Globally-centered autocovariances in MCMC

Medha Agarwal
Department of Mathematics and Statistics
IIT Kanpur
medhaaga@iitk.ac.in

Dootika Vats
Department of Mathematics and Statistics
IIT Kanpur
dootika@iitk.ac.in

August 17, 2020

Abstract

Autocovariances are the fundamental quantity in many features of Markov chain Monte Carlo (MCMC) simulations with autcorrelation function (ACF) plots being often used as a visual tool to ascertain the performance of a Markov chain. Unfortunately, for slow mixing Markov chains, the empirical autocovariance can highly underestimate the truth. For multiple chain MCMC sampling, we propose a globally-centered estimate of the autocovariance that pools information from all Markov chains. We show the impact of these improved estimators in three aspects: (1) acf plots, (2) estimates of the Monte Carlo asymptotic covariance matrix, and (3) estimates of the effective sample size.

1 Introduction

The power of the modern personal computer has made it easy to run parallel Markov chain Monte Carlo (MCMC) implementations. This is particularly useful for slow mixing or multi-modal target distributions, where the starting points of the chains are spread over the state-space in order to capture characteristics of the target distribution more accurately. 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 of interest that drives MCMC output analysis is the autocovariance function (ACvF). From their use in autocorrelation plots to assessing Monte Carlo variability of estimates,

to determining when to stop MCMC simulation, autocovariances drive many visual and quantitative inferences users make from MCMC output. However, tools on estimating ACvF are largely constructed for output from one MCMC chain. a little bit more here.

Let F be the target distribution with mean μ defined on the set $\mathcal{X} \subseteq \mathbb{R}^p$ equipped with a countably generated Borel σ-field $\mathcal{B}(\mathcal{X})$. For $s = 1, \ldots, m$, let $\{X_{st};\ t \in \mathbb{Z}\}$ be the s^{th} F-Harris ergodic stationary 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 depends only on k and is defined as

$$\Gamma(k) = \text{Cov}_F(X_{s1}, X_{s\,1+k}) = \mathbb{E}_F\left[(X_{s1} - \mu)(X_{s\,1+k} - \mu)^T\right].$$

Estimating $\Gamma(k)$ is critical to assessing the quality of the sampler and the quality of sample statistics. Let $\bar{X}_s = n^{-1}\sum_{t=1}^{n} X_{st}$ denote the Monte Carlo estimator of μ from the sth chain. The standard estimator for $\Gamma(k)$ is the sample autocovariance matrix at lag k. For the sth chain, it is given by

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

We need to figure out how we want to define autocovariance: with or without absolute value.

For single-chain MCMC runs, the estimator $\hat{\Gamma}(k)$ is used to construct autocorrelation (ACF) plots, to estimate the long-run variance of Monte Carlo estimators Hannan (1970); Damerdji (1991), and to estimate effective sample size for an estimation problem Kass et al. (1998); Gong and Flegal (2016); Vats et al. (2019a).

However, there is no unified approach to constructing estimators of $\Gamma(k)$ for multiple-chain implementations. Specifically, parallel MCMC chains are often spread across the state space so as to adequately capture high density areas. For slow mixing Markov chains and multi-modal targets, the chains take time to traverse the space so that estimates of μ can be vastly different. The sample ACvF from each chain is typically underestimated yielding false sense of security about the quality of process.

We propose a globally-centered ACvF (G-ACvF) that centers the Markov chain around the overall mean vector from all m chains. We show that the bias under stationarity for G-ACvF is lower than $\hat{\Gamma}(k)$, and through various examples, empirically demonstrate improved estimation. We employ the G-ACvF estimators to construct autocorrelation plots, which leads to remarkable improvements. A demonstrative example is at the end of this Section.

ACvFs are required to estimate the long-run variance of Monte Carlo averages. Specifically, spectral variance estimators are used to estimate $\Sigma = \sum_{k=-\infty}^{\infty}\Gamma(k)$. We replace $\hat{\Gamma}(k)$ with G-ACvF in spectral variance estimators (SVE) and show strong consistency of this estimator under weak conditions. In the spirit of Andrews (1991), we also obtain large-sample bias and variance of the

resulting estimator of Σ. Spectral variance estimators can be slow to prohibitively expensive to calculate (Liu and Flegal, 2018). To relieve the computational burden, we adapt the fast algorithm of Heberle and Sattarhoff (2017) for G-ACvFs that dramatically reduces computation time. The globally-centered spectral variance estimator (G-SVE) is employed in the computation of effective sample size through the estimator given by Vats et al. (2019b). We will show that using G-SVE for estimating Σ safeguards us against early termination of the MCMC process and improves the reliability of estimators.

1.1 Demonstrative example

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

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

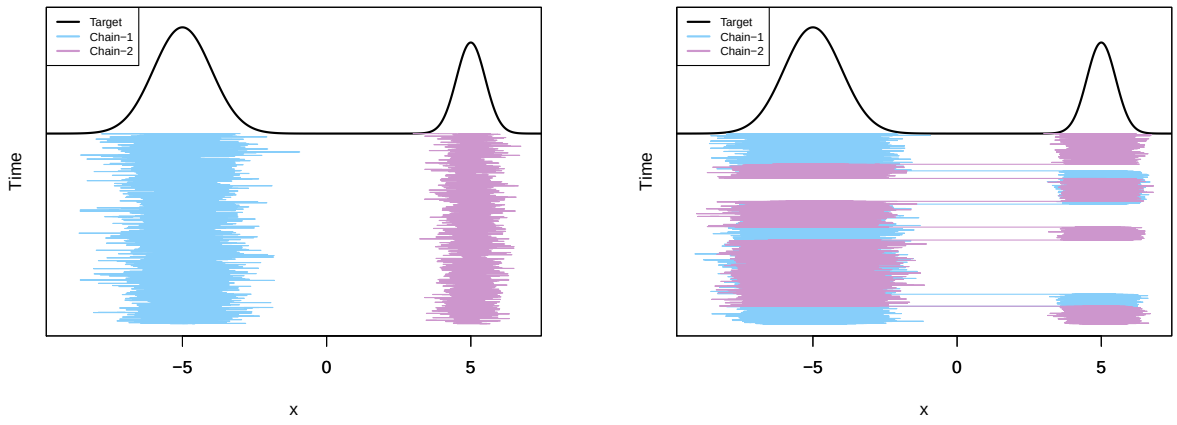
where $f(x; a, b)$ is the density of a normal distribution with mean a and variance b. Two Markov chains are started at each mode. Let n denote the number of samples for each chain. Figure 1 indicates that in the first 10^4 samples, the chains do not jump modes so that both Markov chains yield different estimates of the population mean, μ. At 10^5 sample size, both Markov chains have traversed the state space reasonably and yield similar estimates of μ. We make two points here: first, for small sample size, we observe that the locally-centered ACF gives misleading estimates whereas G-ACF accounts for the discrepancy in sample means between two chains. This is evident from the ACF plots in Figure **??**. Second, for a large sample size, the estimates from G-ACrF as well as ACrF are equivalent. Therefore, G-ACF can be used in any situation with a promise of estimates that are at least as good as empirical ACF estimators.

2 Globally-centered autocovariance

Recall that we run m independent F-ergodic Markov chains denoted by $\{X_{st};\ t \in \mathbb{Z}\}$ for the s^{th} chain. Let the sample mean of the s^{th} Markov chain be $\bar{X}_s = n^{-1} \sum_{t=1}^{n} X_{st}$ and the global mean by $\bar{\bar{X}} = m^{-1} \sum_{s=1}^{m} \bar{X}_s$. The global mean is naturally superior estimator of μ than $\bar{X}_s$. We define the globally-centered autocovariance function (G-ACvF) estimator for the s^{th} Markov chain as

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

In the event that all m Markov chains have been run long enough that $\bar{X}_s \approx \bar{\bar{X}}$ are similar, then $\hat{\Gamma}_s(k) \approx \hat{\Gamma}_{G,s}(k)$. However, for shorter runs or for slow mixing chains, $\Gamma(k)$ is more appropriately



(a) $n = 10^4$ (b) $n = 10^5$

Figure 1: Target density and trace plots for two chains starting at different values.

estimated by $\hat{\Gamma}_{G,s}$ as it utilizes information from all chains and accounts for disparity between estimates of μ. We quantify the gains in bias for the G-ACvF estimator below. Let

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

and let $\Phi^{(1)}$ be denoted by Φ. The proof of the following theorem is available in Appendix 7.2.

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

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

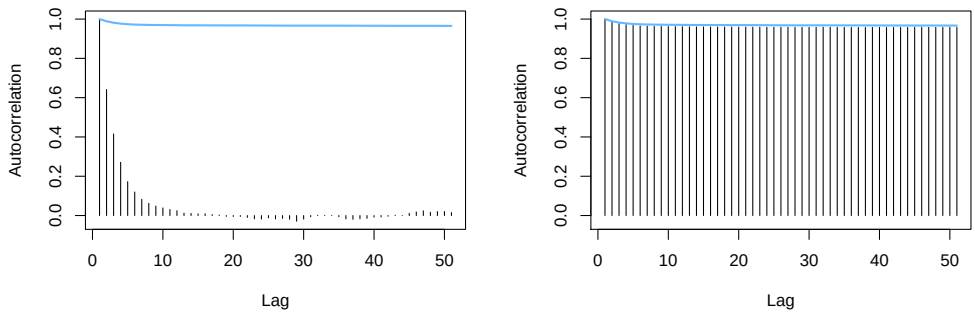
and consequently,

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

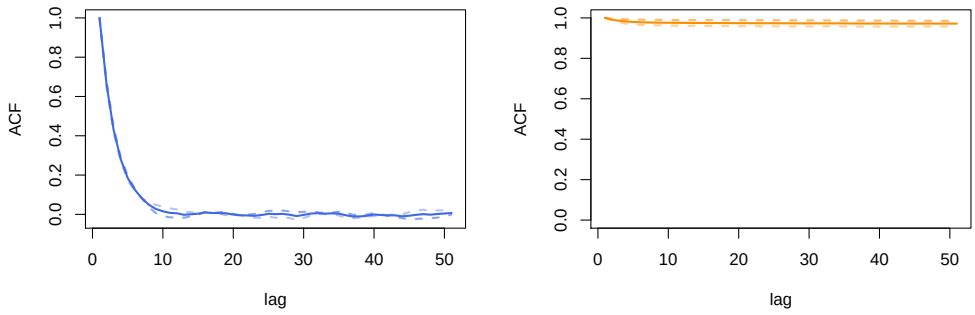
MA: I corrected the RHS for the abover equation. Or did you mean to write $\mathbb{E}_F \left[\hat{\Gamma}_{G,s}(k) - \Gamma(k) \right]$ *asLHS*?

