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

July 29, 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 more accurately capture characteristics of the target distribution. Output from m parallel chains are 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.

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 $\mathcal{B}(\mathcal{X})$. Let $\{X_1, X_2, ..., X_n\}$ be an F-Harris ergodic (aperiodic, F-irreducible, and positive Harris recurrent) Markov chain.

For a stationary Markov chain, the autocovariance function at lag k depends only on k and is defined as

$$\Gamma(k) = \text{Cov}(X_t, X_{t+k}) = \mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)^T].$$

A popular sample autocovariance matrix at lag-k is given by

$$\hat{\Gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (X_t - \bar{X})(X_{(t+|k|)} - \bar{X})^T. \tag{1}$$

It is a known fact that most of the Markov chains obtained from practical MCMC algorithms are not stationary. Regardless, Equation 1 provides the empirical estimator for ACF at a particular lag for the Markov chain with any initial distribution (Geyer (2011)).

Calculation of exact ACF has been done for simple stationary stochastic models like autoregressive (AR) models (under assumption of stationarity), moving average (MA) (Quenouille (1947)), and autoregressive-moving-average (ARMA) (Box et al. (2015)) models. However, for most of the complex sampling algorithms (including MCMC), we rely on sample estimators for ACF. Unfortunately, the estimator given in Equation 1 has been observed to have a some drawbacks. Firstly, it is not immune to outliers (see Ma and Genton (2000)) and secondly, it is usually non-informative in case of slowly mixing Markov chains. We address the second issue in this paper by using multiple representative samples of the target distribution via parallel Markov chains. This approach is especially relevant to MCMC wherein multiple chain sampling is computationally possible and statistically desirable for making informative inferences. Our goal is to pool in the information from all the Markov chains while estimating the autocovariances.

A commonly faced problem with the MCMC models is the slow mixing property. Especially in case of multi modal target distributions, it has been observed that the Markov chains tend to get *stuck* in a particular mode for a long time. In such situations, single chain empirical ACF will severely underestimate the autocovariance. We propose an autocovariance estimator called globally-centered autocovariance function (G-ACF) with a simple fix in centering which has been observed to show drastically better results in many cases. We also prove that our estimator compensates the negative bias of the empirical ACF estimator to give estimates that are closer to reality. We will further

study the impact of G-ACF in calculation of two important applications of ACF (1) asymptotic variance of Markov chain average and (2) effective sample size (ESS).

Let $g : \chi \to \mathbb{R}^d$ be an F-integrable function and suppose we are interested in estimating the expected value $\mu_g = \mathbb{E}_F[g(X)]$. We know that $\sum_{t=1}^{n} g(X_t)/n$ converges to μ_g as $n \to \infty$ almost surely. MCMC ensures asymptotic convergence, however this is not practically very useful where the number of samples is finite and the Markov chain mixes slowly. One of the important criteria to determine the quality of estimation is the Monte Carlo aymptotic covariance matrix which is given by an infinite summation of autocovariance matrices for all lags. For a slowly mixing Markov chain, underestimated autocovariances will deliver underestimated asymptotic variance. There are many techniques for estimating it like batch means (BM), overlapping bath means (OBM), and spectral variance estimator (SVE). Due to its nice statistical properties, we will only focus on SVE in this paper. SVE involves a truncated and weighted sum of autocovariances. We propose a version of SVE that uses our globally centered autocovariances. We call this globally-centered SVE (G-SVE).

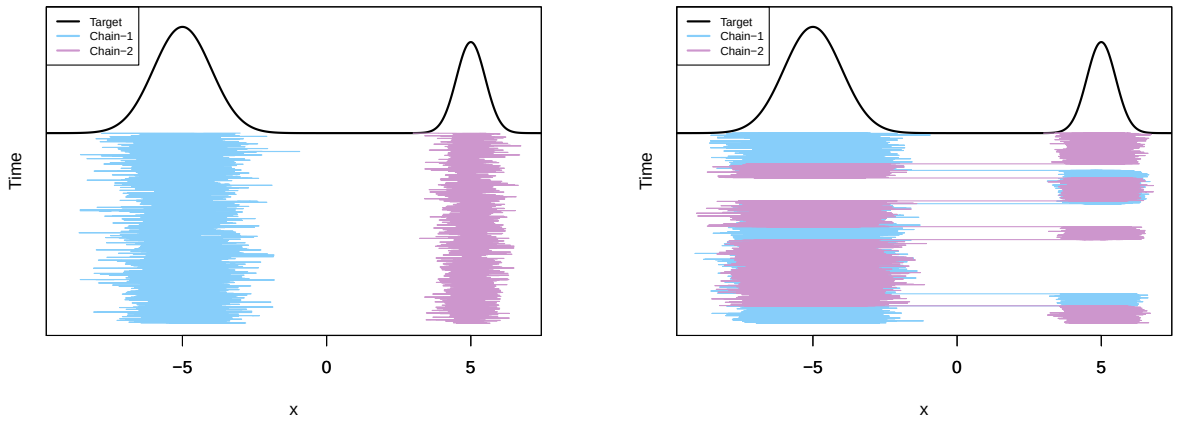
A more powerful estimator of Monte Carlo asymptotic covariance matrix can enable us to estimate the effective sample size for the Markov chain which is closer to reality. There are many criteria for terminating simulation like fixed-time rule, sequencial fixed-volume rule (see Glynn et al. (1992)), and using Gelman Rubin diagnostics (Gelman et al. (1992)). The relative fixed-volume rule has been thoroughly discussed for MCMC by Flegal and Gong (2015) and Gong and Flegal (2016). We will use method of effective sample size (ESS) by Vats et al. (2019) for terminating the simulations. The quality of estimation of asymptotic covariance matrix is critical in estimation of the ESS. We will successfully show that the G-SVE estimates the truth better and prevents the premature stopping of Markov chain.

The rest of the paper is arranged as follows. In Section 2 we introduce the new sample autocovariance estimator and calculate its bias for finite samples. We observe that the bias term for G-ACF is similar to the bias of empirical ACF except for a few small order end-effect terms that vanish as $n \to \infty$. In Section 3, we introduce the globally-centered SVE and give a proof of strong consistency, bias, and variance calculations. We will elaborate on the estimator for ESS using G-SVE in Section 4. Section 5 gives an experimental study on three different examples for the proposed estimators. All the important proofs are given in the Appendix.

Example 1 (Demonstrative example). We demonstrate the striking difference in the ACF plots through mixture of Gaussian target density using a random-walk Metropolis-Hastings sampler. The target density has two well separated and non-equal modes. The target distribution F defined on $\mathbb{R}$ is given by

$$F(x) = 0.7\,\mathrm{N}(-5, 1) + 0.3\,\mathrm{N}(5, 0.5)$$

Two Markov chains are started close to each mode. It can be seen in Figure 1 that for the first 10^4 samples, the chains do not communicate with the other mode. Whereas for a sample size of 10^5, the chains have jumped modes and therefore, a single chain can give reliable information about the distribution. We wish to make two points here: firstly, for small sample size, we observe that ACF gives extremely misleading estimates of autocovariance whereas G-ACF is very close to reality (Figure 2a). Secondly, for a large sample size, the estimates from G-ACF as well as ACF are equivalent. Figure 1b depicts the scenario when convergence is achieved. This shows that G-ACF is at least as good as ACF; hence safe to use in any setting.



