left) and  $n = 10^5$  (right). Bottom: ACF plots for the first chain using local-centering (left) and global-centering (right). Histogram estimates at  $n = 10^4$  with the blue curve being estimates from  $n = 10^5$ .

## 2 Globally-centered autocovariance

Let  $P$  denote the Markov transition kernel that uniquely determines the ACvF under stationarity. Recall that the sample mean of the  $s$ th Markov chain is  $\bar{X}_s$  and denote the global mean vector by  $\bar{\bar{X}} = m^{-1} \sum_{s=1}^m \bar{X}_s$ . The global mean is a naturally superior estimator of  $\mu$  compared to  $\bar{X}_s$ . For  $k \geq 0$ , we define the globally-centered ACvF (G-ACvF) estimator for the  $s$ th Markov chain as

$$\hat{\Gamma}_{G,s}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_{st} - \bar{\bar{X}})(X_{s(t+k)} - \bar{\bar{X}})^T, \quad (2)$$

with  $\hat{\Gamma}_{G,s}(k) = \hat{\Gamma}_{G,s}(-k)^T$  for  $k < 0$ . In the event that all  $m$  Markov chains have been run long enough,  $\bar{X}_s \approx \bar{\bar{X}}$ , and hence  $\hat{\Gamma}_s(k) \approx \hat{\Gamma}_{G,s}(k)$ . However, for shorter runs or for slow mixing chains,  $\Gamma(k)$  is more appropriately estimated by  $\hat{\Gamma}_{G,s}$  as it utilizes information from all chains and accounts for disparity between estimates of  $\mu$ . Let

$$\Phi^{(q)} = \sum_{k=-\infty}^{\infty} |k|^q \Gamma(k),$$

and let  $\Phi^{(1)}$  be denoted by  $\Phi$ . The proof of the following theorem is in Appendix 7.2. Let  $\|\cdot\|$  denote Euclidean norm.

**Theorem 1.** *Let  $\mathbb{E}_F \|X_{11}\|^{2+\delta} < \infty$  for some  $\delta > 0$ . If  $P$  is polynomially ergodic of order  $\xi > (2 + \epsilon)/(1 + 2/\delta)$  for some  $\epsilon > 0$ , then,*

$$\mathbb{E}_F [\hat{\Gamma}_{G,s}(k)] = \left(1 - \frac{|k|}{n}\right) \left(\Gamma(k) - \frac{\Sigma}{mn} - \frac{\Phi}{mn^2}\right) + o(n^{-1}).$$

*Remark 1.* Polynomial ergodicity and the moment conditions are required to ensure  $\Phi$  and  $\Sigma$  are finite. The above result can be stated more generally for  $\alpha$ -mixing processes, but we limit our attention to Markov chains.

When  $m = 1$ ,  $\hat{\Gamma}_{G,s}(k) = \hat{\Gamma}_s(k)$  the bias results for which can be found in Priestley (1981). Since lag-covariances are typically positive in MCMC, Theorem 1 implies that the G-ACvF estimators are asymptotically unbiased and exhibit reduced bias in finite samples compared to  $\hat{\Gamma}_s(k)$ . The consequences of this are particularly pertinent for the diagonals of  $\Gamma(k)$ .

Let  $\Gamma^{ii}(k)$  and  $\hat{\Gamma}_{G,s}^{ii}(k)$  be the  $i$ th component of  $\Gamma(k)$  and  $\hat{\Gamma}_{G,s}(k)$ , respectively. For component  $i$ , the autocorrelation is defined as

$$\rho^{(i)}(k) = \frac{\Gamma^{ii}(k)}{\Gamma^{ii}(0)},$$

and is instrumental in visualizing the serial correlation in the components of the Markov chain. A standard estimator of the autocorrelation is constructed from  $\hat{\Gamma}_s(k)$ . Instead, we advocate for

using G-ACvF estimates so that,

$$\hat{\rho}_{G,s}^{(i)}(k) = \frac{\hat{\Gamma}_{G,s}^{ii}(k)}{\hat{\Gamma}_{G,s}^{ii}(0)}. \quad (3)$$

The globally-centered autocorrelation provides a far more realistic assessment of the correlation structure of the marginal components of the chain as evidenced in Figure 1.

We end this section by noting that we can obtain an average G-ACvF and G-ACF over all  $m$  chains

$$\hat{\Gamma}_G(k) = \frac{1}{m} \sum_{s=1}^m \hat{\Gamma}_{G,s}(k) \quad \text{and} \quad \hat{\rho}_G^{(i)}(k) = \frac{1}{m} \sum_{s=1}^m \hat{\rho}_{G,s}^{(i)}(k).$$

The averaged estimators present a measure of the overall correlation structure induced by the Markov transition  $P$ .

### 3 Long-run variance estimators

A critical use of autocovariances is in the assessment of Monte Carlo variability of estimators. Let  $g : \mathcal{X} \rightarrow \mathbb{R}^p$  be an  $F$ -integrable function so that interest is in estimating  $\mu_g = \mathbb{E}_F[g(X)]$ . Set  $\{Y_{st}\}_{t \geq 1} = \{g(X_{st})\}_{t \geq 1}$  for all  $s = 1, \dots, m$ . Let  $\bar{Y}_s = n^{-1} \sum_{t=1}^n Y_{st}$  and  $\bar{\bar{Y}} = m^{-1} \sum_{s=1}^m \bar{Y}_s$ . By Birkhoff's ergodic theorem,  $\bar{\bar{Y}} \rightarrow \mu_g$  with probability 1 as  $n \rightarrow \infty$ . An asymptotic sampling distribution may be available via a Markov chain central limit theorem (CLT) if there exists a  $p \times p$  positive-definite matrix  $\Sigma$  such that

$$\sqrt{mn}(\bar{\bar{Y}} - \mu_g) \xrightarrow{d} N(0, \Sigma) \quad \text{where,} \quad (4)$$

$$\Sigma = \sum_{k=-\infty}^{\infty} \text{Cov}_F(Y_{11}, Y_{1(1+k)}) := \sum_{k=-\infty}^{\infty} \Upsilon(k). \quad (5)$$

The goal in output analysis for MCMC is to estimate  $\Sigma$  in order to assess variability in  $\bar{\bar{Y}}$  (Flegal et al., 2008; Roy, 2019; Vats et al., 2020). There is a rich literature on estimating  $\Sigma$  for single-chain MCMC implementations. The most common are SV estimators (Andrews, 1991; Vats et al., 2018) and batch means estimators (Chen and Seila, 1987; Vats et al., 2019). Recently, Gupta and Vats (2020) constructed a replicated batch means estimator for estimating  $\Sigma$  from parallel chains. Batch means estimators are computationally more efficient than SV estimators, whereas SV estimators are more reliable (Damerджи, 1995; Flegal and Jones, 2010). Here, we utilize G-ACvF estimators to construct globally-centered SV (G-SV) estimator of  $\Sigma$ . Using the method of Heberle and Sattarhoff (2017), we also provide a computationally efficient algorithm for obtaining the G-SV estimator.

The locally and globally-centered estimators of  $\Upsilon(k)$  are  $\hat{\Upsilon}_s(k)$  and  $\hat{\Upsilon}_{G,s}(k)$ , respectively, where for

$k \geq 0$

$$\hat{\Upsilon}_s(k) = \frac{1}{n} \sum_{t=1}^{n-k} (Y_{st} - \bar{Y}_s)(Y_{st+k} - \bar{Y}_s)^T \quad \text{and} \quad \hat{\Upsilon}_{G,s}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (Y_{st} - \bar{Y})(Y_{st+k} - \bar{Y})^T,$$

and let  $\hat{\Upsilon}_G(k) = m^{-1} \sum_{s=1}^m \hat{\Upsilon}_{G,s}(k)$ . SV estimators are constructed as weighted and truncated sums of estimated ACvFs. For some  $c \geq 1$ , let  $w : \mathbb{R} \rightarrow [-c, c]$  be a lag window function and  $b_n \in \mathbb{N}$  be a bandwidth. The (locally-centered) SV estimator of  $\Sigma$  for chain  $s$  is

$$\hat{\Sigma}_s = \sum_{k=-n+1}^{n-1} w\left(\frac{k}{b_n}\right) \hat{\Upsilon}_s(k). \quad (6)$$

Large-sample properties of  $\hat{\Sigma}_s$  have been widely studied. Vats et al. (2018) provide conditions for strong consistency while Flegal and Jones (2010); Hannan (1970) provide large-sample bias and variance results. These results also extend naturally to an average SV (ASV) estimator defined as  $\hat{\Sigma}_A := m^{-1} \sum_{s=1}^m \hat{\Sigma}_s$ .

### 3.1 Globally-centered spectral variance estimators

We define the G-SV estimator as the weighted and truncated sum of G-ACvFs

$$\hat{\Sigma}_G = \sum_{k=-n+1}^{n-1} w\left(\frac{k}{b_n}\right) \hat{\Upsilon}_G(k).$$

Standard conditions are imposed on the lag window,  $w$ .

*Assumption 1.* The lag window  $w(x)$  is continuous at all but a finite number of points, is a bounded and even function with  $w(0) = 1$ ,  $\int_{-\infty}^{\infty} w^2(x) dx < \infty$ , and  $\sum_{k=-\infty}^{\infty} |w(k/b_n)| < \infty$ .

Assumption 1 is standard (see Anderson, 1971) and is satisfied by most lag windows. We will use the following popular Bartlett lag window in our simulations:

$$w(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

#### 3.1.1 Theoretical results

First, we provide conditions for strong consistency. Strong consistency is particularly important to ensure sequential stopping rules in MCMC yield correct coverage at termination (Flegal and Gong, 2015; Glynn and Whitt, 1992). A critical assumption is that of a strong invariance principle which the following theorem establishes. Let  $B(n)$  denotes a standard  $p$ -dimensional Brownian motion.

**Theorem 2** (Kuelbs (1976); Vats et al. (2018)). *Let  $\mathbb{E}_F \|Y_{11}\|^{2+\delta} < \infty$  for  $\delta > 0$  and let  $P$  be polynomially ergodic of order  $\xi > (q + 1 + \epsilon)/(1 + 2/\delta)$  for  $q \geq 1$ . Then there exists a  $p \times p$  lower triangular matrix  $L$  with  $LL^T = \Sigma$ , a non-negative function  $\psi(n) = n^{1/2-\lambda}$  for some  $\lambda > 0$ , a finite random variable  $D$ , and a sufficiently rich probability space  $\Omega$  such that for all  $n > n_0$ ,*

$$\left\| \sum_{t=1}^n Y_t - n\mu_g - LB(n) \right\| < D\psi(n) \quad \text{with probability 1.}$$

**Theorem 3.** *Let the assumptions of Theorem 2 hold with  $q = 1$ . If  $\hat{\Sigma}_s \xrightarrow{a.s.} \Sigma$  for all  $s$ , and  $n^{-1}b_n \log \log n \rightarrow 0$  as  $n \rightarrow \infty$ , then  $\hat{\Sigma}_G \xrightarrow{a.s.} \Sigma$  as  $n \rightarrow \infty$ .*

Conditions for strong consistency of  $\hat{\Sigma}_s$  for polynomially ergodic Markov chains are in Vats et al. (2018). Typically,  $b_n = \lfloor n^\nu \rfloor$  for some  $0 < \nu < 1$  for which  $n^{-1}b_n \log \log n \rightarrow 0$ , thus Theorem 3 presents no added conditions for strong consistency. Our next two results establish large-sample bias and variance for G-SV and mimic those of  $\hat{\Sigma}_s$  which can be found in Hannan (1970). Let  $\Sigma^{ij}$  and  $\hat{\Sigma}_G^{ij}$  denote the  $ij$ th element of the matrix  $\Sigma$  and  $\hat{\Sigma}_G$ , respectively.