Remark 1. Polynomial ergodicity and the moment conditions are required to ensure Φ and Σ are finite. The above result can be stated more generally for α-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



(a) $n - 10^4$ (histogram) and $n = 5 \times 10^5$ (blue line)



(b) $n = 10^4$

Figure 2: ACF plots for locally-centered ACF (left) and globally centered ACF (right). (a) The histogram plots the estimates for $n = 10^4$ and the blue line plot the estimates for $n = 5 \times 10^4$ for chain-1. (b) $n = 10^4$; the dotted lines are ACF for each chain and the thick solid line is the ACF averaged over both chains.

are asymptotically unbiased and exhibit reduced bias in finite samples compared to locally-centered ACvF estimators. The consequences of this are particularly pertinent for the diagonals of $\Gamma(k)$.

Remark 2. The variance of the target distribution is the lag covariance at lag 0 and is of particular interest. The bias of the estimator reduces by a factor of m:

$$\mathbb{E}\left[\hat{\Gamma}_G(0)\right] = \Gamma(0) - \frac{\Sigma}{mn} - \frac{\Phi}{mn^2} + o(n^{-2})$$

A first use of the ACvFs in MCMC is in constructing ACF plots. Let the target variance of ith component be $\Gamma^{ii}(0)$. For any component i, the autocorrelation is defined as

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

and is instrumental in visualizing the serial correlation in components of the Markov chain. The typical estimate of the autocorrelation is constructed from $\hat{\Gamma}_s(k)$. Instead, we advocate for using G-ACvF estimates so that,

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

The globally-centered autocorrelation provides a far more realistic assessment of the correlation structure of marginal components of the chain as evidenced in Figure **??**.
We end this section by noting that we can obtain an average G-ACvF and G-ACF over all m chains

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

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

3 Variance estimators for multiple Markov chains

A critical use of autocovariances is in the assessment of Monte Carlo variability of estimators. Let $g : \mathcal{X} \longrightarrow \mathbb{R}^p$ 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 $s = 1, \ldots, m$. 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, $\bar{\bar{Y}} \to \mu_g$ as $n \to \infty$ with probability 1.
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 Σ such that

$$\sqrt{mn}(\bar{\bar{Y}} - \mu_g) \xrightarrow{d} N(0, \Sigma) \tag{3}$$

where

$$\Sigma = \sum_{k=-\infty}^{\infty} \mathrm{Cov}_F\left(Y_{11}, Y_{1(1+k)}\right) := \sum_{k=-\infty}^{\infty} \Upsilon(k) \tag{4}$$

The goal in output analysis for MCMC is to estimate Σ 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 Σ for single-chain MCMC implementations. The most common are spectral variance (SV) estimators (Andrews, 1991; Vats et al., 2018) and batch means estimators (Chen and Seila, 1987; Vats et al., 2019a). Recently, Gupta and Vats (2020) constructed a replicated batch means estimator for estimating Σ from parallel chains. Batch means estimators are computationally more efficient than SV estimators, whereas SV estimators are more reliable (Flegal and Jones, 2010). Here, we utilize G-ACvF estimators to construct globally-centered SV estimator (G-SVE) of Σ. Using the methods of Heberle and Sattarhoff (2017), we also provide an efficient algorithm for obtaining the G-SV estimator.
The locally and globally centered estimators for $\Upsilon(k)$ are $\hat{\Upsilon}_s(k)$ and $\hat{\Upsilon}_{G,s}(k)$, respectively, where

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

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

$$\hat{\Sigma}_s = \sum_{k=-n+1}^{n-1} w\left(\frac{k}{b_n}\right)\hat{\Upsilon}_s(k)\,. \tag{5}$$

Large-sample properties of $\hat{\Sigma}_s$ have been studied widely. Vats et al. (2018) provided conditions for strong consistency while Flegal and Jones (2010); Vats and Flegal (2018) provided large-sample bias and variance results. These results also extend naturally to an average SV (ASV) estimator

$$\hat{\Sigma}_A = \frac{1}{m}\sum_{s=1}^{m}\hat{\Sigma}_s\,.$$

3.1 Globally-centered spectral variance estimators

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

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

To accurately estimate Σ, standard conditions are imposed on the lag window, w.

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

Assumption 1 is standard (Anderson, 1971, see) 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

We will present three main results for the G-SV estimator. First, we provide conditions for strong consistency. Strong consistency is particularly important to ensure sequential stopping rules yield correct coverage at termination (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 a polynomially ergodic Markov chain of order $\xi > (q + 1 + \epsilon)/(1 + 2/\delta)$ for $q \geq 1$. Then there exists a $p \times p$ lower triangular matrix L such that $LL^T = \Sigma$, 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 Ω such that for almost all $\omega \in \Omega$ such that for all $n > n_0$, with probability 1,*

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

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

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). However, Hannan (1970) proved the results assuming μ_g was known. Here, we present the results under mild conditions for when μ_g is replaced by $\bar{\bar{Y}}$. Let Σ^{ij} and $\hat{\Sigma}_G^{ij}$ denote the ijth element of the matrix Σ and $\hat{\Sigma}_G$, repectively.

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

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

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

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

3.1.2 Fast implementation

The SV estimator, despite good statistical properties, poses application limitations due to slow computation. From (1) and (5), the complexity of the SV estimator is $\mathcal{O}(n^2p^2)$ in general, and for Bartlett (and other truncated) lag window, the complexity is $\mathcal{O}(b_n np^2)$. In MCMC, and especially for slow mixing Markov chains, n and b_n are large, limiting the use of SV estimators.
To overcome this, we adapt the fast Fourier transform based algorithm of Heberle and Sattarhoff (2017) for our purposes. Suppose $w_k = w(k/b_n)$ and let $T(w)$ be the $n \times n$ Toeplitz matrix of weights with the first column being $(1\ w_1\ w_2\ \ldots,\ w_{n-1})^T$. Notice an alternate formulation of $\hat{\Sigma}_s$

$$\hat{\Sigma}_s = \frac{1}{n} A_s^T T(w) A_s, \qquad \text{where } A_s = \begin{pmatrix} Y_{s1} - \bar{Y}_s & \ldots & Y_{sn} - \bar{Y}_s \end{pmatrix}^T \tag{6}$$

For computational efficiency, Heberle and Sattarhoff (2017) computed $T(w)A_s$ directly using an FFT based algorithm. The algorithm requires embedding $T(w)$ in a $2n \times 2n$ circulation matrix $C(w^*)$. We write the spectral decomposition of $C(w^*)$ as $V\Lambda V^*$ where the eigenvalues of $C(w^*)$ are given by the DFT of the first column of $C(w^*)$. A_s is extrapolated into a $2n \times p$ matrix A_s^* by appending n rows of 0 vector to write Equation 6 in the following form

$$\hat{\Sigma}_s = \frac{1}{n} A_s^T T(w) A_s = \frac{1}{n} A_s^T (C(w^*) A_s^*)_{1:n,:} = \frac{1}{n} A_s^T (V\Lambda V^* A_s^*)_{1:n,:}).$$

Here $M_{1:r,\,1:c}$ denotes the matrix truncation up to first r rows and c columns. For our purpose, we center the chain $\{Y_{st};\ t \in \mathbb{Z}\}$ around $\bar{\bar{Y}}$ instead of $\bar{Y}_s$ in Equation 6 for all $s \in \{1, ..., m\}$. Herberle's algorithm is then applied on the formulation

$$\hat{\Sigma}_{G,s} = \frac{1}{n} B_s^T T(w) B_s \qquad \text{where } B_s = \begin{pmatrix} Y_{s1} - \bar{\bar{Y}} & \ldots & Y_{sn} - \bar{\bar{Y}} \end{pmatrix}^T.$$

The complete algorithm leads to a tremendous reduction in time and storage requirements with a computational complexity of $\mathcal{O}(pn \log n)$. B_s^* is constructed from B_s in the similar way as A_s^*. In Algorithm 1, we use $v^{(i)}$ to denote the i^{th} element of vector v and $M_{(j)}$ to denote the j^{th} column of matrix M. The algorithm is the following.

4 Effective sample size

An important method of quantifying the simulation efforts is to simulate until a desired effective sample size (ESS) is obtained. The key motivation behind using ESS as a termination criteria is to stop the simulations when the variance of Monte Carlo average is small relative to the inherent

Algorithm 1: Herberle's Algorithm

Construct $C(w^*)$ and B_s^* from $\{Y_{st};\, t = 1, ..., n\}$.
Compute Λ by DFT of $C(w^*)_{(1)}$
for $j = 1, 2, ..., p$ **do**
    Calculate $V^* B_s^{*(j)}$ by FFT of $B_{s(j)}^*$.
    Multiply $V^* B_{s(j)}^{*(i)}$ with the eigenvalue λ_i for all $i = 1, ..., 2n$ to construct $\Lambda V^* B_{s(j)}^*$.
    Calculate $C(w^*)B_{s(j)}^* = V\Lambda V^* B_{s(j)}^*$ by inverse FFT of $\Lambda V^* B_{s(j)}^*$.
end
Select the first n rows of $C(w^*)B_s^*$ to form $T(w)B_s$.
Premultiply by B_s^T and divide by n.

variance in g with respect to F by a pre-specified factor. Wilks (1932) suggests the determinant of a covariance matrix as a viable metric to understand the variance of a multivariate process and calls it *generalized variance*. For a p-dimensional stationary Markov chain, let $\Lambda := \mathrm{Var}_F(Y_1))$. Vats et al. (2019b) use the ratio of generalized variances to give the following formulation for multivariate ESS (m-ESS):

$$\mathrm{ESS} = n \left(\frac{|\Lambda|}{|\Sigma|} \right)^{1/p}.$$

The simulation process is stopped once the ESS is greater than a pre-specified lower-bound W_p which is function of dimension p (see Vats et al. (2019b) for details). For a multi-modal target distribution, the single chain estimators for Σ underestimate the true generalized variance; giving an unrealistic assurance of high ESS for a given n.