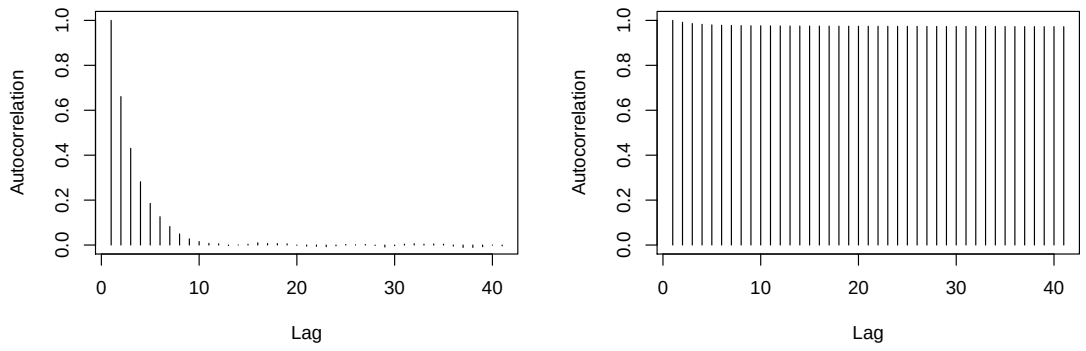
(a) $n = 10^4$ (b) $n = 10^5$

Figure 1: Target density and Density plot for two chains starting at different values for a Gaussian mixture model with mixture probability $= 0.7$, $\mu_1 = -5$, $\mu_2 = 5$, $\sigma_1 = 1$, and $\sigma_2 = 0.5$. (a) $n = 10^4$; chains don't interact. (b) $n = 10^5$; chains interact.

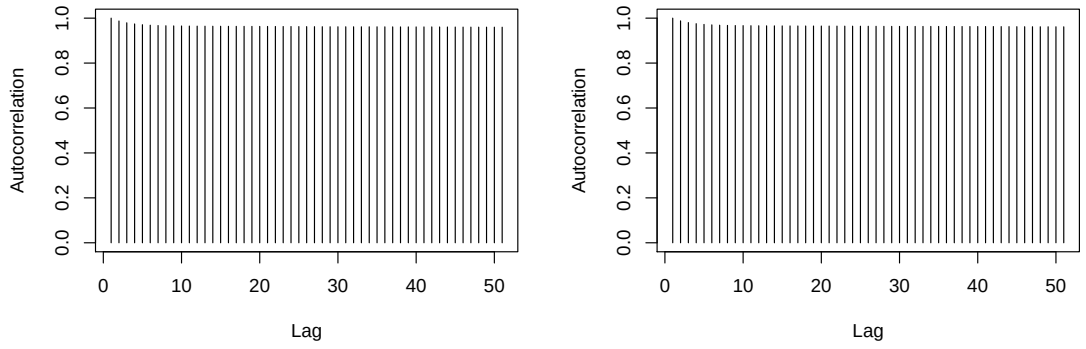
2 Globally-centered autocovariance

Suppose we have m replications of Markov chains for same target distribution and $\{X_{st};\ t \in \mathbb{Z}\}$ denote the s^{th} Markov chain. Let the sample mean of s^{th} Markov chain be denoted by $\bar{X}_s$ and the global mean by $\bar{\bar{X}} = \sum_{s=1}^{m} \bar{X}_s/m$. The global mean provides a better estimate for the expectation value when the number of simulation per chain are not sufficient for the convergence to kick in. The globally-centered autocovariance function (G-ACF) for the s^{th} Markov chain is given by:

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



(a) $n = 10^4$



(b) $n = 10^5$

Figure 2: Autocorrelation plot for locally-centered ACF (left) and globally centered ACF (right) averaged over all chains. (a) $n = 10^4$, not converged yet; (b) $n = 10^5$, converged

and the averaged globally-centered ACF over all m chains is given by

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

If the starting points of these parallel Markov chains are sufficiently dispersed, G-ACF provides more realistic estimation of lag covariances. Priestley (1981) shows that the empirical estimator in Equation 1 is biased with the main bias term equal to $|k|\,\Gamma(k)/n$, ignoring the small order terms arising due to estimation of μ. Though, an unbiased estimator with a divisor of $n-|k|$ instead of n exists, the biased estimator is preferred for its lower mean square error for larger relative k (Priestley (1981))

The bias for G-ACF is similar to the bias of sample autocovariance estimator in case of single chain except for a few small order terms that shall be later addressed in theorem 1. Before getting to the bias results for $\hat{\Gamma}_{RAV}(k)$ for any lag k, we introduce an additional notation. For $q \geq 1$, we define

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

We denote $\Phi^{(1)}$ by Φ.

The proof of the following theorem is available in Appendix A.2.

Fix the proof for this.

Theorem 1. *Under stationarity,*

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

and consequently,

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

When $m = 1$, $\hat{\Gamma}_G(k)$ corresponds to $\hat{\Gamma}(k)$, the expectation of which can be found in Priestley (1981). Theorem 1 implies that the globally-centered autocovariances are asymptotically unbiased and exhibit reduced bias in finite samples compared to locally-centered autocovariances. The consequences of this are particularly pertinent for the diagonals of Γ. Special interest is also in the lag-0 autocovariance, that is, the variance of the target distribution.

Remark 1.

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

Lag-0 autocovariance are particularly useful in calculating the autocorrelations.The variance of i^{th}

component given by $\Gamma^{ii}(0)$. For any component i, the autocorrelation is

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

and is instrumental in visualizing the serial correlation in components of the Markov chain. The typical estimate of the autocorrelation is constructed from locally-centered autocovariance estimates which as was demonstrated in Section 1 can significantly underestimate the autocorrelation. Instead, we calculate the globally-centered autocorrelation using the formula

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

Despite using similar centering in both numerator and denominator, the globally-centered autocorrelation is far greater than locally-centered autocorrelation for each component as is evident from Figure 2.