**Theorem 4.** *Let the assumptions of Theorem 2 hold with  $q$  such that*

$$\lim_{x \rightarrow 0} \frac{1 - w(x)}{|x|^q} = k_q < \infty$$

*and  $b_n^{q+1}/n \rightarrow 0$  as  $n \rightarrow \infty$ . Then  $\lim_{n \rightarrow \infty} b_n^q \mathbb{E} \left[ \hat{\Sigma}_G - \Sigma \right] = -k_q \Phi^{(q)}$ .*

Hannan (1970) proved a similar bias result assuming  $\mu_g$  was known with the rate  $b_n^q/n \rightarrow 0$  as  $n \rightarrow \infty$ . Similar to Anderson (1971) for the univariate case, in Theorem 4, we require  $b_n^{q+1}/n \rightarrow 0$  as  $n \rightarrow \infty$  for the multivariate case for when  $\mu_g$  is replaced by  $\bar{Y}$ .

**Theorem 5.** *Let the assumptions of Theorem 2 hold and let  $\mathbb{E}[D^4] < \infty$  and  $\mathbb{E}\|Y_{11}\|^4 < \infty$ , then  $\lim_{n \rightarrow \infty} b_n^{-1} n \text{Var} \left( \hat{\Sigma}_G^{ij} \right) = [\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2] \int_{-\infty}^{\infty} w(x)^2 dx$ .*

For the Bartlett lag window,  $q = 1$  with  $k_q = 1$  and  $\int w(x)^2 dx = 2/3$ . For other popular lag windows, see Anderson (1971).

### 3.1.2 Fast implementation

The SV estimator, despite having good statistical properties, poses limitations due to slow computation. The complexity of the SV estimator is  $\mathcal{O}(n^2 p^2)$  in general, and for the Bartlett lag window, it is  $\mathcal{O}(b_n n p^2)$ . For slow mixing Markov chains,  $n$  and  $b_n$  can be prohibitively large, limiting the use of SV estimators. To overcome this, we adapt the fast Fourier transform based algorithm of Heberle and Sattarhoff (2017). Suppose  $w_k = w(k/b_n)$  and let  $T(w)$  be the  $n \times n$  Toeplitz matrix

with the first column being  $(1 \ w_1 \ w_2 \ \dots, \ w_{n-1})^T$ . Notice an alternate formulation of  $\hat{\Sigma}_s$

$$\hat{\Sigma}_s = \frac{1}{n} A_s^T T(w) A_s, \quad \text{where } A_s = \begin{pmatrix} Y_{s1} - \bar{Y}_s & \dots & Y_{sn} - \bar{Y}_s \end{pmatrix}^T. \quad (7)$$

Heberle and Sattarhoff (2017) compute the  $n \times 1$  vector  $T(w)A_s$  using an algorithm based on fast Fourier transforms. Let  $w^* = (1 \ w_1 \ w_2 \ \dots, \ w_{n-1}, \ 0, \ w_{n-1}, \dots, w_1)^T$  and set  $C(w^*)$  to be a symmetric circulant matrix such that the matrix truncation  $C(w^*)_{1:n,1:n} = T(w)$ . Let  $M_{(j)}$  denote the  $j$ th column of matrix  $M$  and  $v^{(i)}$  denote the  $i$ th element of the vector  $v$ . With inputs  $C(w^*)$  and  $A_s$ , Algorithm 1 produces  $\hat{\Sigma}_s$  exactly. For more details, see Heberle and Sattarhoff (2017).

---

**Algorithm 1:** Heberle and Sattarhoff (2017) Algorithm

---

**Input:**  $C(w^*)$  and  $A_s$

- 1 Compute eigenvalues  $\lambda_i$  of  $C(w^*)_{(1)}$  using a discrete Fourier transform,  $i = 1, \dots, 2n$
- 2 Construct  $2n \times p$  matrix  $A_s^* = (A_s^T \ 0_{n \times q})^T$
- 3 **for**  $j = 1, 2, \dots, p$  **do**
- 4     Calculate  $V^* A_{s(j)}^*$  by DFT of  $A_{s(j)}^*$ .
- 5     Multiply  $V^* A_{s(j)}^{*(i)}$  with the eigenvalue  $\lambda_i$  for all  $i = 1, \dots, 2n$  to construct  $\Lambda V^* A_{s(j)}^*$ .
- 6     Calculate  $C(w^*) A_{s(j)}^* = V \Lambda V^* A_{s(j)}^*$  by inverse FFT of  $\Lambda V^* A_{s(j)}^*$ .
- 7 **end**
- 8 Select the first  $n$  rows of  $C(w^*) A_s^*$  to form  $T(w) A_s$ .
- 9 Premultiply by  $A_s^T$  and divide by  $n$ .

**Output:**  $\hat{\Sigma}_s$

---

We observe that a similar decomposition is possible for the G-SV estimator. Setting  $B_s = (Y_{s1} - \bar{Y} \ \dots \ Y_{sn} - \bar{Y})^T$  and calling Algorithm 1 with inputs  $C(w^*)$  and  $B_s$  yields  $\hat{\Sigma}_{G,s}$ . Algorithm 1 has complexity  $\mathcal{O}(pn \log n)$  and is thus orders of magnitude faster. Of particular importance is the fact that the bandwidth  $b_n$  has close to no impact on the computation time.

## 4 Effective sample size

An important method of assessing the reliability of  $\bar{\bar{Y}}$  in estimating  $\mu_g$  is to calculate the effective sample size (ESS). ESS are number of independent and identically distributed samples that would yield the same Monte Carlo variability as this correlated sampler for estimating  $\mu_g$  with  $\bar{\bar{Y}}$ . Let  $|\cdot|$  denote determinant. A multiple chain version of the ESS as defined by Vats et al. (2019) is

$$\text{ESS} = mn \left( \frac{|\Upsilon(0)|}{|\Sigma|} \right)^{1/p}.$$

The simulation terminates when the ESS is greater than a pre-specified, theoretically motivated,



lower-bound  $W_p$ . In our setting of  $m$  parallel chains of  $n$  samples each, we estimate ESS with

$$\widehat{\text{ESS}}_G = mn \left( \frac{|\hat{\Upsilon}(0)|}{|\hat{\Sigma}_G|} \right)^{1/p}.$$

We use the locally-centered  $\hat{\Upsilon}(0)$  to estimate  $\Upsilon(0)$  instead of  $\hat{\Upsilon}_G(0)$  when calculating ESS. Both  $\hat{\Upsilon}(0)$  and  $\hat{\Upsilon}_G(0)$  are consistent for  $\Upsilon(0)$ ; however the typical underestimation witnessed  $\Upsilon(0)$  allows a safeguard against early termination.

For the sake of comparison,  $\widehat{\text{ESS}}_A$  is constructed similarly using  $\hat{\Sigma}_A$  instead of  $\hat{\Sigma}_G$  to estimate  $\Sigma$ .

## 5 Examples

In this section we consider three different target distributions and sample  $m$  parallel Markov chains to assess the performance of our proposed estimators. We will make the following three comparisons - (1) locally-centered ACF vs G-ACF estimator, (2) A-SV vs G-SV estimator, and (3)  $\widehat{\text{ESS}}_A$  vs  $\widehat{\text{ESS}}_G$ . The quality of estimation of  $\Sigma$  is studied by coverage probabilities of a 95% Wald confidence region when the true mean  $\mu_g$  is known. The convergence of local and global estimators of  $\Sigma$  and  $\widehat{\text{ESS}}$  as  $n$  increases is studied through two types of running plots (1) logarithm of Frobenius norm of estimated  $\Sigma$ , and (2) logarithm of estimated  $\text{ESS}/mn$ . In all the examples we are interested in estimating the mean of the stationary distribution  $\mu$ , so  $g$  is the identity function. The bandwidth  $b_n$  are the defaults available in the R package `mcmcse` (Flegal et al., 2020).

### 5.1 Vector autoregressive process

Our first example is one where the  $\Upsilon$  and  $\Sigma$  are available in closed-form, so we can compare quality of estimation against a known truth. Consider a  $p$ -dimensional VAR(1) process  $\{X_t\}_{t \geq 1}$  such that

$$X_t = \Phi X_{t-1} + \epsilon_t$$

where  $X_t \in \mathbb{R}^p$ ,  $\Phi$  is a  $p \times p$  matrix,  $\epsilon_t \stackrel{iid}{\sim} N(0, \Omega)$ , and  $\Omega$  is a positive definite  $p \times p$  matrix. We fix  $\Omega$  be a AR correlation matrix with parameter .9. The invariant distribution for this Markov chain is  $N(0, \Psi)$ , where  $\text{vec}(\Psi) = (I_{p^2} - \Phi \times \Phi)^{-1} \text{vec}(W)$ . For  $k \geq 0$ , the lag- $k$  autocovariance is  $\Upsilon(k) = \Gamma(k) = \Phi^k \Psi$ . The process satisfies a CLT for  $\bar{X} = n^{-1} \sum_{t=1}^n X_t$  if the spectral norm of  $\Phi$  is less than 1 (Tjstheim, 1990) and the limiting covariance,  $\Sigma$  is known in closed form (Dai and Jones, 2017). We set  $p = 2$  and set  $\Phi$  such that its eigenvalues are .999 and .001. The resulting process mixes slowly and allows us to highlight the difference in the estimation of ACvF. We further set  $m = 5$  with starting values dispersed across the state space.

We compare the the locally and globally-centered autocorrelations against the truth. In Figure 2 are the estimated ACF plots for the first component of the second chain against the truth in red.

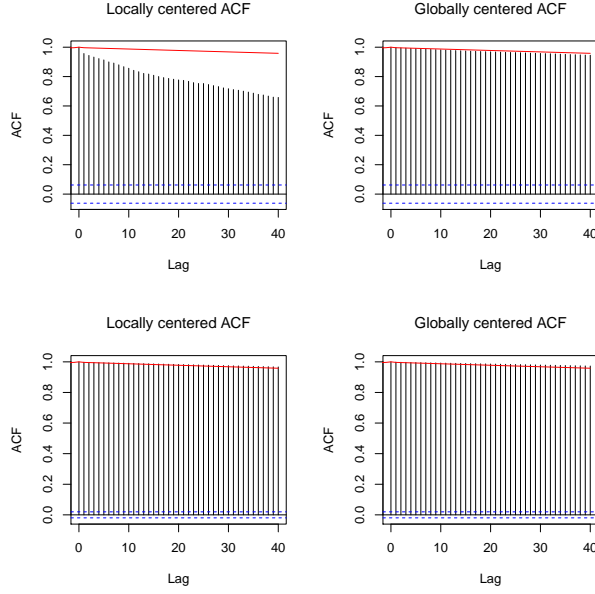


Figure 2: VAR. ACF (left) and G-ACF (right) for the second chain for  $m = 5$ . (Top)  $n = 10^3$  and (bottom)  $n = 10^4$ . The red line is the true ACF.