In our setting of m parallel chains of n samples each, define

$$\hat{\Lambda}_{mn} = \frac{1}{m(n-1)} \sum_{s=1}^{m} \sum_{t=1}^{n} (Y_{st} - \bar{Y}_s)(Y_{st} - \bar{Y}_s)^T.$$

We estimate the m-ESS by using the strongly consistent average G-SV estimator $\hat{\Sigma}_G$ for Σ as

$$\widehat{\mathrm{ESS}}_G = mn \left(\frac{\left|\hat{\Lambda}_{mn}\right|}{\left|\hat{\Sigma}_G\right|} \right)^{1/p}.$$

For the sake of comparison, $\widehat{\mathrm{ESS}}_A$ is constructed similarly using $\hat{\Sigma}_A$ instead of $\hat{\Sigma}_G$ to estimate Σ. We refrain from using the globally centered version of sample covariance matrix, i.e. $\hat{\Upsilon}(0)$ to estimate Λ in both $\widehat{\mathrm{ESS}}_A$ and $\widehat{\mathrm{ESS}}_G$. This is because in practice, it is preferred to underestimate the ESS than do otherwise. Both $\hat{\Upsilon}(0)$ and $\hat{\Lambda}_{mn}$ converge to Λ; however $\left|\hat{\Lambda}_{mn}\right|$ shall underestimate $|\Lambda|$ in the beginning. We will compare $\widehat{\mathrm{ESS}}_G$ to $\widehat{\mathrm{ESS}}_A$ in the following Section.

5 Examples

In this section we consider three different target distributions and sample multiple Markov chains using MCMC methods to experimentally analyse the performance of our globally-centered estimators. We will make the following three comparisons - (1) A-ACF vs G-ACF through ACrF plots; (2) A-SVE vs G-SVE; (3) $\widehat{\text{ESS}}_G$ vs $\widehat{\text{ESS}}_A$. The quality of estimation of Σ is studied by coverage probabilities for a 95% confidence interval in cases where the true mean μ is known. The convergence of local and global estimators as n increases is studied through two types of running plots (1) logarithm of Frobenius norm of estimated Σ denoted by $\|M\|_F$ for a matrix M, and (2) logarithm of estimated ESS/mn. In all the examples we consider the function $g(x) = x$, which implies the Markov chain $\{Y_t\}$ and $\{X_t\}$ are the same.

5.1 Vector Autoregressive Process

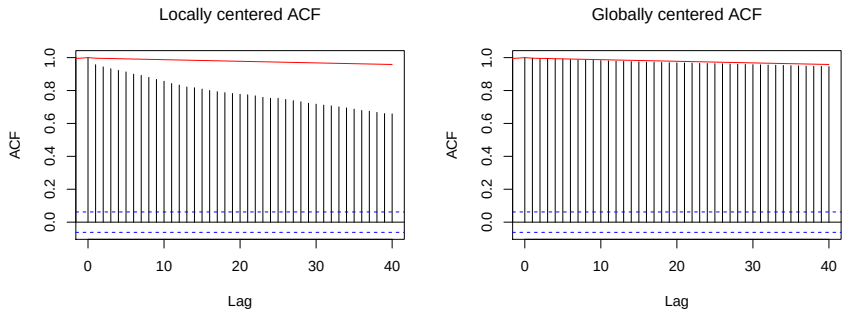
There are only a handful of MCMC methods where the true autocovariance and asymptotic variance are known in closed form. To compare the globally and locally-centered estimators when the truth is known, we use a slowly mixing vector autoregressive process of order 1 (VAR(1)). We want to deliberately make the Markov chains explore the sample space slowly. Consider a p-dimensional VAR(1) process $\{X_t\}_{t\geq 1}$ such that

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

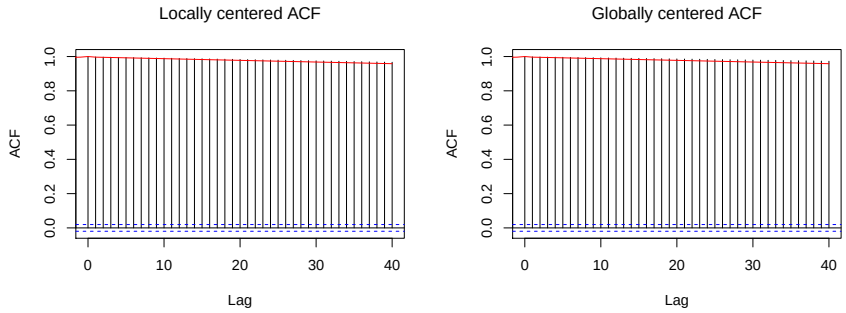
where $X_t \in \mathbb{R}^p$, Φ is a $p \times p$ matrix, $\epsilon_t \overset{i.i.d.}{\sim} N(0, \Omega)$, and Ω is a positive definite $p \times p$ matrix. The invariant distribution for this Markov chain is $N(0, \Psi)$ where $\mathcal{N}(.,.)$ stands for the Gaussian distribution. The lag-k autocovariance $\Gamma(k)$ can be calculated for $k > 0$ as $\Gamma(k) = \Phi^k \Psi$ and $\Gamma(-k) = \Psi(\Phi^T)^k$. The Markov chain is geometrically ergodic when the spectral norm of Φ is less than 1 (Tjstheim (1990)). The CLT holds for the invariant distribution, therefore, Σ exists and is known in closed form.

Let $\phi_{\max}$ be the largest absolute eigenvalue of Φ such that $|\phi_{\max}| < 1$. The larger it is, the slower the Markov chain mixes. For our case, we require a slowly mixing VAR(1) process. We consider a bivariate example with $\phi_{\max} = 0.999$. We run five parallel deterministically initialized Markov chains with their starting points evenly distributed in the sample space. The correlation between the two components in the target distribution is very high for the chosen value of Φ and Ω. As a result, it takes a long time for the Markov chain to explore the sample space significantly.

The locally and globally-centered autocorrelations for the second chain have been shown in Figure 3 for two different simulation sizes where the true ACF is shown by the red line. When the individual chains have not explored well, the local sample means have not converged to μ. As a consequence, the single chain empirical ACF estimator severely underestimate the truth. Due to sufficiently



(a) $n = 10^3$



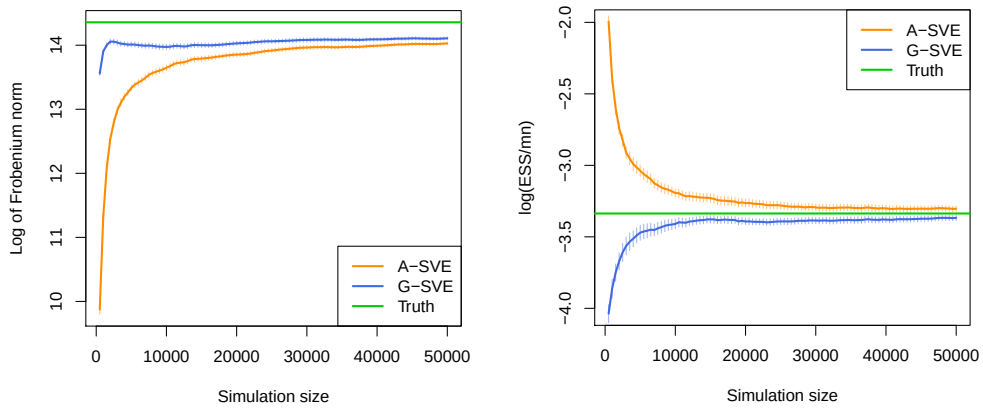
(b) $n = 10^4$

Figure 3: VAR(1). ACF (left) and G-ACF (right) for the second chain for $m = 5$. (a): $n = 10^3$, not converged yet; (b): $n = 10^4$, converged. The red line marks the true ACF.

dispersed starting values, G-ACF is closer to reality. In subfigure 3b, the convergence property of sample means has kicked in for each chain. Therefore, both the estimators provide equivalent results. Note that G-ACF reveals the true nature of ACF for almost ten times lesser chain length n.

The quality of estimation of Σ and ESS/mn has been studied through convergence of - (a) $\log(\|\hat{\Sigma}_A\|_F)$ and $\log(\|\hat{\Sigma}_G\|_F)$ as n increases and (b) $\log(\widehat{\text{ESS}}_A/mn)$ and $\log(\widehat{\text{ESS}}_G/mn)$ as n increases. In practice, it is preferred to slightly overestimate the variance and underestimate the ESS than do otherwise. In Figure 4, both the running plots are shown along with the green horizontal line which marks the truth. We run the simulations for 50 replications for each value of n and plot the average as well as standard errors. We run the simulations for a maximum chain length of 5×10^4 when the estimators seem to have almost converged. From Figure 4a, it can be seen that for first 10000 samples, A-SV estimates can be absurdly small giving false confidence in single chain estimates whereas G-SV gives a better understanding of reality. Similar comparative argument can can given for ESS.

We also report the coverage probabilities for a 95% confidence interval constructed using $\hat{\Sigma}_A$ and $\hat{\Sigma}_G$ because the true mean is known here. Table 1 shows that for each sample size, the globally-centered



(a) $log(\|\hat{\Sigma}\|_F)$ (b) $\log(\widehat{\text{ESS}}/mn)$

Figure 4: VAR(1). (a) Running plot for logarithm of Frobenius norm of A-SVE and G-SVE from $n = 500$ to $n = 10^5$. (b) Running plot for logarithm of $\widehat{\text{ESS}}/mn$ calculated using A-SVE and G-SVE from $n = 500$ to $n = 10^5$.

SV estimator for variance gives better coverage than A-SVE. At $n = 10^5$, both the estimators have converged and give equivalently good coverage probability.

5.2 Boomerang Distribution

Our globally-centered estimators perform exceptionally better than the single-chain estimators in case of multimodal target distribution. To that end, we will use a bivariate bi-modal distribution introduced by Gelman and Meng (1991) which has Gaussian conditional distributions in both directions. This allows us to sample parallel Markov chains using the Gibbs sampler. Let x and y be two random variable that are jointly distributed as

$$f(x, y) \propto \exp\left(-\frac{1}{2}\left[Ax^2y^2 + x^2 + y^2 - 2Bxy - 2C_1x - 2C_2y\right]\right)$$

The conditional distribution of x given y and vice versa is a normal distribution given by:

$$x_1 \mid x_2 \sim N\left(\frac{Bx_2 + C_1}{Ax_2^2 + 1}, \frac{1}{Ax_2^2 + 1}\right)$$
$$x_2 \mid x_1 \sim N\left(\frac{Bx_1 + C_2}{Ax_1^2 + 1}, \frac{1}{Ax_1^2 + 1}\right)$$

We report results for two different bi-modal parameter settings in this example. Firstly, we use a carefully chosen parameterization of $A = 1,\ B = 3,\ C = 8$ wherein the two modes are well-

n	A-SVE	G-SVE
1000	0.71	0.956
5000	0.843	0.937
10,000	0.885	0.924
50,000	0.928	0.945
100,000	0.944	0.952

Table 1: VAR(1). Coverage probabilities for a 95 percent confidence interval of global mean constructed using A-SVE and G-SVE as variance estimates around the known mean.