3 Variance Estimators for multiple Markov chains

Let $g : \chi \longrightarrow \mathbb{R}^d$ be an F-integrable function. We are interested in the expected value

$$\mu_g = \mathbb{E}_F[g(X)] = \int_\chi g(x)F(dx)\ .$$

Let Y denote the random variable $g(X)$ where $X \sim F$. Since we have m chains of n samples each, let us make following notations: (1) $Y_{st} = g(X_{st})$ for $t \in \{1, 2, ..., n\}$ and $s \in \{1, 2, ..., m\}$; (2) $\bar{Y}_s = \sum_{t=1}^{n} g(X_{st})/n$; and (3) $\bar{\bar{Y}} = \sum_{s=1}^{m} \bar{Y}_s/m$. To assess the reliability of our simulations, we are interested in Monte Carlo error, i.e. $\bar{Y}_s - \mu$ for the s^{th} chain. The sampling distribution for Monte Carlo error is available via Markov chain CLT (Jones et al. (2004)) if there exists a positive definite matrix Σ such that

$$\sqrt{n}(\bar{X}_s - \mu) \xrightarrow{d} N(0, \Sigma)$$

Given m parallel Markov chains, we have m i.i.d. samples of $\bar{X}_s$, $s \in \{1, ..., m\}$. The Markov Chain standard error (MCSE) is given by Σ/mn. We are interested in consistent estimates of Σ for two reasons. Firstly, to construct asymptotically valid confidence intervals and secondly, to determine when to stop the simulations. For this purpose, we use the class of Multivariate Spectral Variance Estimators (MSVE). The most common practice would be to combine the spectral variance estimates from m chains by averaging over them. We call this technique average spectral variance estimator (A-SVE) and denote it by $\hat{\Sigma}_A$. Let $\hat{\Sigma}_{SV,s}$ denote the SVE for the s^{th} chain, then the A-SVE is given by

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

We propose a globally-centered spectral variance estimator (G-SVE) wherein we use G-ACF to estimate the autocovariance. Strong consistency for MSVE has been addressed by Vats et al. (2018). Bias and variance calculations for a variant of SVE with known mean is done by Hannan (2009). In this paper, we will address the asymptotic properties (strong consistency), bias and variance calculations for G-SVE.

3.1 Globally-centered spectral variance estimators

Let $w_n : \mathbb{R} \to [-z, z]$ be a lag window function and let $b_n \in \mathbb{N}$ be a truncation point. The usual SV estimator is the weighted and truncated sum of lag-k autocovariances:

$$\hat{\Sigma}_{SV} = \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right)\hat{\Gamma}_n(k) \tag{2}$$

Consider the globally-centered SV estimator (G-SVE) for the sth chain using G-ACF:

$$\hat{\Sigma}_{G,s} = \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right)\hat{\Gamma}_{G,s}(k)\,,$$

and the combined G-SVE is

$$\hat{\Sigma}_G = \frac{1}{m}\sum_{s=1}^{m} \hat{\Sigma}_{G,s}\,.$$

Define

$$W_n = \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right).$$

We will consider a range of lag windows $w(x)$; where $w(x)$ is a continuous, even function with $w(0) = 1, \big|w(x)\big| < c$, $\int_{-\infty}^{\infty} w^2(x)dx < \infty$, and $W_n < \infty$.

3.1.1 Theoretical results

The G-SVE estimator centers the data around the global mean. We are interested in proving the strong consistency, finding the bias and variance of G-SVE.

Assumption 1 (Strong Invariance Principle (SIP)). We assume the SIP holds for the Markov chains $\{Y_{st}\}_{t\geq 1}\ \forall s$. Here $\{Y_t\}_{t\geq 1}$ is the stochastic process that follows SIP with mean μ_g. Let $\|.\|$ denote the Euclidean norm. Through SIP, there exists a $p \times p$ lower triangular matrix L, an increasing function ψ on the integers, a finite random variable D, and a sufficiently rich probability space such

that, with probability 1,

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

Let $\{B(n)\}_{n\geq 0}$ denotes a standard p-dim Brownian motion and $B^{(i)}$ denotes its i^{th} component. As shown in the results Kuelbs and Philipp (1980), SIP will hold for $\psi(n) = n^{1/2-\lambda,\ \lambda>0}$ for Markov chains commonly encountered in MCMC settings. We know that $\psi(n)$ is an inherent property of the stochastic process that satisfies the CLT. This implies that $\psi(n)$ is at least $o(\sqrt{n})$. Using Law of Iterative Logarithms (LIL), a tighter bound for $\psi(n)$ is given by Strassen (1964) as $\mathcal{O}(\sqrt{n \log\log n})$.

Theorem 2. *Let the Assumptions 1 hold,* $\hat{\Sigma}_{SV,s} \xrightarrow{a.s.} \Sigma\ \forall s$, *and* $n^{-1}b_n \log\log n \to 0$ *as* $n \to \infty$, *then* $\hat{\Sigma}_G \to \Sigma$ *with probability 1, as* $n \to \infty$.

We are interested in the finding the bias for G-SVE estimator as a function of n.

Assumption 2. Let $\Gamma_s(k)$ be the lag-k autocovariance for s^{th} Markov chain and $w(x)$ be the lag window. We assume that there exists a $q \geq 0$ such that for all s

a. $\sum_{-\infty}^{\infty} |k|^q \|\Gamma_s(k)\| < \infty$

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

c. $\dfrac{b_n^q}{n} \to 0$ as $n \to \infty$

If k_q is finite for some q_o, then it is zero for $q < q_0$. q is taken to be the largest number satisfying the first two conditions above.

Theorem 3. *Let the Assumption 2 hold, then the limiting bias of G-SVE is given by:*

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

Theorem 4. *Let Assumption 1 holds,* $\mathbb{E}[D^4] < \infty$ *where D is a finite random variable associated with the SIP, and* $\mathbb{E}[Y^4] < \infty$, *then* $b_n^{-1} n\, Var\left(\hat{\Sigma}_G^{ij}\right) = [\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2] \int_{-\infty}^{\infty} w(x)^2 dx + o(1)$.

3.1.2 Fast implementation

The multivariate spectral variance estimator (MSVE) despite having good statistical properties poses application limitations due to slow computation and high storage demands. Observe from equation 1 and 2, the computation of spectral variance estimator has a complexity of $\mathcal{O}(b_n np)$ where n is the MCMC sample size, p is the dimensionality and b_n is the truncation point. To overcome this limitation for our experimental purposes, we used a fast Fourier transform based

algorithm presented by Heberle and Sattarhoff (2017) for calculating the alternate formulation of MSVE given by Kyriakoulis (2005).

Suppose $\{X_t\}_{t=1}^n$ is the Markov chain of n samples where $\overline{X}_n = n^{-1}\sum_{t=1}^n X_t$ is the MCMC average. Let w_k denote the lag-window at lag-k, i.e. $w_k = w_n(k/b_n)$. The alternate formulation for MSVE is given by

$$\hat{\Sigma}_{SV} = \frac{1}{n}A^T T(w) A, \qquad \text{where } A = \begin{pmatrix} X_1 - \overline{X}_n & \dots & X_n - \overline{X}_n \end{pmatrix}^T$$

and $T(w)$ is the $n \times n$ Toeplitz matrix of weights with the first column being $(1\ w_1\ w_2\ \dots,\ w_{n-1})^T$ $T(w)$ is an $n \times n$ matrix which can be difficult to store. Therefore, Heberle and Sattarhoff (2017) computed $T(w)A$ 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 $V\Lambda V^*$. A is extrapolated into a $2n \times p$ matrix A^* such that

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

The complete algorithm is as follows

Algorithm 1: Herberle's Algorithm

1 Construct $C(w^*)$ and A^* from the MCMC samples.
2 Compute the DFT of the first column of $C(w^*)$. This gives the eigenvalues of $C(w^*)$.
3 **for** $j = 1, 2, ..., p$ **do**
4 Calculate $V^* A_j^*$ by FFT of A_j
5 Multiply the i^{th} entry of the vector $V^* A_j^*$ with the eigenvalue λ_i for all $i = 1, 2, ..., 2n$ in order to construct $\Lambda V^* A_j^*$.
6 Determine $C(w^*)A^* = V\Lambda V^* A_j^*$ by inverse FFT of $\Lambda V^* A_j^*$.
7 **end**
8 Select the first n rows of $C(w^*)A^*$ to form $T(w)A$.
9 Premultiply by A^T and divide by n.