For a run length of  $10^3$  (top row), the commonly used locally centered ACF underestimates the true correlation structure giving a false sense of security about the mixing of the chain. The G-ACF estimator, on the other hand, is able to more accurately measure the ACF of the chain. This difference is negligible at the larger run length of  $n = 10^4$  when each of the 5 chains have sufficiently explored the state space.

Since  $\mu$  is known, we assess the performance of A-SV and G-SV estimations by assessing coverage probabilities of 95% Wald confidence regions over 1000 replications. Table 1 shows that for each sample size, the  $\hat{\Sigma}_G$  results in close to nominal coverage probability, whereas  $\hat{\Sigma}_A$  yields critically low coverage. The low coverage is a consequence of underestimating the autocovariances. It is only at the sample size of  $n = 10^5$  that  $\hat{\Sigma}_A$  yields close to nominal coverage.

$n$	1000	5000	10000	50000	100000
A-SV	0.710	0.843	.885	.928	.944
G-SV	0.956	0.937	.924	.945	.952

Table 1: VAR. Coverage probabilities at 95% nominal level. Replications =  $10^3$ .

The quality of estimation of  $\Sigma$  and ESS is assessed by running plots from 50 replications of run length 50000. In Figure 3, we present running plots of  $\log(\|\hat{\Sigma}\|_F)$  for both  $\hat{\Sigma}_G$  and  $\hat{\Sigma}_A$  and the running plots of  $\log(\widehat{\text{ESS}})/mn$  for both  $\widehat{\text{ESS}}_G$  and  $\widehat{\text{ESS}}_A$ . It is evident that  $\hat{\Sigma}_A$  severely underestimates the truth, leading to an overestimation of ESS. The G-SV estimator is able to estimate  $\Sigma$  more

accurately early on, safeguarding against early termination using ESS.

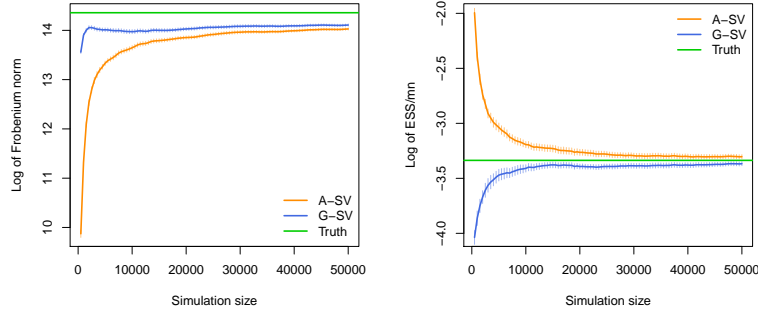


Figure 3: VAR(1). (Left) Running plot for logarithm of Frobenius norm of A-SV and G-SV estimator. (Right) Running plot for logarithm of  $\bar{\text{ESS}}/mn$  using A-SV and G-SV estimator.

## 5.2 Boomerang Distribution

Consider the following family of bimodal bivariate distributions introduced by Gelman and Meng (1991), which we term as a boomerang distribution. For  $A \geq 0$  and  $B, C \in \mathbb{R}$ , the target density is

$$f(x, y) \propto \exp \left( -\frac{1}{2} \left[ Ax^2y^2 + x^2 + y^2 - 2Bxy - 2Cx - 2Cy \right] \right).$$

We run a deterministic scan Gibbs sampler to sample from this target using the following full conditional densities

$$\begin{aligned} x | y &\sim N \left( \frac{By + C}{Ay^2 + 1}, \frac{1}{Ay^2 + 1} \right) \\ y | x &\sim N \left( \frac{Bx + C}{Ax^2 + 1}, \frac{1}{Ax^2 + 1} \right). \end{aligned}$$

We consider two settings; in setting 1,  $A = 1$ ,  $B = 3$ ,  $C = 8$  which results in well-separated modes. In setting 2, we let  $A = 1$ ,  $B = 10$ ,  $C = 7$  which yields a boomerang shape for the contours. The contour plots for these two settings are in the left in Figure 4, overlaid with scatter plots of two parallel runs of the Gibbs sampler. Setting 2 is chosen specifically to illustrate that the locally and globally-centered ACvFs perform similarly when the Markov chain moves freely in the state space. We run  $m = 5$  parallel Markov chains with starting points evenly distributed across the state space. For setting 1, when the Markov chains have not been able to jump modes, locally-centered ACF severely underestimates autocorrelation. This is seen in the top-middle plot of Figure 4, where the locally-centered autocorrelations at  $n = 1000$  (histogram) are drastically different from the locally-centered autocorrelations at  $n = 10^4$  (overlaid blue line). Somewhere between  $n = 1000$  and

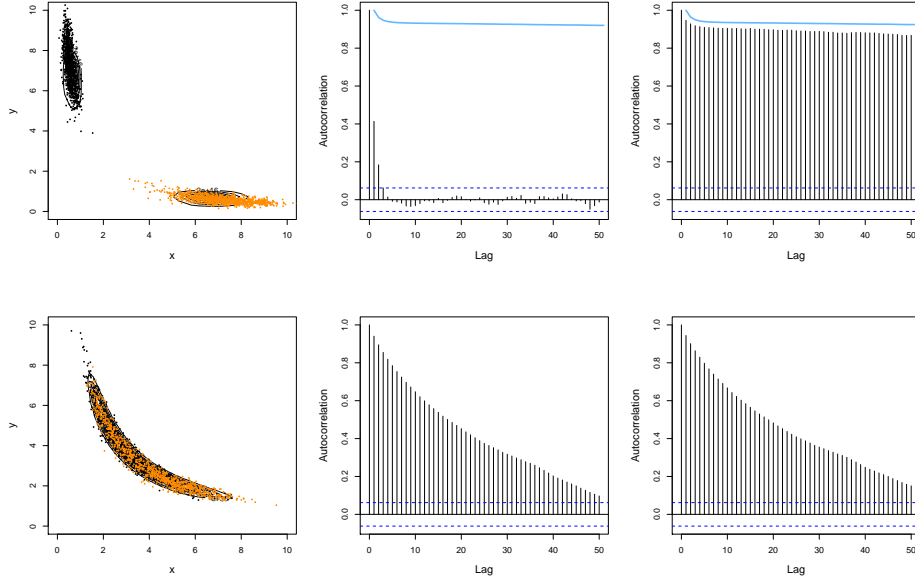


Figure 4: Boomerang. Top row is Setting 1 and bottom row is Setting 2. Contour plots of the target distributions overlaid with scatter plots for two chains for both settings. Locally-centered ACF (middle) and G-ACF (right) plots for one chain for  $n = 1000$  with the blue line being the respective ACF at  $n = 10^4$ .

$n = 10^4$ , the Markov chains jump modes and it is only then that the locally-centered ACFs provide better estimates. The G-ACFs, on the other hand, produce similar ACF estimates at  $n = 1000$  and  $n = 10^4$  by measuring deviations about the global mean. For setting 2, both methods yield similar ACFs reinforcing our claim that there is much to gain by using G-ACvF and nothing to lose. **At what sample size?**

$n$	$m = 2$		$m = 5$	
	A-SV	G-SV	A-SV	G-SV
5000	0.595	0.700	0.402	0.640
10000	0.563	0.665	0.59	0.739
50000	0.775	0.814	0.807	0.864
100000	0.847	0.864	0.884	0.902

Table 2: Setting 1: Coverage probabilities from  $10^3$  replications.

$n$	$m = 2$		$m = 5$	
	A-SV	G-SV	A-SV	G-SV
1000	0.856	0.868	0.895	0.910
5000	0.921	0.925	0.910	0.915
10000	0.928	0.93	0.919	0.926
50000	0.943	0.944	0.951	0.952

Table 3: Setting 2: Coverage probabilities from  $10^3$  replications.

The true mean of the target distribution can be obtained using numerical approximation. Using the A-SV and G-SV estimators, we construct 95% confidence regions and report coverage probabilities for 1000 replications for both  $m = 2$  and  $m = 5$ . Tables 2 and 3 reports all results. In setting

1, systematically, the G-SV estimator yields far superior coverage than the A-SV estimator for all values of  $n$ . Whereas for setting 2, the results are almost similar indicating the equivalence of A-SV and G-SV estimator for fast mixing Markov chains.

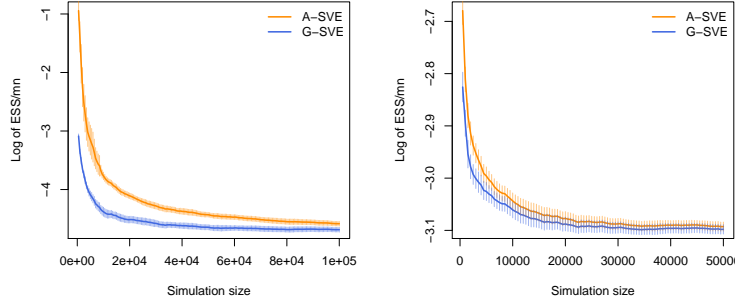


Figure 5: Boom: Running plot of  $\log(\widehat{\text{ESS}}_A)/mn$  and  $\log(\widehat{\text{ESS}}_G)/mn$  with  $m = 5$  for setting 1 (left) and setting 2 (right).

In Figure 5 we present running plots of estimates of  $\log(\text{ESS})/mn$  for both setting 1 and setting 2. As the sample size increases,  $\widehat{\text{ESS}}_G$  and  $\widehat{\text{ESS}}_A$  become closer, but early on for setting 1,  $\widehat{\text{ESS}}_G$  estimates are much smaller, safeguarding users against early termination.