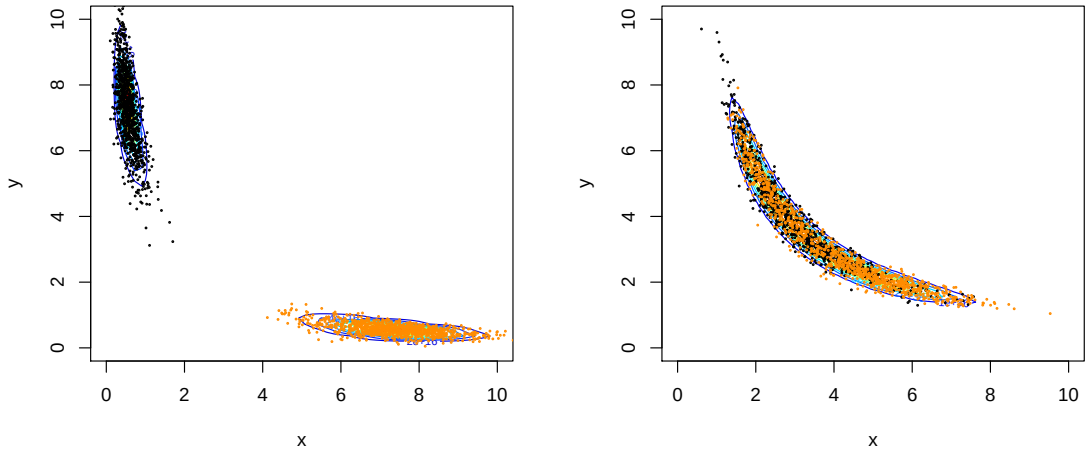
separated with a very low connection (Setting-1, Figure 5a). Secondly, to examine the performance of G-SVE for a nicely mixing Markov chain, we use the parametrization of $A = 1,\ B = 10,\ C = 7$ wherein the two modes are closer and well-connected (Setting-2, Figure 5b). We will make another point through this example. For a fast mixing Markov chain, the globally-centered and locally-centered estimators give equivalent results. Therefore, it is a better practice to globally center the chains in any situation. We sample five parallel deterministically initialized chains with starting points dispersed uniformly in the first quadrant. Finding the actual mean of this distribution in closed form is difficult. Therefore, we use numerical integration with fine tuning to calculate it.

To visualize the sticky nature of Markov chains in Setting-1, we show a scatter plot of two Markov chains starting near the two modes for $n = 1000$ in Figure 5. For the first thousand samples, both the chains are oblivious of the existence of another mode in setting-1. Whereas in setting-2, the chains have mixed fairly well for just $n = 1000$.

In Figure 6, the autocorrelations are severely underestimated by A-ACF for $n = 1000$ because the chains have not jumped modes. Whereas, for $n = 10000$, both ACF and G-ACF give almost indistinguishable results indicating that A-ACF and G-ACF have converged to the truth. We can observe in Figure 7 that A-ACF and G-ACF give similar results for just $n = 1000$ under simulation setting-2.

The convergence of $\log \widehat{\text{ESS}}_G/mn$ and $\log \widehat{\text{ESS}}_A/mn$ as n increases can be seen in Figure 8. We can see that in the beginning, A-SVE gives misleadingly higher estimates of ESS/mn than G-SVE. This can cause us to stop the sampling before each chain jumps to the other mode. As a consequence, each chain will correspond to a uni-modal target distribution. Whereas, as expected, the ESS estimates for setting-2 are almost same. Similarly, $\|\hat{\Sigma}_G\|_F$ reaches the convergence point in less than half the time as $\|\hat{\Sigma}_G\|_F$; as evidenced in Figure 9a. No significant difference can be observed for setting-2 (Figure 9b).

Using the numerical approximation of the true mean, we construct the 95% confidence interval using A-SVE and G-SVE. We report the coverage probabilities for a 1000 replications at different



(a) Setting-1 (b) Setting-2

Figure 5: Boomerang. Contour plots of the target distribution for both settings. Overlayed scatter plot for two chains (in different colors) starting near each mode.

n	$m = 2$		$m = 5$	
	A-SVE	G-SVE	A-SVE	G-SVE
5000	0.595	0.7	0.402	0.64
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

n	$m = 2$		$m = 5$	
	A-SVE	G-SVE	A-SVE	G-SVE
1000	0.856	0.868	0.895	0.9100
5000	0.921	0.925	0.91	0.915
10000	0.928	0.93	0.919	0.926
50000	0.943	0.944	0.951	0.952

Table 3: Setting-2

sample sizes n. For both the simulation settings, we provide coverage probabilities for $m = 2$ and $m = 5$. Table 2, 3 reports all these results. In setting-1, it can be observed that G-SVE gives higher coverage probability than A-SVE for all values of n. Whereas for setting-2, the results are almost similar indicating the equivalence of A-SVE and G-SVE.

5.3 Sensor Network Localization

For our third example, we consider a real-life example of sensor locations previously discussed by Ihler et al. (2005). The goal is to identify unknown sensor locations using noisy distance data. This problem is specifically interesting in our case because the marginal posterior distribution for missing sensor locations is multi-modal for all locations.

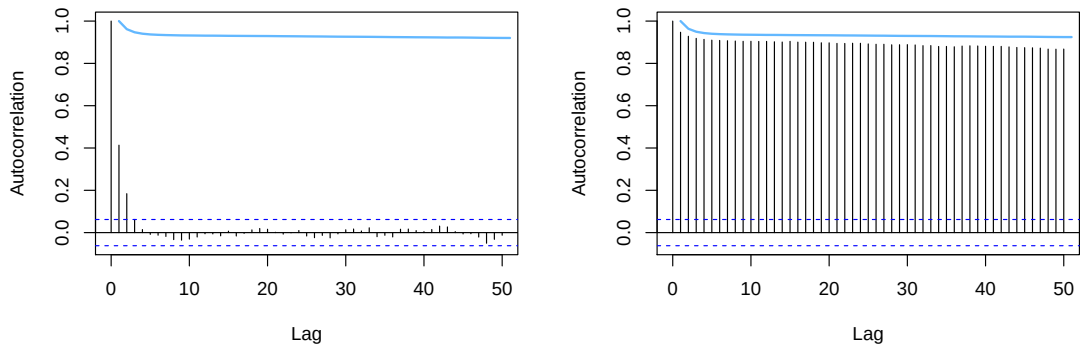


Figure 6: Boomerang (setting-I). A-ACF (left) and G-ACF (right) of component-1 for $n = 1000$ (histogram) and $n = 10000$ (blue line)

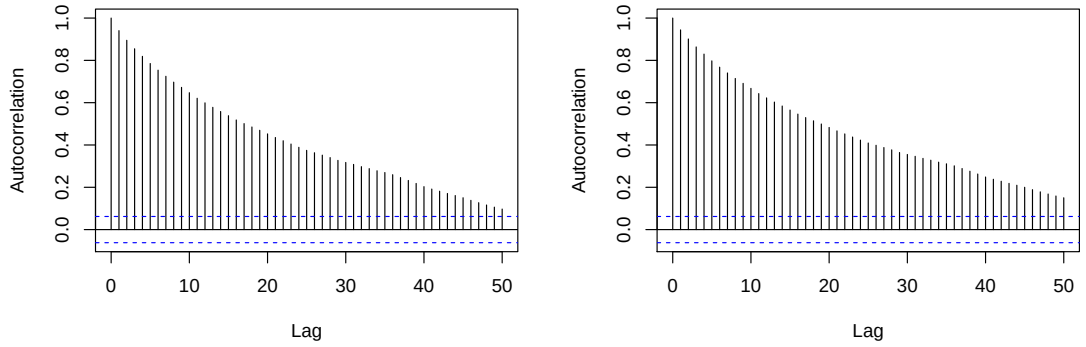
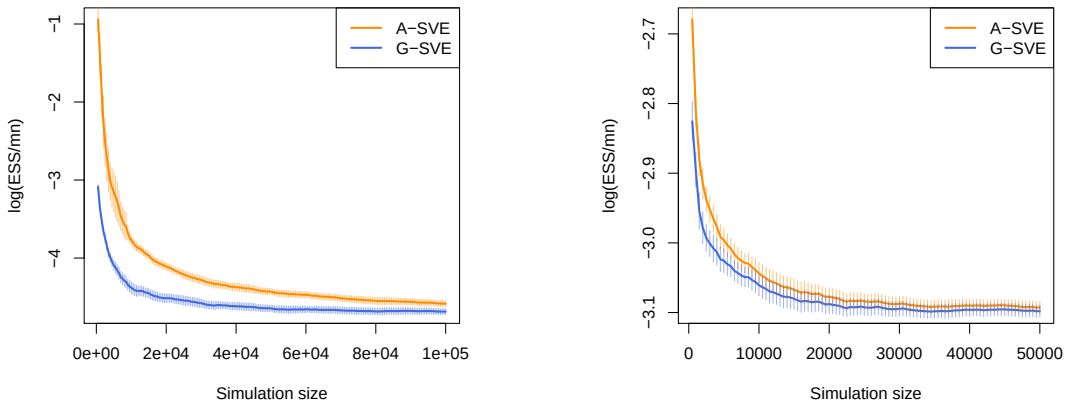


Figure 7: Boomerang. A-ACF (left) and G-ACF (right) of component-1 for $n = 1000$

Following Tak et al. (2018), we assume that there are six sensors scattered on a planar region where $x_i = (x_{i1}, x_{i2})^T$ denote the $2d$ coordinates of i^{th} sensor. Let $y_{ij} = (y_{ji})$ denote the distance between the sensors x_i and x_j. The distance between x_i and x_j is observed with probability $\pi(x_i, x_j) = \exp \|x_i - x_j\|^2/2R^2$ and with a Gaussian measurement error of σ^2. Let z_{ij} denote the indicator variable which is equal to 1 when the distance between x_i and x_j is observed. The probability model is then,

$$z_{ij} \mid x_1, ..., x_6 \sim \text{Bernoulli}\left(\exp\left(\frac{-\|x_i - x_j\|^2}{2R^2}\right)\right)$$

$$y_{ij} \mid w_{ij} = 1, x_1, ..., x_6 \sim N(\|x_i - x_j\|^2, \sigma^2)$$



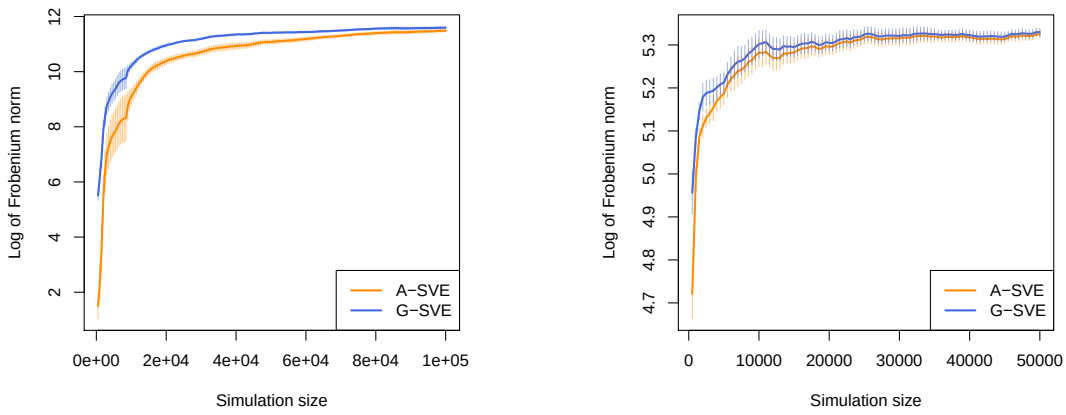
(a) Setting-1 (b) Setting-2

Figure 8: Boomerang. Running plot for $\log \hat{\mathrm{ESS}}_A/mn$ and $\log \hat{\mathrm{ESS}}_G/mn$ for $m = 5$.