The algorithm has been implemented in the `R` package `mcmcse` for fast implementation of spectral variance estimators. We do a slight variation in the alternate formulation of SVE proposed by Kyriakoulis (2005) by centering the chain $\{X_{st};\ t \in \mathbb{Z}\}$ around $\overline{\overline{X}}$ instead of $\overline{X}_s$ for all $s \in \{1, ..., m\}$. Herberle's algorithm is then applied on the formulation

$$\hat{\Sigma}_{S,s} = \frac{1}{n}B^T T(w) B \qquad \text{where } B = \begin{pmatrix} X_{s1} - \overline{\overline{X}} & \dots & X_{sn} - \overline{\overline{X}} \end{pmatrix}^T$$

.

4 Effective sample size

Estimating the MCSE is crucial for determining when to terminate the simulations. The existing stopping rules rely on the strong consistency of the estimator of Σ. The determinant of Monte Carlo standard error is called generalized variance in Wilks (1932) and gives a metric for volume of confidence interval.

We will use the multivariate effective sample size (m-ESS) by Vats et al. (2019) to understand the effectiveness of simulations so far. m-ESS requires a strongly consistent estimator of Σ. G-SVE better captures the MCSE by considering the global sample mean across the Markov chains; which might otherwise get lost when considering a single localized slowly mixing Markov chain. We therefore define our ESS estimator as

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

where Λ_{mn} is the sample covariance of mn samples. When there is no correlation in Markov chain $\hat{\Sigma}_G = \Lambda_{mn}$ and therefore, ESS $= n$.

5 Examples

There are only a handful of stochastic processes where the true autocovariance and Σ are known. To compare the estimators when the truth is known, we use a slowly mixing vector autoregressive process of order 1 (VAR(1)). A more promising advantage of our estimators is observed when the target distribution is multi-modal. For this purpose we use a bivariate bi-modal probability distribution introduced by Gelman and Meng (1991) in section 5.2. Through this example, we will also show that in case of nicely mixing Markov chains, these estimators give almost equivalent results. As a consequence, while out replicated version of variance estimator is better in cases where the Markov chains do not explore the entire sample space in finite time, it is harmless to be used in cases where they do. Lastly, we consider a real-life example of finding unknown locations of sensors using noisy distance data in section 3. The posterior distribution is this case is marginally bimodal in all dimensions. All the examples have been selected to display some kind of *sticky* nature in the Markov chains. In such scenario, we successfully show that our globally-centered estimators yield better results as compared to the classical autocovariance function, spectral variance estimator, and effective sample size.

5.1 Vector Autoregressive Process

We examine the vector autoregressive process of order 1 (VAR(1)) where the true autocovariance in known in closed form. 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 $\vec{\Psi} = (I_{p^2} - \Phi \otimes \Phi)\vec{\Omega}$. The lag-k autocovariance can be calculated easily for $k > 0$ as

$$\Gamma(k) = \Phi^k \Psi$$
$$\Gamma(-k) = \Psi(\Phi^T)^k$$

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 given by

$$\Sigma = (1 - \Phi)^{-1}\Psi + \Psi(1 - \Phi^T)^{-1} - \Psi$$

Let ϕ_{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 $\phi_{max} = 0.9999$. We run five parallel Markov chains with their starting points evenly distributed about the center of invariant distribution. Figure **??** shows that for the first 500 samples, the chains have not explored the sample space well unlike Figure **??** with $5 \cdot 10^4$ samples.

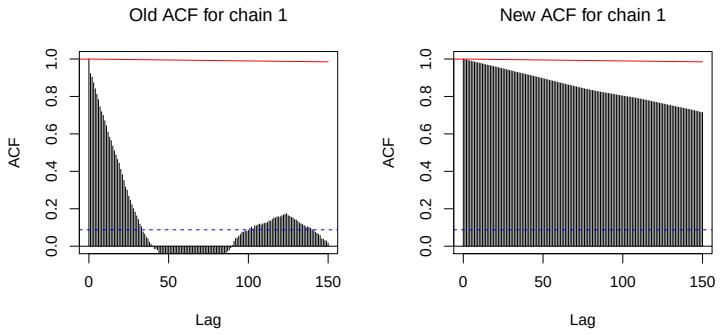
5.2 Boomerang Distribution

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

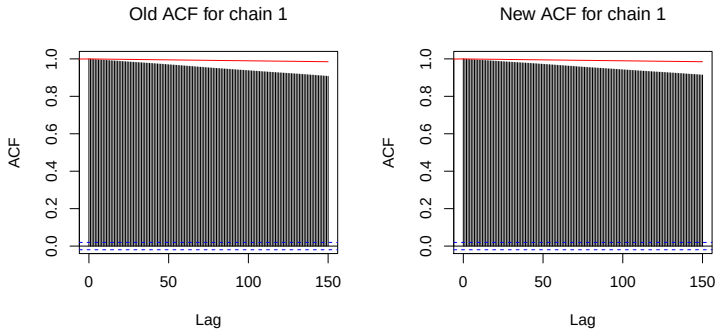
$$f(x, y) \propto \exp\left(-\frac{1}{2}\left[Ax^2y^2 + x^2 + y^2 - 2Bxy - 2C_1x - 2C_2y\right]\right)$$

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

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



(a) $n = 500$



(b) $n = 10^4$

Figure 3: Autocorrelation plot for ACF and G-ACF for first chain out of two parallel Markov chains. First column corresponds to ACF calculated using arithmetic mean of first Markov chain and the second column corresponds to the one calculated using global mean of two Markov chains.(a): $n = 500$, not converged yet; (b): $n = 10^4$, converged

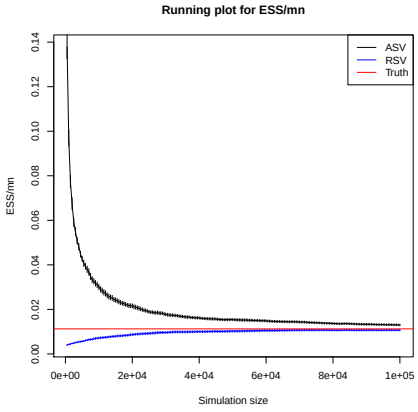
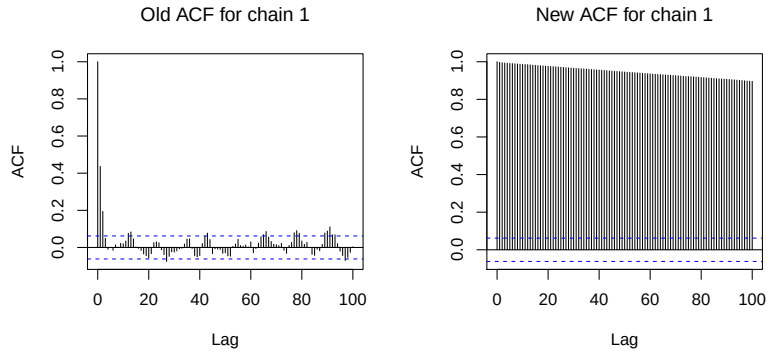