### 5.3 Sensor network localization

Consider the sensor network localization problem Overleaf Ihler et al. (2005) where the goal is to identify unknown sensor locations using noisy distance data. We use the data and the model as specified by Tak et al. (2018). There are four sensors scattered on a planar region where  $x_i = (x_{i1}, x_{i2})^T$  denotes the coordinates of the  $i$ th sensor. Let  $y_{ij}$  denote the distance between the sensors  $x_i$  and  $x_j$  and if observed, yields  $z_{ij} = 1$  otherwise,  $z_{ij} = 0$ . The complete model is then,

$$z_{ij} \mid x_1, \dots, x_4 \sim \text{Bernoulli} \left( \exp \left( \frac{-\|x_i - x_j\|^2}{2R^2} \right) \right)$$

$$y_{ij} \mid z_{ij} = 1, x_i, x_j \sim N \left( \|x_i - x_j\|^2, \sigma^2 \right).$$

Tak et al. (2018) set  $R = 0.3$  and  $\sigma = 0.02$  and use independent  $N(0, 100I_2)$  priors on the locations. Distance  $y_{ij}$  is specified only if  $z_{ij} = 1$ . The 8-dimensional posterior of  $(x_1, x_2, x_3, x_4)$  is intractable with unknown full conditionals. A Metropolis-within-Gibbs type sampler is implemented with each full conditional employing the repelling attractive Metropolis (RAM) algorithm of Tak et al. (2018). The RAM algorithm runs Markov chains with higher jumping frequency between the modes. We run  $m = 5$  parallel Markov chains with well-separated starting points. Coverage probabilities are not estimable since the true posterior mean is unknown. Trace plot of  $x_{11}$  for two Markov chains

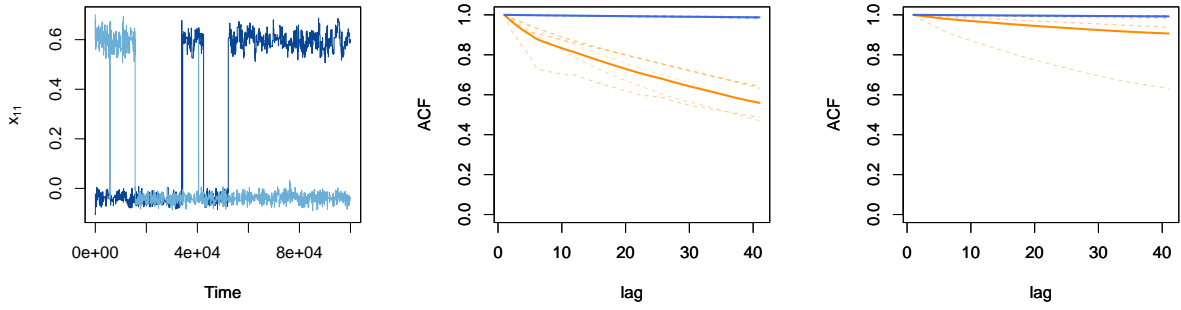


Figure 6: Sensor: Trace plot of  $x_{11}$  (left) for two parallel chains. Average locally-centered ACF (solid orange) and G-ACF (solid blue) at  $n = 5000$  (middle) and  $n = 50000$  (right). Dashed lines are individual chain estimates.

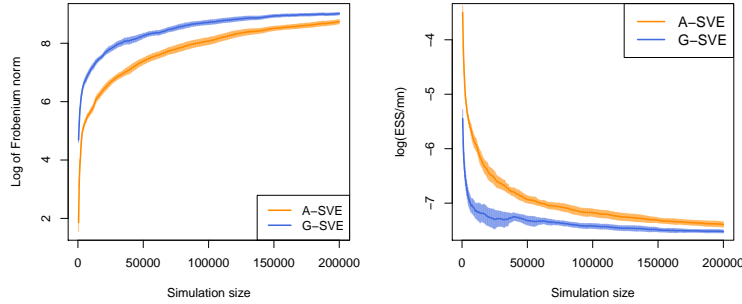


Figure 7: Sensor: Running plot of  $\log(\|\Sigma\|_F)$  (left) and  $\log(\text{ESS})/mn$  (right) estimated using A-SV and G-SV along with standard errors for 10 replications.

is shown in Figure 6. Figure 6 also plots locally and globally-centered ACFs. As seen in previous examples, early estimates of locally-centered ACFs are much smaller than later estimates. The globally-centered ACFs, on the other hand are similar at small and large sample sizes. Figure 7 presents the running plot of  $\log \|\Sigma\|_F$  and  $\log \text{ESS}/mn$  using A-SV and G-SV estimators along with standard errors from 10 replications. In both the plots, G-SV estimator and  $\widehat{\text{ESS}}_G/mn$  reach stability significantly earlier than A-SV estimator and  $\widehat{\text{ESS}}_A/mn$ .

## 6 Discussion

For slow mixing Markov chains, a naive average of locally-centered ACvFs can dramatically underestimate the truth. This has a severe impact on ACF plots, Monte Carlo variance, and stopping time of MCMC algorithms. We provide a globally-centered estimate of the ACvF that leads to improvements in all three aspects of MCMC output analysis.

In our simulations, we use the Bartlett lag window, which is convenient but not optimal. The untruncated quadratic spectral lag window of Andrews (1991) is known to be mean-squared optimal and could easily be used here to obtain  $\hat{\Sigma}_G$ . The quadratic spectral lag window is typically slow to compute, but by adapting Heberle and Sattarhoff's algorithm, the implementation is similar to that of Bartlett's.

Another class of estimators for asymptotic variance of Monte Carlo averages are the multivariate initial sequence (mIS) estimators (Dai and Jones, 2017). Similar to SV estimator, ACvFs are a fundamental part of mIS estimators and underestimation in ACvFs yields underestimates of  $\Sigma$ . It is easy to show that  $\hat{\Upsilon}_{G,s}(k)$  is a strongly consistent estimators for  $\Upsilon(k)$ . Replacing  $\hat{\Upsilon}_s(k)$  for  $\hat{\Upsilon}_{G,s}(k)$  and following the proof of Theorem 2 in (Dai and Jones, 2017) will yield consistent overestimation of the generalized variance  $|\Sigma|$  estimated using a globally-centered version of the mIS estimator. A valuable line of research would be to carry out the theoretical and experimental analysis of mIS estimators using G-ACvF.

## 7 Appendix

### 7.1 Preliminaries

Let  $B^{(i)}$  denote the  $i$ th component of the  $p$ -dimensional standard Brownian motion  $B$ .

**Lemma 1.** (*Csörgö and Révész (1981)*). *Let  $b_n$  be a sequence such that  $b_n$  and  $n/b_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Then for all  $\epsilon > 0$  and for almost all sample paths, there exists  $n_0(\epsilon)$  such that  $\forall n \geq n_0(\epsilon)$*

$$\begin{aligned} \sup_{0 \leq t \leq n-b_n} \sup_{0 \leq s \leq b_n} \left| B^{(i)}(t+s) - B^{(i)}(t) \right| &< (1+\epsilon) \left( 2b_n \left( \log \frac{n}{b_n} + \log \log n \right) \right)^{1/2}, \\ \sup_{0 \leq s \leq b_n} \left| B^{(i)}(n) - B^{(i)}(n-s) \right| &< (1+\epsilon) \left( 2b_n \left( \log \frac{n}{b_n} + \log \log n \right) \right)^{1/2}, \text{ and} \\ \left| B^{(i)}(n) \right| &< (1+\epsilon) \sqrt{2n \log \log n}. \end{aligned}$$

In order to generalize summations for positive and negative lags in the following proofs, we define the sets  $I_k, J_{k1}$  and  $J_{k2}$  as  $I_k := \{1, \dots, n-k\}, J_{k1} := \{n-k+1, \dots, n\}, J_{k2} := \{1, \dots, k\}$  for  $k > 0$  and  $I_k := \{1-k, \dots, n\}, J_{k1} := \{1, \dots, -k\}, J_{k2} := \{n+k+1, \dots, n\}$  for  $k < 0$ . For  $k = 0$ ,  $J_{k1}$  and  $J_{k2}$  are empty sets and  $I_k := \{1, \dots, n\}$ . Notice that the empirical autocovariance estimator at lag- $k$  requires a summation over  $I_k$ .

## 7.2 Proof of Theorem 1

We can break  $\hat{\Gamma}_{G,s}$  into four parts for all  $k \geq 1$  as:

$$\begin{aligned}
\hat{\Gamma}_{G,s}(k) &= \frac{1}{n} \sum_{t \in I_k} (X_{st} - \bar{X}) (X_{s(t+k)} - \bar{X})^T \\
&= \left[ \frac{1}{n} \sum_{t \in I_k} (X_{st} - \bar{X}_s) (X_{s(t+k)} - \bar{X}_s)^T \right] + \left[ \frac{1}{n} \sum_{t \in J_{k1}} (\bar{X}_s - X_{st}) (\bar{X}_s - \bar{X})^T \right] \\
&\quad + \left[ \frac{1}{n} \sum_{t \in J_{k2}} (\bar{X}_s - \bar{X}) (\bar{X}_s - X_{st})^T \right] + \left[ \frac{n-|k|}{n} (\bar{X}_s - \bar{X}) (\bar{X}_s - \bar{X})^T \right] \\
&= \hat{\Gamma}_s(k) - \frac{1}{n} \sum_{t \in J_{k1}} A_{st} - \frac{1}{n} \sum_{t \in J_{k2}} A_{st}^T + \frac{n-|k|}{n} (\bar{X}_s - \bar{X}) (\bar{X}_s - \bar{X})^T, \tag{8}
\end{aligned}$$

where  $A_{st} = (X_{st} - \bar{X}_s)(\bar{X}_s - \bar{X})^T$ . Under the assumption of stationarity, we will study the expectations of each of the above terms. Without loss of generality, consider  $A_{11}$ ,

$$\begin{aligned}
\mathbb{E}[A_{11}] &= \mathbb{E} \left[ (X_{11} - \bar{X}_1) (\bar{X}_1 - \bar{X})^T \right] \\
&= \mathbb{E} \left[ X_{11} \bar{X}_1^T \right] - \frac{1}{m} \mathbb{E} \left[ X_{11} \bar{X}_1^T \right] - \frac{m-1}{m} \mathbb{E} \left[ X_{11} \bar{X}_2^T \right] + \frac{1}{m} \mathbb{E} \left[ \bar{X}_1 \bar{X}_1^T \right] + \frac{m-1}{m} \mathbb{E} \left[ \bar{X}_1 \bar{X}_2^T \right] - \mathbb{E} \left[ \bar{X}_1 \bar{X}_1^T \right] \\
&= \frac{m-1}{m} \left( \mathbb{E} \left[ X_{11} \bar{X}_1^T \right] - \mathbb{E} \left[ X_{11} \bar{X}_2^T \right] + \mathbb{E} \left[ \bar{X}_1 \bar{X}_2^T \right] - \mathbb{E} \left[ \bar{X}_1 \bar{X}_1^T \right] \right) \\
&= \frac{m-1}{m} \left( \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[ X_{11} X_{1t}^T \right] - \mathbb{E} \left[ X_{11} \right] \mathbb{E} \left[ \bar{X}_2^T \right] + \mathbb{E} \left[ \bar{X}_1 \right] \mathbb{E} \left[ \bar{X}_2^T \right] - \text{Var} \left[ \bar{X}_1 \right] - \mathbb{E} \left[ \bar{X}_1 \right] \mathbb{E} \left[ \bar{X}_1^T \right] \right) \\
&= \frac{m-1}{mn} \left( \sum_{k=0}^{n-1} \Gamma(k) - n \text{Var} \left[ \bar{X}_1 \right] \right). \tag{9}
\end{aligned}$$

Similarly,

$$\mathbb{E} \left[ A_{11}^T \right] = \mathbb{E} \left[ A_{11} \right]^T = \frac{m-1}{mn} \left( \sum_{k=0}^{n-1} \Gamma(k)^T - n \text{Var} \left[ \bar{X}_1 \right] \right). \tag{10}$$

Further,

$$\begin{aligned}
\mathbb{E} \left[ (\bar{X}_1 - \bar{X}) (\bar{X}_1 - \bar{X})^T \right] &= \mathbb{E} \left[ \bar{X}_1 \bar{X}_1^T - \bar{X}_1 \bar{X}^T - \bar{X} \bar{X}_1^T + \bar{X} \bar{X}^T \right] \\
&= \left( \text{Var}(\bar{X}_1) + \mu \mu^T - \text{Var}(\bar{X}) - \mu \mu^T \right) = \frac{m-1}{m} \text{Var}(\bar{X}_1). \tag{11}
\end{aligned}$$