Ahn et al. (2013) suggested the value of $R = 0.3$ and $\sigma = 0.02$. We use a Gaussian prior for the unknown locations with mean equal to $(0, 0)$ and covariance matrix equal to $100I_2$. y_{ij} is specified only if $w_{ij} = 1$. We follow the Markov chain structure as described by Tak et al. (2018) and sample from the four bivariate conditionals for each sensor location using a Gibbs sampler. In their paper on Repelling Attractive Metropolis (RAM) algorithms, Tak et al. (2018) compare the performance of different sampling techniques and show that RAM improves the acceptance rate by a factor of at least 5.5 over Metropolis using the same jumping scale. RAM algorithm supplies Markov chains with higher jumping frequency between the modes.
This is an 8-dimensional estimation problem where the components corresponds to x and y coordinates of four unknown sensor locations. We will use the RAM algorithm with a jumping scale equal to 0.5 to sample five parallel Markov chains with well-separated starting points. The total simulation size for each chain is fixed at $100,000$. Since the truth about actual mean and asymptotic variance is not known in this case, we are not interested in coverage probabilities.

To illustrate the *sticky* nature of the Markov chains, we plot the evolution of two chains (starting near different modes) with time. Figure 10 shows the trace plot of x and y coordinate of sensor location-1. It is evident that the marginal posterior for both the components is bimodal. Similarly, the other three locations also have a bimodal marginal distribution in each component. Observe that each chain explores one particular mode for a long time before jumping to another mode.

Figure 11 plots the locally and globally centered ACF estimators for individual chains as well as the average over all the chains. The thick solid line plots the averaged estimator and the thin



(a) Setting-1

(b) Setting-2

Figure 9: Boomerang. Running plot for $\log \|\hat{\Sigma}_A\|_F$ and $\log \|\hat{\Sigma}_G\|_F$ for $m = 5$.

dotted lines correspond to ACF calculated using single chains. Figure 12 presents the running plot of $\log \|\Sigma\|_F$ and $\log \mathrm{ESS}/mn$ estimated using A-SVE and G-SVE along with standard errors from 10 replications for each value of n. In both the plots, G-SVE and $\widehat{\mathrm{ESS}}_G/mn$ reach the convergence value faster.

6 Discussion

In case of slowly mixing Markov chains or multimodal target distributions, the traditional single chain empirical estimator for ACvF can excruciatingly underestimate the truth. This will have serious impact on estimation of lag covariances, large sample variance of Monte Carlo averages, and determining when to stop the simulation process. It has been concluded in the paper that all of this can be significantly avoided by sampling multiple chains starting from dispersed points and pooling in the information from all chains to build consistent estimators for $\Gamma(k)$, Σ, and ESS. We have derived the bias of G-ACvF. We have shown that not only G-ACvF is asymptotically unbiased but also counters the negative bias of single chain ACF estimator by an amount proportional to the number of chains m. We have also proved the strong consistency of G-SVE and derived its limiting bias and variance results. Although the asymptotic results for G-SVE match those for SVE in Andrews (1991), finite sample advantages are observed in terms of better coverage probability and faster convergence to truth.

For the purpose of experimentation in Section 5, Bartlett lag window has been used throughout. Lag window comparisons done by Andrews (1991) reveal that quadratic spectral (QS) lag window gives consistently better coverage probabilities for any confidence interval. However, SVE gives

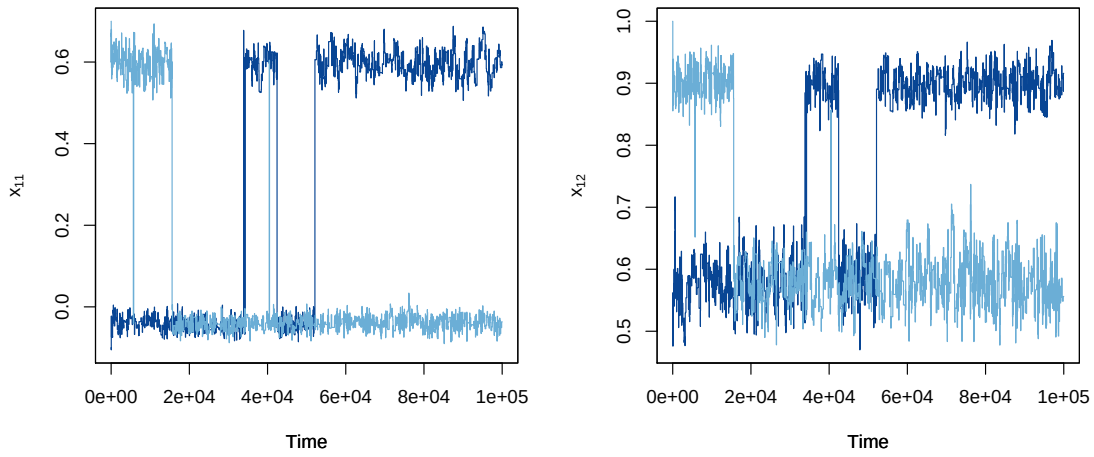
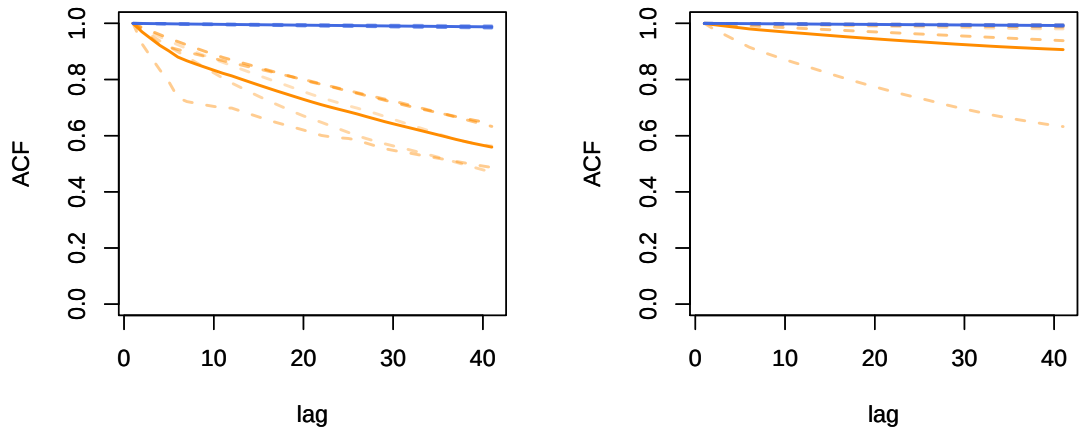


Figure 10: Sensor Network. Trace plot of two chains for x (left) and y (right) coordinate of sensor location-1.

poor coverage probability for all lag windows when the amount of autocorrelation is high in the Markov chain. Recently, *lugsail* lag window has been proposed by Vats and Flegal (2018) to deal with this issue by deliberately accentuating the autocovariance at smaller lags. We have refrained from doing lag window comparisons for G-SVE in our examples but we believe it can help improve the estimation quality significantly.

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, lag-covariances form the fundamental part of mIS estimators. In case of multi-modal target distributions, single chain mIS estimator will suffer from the same drawbacks as SVE. In this paper, we have chiefly focused on the usage of G-ACvF for estimating autocovariances in SVE; however the argument for a similar usage in mIS is straightforward. Following Dai and Jones (2017), generalised variance estimates obtained from mIS consistently overestimate $|\Sigma|$. From Theorem 1, one can prove that, on average, the eigenvalues of $\hat{\Gamma}_{G,s}(k)$ are greater than the eigenvalues of $\hat{\Gamma}_s(k)$. This supplies readers with sufficient motivation to believe that the generalized variance for mIS constructed from $\hat{\Gamma}_G(k)$ not only consistently overestimates $|\Sigma|$ but also upper-bounds the mIS from empirical single chain estimator of $\Gamma(k)$ for finite samples. An valuable line of research would be to prove these observations and carry out the theoretical analysis of mIS estimators using G-ACvF.

The choice of starting vector is crucial for multiple chain sampling. In case the chains are started very near to each other, all the chains will explore nearly the same neighborhood in the state space. As a consequence, the G-ACvF will be equivalent to A-ACvF, and no significant improvement will be observed in estimation quality. We have used a deterministic initialization method in



(a) $n = 5 \times 10^3$ (b) $n = 5 \times 10^4$

Figure 11: Sensor Network. ACrF (orange) and G-ACrF (blue) for individual chains (dotted lines) and average over all chains (thick solid line).

our examples using our prior knowledge about the location of modes in the target distribution. Comparing existing initialization methods MA: maybe cite a few and developing new ones could be an interesting direction for future research.