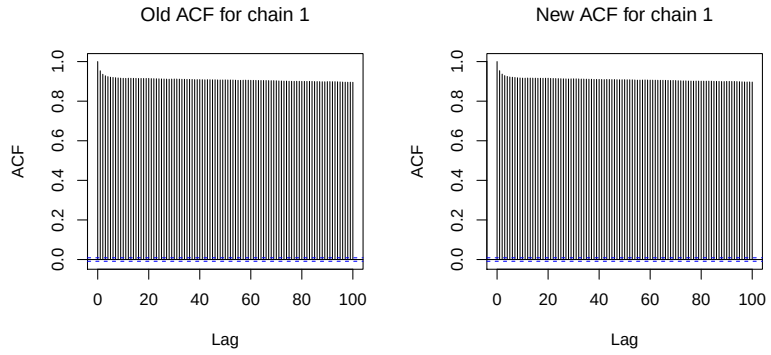
Figure 4: Running plots for $\hat{\text{E}}\text{SS}/mn$ calculated using ASV and G-SVE with the truth.

We use a carefully chosen parameterization of $A = 1,\ B = 3,\ C = 8$ which ensures bimodality for our purpose. Let n be the number of samples in each chain and m be the number of Markov chain replications. Finding the actual mean of this distribution in closed form is difficult. Therefore, we use numerical integration with fine tuning to calculate it. We sample two parallel Markov chains with well-separated starting values.

In Figure **??**, we demonstrate the "sticky" nature of the Markov chains. For the first thousand samples, both the chains are oblivious of the existence of another mode. By ten thousand samples, both the Markov chains have explored the two modes. This property will affect the estimation of ACF from a single chain. In Figure 5a, the ACF is severely underestimated because the first chain has not jumped to the other mode. Whereas, in Figure 5b, both ACF and G-ACF give almost similar results indicating that both modes have been explored by chain-1.



(a) $n = 1000$



(b) $n = 50000$

Figure 5: ACF and G-ACF for component-1 of chain-1 at two different number of Monte Carlo samples.

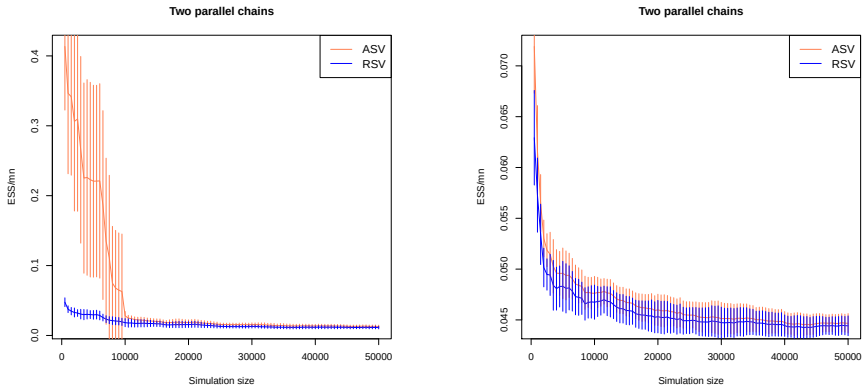
We also run five parallel Markov chains with well-separated starting points. In Table 1, we give the coverage probabilities for 95% confidence interval for both the estimator. We can see that the G-SVE gives better coverage probabilities for all the n. For smaller sample size, the coverage

probability of G-SVE is significantly higher than A-SVE. As the number of samples per chain increases, they start coming closer due to the strong consistency of both the estimators.

n	$m = 2$		$m = 5$	
	ASV	G-SVE	ASV	G-SVE
1000	0.612	0.689	0.602	0.753
2000	0.693	0.751	0.735	0.827
5000	0.826	0.854	0.847	0.880
10000	0.862	0.868	0.884	0.907
20000	0.899	0.906	0.922	0.934

Table 1: Coverage probabilities for parameter values $A = 1,\ B = 3,\ C = 8$.

A good estimate of ESS is crucial to determine when to stop the simulations. We can see in Figure **??** that for the first few thousand samples, A-SVE gives misleadingly higher $E\hat{S}S/mn$ than G-SVE. This can cause us to stop the sampling before the sample space has been explored by the chains. The inferences derived from this chain would then be severely non-informative.



(a) $A = 1,\ B = 3,\ C = 8$ (b) $A = 1,\ B = 10,\ C = 7$

Figure 6: Running plot for $\hat{\text{ESS}}/mn$ using ASV and G-SVE for $m = 2$.

To examine the performance of G-SVE for a nicely mixing Markov chain, we use the parametrization of $A = 1,\ B = 10,\ C = 7$. This is also a bimodal distribution, however, the two modes interact well with each other (Figure **??**). In Figure **??**, we observe that both G-SVE and A-SVE give almost the same estimates for ESS. Similar is the case with coverage probabilities in Table 2 for a 95% confidence region.

n	$m = 2$		$m = 5$	
	ASV	G-SVE	ASV	G-SVE
10000	0.873	0.884	0.875	0.898
2000	0.881	0.888	0.897	0.914
5000	0.902	0.908	0.919	0.929
10000	0.923	0.927	0.925	0.926
50000	0.951	0.952	0.939	0.939

Table 2: Coverage probabilities for parameter values $A = 1,\ B = 10,\ C = 7$.

5.3 Sensor Network Localization

For our third example, we consider a real-life problem 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 joint posterior distribution for missing sensor locations is multi-modal.

Tak et al. (2018) modified the simulation settings of Lan et al. (2014) to include only six sensor locations (four unknown and two unknown) out of eleven locations (eight known and three unknown). 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 zij 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,

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

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

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 $w_{ij} = 1$. The likelihood function is then,

$$L(x_1, x_2, x_3, x_4) \propto$$

$$\prod_{j>i}\left[\left(\exp\left(\frac{-\|x_i-x_j\|^2}{2\times 0.3^2}\right)\right)^{w_{ij}}\left(1-\exp\left(\frac{-\|x_i-x_j\|^2}{2\times 0.3^2}\right)\right)^{1-w_{ij}}\exp\left(-\frac{y_{ij}-\|x_i-x_j\|^2}{2\times 0.02^2}\right)\right]$$

The full posterior distribution with Gaussian prior is given by,

$$\pi(x_1,x_2,x_3,x_4|y,w)\propto L(x_1,x_2,x_3,x_4)\times\exp\left(-\frac{\sum_{i=1}^{4}x_i^Tx_i}{2\times 0.02^2}\right) \tag{3}$$

where $y=(y_{ij},j>i)$ and $w=(w_{ij},j>i)$. 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 AttG-tive Metropolis (RAM) algorithms, Tak et al. (2018) compare the performance of three different sampling techniques namely - Metropolis, RAM and Tempered Transitions (TT). RAM is shown to improve 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 of a multimodal target distribution.