Additionally,  $\hat{\Gamma}_s(k)$  exhibits the following expectation (from Priestley (1981))

$$\mathbb{E}[\hat{\Gamma}_s(k)] = \left(1 - \frac{|k|}{n}\right) (\Gamma(k) - \text{Var}(\bar{X}_s)) . \quad (12)$$

Using (9), (10), (11), and (12) in (8),

$$\begin{aligned} & \mathbb{E}[\hat{\Gamma}_{G,s}(k)] \\ &= \mathbb{E}[\hat{\Gamma}_s(k)] - \frac{1}{n} \left( \sum_{t \in J_{k1}} \mathbb{E}[A_{1t}] + \sum_{t \in J_{k2}} \mathbb{E}[A_{1t}^T] \right) + \left(1 - \frac{|k|}{n}\right) \left(1 - \frac{1}{m}\right) \text{Var}(\bar{X}_1) \\ &= \mathbb{E}[\hat{\Gamma}_s(k)] - \frac{|k|}{n} \left(1 - \frac{1}{m}\right) \left( \frac{1}{n} \sum_{h=0}^{n-1} \Gamma(h) + \frac{1}{n} \sum_{h=0}^{n-1} \Gamma(h)^T - 2\text{Var}(\bar{X}_1) \right) + \left(1 - \frac{|k|}{n}\right) \left(1 - \frac{1}{m}\right) \text{Var}(\bar{X}_1) \\ &= \left(1 - \frac{|k|}{n}\right) \left( \Gamma(k) - \frac{\text{Var}(\bar{X}_1)}{m} \right) + o(n^{-1}) \end{aligned}$$

By (Song and Schmeiser, 1995, Proposition 1),

$$\text{Var}(\bar{X}_s) = \frac{\Sigma}{n} + \frac{\Phi}{n^2} + o(n^{-2}) .$$

As a consequence, we get the result

$$\mathbb{E}[\hat{\Gamma}_{G,s}(k)] = \left(1 - \frac{|k|}{n}\right) \left( \Gamma(k) - \frac{\Sigma}{mn} - \frac{\Phi}{mn^2} \right) + o(n^{-1}) .$$

Although this completes the proof of Theorem 1, we will require the following decomposition

$$\mathbb{E}[\hat{\Gamma}_{G,s}(k)] = \left(1 - \frac{|k|}{n}\right) \Gamma(k) + O_1 + O_2 . \quad (13)$$

where,

$$\begin{aligned} O_1 &= -\frac{|k|}{n} \left[ \left(1 - \frac{1}{m}\right) \left( \frac{1}{n} \sum_{h=0}^{n-1} \Gamma(h)^T + \frac{1}{n} \sum_{h=0}^{n-1} \Gamma(h) \right) - \left(2 - \frac{1}{m}\right) \left( \frac{\Sigma}{n} + \frac{\Phi}{n^2} \right) \right] + o(n^{-2}) , \\ O_2 &= -\frac{1}{m} \left( \frac{\Sigma}{n} + \frac{\Phi}{n^2} \right) + o(n^{-2}) \end{aligned}$$

### 7.3 Strong consistency of G-SV estimator

Consider pseudo autocovariance and spectral variance estimators for the  $s$ th chain, denoted by  $\tilde{\Upsilon}_s(k)$  and  $\tilde{\Sigma}_s$  that use data centered around the unobserved actual mean  $\mu_g$ :

$$\tilde{\Upsilon}_s(k) = \frac{1}{n} \sum_{t \in I_k} (Y_{st} - \mu_g)(Y_{s(t+k)} - \mu_g)^T \quad \text{and} \quad \tilde{\Sigma}_s = \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \tilde{\Upsilon}_s(k).$$

The average pseudo spectral variance estimator is  $\tilde{\Sigma}_A = m^{-1} \sum_{s=1}^m \tilde{\Sigma}_s$ . Further, let

$$M_1 = \frac{1}{m} \sum_{s=1}^m \left\{ \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \sum_{t \in I_k} \frac{1}{n} \left[ (Y_{st} - \mu_g)_i (\mu_g - \bar{Y})_j + (\mu_g - \bar{Y})_i (Y_{s(t+k)} - \mu_g)_j \right] \right\},$$

$$M_2 = (\mu_g - \bar{Y})_i (\mu_g - \bar{Y})_j \sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n}\right) w\left(\frac{k}{b_n}\right).$$

**Lemma 2.** *Let Assumption 1 holds, then for the G-SV estimator,  $\hat{\Sigma}_G^{ij} = \tilde{\Sigma}_A^{ij} + M_1 + M_2$  and*

$$|M_1 + M_2| \leq D^2 g_1(n) + D g_2(n) + g_3(n),$$

where for some constant  $C$ ,

$$g_1(n) = (4c + C) \frac{b_n \psi^2(n)}{n^2} - 4c \frac{\psi^2(n)}{n^2} \rightarrow 0$$

$$g_2(n) = 2\sqrt{2} \|L\| p^{1/2} (1 + \epsilon) \left[ (4c + C) \frac{b_n \psi(n) \sqrt{n \log \log n}}{n^2} - 4c \frac{\psi(n) \sqrt{n \log \log n}}{n^2} \right] \rightarrow 0$$

$$g_3(n) = \|L\|^2 p (1 + \epsilon)^2 \left[ (4c + C) \frac{b_n \log \log n}{n} - 4c \frac{\log \log n}{n} \right] \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

*Proof.* The proof follows from standard algebraic calculations and is presented here for completeness. Consider,

$$\begin{aligned} \hat{\Sigma}_G^{ij} &= \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t \in I_k} (Y_{st} - \bar{Y})_i (Y_{s(t+k)} - \bar{Y})_j \\ &= \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t \in I_k} \left[ (Y_{st} - \mu_g)_i (Y_{s(t+k)} - \mu_g)_j + (Y_{st} - \mu_g)_i (\mu_g - \bar{Y})_j \right. \\ &\quad \left. + (\mu_g - \bar{Y})_i (Y_{s(t+k)} - \mu_g)_j + (\mu_g - \bar{Y})_i (\mu_g - \bar{Y})_j \right] \\ &= \tilde{\Sigma}_A^{ij} + \left[ (\mu_g - \bar{Y})_i (\mu_g - \bar{Y})_j \sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n}\right) w\left(\frac{k}{b_n}\right) \right] \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \sum_{t \in I_k} \left[ \frac{1}{n} (Y_{st} - \mu_g)_i (\mu_g - \bar{Y})_j + \frac{1}{n} (\mu_g - \bar{Y})_i (Y_{s(t+k)} - \mu_g)_j \right] \\
& = \tilde{\Sigma}_A^{ij} + M_1 + M_2.
\end{aligned}$$

Consequently

$$|\hat{\Sigma}_G^{ij} - \tilde{\Sigma}_A^{ij}| = |M_1 + M_2| \leq |M_1| + |M_2|.$$

We first present a result which will be useful later. For any Markov chain  $s$ ,

$$\begin{aligned}
\|\bar{Y}_s - \mu_g\|_\infty & \leq \|\bar{Y}_s - \mu_g\| = \frac{1}{n} \left\| \sum_{t=1}^n Y_{st} - n\mu_g \right\| \\
& \leq \frac{1}{n} \left\| \sum_{t=1}^n Y_{st} - n\mu_g - LB(n) \right\| + \frac{\|LB(n)\|}{n} \\
& < \frac{D\psi(n)}{n} + \frac{\|LB(n)\|}{n} \\
& < \frac{D\psi(n)}{n} + \frac{1}{n} \|L\| \left( \sum_{i=1}^p |B^{(i)}(n)|^2 \right)^{1/2} \\
& \leq \frac{D\psi(n)}{n} + \frac{1}{n} \|L\| p^{1/2} (1 + \epsilon) \sqrt{2n \log \log n}. \tag{14}
\end{aligned}$$

Similarly,

$$\|\bar{Y} - \mu_g\|_\infty \leq \frac{D\psi(n)}{n} + \frac{1}{n} \|L\| p^{1/2} (1 + \epsilon) \sqrt{2n \log \log n}. \tag{15}$$

Now consider,

$$\begin{aligned}
& |M_1| \\
& \leq \frac{1}{m} \sum_{s=1}^m \left\{ \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \left[ \frac{1}{n} \left| \sum_{t \in I_k} (Y_{st} - \mu_g)_i \right| |(\mu_g - \bar{Y})_j| + \frac{1}{n} |(\mu_g - \bar{Y})_i| \left| \sum_{t \in I_k} (Y_{s(t+k)} - \mu_g)_j \right| \right] \right\} \\
& \leq \frac{c\|(\bar{Y} - \mu_g)\|_\infty}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left[ \frac{1}{n} \left\| \sum_{t \in I_k} (Y_{st} - \mu_g) \right\|_\infty + \frac{1}{n} \left\| \sum_{t \in I_k} (Y_{s(t+k)} - \mu_g) \right\|_\infty \right] \\
& \leq \frac{c\|(\bar{Y} - \mu_g)\|_\infty}{m} \\
& \quad \times \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left[ \frac{1}{n} \left\| \sum_{t \in J_{k1}} (Y_{st} - \mu_g) - n(\bar{Y}_s - \mu_g) \right\|_\infty + \frac{1}{n} \left\| \sum_{t \in J_{k2}} (Y_{st} - \mu_g) - n(\bar{Y}_s - \mu_g) \right\|_\infty \right]
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{c\|(\bar{Y} - \mu_g)\|_\infty}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left[ \frac{1}{n} \left\| \sum_{t \in J_{k1}} (Y_{st} - \mu_g) \right\|_\infty + \frac{1}{n} \left\| \sum_{t \in J_{k2}} (Y_{st} - \mu_g) \right\|_\infty + 2\|\bar{Y}_s - \mu_g\|_\infty \right] \\
&\leq \frac{c\|(\bar{Y} - \mu_g)\|_\infty}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \frac{1}{n} \left[ \left\| \sum_{t \in J_{k1}} (Y_{st} - \mu_g) \right\|_\infty + \left\| \sum_{t \in J_{k2}} (Y_{st} - \mu_g) \right\|_\infty \right] \\
&\quad + 2c(2b_n - 1)\|\bar{Y} - \mu_g\|_\infty \|\bar{Y}_h - \mu_g\|_\infty \text{ for some } h \in \{1, \dots, m\}.
\end{aligned}$$

Using SIP on summation of  $k$  terms, we obtain the following upper bound for  $|M_1|$

$$\begin{aligned}
|M_1| &< 2c\|(\bar{Y} - \mu_g)\|_\infty \left[ \sum_{k=-b_n+1}^{b_n-1} \left[ \frac{D\psi(k)}{n} + \frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2k \log \log k}}{n} \right] + (2b_n - 1)\|\bar{Y}_h - \mu_g\|_\infty \right] \\
&\leq 2c(2b_n - 1)\|(\bar{Y} - \mu_g)\|_\infty \left[ \frac{D\psi(n)}{n} + \frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{n \log \log n}}{n} + \|\bar{Y}_h - \mu_g\|_\infty \right] \\
&\leq 4c(2b_n - 1) \left[ \frac{D\psi(n)}{n} + \frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{n \log \log n}}{n} \right]^2 \quad (\text{by (14) and (15)}). \tag{16}
\end{aligned}$$