7 Appendix

7.1 Preliminaries

Lemma 1. *(Csörgo and Révész (2014)). Suppose the conditions of Theorem 2 hold, then for all* $\epsilon > 0$ *and for almost all sample paths, there exists* $n_0(\epsilon)$ *such that* $\forall n \geq n_0$ *and* $\forall i = 1, ..., p$

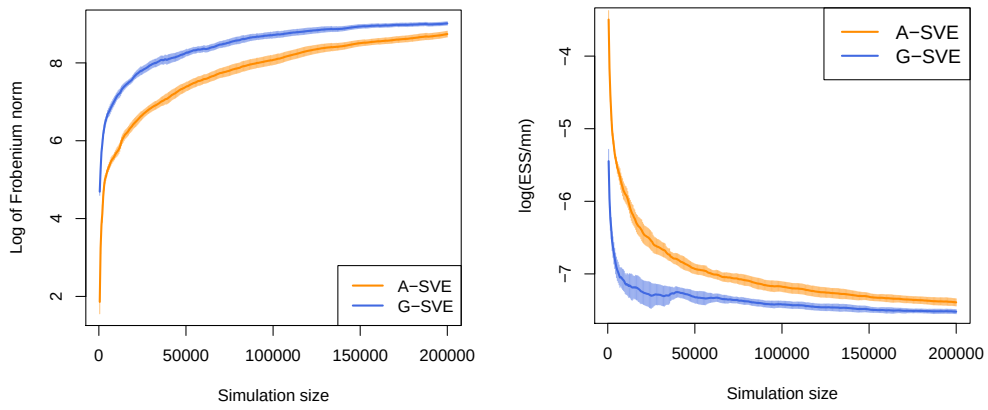
$$\sup_{0 \leq t \leq n - b_n} \sup_{0 \leq s \leq b_n} \left| B^{(i)}(t+s) - B^{(i)}(t) \right| < (1+\epsilon) \left(2b_n \left(\log \frac{n}{b_n} + \log \log n \right) \right)^{1/2},$$

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

$$\left| B^{(i)}(n) \right| < (1+\epsilon) \sqrt{2n \log \log n}.$$

7.2 Proof of Theorem 1

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



(a) log Frobenius norm

(b) $\log(\widehat{\mathrm{ESS}}/mn)$

Figure 12: Sensor Network. Running plot for $\log \|\Sigma\|_F$ (left) and $\log \mathrm{ESS}/mn$ (right) estimated using A-SVE and G-SVE along with standard errors for 10 replications.

$$
\begin{aligned}
\hat{\Gamma}_{G,s}(k) &= \frac{1}{n}\sum_{t=1}^{n-|k|}\left(X_{st}-\bar{\bar{X}}\right)\left(X_{s(t+k)}-\bar{\bar{X}}\right)^T \\
&= \left[\frac{1}{n}\sum_{t=1}^{n-|k|}\left(X_{st}-\bar{X}_s\right)\left(X_{s(t+k)}-\bar{X}_s\right)^T\right] + \left[\frac{1}{n}\sum_{t=1}^{|k|}\left(\bar{X}_s-\bar{\bar{X}}\right)\left(\bar{X}_s-X_{st}\right)^T\right] \\
&\quad + \left[\frac{1}{n}\sum_{t=n-|k|+1}^{n}\left(\bar{X}_s-X_{st}\right)\left(\bar{X}_s-\bar{\bar{X}}\right)^T\right] + \left[\frac{n-|k|}{n}\left(\bar{X}_s-\bar{\bar{X}}\right)\left(\bar{X}_s-\bar{\bar{X}}\right)^T\right] \\
&= \hat{\Gamma}_s(k) - \frac{1}{n}\sum_{t=1}^{|k|}A_{st}^T - \frac{1}{n}\sum_{t=n-|k|+1}^{n}A_{st} + \frac{n-|k|}{n}\left(\bar{X}_s-\bar{\bar{X}}\right)\left(\bar{X}_s-\bar{\bar{X}}\right)^T, \qquad (7)
\end{aligned}
$$

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

$$
\begin{aligned}
&\mathbb{E}\left[A_{11}\right] \\
&= \mathbb{E}\left[\left(X_{11}-\bar{X}_1\right)\left(\bar{X}_1-\bar{\bar{X}}\right)^T\right] \\
&= \mathbb{E}\left[X_{11}\bar{X}_1^T\right] - \frac{1}{m}\mathbb{E}\left[X_{11}\bar{X}_1^T\right] - \frac{m-1}{m}\mathbb{E}\left[X_{11}\bar{X}_2^T\right] + \frac{1}{m}\mathbb{E}\left[\bar{X}_1\bar{X}_1^T\right] + \frac{m-1}{m}\mathbb{E}\left[\bar{X}_1\bar{X}_2^T\right] - \mathbb{E}\left[\bar{X}_1\bar{X}_1^T\right] \\
&= \frac{m-1}{m}\left(\mathbb{E}\left[X_{11}\bar{X}_1^T\right] - \mathbb{E}\left[X_{11}\bar{X}_2^T\right] + \mathbb{E}\left[\bar{X}_1\bar{X}_2^T\right] - \mathbb{E}\left[\bar{X}_1\bar{X}_1^T\right]\right)
\end{aligned}
$$

$$= \frac{m-1}{m}\left(\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[X_{11}X_{1t}^T\right] - \mathbb{E}\left[X_{11}\right]\mathbb{E}\left[\bar{X}_2^T\right] + \mathbb{E}\left[\bar{X}_1\right]\mathbb{E}\left[\bar{X}_2^T\right] - \mathrm{Var}\left[\bar{X}_1\right] - \mathbb{E}\left[\bar{X}_1\right]\mathbb{E}\left[\bar{X}_1^T\right]\right)$$

$$= \frac{m-1}{mn}\left(\sum_{k=0}^{n-1}\Gamma(k) - n\mathrm{Var}\left[\bar{X}_1\right]\right). \tag{8}$$

Similarly,

$$\mathbb{E}\left[A_{11}^T\right] = \mathbb{E}\left[A_{11}\right]^T = \frac{m-1}{mn}\left(\sum_{k=0}^{n-1}\Gamma(k)^T - n\mathrm{Var}\left[\bar{X}_1\right]\right). \tag{9}$$

Further,

$$\begin{aligned}\mathbb{E}\left[\left(\bar{X}_1 - \bar{\bar{X}}\right)\left(\bar{X}_1 - \bar{\bar{X}}\right)^T\right] &= \mathbb{E}\left[\bar{X}_1\bar{X}_1^T - \bar{X}_1\bar{\bar{X}}^T - \bar{\bar{X}}\bar{X}_1^T + \bar{\bar{X}}\bar{\bar{X}}^T\right] \\ &= \left(\mathrm{Var}(\bar{X}_1) + \mu\mu^T - \mathrm{Var}(\bar{\bar{X}}) - \mu\mu^T\right) \\ &= \frac{m-1}{m}\mathrm{Var}(\bar{X}_1)\,. \end{aligned} \tag{10}$$

Additionally, the locally-centered autocovariance exhibits the following expectation (from Priestley (1981))

$$\mathbb{E}[\hat{\Gamma}(k)] = \left(1 - \frac{|k|}{n}\right)\left(\Gamma(k) - \mathrm{Var}(\bar{X})\right). \tag{11}$$

As a consequence, if $\mathrm{Var}(\bar{X})$ is finite, then $\mathrm{Var}(\bar{X}) \to 0$ as $n \to \infty$. Using (11), (8), (9), and (10) in (7),

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

We can use the results of Song and Schmeiser (1995) to expand $\mathrm{Var}(\bar{X}_1)$. By proposition 1 in Song and Schmeiser (1995)

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

As a consequence, if $\mathrm{Var}(\bar{X})$ is finite, then $\mathrm{Var}(\bar{X}) \to 0$ as $n \to \infty$. Expectation of $\hat{\Gamma}_{G,s}(k)$ can then broken down as following,

$$\mathbb{E}\left[\hat{\Gamma}_{G,s}(k)\right] = \left(1 - \frac{|k|}{n}\right)\Gamma(k) + O_1 + O_2\,. \tag{12}$$

where,

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

We observe that both O_1 and O_2 are small order terms that converge to 0 as $n \to \infty$. Here, $O_1 = (-|k|/n)\mathcal{O}(1/n)$ and $O_2 = \mathcal{O}(1/n)$. For a diagonal element of Γ,

$$\mathbb{E}\left[\hat{\Gamma}^{ii}_{G,s}\right]$$
$$= \mathbb{E}\left[\hat{\Gamma}^{ii}_{s}(k)\right] - \frac{|k|}{n}\left[\left(1-\frac{1}{m}\right)\left(\frac{1}{n}\sum_{h=0}^{n-1}\left[\Gamma^{ii}(h)^T + \Gamma^{ii}(h)\right]\right) - \left(2-\frac{1}{m}\right)\mathrm{Var}(\bar{X}_1)^{ii}\right] - \frac{\mathrm{Var}(\bar{X}_1)^{ii}}{m}\,.$$

In the presence of positive correlation, the leftover term is positive.

7.3 Strong consistency argument

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

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

The average pseudo spectral variance estimator is

$$\tilde{\Sigma} = \frac{1}{m}\sum_{s=1}^{m}\tilde{\Sigma}_s$$

Further, let

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

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

Lemma 2. *For the G-SVE estimator,* $\hat{\Sigma}_G^{ij} = \tilde{\Sigma}^{ij} + M_1 + M_2$ *and*

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

where

$$g_1(n) = (4+C_1)\frac{b_n\psi^2(n)}{n^2} - 4\frac{\psi^2(n)}{n^2} \to 0$$

$$g_2(n) = 2\sqrt{2}\|L\|p^{1/2}(1+\epsilon)\left[(4+C_1)\frac{b_n\psi(n)\sqrt{n\log\log n}}{n^2} - 4\frac{\psi(n)\sqrt{n\log\log n}}{n^2}\right]\to 0$$

$$g_3(n) = \|L\|^2 p(1+\epsilon)^2\left[(4+C_1)\frac{b_n\log\log n}{n} - 4\frac{\log\log n}{n}\right]\to 0.$$

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

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

Consequently

$$\left|\hat{\Sigma}_G^{ij}-\tilde{\Sigma}^{ij}\right| = |M_1+M_2| \le |M_1|+|M_2|.$$

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

$$\begin{aligned}
\|\bar{Y}_s - \mu_g\|_\infty \leq \|\bar{Y}_s - \mu_g\| &= \frac{1}{mn}\left\|\sum_{t=1}^{n} Y_{st} - n\mu_g\right\| \\
&\leq \frac{1}{n}\left\|\sum_{t=1}^{n} Y_{st} - n\mu_g - LB(n)\right\| + \frac{\|LB(n)\|}{n} \\
&< \frac{D\psi(n)}{n} + \frac{\|LB(n)\|}{n} \\
&< \frac{D\psi(n)}{n} + \frac{1}{n}\|L\|\left(\sum_{i=1}^{p}|B^{(i)}(n)|^2\right)^{1/2} \\
&\leq \frac{D\psi(n)}{n} + \frac{1}{n}\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}\,. \qquad (13)
\end{aligned}$$

Similarly,

$$\|\bar{\bar{Y}} - \mu_g\|_\infty \leq \frac{D\psi(n)}{n} + \frac{1}{n}\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}\,. \qquad (14)$$