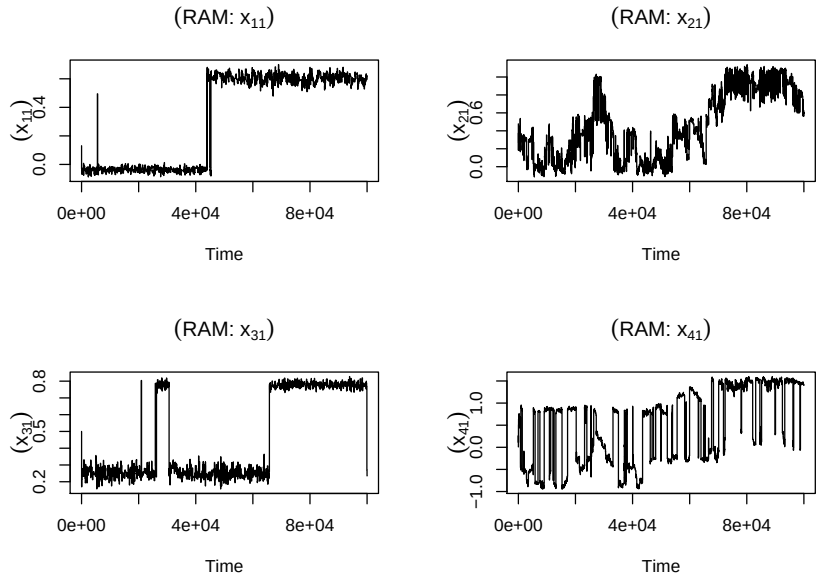


Figure 7: trace plots for the x-component of all four locations of the third chain.

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 $100,000$. Figure 7 demonstrates the trace plots of the first chain for the x-component of all four unknown locations. We observe the bi-modal nature of the marginal distribution for each component. The Markov chains get stuck at a particular mode for a long time before jumping to the other

mode. Figure **??** shows the effect of this "sticky" nature of Markov chains over the autocorrelations centered around the local and global mean. For a small sample size of 1000, it can be seen that the third chain has not explored the sample space well. As a consequence, the local mean differs a lot from the global mean (averaged over all five chains).

6 Discussion

A Appendix

A.1 Preliminaries

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

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

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

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

A.2 Proof of Theorem 1

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

$$\begin{aligned}
\hat{\Gamma}_{G,s}(k) &= \frac{1}{n}\sum_{t=1}^{n-|k|}\left(X_{st} - \bar{\bar{X}}\right)\left(X_{s(t+k)} - \bar{\bar{X}}\right)^T \\
&= \left[\frac{1}{n}\sum_{t=1}^{n-|k|}\left(X_{st} - \bar{X}_s\right)\left(X_{s(t+k)} - \bar{X}_s\right)^T\right] + \left[\frac{1}{n}\sum_{t=1}^{|k|}\left(\bar{X}_s - \bar{\bar{X}}\right)\left(\bar{X}_s - X_{st}\right)^T\right] \\
&\quad + \left[\frac{1}{n}\sum_{t=n-|k|+1}^{n}\left(\bar{X}_s - X_{st}\right)\left(\bar{X}_s - \bar{\bar{X}}\right)^T\right] + \left[\frac{n-|k|}{n}\left(\bar{X}_s - \bar{\bar{X}}\right)\left(\bar{X}_s - \bar{\bar{X}}\right)^T\right] \\
&= \hat{\Gamma}_s(k) - \frac{1}{n}\sum_{t=1}^{|k|}A_{st}^T - \frac{1}{n}\sum_{t=n-|k|+1}^{n}A_{st} + \frac{n-|k|}{n}\left(\bar{X}_s - \bar{\bar{X}}\right)\left(\bar{X}_s - \bar{\bar{X}}\right)^T, \qquad (4)
\end{aligned}$$

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

$$
\begin{aligned}
&\mathbb{E}\left[A_{11}\right] \\
&= \mathbb{E}\left[\left(X_{11} - \bar{X}_1\right)\left(\bar{X}_1 - \bar{\bar{X}}\right)^T\right] \\
&= \mathbb{E}\left[X_{11}\bar{X}_1^T\right] - \frac{1}{m}\mathbb{E}\left[X_{11}\bar{X}_1^T\right] - \frac{m-1}{m}\mathbb{E}\left[X_{11}\bar{X}_2^T\right] + \frac{1}{m}\mathbb{E}\left[\bar{X}_1\bar{X}_1^T\right] + \frac{m-1}{m}\mathbb{E}\left[\bar{X}_1\bar{X}_2^T\right] - \mathbb{E}\left[\bar{X}_1\bar{X}_1^T\right] \\
&= \frac{m-1}{m}\left(\mathbb{E}\left[X_{11}\bar{X}_1^T\right] - \mathbb{E}\left[X_{11}\bar{X}_2^T\right] + \mathbb{E}\left[\bar{X}_1\bar{X}_2^T\right] - \mathbb{E}\left[\bar{X}_1\bar{X}_1^T\right]\right) \\
&= \frac{m-1}{m}\left(\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[X_{11}X_{1t}^T\right] - \mathbb{E}\left[X_{11}\right]\mathbb{E}\left[\bar{X}_2^T\right] + \mathbb{E}\left[\bar{X}_1\right]\mathbb{E}\left[\bar{X}_2^T\right] - \mathrm{Var}\left[\bar{X}_1\right] - \mathbb{E}\left[\bar{X}_1\right]\mathbb{E}\left[\bar{X}_1^T\right]\right) \\
&= \frac{m-1}{mn}\left(\sum_{k=0}^{n-1}\Gamma(k) - n\mathrm{Var}\left[\bar{X}_1\right]\right) . \qquad (5)
\end{aligned}
$$

Similarly,

$$
\mathbb{E}\left[A_{11}^T\right] = \mathbb{E}\left[A_{11}\right]^T = \frac{m-1}{mn}\left(\sum_{k=0}^{n-1}\Gamma(k)^T - n\mathrm{Var}\left[\bar{X}_1\right]\right) . \qquad (6)
$$

Further,

$$
\begin{aligned}
\mathbb{E}\left[\left(\bar{X}_1 - \bar{\bar{X}}\right)\left(\bar{X}_1 - \bar{\bar{X}}\right)^T\right] &= \mathbb{E}\left[\bar{X}_1\bar{X}_1^T - \bar{X}_1\bar{\bar{X}}^T - \bar{\bar{X}}\bar{X}_1^T + \bar{\bar{X}}\bar{\bar{X}}^T\right] \\
&= \left(\mathrm{Var}(\bar{X}_1) + \mu\mu^T - \mathrm{Var}(\bar{\bar{X}}) - \mu\mu^T\right) \\
&= \frac{m-1}{m}\mathrm{Var}(\bar{X}_1) . \qquad (7)
\end{aligned}
$$

Additionally, the locally-centered autocovariance exhibits the following expectation

$$
\mathbb{E}[\hat{\Gamma}(k)] = \left(1 - \frac{|k|}{n}\right)\left(\Gamma(k) - \mathrm{Var}(\bar{X})\right) . \qquad (8)
$$