For  $M_2$ ,

$$\begin{aligned}
|M_2| &= \left| \frac{1}{m} \sum_{s=1}^m \left\{ (\mu_g - \bar{Y})_i (\mu_g - \bar{Y})_j \sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n}\right) w\left(\frac{k}{b_n}\right) \right\} \right| \\
&\leq \|\bar{Y} - \mu_g\|_\infty^2 \left| \sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n}\right) w\left(\frac{k}{b_n}\right) \right| < \|\bar{Y} - \mu_g\|_\infty^2 \left[ \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \right] \\
&\leq b_n \|\bar{Y} - \mu_g\|_\infty^2 \int_{-\infty}^{\infty} |w(x)| dx \\
&\leq Cb_n \left[ \frac{D\psi(n)}{n} + \frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{n \log \log n}}{n} \right]^2 \quad \text{for some constant } C \text{ (by (15))}. \tag{17}
\end{aligned}$$

Using (16) and (17),

$$|M_1 + M_2| \leq |M_1| + |M_2| = D^2 g_1(n) + D g_2(n) + g_3(n),$$

where

$$g_1(n) = (8c + C) \frac{b_n \psi^2(n)}{n^2} - 4c \frac{\psi^2(n)}{n^2}$$

$$g_2(n) = 2\sqrt{2}\|L\|p^{1/2}(1+\epsilon) \left[ (8c+C) \frac{b_n\psi(n)\sqrt{n\log\log n}}{n^2} - 4c \frac{\psi(n)\sqrt{n\log\log n}}{n^2} \right]$$

$$g_3(n) = \|L\|^2 p(1+\epsilon)^2 \left[ (8c+C) \frac{b_n \log\log n}{n} - 4c \frac{\log\log n}{n} \right].$$

Under the assumptions of Theorem 2,  $\psi(n) = o(n^{1/2})$ . Using the law of iterative logarithms (LIL), a tighter bound for  $\psi(n)$  is given by Strassen (1964) as  $o(\sqrt{n\log\log n})$ . Since  $b_n \log\log n/n \rightarrow 0$  as  $n \rightarrow \infty$ , consequently,  $b_n\psi^2(n)/n^2 \rightarrow 0$ ,  $\psi^2(n)/n^2 \rightarrow 0$ ,  $b_n\psi(n)\sqrt{n\log\log n}/n^2 \rightarrow 0$ , and  $\psi(n)\sqrt{n\log\log n}/n^2 \rightarrow 0$ . Thus,  $g_1(n), g_2(n)$  and  $g_3(n) \rightarrow 0$  as  $n \rightarrow \infty$ .  $\square$

*Proof of Theorem 3.* We have the following decomposition,

$$\begin{aligned} \hat{\Sigma}_A^{ij} &= \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t \in I_k} (Y_{st} \pm \bar{Y}_s - \mu_g)_i (Y_{s(t+k)} \pm \bar{Y}_s - \mu_g)_j \\ &= \hat{\Sigma}_A^{ij} + \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t \in I_k} \left[ (Y_{st} - \bar{Y}_s)_i (\bar{Y}_s - \mu_g)_j + (\bar{Y}_s - \mu_g)_i (Y_{s(t+k)} - \bar{Y}_s)_j \right] \\ &\quad + \left[ \frac{1}{m} \sum_{s=1}^m (\bar{Y}_s - \mu_g)_i (\bar{Y}_s - \mu_g)_j \right] \left[ \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \left(1 - \frac{|k|}{b_n}\right) \right] \\ &= \hat{\Sigma}_A^{ij} + N_1 + N_2, \end{aligned}$$

where

$$N_1 = \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t \in I_k} \left[ (Y_{st} - \bar{Y}_s)_i (\bar{Y}_s - \mu_g)_j + (\bar{Y}_s - \mu_g)_i (Y_{s(t+k)} - \bar{Y}_s)_j \right]$$

$$N_2 = \left[ \frac{1}{m} \sum_{s=1}^m (\bar{Y}_s - \mu_g)_i (\bar{Y}_s - \mu_g)_j \right] \left[ \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \left(1 - \frac{|k|}{b_n}\right) \right].$$

Using the above breakdown and Lemma 2,

$$\left| \hat{\Sigma}_G^{ij} - \Sigma^{ij} \right| = \left| \hat{\Sigma}_A^{ij} - \Sigma^{ij} + N_1 + N_2 + M_1 + M_2 \right| \leq \left| \hat{\Sigma}_A^{ij} - \Sigma^{ij} \right| + |N_1| + |N_2| + |M_1 + M_2| \quad (18)$$

By the strong consistency of single-chain SV estimator, the first term goes to 0 with probability 1 and by Lemma 2, the third term goes to 0 with probability 1 as  $n \rightarrow \infty$ . It is left to show that  $|N_1| \rightarrow 0$  and  $|N_2| \rightarrow 0$  with probability 1. Consider,

$$|N_1| = \left| \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t \in I_k} \left[ (Y_{st} - \bar{Y}_s)_i (\bar{Y}_s - \mu_g)_j + (\bar{Y}_s - \mu_g)_i (Y_{s(t+k)} - \bar{Y}_s)_j \right] \right|$$

$$\begin{aligned}
&\leq \left| \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t \in I_k} \left[ (Y_{st} - \bar{Y}_s)_i (\bar{Y}_s - \mu_g)_j \right] \right| \\
&\quad + \left| \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t \in I_k} \left[ (\bar{Y}_s - \mu_g)_i (Y_{s(t+k)} - \bar{Y}_s)_j \right] \right|.
\end{aligned}$$

We will show that the first term goes to 0 and the proof for the second term is similar. Consider

$$\begin{aligned}
&\left| \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t \in I_k} \left[ (Y_{st} - \bar{Y}_s)_i (\bar{Y}_s - \mu_g)_j \right] \right| \\
&\leq \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \left| \frac{(\bar{Y}_s - \mu_g)_j}{n} \right| \left| \sum_{t \in J_{k1}} (\mu_g - Y_{st})_i + |k| (\bar{Y}_s - \mu_g)_i \right| \\
&\leq \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \frac{\|\bar{Y}_s - \mu_g\|_\infty}{n} \left\| \sum_{t \in J_{k1}} (\mu_g - Y_{st}) + |k| (\bar{Y}_s - \mu_g) \right\|_\infty \\
&\leq \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \frac{\|\bar{Y}_s - \mu_g\|_\infty}{n} \left( \left\| \sum_{t \in J_{k1}} (Y_{st} - \mu_g) \right\|_\infty + |k| \|\bar{Y}_s - \mu_g\|_\infty \right) \\
&\leq \frac{c}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \frac{\|\bar{Y}_s - \mu_g\|_\infty}{n} \left\| \sum_{t \in J_{k1}} (Y_{st} - \mu_g) \right\|_\infty + \frac{c}{m} \sum_{s=1}^m \frac{b_n(b_n-1)}{n} \|\bar{Y}_s - \mu_g\|_\infty^2.
\end{aligned}$$

Using SIP on the summation of  $k$  terms,

$$\begin{aligned}
&\left| \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-k} \left[ (Y_{st} - \bar{Y}_s)_i (\bar{Y}_s - \mu_g)_j \right] \right| \\
&< \frac{c}{m} \sum_{s=1}^m \|\bar{Y}_s - \mu_g\|_\infty \sum_{k=-b_n+1}^{b_n-1} \left[ \frac{D\psi(k)}{n} + \frac{\|L\| p^{1/2} (1+\epsilon) \sqrt{2k \log \log k}}{n} \right] + \frac{c}{m} \sum_{s=1}^m \frac{b_n(b_n-1)}{n} \|\bar{Y}_s - \mu_g\|_\infty^2 \\
&< \frac{c(2b_n-1)}{m} \sum_{s=1}^m \|\bar{Y}_s - \mu_g\|_\infty \left[ \frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2} (1+\epsilon) \sqrt{2n \log \log n}}{n} \right] + \frac{c}{m} \sum_{s=1}^m \frac{b_n(b_n-1)}{n} \|\bar{Y}_s - \mu_g\|_\infty^2 \\
&\leq c \left( 2b_n - 1 + \frac{b_n^2}{n} - \frac{b_n}{n} \right) \left[ \frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2} (1+\epsilon) \sqrt{2n \log \log n}}{n} \right]^2 \rightarrow 0. \quad (\text{by (14)})
\end{aligned}$$

Similarly, the second part of  $N_1 \rightarrow 0$  with probability 1. Following the steps in (17),

$$|N_2| \leq C b_n \left[ \frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2} (1+\epsilon) \sqrt{2n \log \log n}}{n} \right]^2 \rightarrow 0.$$

Thus, in (18), every term goes to 0 and  $\hat{\Sigma}_G^{ij} \rightarrow \Sigma^{ij}$  with probability 1 as  $n \rightarrow \infty$ .  $\square$

#### 7.4 Proof of Theorem 4

By Equation 13,

$$\mathbb{E} \left[ \hat{\Upsilon}_G(k) \right] = \left( 1 - \frac{|k|}{n} \right) \Upsilon(k) + O_1 + O_2.$$

where both  $O_1$  and  $O_2$  are the small order terms where  $O_1 = n^{-1}|k| \mathcal{O}(n^{-1})$  and  $O_2 = \mathcal{O}(n^{-1})$ . By our assumptions,  $\sum_{k=-\infty}^{\infty} \Upsilon(k) < \infty$ . For a truncation point  $b_n$ ,  $w(k/b_n) = 0$  for all  $|k| > b_n$ . Therefore, SV estimator can be written as a weighted sum of estimated autocovariances from lag  $-n+1$  to  $n-1$ . Consider the G-SV estimator,

$$\begin{aligned} \mathbb{E} \left[ \hat{\Sigma}_G - \Sigma \right] &= \sum_{k=-n+1}^{n-1} w \left( \frac{k}{b_n} \right) \mathbb{E} \left[ \hat{\Upsilon}_G(k) \right] - \sum_{k=-\infty}^{\infty} \Upsilon(k) \\ &= \sum_{k=-n+1}^{n-1} w \left( \frac{k}{b_n} \right) \left[ \left( 1 - \frac{|k|}{n} \right) \Upsilon(k) + O_1 + O_2 \right] - \sum_{k=-\infty}^{\infty} \Upsilon(k) \\ &= \sum_{k=-n+1}^{n-1} \left[ w \left( \frac{k}{b_n} \right) \left( 1 - \frac{|k|}{n} \right) \Upsilon(k) \right] - \sum_{k=-\infty}^{\infty} \Upsilon(k) + \sum_{k=-n+1}^{n-1} \left[ w \left( \frac{k}{b_n} \right) (O_1 + O_2) \right] \\ &= P_1 + P_2, \end{aligned} \tag{19}$$

where

$$\begin{aligned} P_1 &= \sum_{k=-n+1}^{n-1} \left[ w \left( \frac{k}{b_n} \right) \left( 1 - \frac{|k|}{n} \right) \Upsilon(k) \right] - \sum_{k=-\infty}^{\infty} \Upsilon(k) \text{ and} \\ P_2 &= \sum_{k=-n+1}^{n-1} \left[ w \left( \frac{k}{b_n} \right) (O_1 + O_2) \right]. \end{aligned}$$