Now consider,

$$\begin{aligned}
&|M_1| \\
&\leq \frac{1}{m}\sum_{s=1}^{m}\left\{\sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right)\left[\frac{1}{n}\left|\sum_{t=1}^{n-|k|}(Y_{st}-\mu_g)_i\right|\left|(\mu_g-\bar{\bar{Y}})_j\right| + \frac{1}{n}\left|(\mu_g-\bar{\bar{Y}})_i\right|\left|\sum_{t=1}^{n-|k|}(Y_{j(t+k)}-\mu_g)_j\right|\right]\right\} \\
&\leq \frac{\|(\bar{\bar{Y}}-\mu_g)\|_\infty}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1}\left[\frac{1}{n}\left\|\sum_{t=1}^{n-|k|}(Y_{st}-\mu_g)\right\|_\infty + \frac{1}{n}\left\|\sum_{t=1}^{n-|k|}(Y_{s(t+k)}-\mu_g)\right\|_\infty\right] \\
&\leq \frac{\|(\bar{\bar{Y}}-\mu_g)\|_\infty}{m} \\
&\quad\times\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1}\left[\frac{1}{n}\left\|\sum_{t=n-|k|+1}^{n}(Y_{st}-\mu_g) - n(\bar{Y}_s-\mu_g)\right\|_\infty + \frac{1}{n}\left\|\sum_{t=1}^{|k|}(Y_{st}-\mu_g) - n(\bar{Y}_s-\mu_g)\right\|_\infty\right] \\
&\leq \frac{\|(\bar{\bar{Y}}-\mu_g)\|_\infty}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1}\left[\frac{1}{n}\left\|\sum_{t=n-|k|+1}^{n}(Y_{st}-\mu_g)\right\|_\infty + \frac{1}{n}\left\|\sum_{t=1}^{|k|}(Y_{st}-\mu_g)\right\|_\infty + 2\|\bar{Y}_s-\mu_g\|_\infty\right] \\
&\leq \frac{\|(\bar{\bar{Y}}-\mu_g)\|_\infty}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1}\frac{1}{n}\left[\left\|\sum_{t=n-|k|+1}^{n}(Y_{st}-\mu_g)\right\|_\infty + \left\|\sum_{t=1}^{|k|}(Y_{st}-\mu_g)\right\|_\infty\right] \\
&\quad + 2(2b_n-1)\|\bar{\bar{Y}}-\mu_g\|_\infty\|\bar{Y}_1-\mu_g\|_\infty\,.
\end{aligned}$$

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

$$
\begin{aligned}
|M_1| &< 2\|(\bar{\bar{Y}} - \mu_g)\|_\infty \left[\sum_{k=-b_n+1}^{b_n-1} \left[\frac{D\psi(k)}{n} + \frac{\|L\| p^{1/2}(1+\epsilon)\sqrt{2k \log\log k}}{n} \right] \right] + 2(2b_n - 1)\|\bar{Y}_1 - \mu_g\|_\infty \\
&\leq 2(2b_n - 1)\|(\bar{\bar{Y}} - \mu_g)\|_\infty \left[\frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2}(1+\epsilon)\sqrt{n \log\log n}}{n} + \|\bar{Y}_1 - \mu_g\|_\infty \right] \\
&\leq 4(2b_n - 1) \left[\frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2}(1+\epsilon)\sqrt{n \log\log n}}{n} \right]^2 \quad \text{(by (13) and (14))}\,. \qquad (15)
\end{aligned}
$$

For M_2,

$$
\begin{aligned}
|M_2| &= \left| \frac{1}{m} \sum_{s=1}^{m} \left\{ \left(\mu_g - \bar{\bar{Y}}\right)_i \left(\mu_g - \bar{\bar{Y}}\right)_j \sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n}\right) w\left(\frac{k}{b_n}\right) \right\} \right| \\
&\leq \|\bar{\bar{Y}} - \mu_g\|_\infty^2 \left[\sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n}\right) w\left(\frac{k}{b_n}\right) \right] < \|\bar{\bar{Y}} - \mu_g\|_\infty^2 \left[\sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \right] \\
&\leq b_n \|\bar{\bar{Y}} - \mu_g\|_\infty^2 \int_{-\infty}^{\infty} |w(x)| dx \\
&\leq C b_n \left[\frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2}(1+\epsilon)\sqrt{n \log\log n}}{n} \right]^2 \quad \text{(by (14))}\,. \qquad (16)
\end{aligned}
$$

Using (15) and (16),

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

where

$$
\begin{aligned}
g_1(n) &= (8 + C)\frac{b_n \psi^2(n)}{n^2} - 4\frac{\psi^2(n)}{n^2} \\
g_2(n) &= 2\sqrt{2}\|L\| p^{1/2}(1+\epsilon) \left[(8 + C)\frac{b_n \psi(n)\sqrt{n \log\log n}}{n^2} - 4\frac{\psi(n)\sqrt{n \log\log n}}{n^2} \right] \\
g_3(n) &= \|L\|^2 p (1+\epsilon)^2 \left[(8 + C)\frac{b_n \log\log n}{n} - 4\frac{\log\log n}{n} \right]\,.
\end{aligned}
$$

Under our assumptions, $b_n \log\log n/n \to 0$ and $\psi(n) = o(\sqrt{n \log\log n})$. Consequently, $b_n\psi^2(n)/n^2 \to 0$, $\psi^2(n)/n^2 \to 0$, $b_n\psi(n)\sqrt{n \log\log n}/n^2 \to 0$, and $\psi(n)\sqrt{n \log\log n}/n^2 \to 0$. Thus, $g_1(n), g_2(n)$ and $g_3(n) \to 0$ as $n \to \infty$. MA: Should we add a footnote here telling why $\phi(n) = o(\sqrt{n \log\log n})$? It was earlier stated explicitly in Assumption-1 which is not Theorem-1.

□

Proof of theorem 3. We have the following decomposition,

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

where

$$\begin{aligned}
N_1 &= \frac{1}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right)\frac{1}{n}\sum_{t=1}^{n-|k|}\left[\left(Y_{st}-\bar{Y}_s\right)_i\left(\bar{Y}_s-\mu_g\right)_j+\left(\bar{Y}_s-\mu_g\right)_i\left(Y_{s(t+k)}-\bar{Y}_s\right)_j\right] \\
N_2 &= \left[\frac{1}{m}\sum_{s=1}^{m}\left(\bar{Y}_s-\mu_g\right)_i\left(\bar{Y}_s-\mu_g\right)_j\right]\left[\sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right)\left(1-\frac{|k|}{b_n}\right)\right].
\end{aligned}$$

Using the above and Lemma 2,

$$\left|\hat{\Sigma}^{ij}_G-\Sigma^{ij}\right| = \left|\hat{\Sigma}^{ij}_{SV}-\Sigma^{ij}+N_1+N_2+M_1+M_2\right| \le \left|\hat{\Sigma}^{ij}_{SV}-\Sigma^{ij}\right|+|N_1|+|N_2|+|M_1+M_2| \quad (17)$$

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 $n\to\infty$. It is left to show that $|N_1|\to 0$ and $|N_2|\to 0$ with probability 1

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

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

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

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

Using SIP on the summation of k terms,

$$
\begin{aligned}
&\left|\frac{1}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1}w\left(\frac{k}{b_n}\right)\frac{1}{n}\sum_{t=1}^{n-|k|}\left[\left(Y_{st}-\bar{Y}_s\right)_i\left(\bar{Y}_s-\mu_g\right)_j\right]\right|\\
&<\frac{1}{m}\sum_{s=1}^{m}\|\bar{Y}_s-\mu_g\|_\infty\sum_{k=-b_n+1}^{b_n-1}\left[\frac{D\psi(k)}{n}+\frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2k\log\log k}}{n}\right]+\frac{1}{m}\sum_{s=1}^{m}\frac{b_n(b_n-1)}{n}\|\bar{Y}_s-\mu_g\|_\infty^2\\
&<\frac{(2b_n-1)}{m}\sum_{s=1}^{m}\|\bar{Y}_s-\mu_g\|_\infty\left[\frac{D\psi(n)}{n}+\frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}}{n}\right]+\frac{1}{m}\sum_{s=1}^{m}\frac{b_n(b_n-1)}{n}\|\bar{Y}_s-\mu_g\|_\infty^2\\
&\leq\left(2b_n-1+\frac{b_n^2}{n}-\frac{b_n}{n}\right)\left[\frac{D\psi(n)}{n}+\frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}}{n}\right]^2\to 0\,. \quad \text{(by (13))}
\end{aligned}
$$

Similarly, the second part of $N_1 \to 0$ with probability 1. Following the steps in (16),

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

Thus, in (17), every term goes to 0 and $\hat{\Sigma}_G^{ij}\to\Sigma^{ij}$ with probability 1 as $n\to\infty$. □

7.4 Proof of Theorem 4

By Equation 12,

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

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, $\sum_{h=-\infty}^{\infty} \Upsilon(h) < \infty$. Consider the G-SVE estimator,

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

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 (2009), 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,

$$P_1 = -\sum_{|k|\geq n} \Upsilon(k) - \sum_{k=-n+1}^{n-1} w\left(\frac{|k|}{n}\right)\frac{|k|}{n}\Upsilon(k) - \sum_{k=-n+1}^{n-1}\left(1 - w\left(\frac{|k|}{n}\right)\right)\Upsilon(k)\,. \tag{18}$$

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

$$-\sum_{|k|\geq n} \Upsilon(k) \leq \sum_{|k|\geq n}\left|\frac{k}{n}\right|^q \Upsilon(k) = \frac{1}{b_n^q}\left|\frac{b_n}{n}\right|^q \sum_{|k|\geq n} |k|^q \Upsilon(k) = o\left(\frac{1}{b_n^q}\right), \tag{19}$$

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,

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

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

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

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

$$\begin{aligned}\sum_{k=-n+1}^{n-1}\left(1-w\left(\frac{k}{b_n}\right)\right)\Upsilon(k) &= -\frac{1}{b_n^q}\sum_{k=-n+1}^{n-1}\left(\frac{|k|}{b_n}\right)^{-q}\left(1-w\left(\frac{|k|}{b_n}\right)\right)|k|^{q}\Upsilon(k)\\ &= -\frac{1}{b_n^q}\sum_{k=-n+1}^{n-1}\left[k_q+o(1)\right]|k|^{q}\Gamma(k)\\ &= -\frac{k_q\Phi^{(q)}}{b_n^q} + o\left(\frac{1}{b_n^q}\right).\end{aligned} \tag{21}$$

Finally, we will solve for P_2. Note that O_2 is independent of k. We will write O_1 as $(|k|/n)\mathcal{O}(1/n)$ and O_2 as $\mathcal{O}(1/n)$. We will find an upper bound and prove that it is $\mathcal{O}\left(n^{-1}\right)$:

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

Using (19), (20), and (21) in (18), we get

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

which completes the result.