As a consequence, if $\mathrm{Var}(\bar{X})$ is finite, then $\mathrm{Var}(\bar{X}) \to 0$ as $n \to \infty$. Using (8), (5), (6), and (7) in (4),

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

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

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

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,

$$\mathbb{E}\left[\hat{\Gamma}_{G,s}(k)\right] = \left(1-\frac{|k|}{n}\right)\Gamma(k) + O_1 + O_2\,. \tag{9}$$

where,

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

$$O_2 = -\frac{1}{m}\left(\frac{\Sigma}{n}+\frac{\Phi}{n^2}\right) + o(n^{-2})$$

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,

$$\mathbb{E}\left[\hat{\Gamma}^{ii}_{G,s}\right]$$

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

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

A.3 Strong consistency argument

Consider pseudo spectral variance estimators for the sth chain, denoted by $\tilde{\Sigma}_s$ which uses data centered around the unobserved actual mean μ:

$$\tilde{\Gamma}_s(k) = \frac{1}{n}\sum_{t=1}^{n-|k|}(X_{st}-\mu)(X_{s(t+k)}-\mu)^T$$

$$\tilde{\Sigma}_s = \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right)\tilde{\Gamma}_s(k)\,.$$

The average pseudo spectral variance estimator is

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

Further, let

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

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

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,

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

$$+\frac{1}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right)\sum_{t=1}^{n-|k|}\left[\frac{1}{n}\left(X_{st}-\mu\right)_i\left(\mu-\bar{\bar{X}}\right)_j+\frac{1}{n}\left(\mu-\bar{\bar{X}}\right)_i\left(X_{s(t+k)}-\mu\right)_j\right]$$
$$=\tilde{\Sigma}^{ij}+M_1+M_2\,.$$

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,

$$\begin{aligned}\|\bar{X}_s-\mu\|_\infty\le\|\bar{X}_s-\mu\|&=\frac{1}{mn}\left\|\sum_{t=1}^{n}X_{st}-n\mu\right\|\\&\le\frac{1}{n}\left\|\sum_{t=1}^{n}X_{st}-n\mu-LB(n)\right\|+\frac{\|LB(n)\|}{n}\\&<\frac{D\psi(n)}{n}+\frac{\|LB(n)\|}{n}\\&<\frac{D\psi(n)}{n}+\frac{1}{n}\|L\|\left(\sum_{i=1}^{p}|B^{(i)}(n)|^2\right)^{1/2}\\&\le\frac{D\psi(n)}{n}+\frac{1}{n}\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}\,.\end{aligned}\tag{10}$$

Similarly,

$$\|\bar{\bar{X}}-\mu\|_\infty\le\frac{D\psi(n)}{n}+\frac{1}{n}\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}\,.\tag{11}$$

Now consider,

$$\begin{aligned}&|M_1|\\&\le\frac{1}{m}\sum_{s=1}^{m}\left\{\sum_{k=-b_n+1}^{b_n-1}w\left(\frac{k}{b_n}\right)\left[\frac{1}{n}\left|\sum_{t=1}^{n-|k|}(X_{st}-\mu)_i\right|\left|(\mu-\bar{\bar{X}})_j\right|+\frac{1}{n}\left|(\mu-\bar{\bar{X}})_i\right|\left|\sum_{t=1}^{n-|k|}(X_{j(t+k)}-\mu)_j\right|\right]\right\}\\&\le\frac{\|(\bar{\bar{X}}-\mu)\|_\infty}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1}\left[\frac{1}{n}\left\|\sum_{t=1}^{n-|k|}(X_{st}-\mu)\right\|_\infty+\frac{1}{n}\left\|\sum_{t=1}^{n-|k|}(X_{s(t+k)}-\mu)\right\|_\infty\right]\\&\le\frac{\|(\bar{\bar{X}}-\mu)\|_\infty}{m}\\&\quad\times\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1}\left[\frac{1}{n}\left\|\sum_{t=n-|k|+1}^{n}(X_{st}-\mu)-n(\bar{X}_s-\mu)\right\|_\infty+\frac{1}{n}\left\|\sum_{t=1}^{|k|}(X_{st}-\mu)-n(\bar{X}_s-\mu)\right\|_\infty\right]\end{aligned}$$

$$\leq \frac{\|(\bar{\bar{X}}-\mu)\|_\infty}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1}\left[\frac{1}{n}\left\|\sum_{t=n-|k|+1}^{n}(X_{st}-\mu)\right\|_\infty+\frac{1}{n}\left\|\sum_{t=1}^{|k|}(X_{st}-\mu)\right\|_\infty+2\|\bar{X}_s-\mu\|_\infty\right]$$

$$\leq \frac{\|(\bar{\bar{X}}-\mu)\|_\infty}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1}\frac{1}{n}\left[\left\|\sum_{t=n-|k|+1}^{n}(X_{st}-\mu)\right\|_\infty+\left\|\sum_{t=1}^{|k|}(X_{st}-\mu)\right\|_\infty\right]$$
$$+2(2b_n-1)\|\bar{\bar{X}}-\mu\|_\infty\|\bar{X}_1-\mu\|_\infty\,.$$

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

$$|M_1|<2\|(\bar{\bar{X}}-\mu)\|_\infty\left[\sum_{k=-b_n+1}^{b_n-1}\left[\frac{D\psi(k)}{n}+\frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2k\log\log k}}{n}\right]\right]+2(2b_n-1)\|\bar{X}_1-\mu\|_\infty$$
$$\leq 2(2b_n-1)\|(\bar{\bar{X}}-\mu)\|_\infty\left[\frac{D\psi(n)}{n}+\frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{n\log\log n}}{n}+\|\bar{X}_1-\mu\|_\infty\right]$$
$$\leq 4(2b_n-1)\left[\frac{D\psi(n)}{n}+\frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{n\log\log n}}{n}\right]^2\quad\text{(by (10) and (11))}\,. \tag{12}$$

For M_2,

$$|M_2|=\left|\frac{1}{m}\sum_{s=1}^{m}\left\{\left(\mu-\bar{\bar{X}}\right)_i\left(\mu-\bar{\bar{X}}\right)_j\sum_{k=-b_n+1}^{b_n-1}\left(1-\frac{|k|}{n}\right)w\left(\frac{k}{b_n}\right)\right\}\right|$$
$$\leq \|\bar{\bar{X}}-\mu\|_\infty^2\left[\sum_{k=-b_n+1}^{b_n-1}\left(1-\frac{|k|}{n}\right)w\left(\frac{k}{b_n}\right)\right]<\|\bar{\bar{X}}-\mu\|_\infty^2\left[\sum_{k=-b_n+1}^{b_n-1}\left|w\left(\frac{k}{b_n}\right)\right|\right]$$
$$\leq b_n\|\bar{\bar{X}}-\mu\|_\infty^2\int_{-\infty}^{\infty}|w(x)|dx$$
$$\leq Cb_n\left[\frac{D\psi(n)}{n}+\frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{n\log\log n}}{n}\right]^2\quad\text{(by (11))}\,. \tag{13}$$

Using (12) and (13),

$$|M_1+M_2\leq|M_1|+|M_2|=D^2g_1(n)+Dg_2(n)+g_3(n)\,,$$

where

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

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

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 $n \to \infty$.