Similar to Hannan (1970), we break  $P_1$  into three parts. Note that notation  $A = o(z)$  for matrix  $A$  implies  $A^{ij} = o(z)$  for every  $(i, j)$ th element of the matrix  $A$ . Consider,

$$P_1 = - \sum_{|k| \geq n} \Upsilon(k) - \sum_{k=-n+1}^{n-1} w \left( \frac{k}{b_n} \right) \frac{|k|}{n} \Upsilon(k) - \sum_{k=-n+1}^{n-1} \left( 1 - w \left( \frac{k}{b_n} \right) \right) \Upsilon(k). \tag{20}$$

We deal with the three subterms of term  $P_1$  individually. First,

$$- \sum_{|k| \geq n} \Upsilon(k) \leq \sum_{|k| \geq n} \left| \frac{k}{n} \right|^q \Upsilon(k) = \frac{1}{b_n^q} \left| \frac{b_n}{n} \right|^q \sum_{k \geq n} |k|^q \Upsilon(k) = o \left( \frac{1}{b_n^q} \right), \tag{21}$$

since  $\sum_{|k| \geq n} |k|^q \Upsilon(k) < \infty$ . Next,

$$\sum_{k=-n+1}^{n-1} w\left(\frac{k}{n}\right) \frac{|k|}{n} \Upsilon(k) \leq \frac{c}{n} \sum_{k=-n+1}^{n-1} |k| \Upsilon(k).$$

For  $q \geq 1$ ,

$$\frac{c}{n} \sum_{k=-n+1}^{n-1} |k| \Upsilon(k) \leq \frac{c}{n} \sum_{k=-n+1}^{n-1} |k|^q \Upsilon(k) = \frac{1}{b_n^q} \frac{b_n^q}{n} c \sum_{k=-n+1}^{n-1} |k|^q \Upsilon(k) = o\left(\frac{1}{b_n^q}\right).$$

For  $q < 1$ ,

$$\frac{c}{n} \sum_{k=-n+1}^{n-1} |k| \Upsilon(k) \leq c \sum_{k=-n+1}^{n-1} \left|\frac{k}{n}\right|^q \Upsilon(k) = \frac{1}{b_n^q} \frac{b_n^q}{n^q} c \sum_{k=-n+1}^{n-1} |k|^q \Upsilon(k) = o\left(\frac{1}{b_n^q}\right).$$

So,

$$\sum_{k=-n+1}^{n-1} w\left(\frac{k}{n}\right) \frac{|k|}{n} \Upsilon(k) = o\left(\frac{1}{b_n^q}\right) \quad (22)$$

Lastly, by our assumptions, for  $x \rightarrow 0$

$$\frac{1 - w(x)}{|x|^q} = k_q + o(1).$$

For  $x = k/b_n$ ,  $|k/b_n|^{-q} (1 - w(k/b_n))$  converges boundedly to  $k_q$  for each  $k$ . So,

$$\begin{aligned} \sum_{k=-n+1}^{n-1} \left(1 - w\left(\frac{k}{b_n}\right)\right) \Upsilon(k) &= -\frac{1}{b_n^q} \sum_{k=-n+1}^{n-1} \left(\frac{|k|}{b_n}\right)^{-q} \left(1 - w\left(\frac{k}{b_n}\right)\right) |k|^q \Upsilon(k) \\ &= -\frac{1}{b_n^q} \sum_{k=-n+1}^{n-1} [k_q + o(1)] |k|^q \Upsilon(k) \\ &= -\frac{k_q \Phi^{(q)}}{b_n^q} + o\left(\frac{1}{b_n^q}\right). \end{aligned} \quad (23)$$

Finally, we will solve for  $P_2$ . Note that  $O_1 = (|k|/n) \mathcal{O}(1/n)$  and  $O_2 = \mathcal{O}(1/n)$ . So,

$$\begin{aligned} P_2 &\leq \sum_{k=-n+1}^{n-1} \left| w\left(\frac{k}{b_n}\right) O_1 \right| + \sum_{k=-n+1}^{n-1} \left| w\left(\frac{k}{b_n}\right) O_2 \right| \\ &= \mathcal{O}\left(\frac{1}{n}\right) \sum_{k=-n+1}^{n-1} \frac{|k|}{n} \left| w\left(\frac{k}{b_n}\right) \right| + \mathcal{O}\left(\frac{1}{n}\right) \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \end{aligned}$$



$$\begin{aligned}
&= \mathcal{O}\left(\frac{b_n}{n}\right) \frac{1}{b_n} \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \\
&= \mathcal{O}\left(\frac{b_n}{n}\right) = o\left(\frac{b_n}{b_n^{q+1}}\right) = o\left(\frac{1}{b_n^q}\right).
\end{aligned} \tag{24}$$

Using (20), (21), (22), (23), and (24) in (19),

$$\mathbb{E} \left[ \hat{\Sigma}_G - \Sigma \right] = -\frac{k_q \Phi^{(q)}}{b_n^q} + o\left(\frac{1}{b_n^q}\right),$$

which completes the result.

## 7.5 Proof of Theorem 5

By Lemma 2,

$$\begin{aligned}
&\left| \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right| \\
&\leq \frac{1}{m} \sum_{s=1}^m \left| \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \sum_{t \in I_k} \left[ \left( \frac{(Y_{st} - \mu_g)_i (\mu_g - \bar{Y})_j}{n} \right) + \left( \frac{(\mu_g - \bar{Y})_i (Y_{s(t+k)} - \mu_g)_j}{n} \right) \right] \right| \\
&\quad + (\mu_g - \bar{Y})(\mu_g - \bar{Y})^T \sum_{k=-b_n+1}^{b_n-1} \left( \frac{n - |k|}{n} \right) w\left(\frac{k}{n}\right) \Big| < D^2 g_1(n) + D g_2(n) + g_3(n),
\end{aligned}$$

where  $g_1(n), g_2(n), g_3(n) \rightarrow 0$  as  $n \rightarrow \infty$  as defined in Lemma 2. Then there exists an  $N_0$  such that

$$\begin{aligned}
\left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right)^2 &= \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right)^2 I(0 \leq n \leq N_0) + \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right)^2 I(n > N_0) \\
&\leq \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right)^2 I(0 \leq n \leq N_0) + \left( D^2 g_1(n) + D g_2(n) + g_3(n) \right)^2 I(n > N_0) \\
&:= g_n^*(Y_{11}, \dots, Y_{1n}, \dots, Y_{m1}, \dots, Y_{mn}).
\end{aligned}$$

But since by assumption  $\mathbb{E} D^4 < \infty$  and the fourth moment is finite,

$$\mathbb{E} |g_n^*| \leq \mathbb{E} \left[ \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}_A^{ij} \right)^2 \right] + \mathbb{E} \left[ \left( D^2 g_1(n) + D g_2(n) + g_3(n) \right)^2 \right] < \infty.$$

Thus,  $\mathbb{E} |g_n^*| < \infty$  and further as  $n \rightarrow \infty$ ,  $g_n \rightarrow 0$  under the assumptions. Since  $g_1, g_2, g_3 \rightarrow 0$ ,  $\mathbb{E} g_n^* \rightarrow 0$ . By the majorized convergence theorem (Zeidler, 2013), as  $n \rightarrow \infty$ ,

$$\mathbb{E} \left[ \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right)^2 \right] \rightarrow 0. \tag{25}$$

We will use (25) to show that the variances are equivalent. Define,

$$\xi \left( \hat{\Sigma}_G^{ij}, \tilde{\Sigma}^{ij} \right) = \text{Var} \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right) + 2\mathbb{E} \left[ \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right) \left( \tilde{\Sigma}^{ij} - \mathbb{E} \left( \tilde{\Sigma}^{ij} \right) \right) \right].$$

We will show that the above is  $o(1)$ . Using Cauchy-Schwarz inequality followed by (25),

$$\begin{aligned} \left| \xi \left( \hat{\Sigma}_G^{ij}, \tilde{\Sigma}^{ij} \right) \right| &\leq \left| \text{Var} \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right) \right| + \left| 2\mathbb{E} \left[ \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right) \left( \tilde{\Sigma}^{ij} - \mathbb{E} \left( \tilde{\Sigma}^{ij} \right) \right) \right] \right| \\ &\leq \mathbb{E} \left[ \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right)^2 \right] + 2 \left| \left( \mathbb{E} \left[ \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right)^2 \right] \text{Var} \left( \tilde{\Sigma}^{ij} \right) \right)^{1/2} \right| \\ &= o(1) + 2 \left( o(1) \left( O \left( \frac{b_n}{n} \right) + o \left( \frac{b_n}{n} \right) \right) \right) = o(1). \end{aligned}$$

Finally,

$$\begin{aligned} \text{Var} \left( \hat{\Sigma}_G^{ij} \right) &= \mathbb{E} \left[ \left( \hat{\Sigma}_G^{ij} - \mathbb{E} \left[ \hat{\Sigma}_G^{ij} \right] \right)^2 \right] \\ &= \mathbb{E} \left[ \left( \hat{\Sigma}_G^{ij} \pm \tilde{\Sigma}^{ij} \pm \mathbb{E} \left[ \tilde{\Sigma}^{ij} \right] - \mathbb{E} \left[ \hat{\Sigma}_G^{ij} \right] \right)^2 \right] \\ &= \mathbb{E} \left[ \left( \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right) + \left( \tilde{\Sigma}^{ij} - \mathbb{E} \left[ \tilde{\Sigma}^{ij} \right] \right) + \left( \mathbb{E} \left[ \tilde{\Sigma}^{ij} \right] - \mathbb{E} \left[ \hat{\Sigma}_G^{ij} \right] \right) \right)^2 \right] \\ &= \mathbb{E} \left[ \left( \tilde{\Sigma}^{ij} - \mathbb{E} \left[ \tilde{\Sigma}^{ij} \right] \right)^2 \right] + \mathbb{E} \left[ \left( \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right) + \left( \mathbb{E} \left[ \tilde{\Sigma}^{ij} \right] - \mathbb{E} \left[ \hat{\Sigma}_G^{ij} \right] \right) \right)^2 \right] \\ &\quad + 2\mathbb{E} \left[ \left( \tilde{\Sigma}^{ij} - \mathbb{E} \left[ \tilde{\Sigma}^{ij} \right] \right) \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right) + 2 \left( \tilde{\Sigma}^{ij} - \mathbb{E} \left[ \tilde{\Sigma}^{ij} \right] \right) \left( \mathbb{E} \left[ \tilde{\Sigma}^{ij} \right] - \mathbb{E} \left[ \hat{\Sigma}_G^{ij} \right] \right) \right] \\ &= \text{Var} \left( \tilde{\Sigma}^{ij} \right) + \text{Var} \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right) + 2\mathbb{E} \left[ \left( \hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij} \right) \left( \tilde{\Sigma}^{ij} - \mathbb{E} \left( \tilde{\Sigma}^{ij} \right) \right) \right] + o(1) \\ &= \text{Var} \left( \tilde{\Sigma}^{ij} \right) + o(1). \end{aligned}$$

Hannan (1970) has given the calculations for variance of  $\tilde{\Sigma}$  as

$$\frac{n}{b_n} \text{Var}(\tilde{\Sigma}^{ij}) = \left[ \Sigma^{ii} \Sigma^{jj} + \left( \Sigma^{ij} \right) \right] \int_{-\infty}^{\infty} w^2(x) dx + o(1). \quad (26)$$

Plugging (26) into variance of  $\hat{\Sigma}_G$  gives the result of the theorem.