7.5 Proof of Theorem 5

Due to the strong consistency proof from theorem 3, as $n \to \infty$,

$$\left|\hat{\Sigma}_G - \tilde{\Sigma}\right| \to 0 \text{ with probability } 1\,. \tag{22}$$

Further, we have defined $g_1(n), g_2(n), g_3(n)$ such that as $n \to \infty$,

$$\begin{aligned}
g_1(n) &= (4 + C_1)\frac{b_n \psi^2(n)}{n^2} - 4\frac{\psi^2(n)}{n^2} \to 0 \\
g_2(n) &= 2\sqrt{2}\|L\|p^{1/2}(1+\epsilon)\left[(4 + C_1)\frac{b_n \psi(n)\sqrt{n \log\log n}}{n^2} - 4\frac{\psi(n)\sqrt{n\log\log n}}{n^2}\right] \to 0 \\
g_3(n) &= \|L\|^2 p(1+\epsilon)^2\left[(4 + C_1)\frac{b_n \log\log n}{n} - 4\frac{\log\log n}{n}\right] \to 0\,.
\end{aligned}$$

We have shown from the proof of strong consistency that,

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

By (22), there exists an N_0 such that

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

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

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

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

$$\mathbb{E}\left[\left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right)^2\right] \to 0\,. \tag{23}$$

We will use (23) to show that the variances are equivalent. Define,

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

We will show that the above is $o(1)$. Using Cauchy-Schwarz inequality followed by (23),

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

Finally,

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

$$= \text{Var}\left(\tilde{\Sigma}^{ij}\right) + \text{Var}\left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right) + 2\mathbb{E}\left[\left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right)\left(\tilde{\Sigma}^{ij} - \mathbb{E}\left(\tilde{\Sigma}^{ij}\right)\right)\right] + o(1)$$
$$= \text{Var}\left(\tilde{\Sigma}^{ij}\right) + o(1)\,.$$

Hannan (2009) has given the calculations for variance of $\tilde{\Sigma}$ as

$$\frac{n}{b_n}\text{Var}(\tilde{\Sigma}^{ij}) = \left[\Sigma^{ii}\Sigma^{jj} + \left(\Sigma^{ij}\right)\right]\int_{-\infty}^{\infty} w^2(x)dx + o(1) \tag{24}$$

Plugging (24) into variance of $\hat{\Sigma}_G$ gives the result of the theorem.

8 Additional Examples

We present two additional real-world examples to illustrate the advantage of global autocorrelation estimator using ACF plots. For both the examples, we use a pre-established Markov chain structure and implement our estimator on the Markov chains obtained.

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:

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

This is a 7-dimensional estimation problem wherein the marginal distribution of majority of components display a multimodal nature. Figure 13 shows the evolution of two chains started from random points with time for the second component. We will report the ACF plots for component-2. Similar behavior is observed for ACF plots of other components as well.

Figure 14 demonstrates a striking advantage of G-ACF in estimating the true autocorrelations. G-ACF gives a more realistic and accurate estimation of autocorrelations in almost ten times lesser chain length.

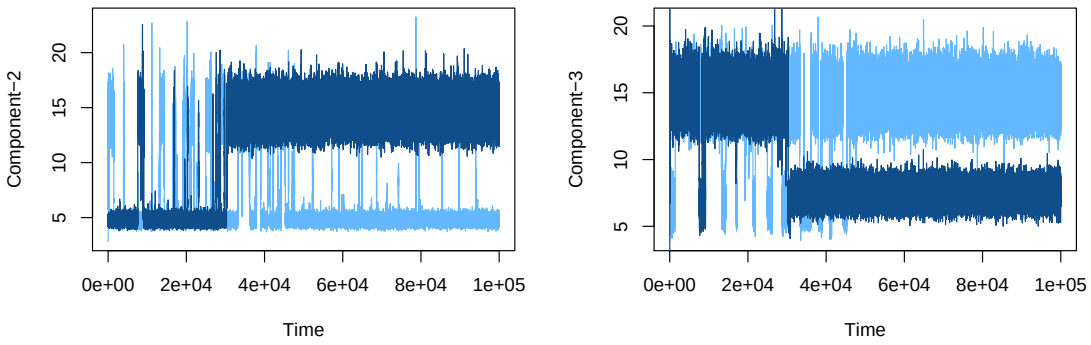


Figure 13: Trace plot for second (left) and third (right) component of two chains started from random points.The two colors denote the two different chains.

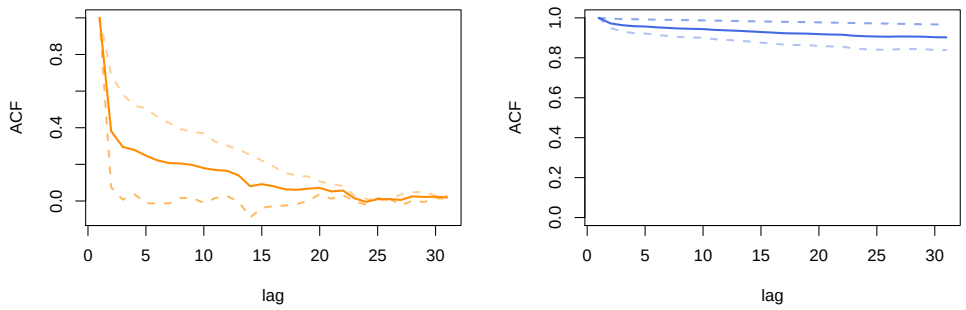
8.2 Network crawling

The `faux.magnolia.high` dataset available in the `ergm R` package represents a simulated within school 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).

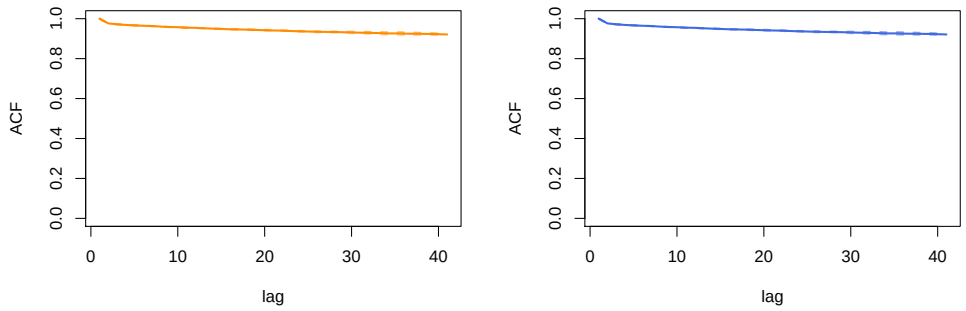
We believe that the students from different races engage in a "selective networking" which might cause formation of clusters in the network wherein students within a cluster engage more with each other than with the students outside it. We sample two parallel Markov chains starting from two students belonging to different races and study its impacts on the average features of their immediate social group. Figure 15 shows the ACF and G-ACF plots for the third feature at two different simulation sizes. As evident, G-ACF displays a clear advantage over ACF in just $n = 100$ samples.

References

Ahn, S., Chen, Y., and Welling, M. (2013). Distributed and adaptive darting monte carlo through regenerations. In *Artificial Intelligence and Statistics*, pages 108–116.



(a) $n = 1000$



(b) $n = 10000$

Figure 14: Poisson change point. ACF (left) and G-ACF (right) for individual chains and average over m chains for two chain lengths.

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.

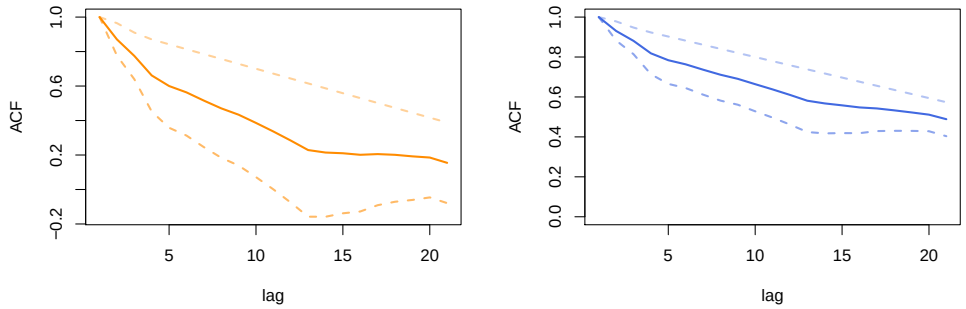
Chen, D.-F. R. and Seila, A. F. (1987). Multivariate inference in stationary simulation using batch means. In *Proceedings of the 19th conference on Winter simulation*, pages 302–304. ACM.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of econometrics*, 86(2):221–241.

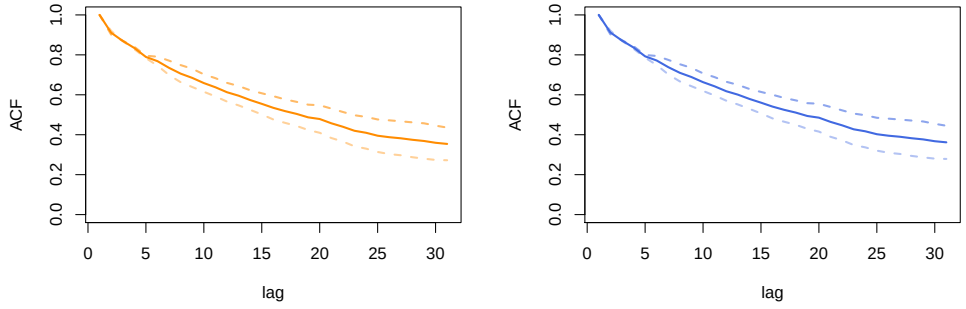
Csörgo, M. and Révész, P. (2014). *Strong approximations in probability and statistics.* Academic Press.

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.



(a) $n = 100$



(b) $n = 1000$

Figure 15: Network crawling. ACF (left) and G-ACF (right) for individual chains and average over m chains for two chain lengths.

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. 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.

Hannan, E. J. (2009). *Multiple time series*, volume 38. John Wiley & Sons.

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.

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. and Flegal, J. M. (2018). Lugsail lag windows and their application to MCMC. *ArXiv e-prints*.

Vats, D. and Flegal, J. M. (2018). Lugsail lag windows and their application to mcmc. *arXiv preprint arXiv:1809.04541*.

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. (2019a). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika*, 106:321–337.

Vats, D., Flegal, J. M., and Jones, G. L. (2019b). Multivariate output analysis for markov chain monte carlo. *Biometrika*, 106(2):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.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, pages 471–494.

Zeidler, E. (2013). *Nonlinear functional analysis and its applications: III: variational methods and optimization*. Springer Science & Business Media.