## 8 Additional Examples

We present two additional examples to illustrate the advantage of the G-ACF estimator.

### 8.1 Bayesian Poisson Change Point Model

Consider the militarized interstate dispute (MID) data of Martin et al. (2011) which describes the annual number of military conflicts in the United States. In order to detect the number and timings of the cyclic phases in international conflicts, we fit a Bayesian Poisson change-point model. Following Martin et al. (2011), we will use `MCMCpoissonChange` from `MCMCpack` to fit the model with six change-points which samples the latent states based on the algorithm in Chib (1998). The Poisson change-point model in `MCMCpoissonChange` uses conjugate priors and is the following:

$$\begin{aligned} y_t &\sim \text{Poisson}(\lambda_i), & i = 1, \dots, k \\ \lambda_i &\sim \text{Gamma}(c_o, d_o), & i = 1, \dots, k \\ p_{ii} &\sim \text{Beta}(\alpha, \beta), & i = 1, \dots, k. \end{aligned}$$

This yields a 7-dimensional posterior distribution in  $\lambda = (\lambda_1, \dots, \lambda_7)^T$ . Figure 8 shows the evolution of two chains started from random points for the second and third component. We report the ACF plots for the second component component only, however, similar behavior is observed in ACF plots of other components as well.

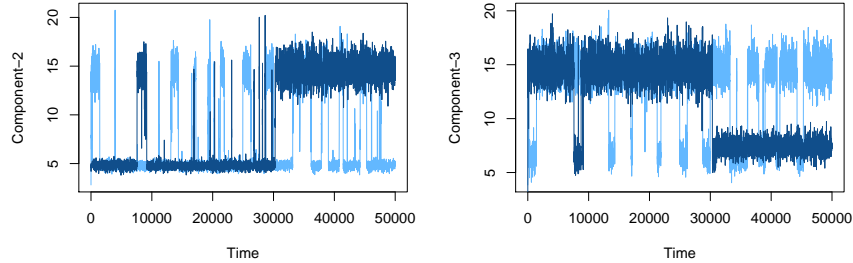


Figure 8: Trace plot for second (left) and third (right) component of two chains.

Figure 9 demonstrates a striking advantage of G-ACF against locally-centered ACF in estimating the autocorrelations. For a chain length of  $n = 1000$ , locally-centered ACF gives a false sense of security about the Markov chains whereas in reality, the process is highly correlated. G-ACF gives a more accurate estimation of autocorrelations even at short run lengths.

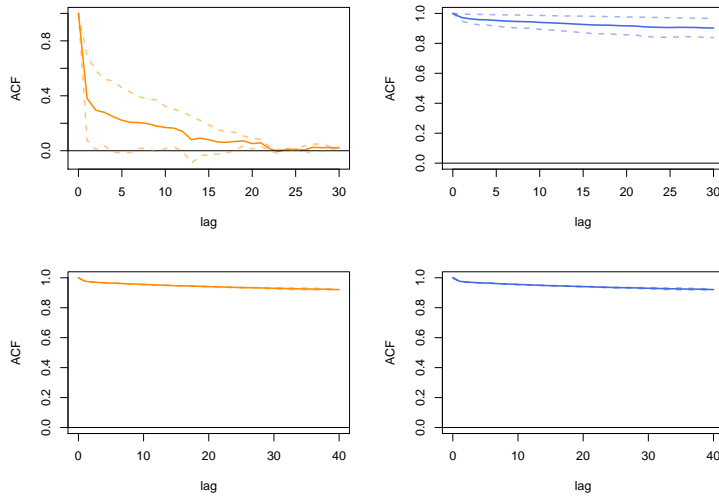


Figure 9: Change point: ACF plots using local (left) and global (right) centering for  $m = 2$  parallel Markov chains. The individual chains are shown through dashed lines and average over  $m$  chains is the solid line. The chain lengths are  $n = 1000$  (top row) and  $n = 10000$  (bottom row).

## 8.2 Network crawling

The `faux.magnolia.high` dataset available in the `ergm` R package represents a simulated friendship network based on Ad-Health data (Resnick et al. (1997)). The school communities represented by the network data are located in the southern United States. Each node represents a student and each edge represents a friendship between the nodes it connects.

The goal is to draw nodes uniformly from the network by using a network crawler. Nilakanta et al. (2019) modified the data by removing 1,022 out of 1,461 nodes to obtain a well-connected graph. This resulting social network has 439 nodes and 573 edges. We use a Metropolis-Hastings algorithm with a simple random-walk proposal suggested by Gjoka et al. (2011). Each node is associated with five features namely - degree of connection, cluster coefficient, grade, binary sex indicator (1 for female, 0 for male), and binary race indicator (1 for white, 0 for others)

We sample two parallel Markov chains starting from two students belonging to different races. Figure 10 shows the ACF plots for the *grade* feature at two different simulation sizes for  $m = 2$  parallel Markov chains. For a shorter chain length, both the chains have explored different clusters and as a consequence, the local and global mean do not agree. Regardless, G-ACF displays a clear advantage over locally-centered ACF for a chain length of  $n = 100$  samples whereas the latter takes  $n = 1000$  samples to reach the truth.

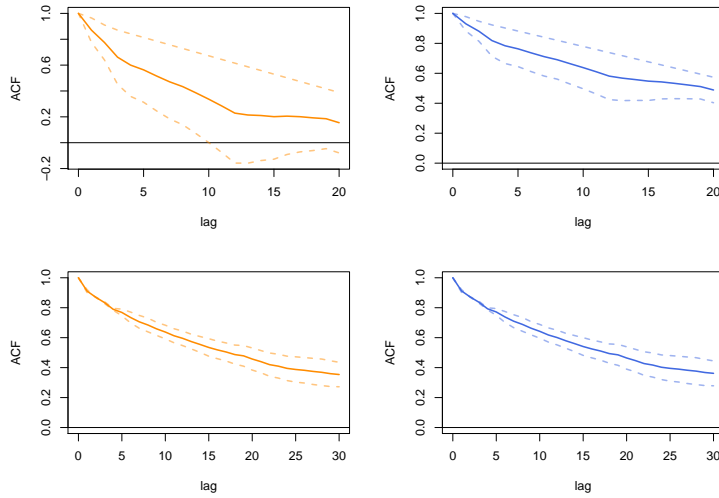


Figure 10: Crawling: ACF plots using local-centering (left) and global-centering (right) for  $m = 2$  parallel Markov chains. The individual chains are shown through dashed lines and average over  $m$  chains is the solid line. The chain lengths are  $n = 100$  (top row) and  $n = 1000$  (bottom row).

## References

- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. John Wiley & Son, New York.
- Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.
- Chen, D.-F. R. and Seila, A. F. (1987). Multivariate inference in stationary simulation using batch means. In *Proceedings of the 19th conference on Winter simulation*, pages 302–304. ACM.
- Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of econometrics*, 86(2):221–241.
- Csörgö, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Probability and Mathematical Statistics : a series of monographs and textbooks. Academic Press.
- Dai, N. and Jones, G. L. (2017). Multivariate initial sequence estimators in Markov chain Monte Carlo. *Journal of Multivariate Analysis*, 159:184–199.
- Damerdji, H. (1991). Strong consistency and other properties of the spectral variance estimator. *Management Science*, 37:1424–1440.
- Damerdji, H. (1995). Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Operations Research*, 43(2):282–291.
- Flegal, J. M. and Gong, L. (2015). Relative fixed-width stopping rules for Markov chain Monte Carlo simulations. *Statistica Sinica*, 25:655–676.

- Flegal, J. M., Haran, M., and Jones, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, 23:250–260.
- Flegal, J. M., Hughes, J., Vats, D., and Dai, N. (2020). *mcmcse: Monte Carlo Standard Errors for MCMC*. Riverside, CA, Denver, CO, Coventry, UK, and Minneapolis, MN. R package version 1.4-1.
- Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38:1034–1070.
- Gelman, A. and Meng, X.-L. (1991). A note on bivariate distributions that are conditionally normal. *The American Statistician*, 45(2):125–126.
- Gjoka, M., Kurant, M., Butts, C. T., and Markopoulou, A. (2011). Practical recommendations on crawling online social networks. *IEEE Journal on Selected Areas in Communications*, 29(9):1872–1892.
- Glynn, P. W. and Whitt, W. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability*, 2:180–198.
- Gong, L. and Flegal, J. M. (2016). A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 25:684–700.
- Gupta, K. and Vats, D. (2020). Estimating Monte Carlo variance from multiple Markov chains. *arXiv preprint arXiv:2007.04229*.
- Hannan, E. J. (1970). Multiple time series: Wiley series in probability and mathematical statistics.
- Heberle, J. and Sattarhoff, C. (2017). A fast algorithm for the computation of hac covariance matrix estimators. *Econometrics*, 5(1):9.
- Ihler, A. T., Fisher, J. W., Moses, R. L., and Willsky, A. S. (2005). Nonparametric belief propagation for self-localization of sensor networks. *IEEE Journal on Selected Areas in Communications*, 23(4):809–819.
- Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998). Markov chain Monte Carlo in practice: a roundtable discussion. *The American Statistician*, 52:93–100.
- Kuelbs, J. (1976). A strong convergence theorem for Banach space valued random variables. *The Annals of Probability*, 4:744–771.
- Liu, Y. and Flegal, J. M. (2018). Weighted batch means estimators in Markov chain Monte Carlo. *Electronic Journal of Statistics*, 12:3397–3442.
- Martin, A. D., Quinn, K. M., and Park, J. H. (2011). MCMCpack: Markov chain Monte Carlo in R.
- Meyn, S. P. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press.
- Nilakanta, H., Almquist, Z. W., and Jones, G. L. (2019). Ensuring reliable Monte Carlo estimates of network properties. *arXiv preprint arXiv:1911.08682*.

- Priestley, M. B. (1981). *Spectral analysis and time series: probability and mathematical statistics*. Number 04; QA280, P7.
- Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., Tabor, J., Beuhring, T., Sieving, R. E., Shew, M., et al. (1997). Protecting adolescents from harm: findings from the national longitudinal study on adolescent health. *Jama*, 278(10):823–832.
- Roy, V. (2019). Convergence diagnostics for Markov chain Monte Carlo. *Annual Review of Statistics and Its Application*, 7.
- Song, W. T. and Schmeiser, B. W. (1995). Optimal mean-squared-error batch sizes. *Management Science*, 41(1):110–123.
- Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 3:211–226.
- Tak, H., Meng, X.-L., and van Dyk, D. A. (2018). A repelling–attracting metropolis algorithm for multimodality. *Journal of Computational and Graphical Statistics*, 27(3):479–490.
- Tjstheim, D. (1990). Non-linear time series and Markov chains. *Advances in Applied Probability*, 22(3):587–611.
- Vats, D., Flegal, J. M., and Jones, G. L. (2018). Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Bernoulli*, 24:1860–1909.
- Vats, D., Flegal, J. M., and Jones, G. L. (2019). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika*, 106:321–337.
- Vats, D., Robertson, N., Flegal, J. M., and Jones, G. L. (2020). Analyzing Markov chain Monte Carlo output. *Wiley Interdisciplinary Reviews: Computational Statistics*, 12:e1501.
- Zeidler, E. (2013). *Nonlinear functional analysis and its applications: III: variational methods and optimization*. Springer Science & Business Media.