□

Proof of theorem 2. We have the following decomposition,

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

where

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

Using the above and Lemma 2,

$$\left|\hat{\Sigma}^{ij}_G - \Sigma^{ij}\right| = \left|\hat{\Sigma}^{ij}_{SV} - \Sigma^{ij} + N_1 + N_2 + M_1 + M_2\right| \le \left|\hat{\Sigma}^{ij}_{SV} - \Sigma^{ij}\right| + |N_1| + |N_2| + |M_1 + M_2| \quad (14)$$

By Assumption **??**, 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

$$|N_1| = \left|\frac{1}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right)\frac{1}{n}\sum_{t=1}^{n-|k|}\left[\left(X_{st}-\bar{X}_s\right)_i\left(\bar{X}_s-\mu\right)_j + \left(\bar{X}_s-\mu\right)_i\left(X_{s(t+k)}-\bar{X}_s\right)_j\right]\right|$$

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

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

$$\left| \frac{1}{m} \sum_{s=1}^{m} \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} \left[\left(X_{st} - \bar{X}_s\right)_i \left(\bar{X}_s - \mu\right)_j \right] \right|$$
$$\leq \frac{1}{m} \sum_{s=1}^{m} \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \frac{\left| \left(\bar{X}_s - \mu\right)_j \right|}{n} \left[\left| \sum_{t=1}^{|k|} \left(\mu - X_{st}\right)_i \right| + |k| \left| \left(\bar{X}_s - \mu\right)_i \right| \right]$$
$$\leq \frac{1}{m} \sum_{s=1}^{m} \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \frac{\|\bar{X}_s - \mu\|_\infty}{n} \left\| \sum_{t=1}^{|k|} \left(\mu - X_{st}\right) + |k| \left(\bar{X}_s - \mu\right) \right\|_\infty$$
$$\leq \frac{1}{m} \sum_{s=1}^{m} \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \frac{\|\bar{X}_s - \mu\|_\infty}{n} \left(\left\| \sum_{t=1}^{|k|} \left(X_{st} - \mu\right) \right\|_\infty + |k| \left\| \bar{X}_s - \mu \right\|_\infty \right)$$
$$\leq \frac{1}{m} \sum_{s=1}^{m} \sum_{k=-b_n+1}^{b_n-1} \frac{\|\bar{X}_s - \mu\|_\infty}{n} \left\| \sum_{t=1}^{|k|} \left(X_{st} - \mu\right) \right\| + \frac{1}{m} \sum_{s=1}^{m} \frac{b_n(b_n-1)}{n} \left\| \bar{X}_s - \mu \right\|_\infty^2 .$$

Using SIP on the summation of k terms,

$$\left| \frac{1}{m} \sum_{s=1}^{m} \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} \left[\left(X_{st} - \bar{X}_s\right)_i \left(\bar{X}_s - \mu\right)_j \right] \right|$$
$$< \frac{1}{m} \sum_{s=1}^{m} \|\bar{X}_s - \mu\|_\infty \sum_{k=-b_n+1}^{b_n-1} \left[\frac{D\psi(k)}{n} + \frac{\|L\| p^{1/2} (1+\epsilon) \sqrt{2k \log\log k}}{n} \right] + \frac{1}{m} \sum_{s=1}^{m} \frac{b_n(b_n-1)}{n} \|\bar{X}_s - \mu\|_\infty^2$$
$$< \frac{(2b_n - 1)}{m} \sum_{s=1}^{m} \|\bar{X}_s - \mu\|_\infty \left[\frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2} (1+\epsilon) \sqrt{2n \log\log n}}{n} \right] + \frac{1}{m} \sum_{s=1}^{m} \frac{b_n(b_n-1)}{n} \|\bar{X}_s - \mu\|_\infty^2$$
$$\leq \left(2b_n - 1 + \frac{b_n^2}{n} - \frac{b_n}{n} \right) \left[\frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2} (1+\epsilon) \sqrt{2n \log\log n}}{n} \right]^2 \to 0 . \quad \text{(by (10))}$$

Similarly, the second part of $N_1 \to 0$ with probability 1. Following the steps in (13),

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

Thus, in (14), every term goes to 0 and $\hat{\Sigma}_G^{ij} \to \Sigma^{ij}$ with probability 1 as $n \to \infty$. □

A.4 G-SVE Bias

Proof of theorem 3. The convoluted end effect terms in theorem 1 are all essentially $\mathcal{O}(1/n)$. We are interested in finding the asymptotic bias here. Therefore we can write the expectation of replicated autocovariance in form of equation 9 as,

$$\begin{aligned}
&\mathbb{E}\left[\hat{\Gamma}_{G-}(k)\right] \\
&= \left(1 - \frac{|k|}{n}\right)\Gamma(k) - \frac{|k|}{n}\left[\left(1 - \frac{1}{m}\right)\left(\frac{1}{n}\sum_{h=0}^{n-1}\Gamma(h)^T + \frac{1}{n}\sum_{h=0}^{n-1}\Gamma(h)\right) - \left(2 - \frac{1}{m}\right)\mathrm{Var}(\bar{X}_1)\right] - \frac{\mathrm{Var}(\overline{X}_1)}{m} \\
&= \left(1 - \frac{|k|}{n}\right)\Gamma(k) + O_1 + O_2\,. \qquad (15)
\end{aligned}$$

where both O_1 and O_2 are $\mathcal{O}(n^{-1})$ terms. By our assumptions, $\sum_{k=-\infty}^{\infty}|k|^q \left\|\Gamma(h)\right\| < \infty$. Consider the G-SVE estimator,

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

where

$$P_1 = \sum_{k=-n+1}^{n-1}\left[w\left(\frac{k}{b_n}\right)\left(1 - \frac{|k|}{n}\right)\Gamma(k)\right] - \sum_{k=-\infty}^{\infty}\Gamma(k)\,.P_2 \quad = \sum_{k=-n+1}^{n-1}\left[w\left(\frac{k}{b_n}\right)(O_1 + O_2)\right].$$

Dootika: Add something about P_2 here as well. And I don't think we need the $W_n < \infty$ assumption.

We first solve for term P_1 by breaking it into three parts as in Hannan (2009). Note that notation $A = o(z)$ for matrix A implies $A^{ij} = o(z)$ for each element of the matrix A. Consider,

$$P_1 = -\sum_{|k|\geq n}\Gamma(k) - \sum_{k=-n+1}^{n-1} w\left(\frac{|k|}{n}\right)\frac{|k|}{n}\Gamma(k) - \sum_{k=-n+1}^{n-1}\left(1 - w\left(\frac{|k|}{n}\right)\right)\Gamma(k)\,. \qquad (16)$$

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

$$-\sum_{|k|\geq n}\Gamma(k)\leq\sum_{|k|\geq n}\left|\frac{k}{n}\right|^q\Gamma(k)=\frac{1}{b_n^q}\left|\frac{b_n}{n}\right|^q\sum_{|k|\geq n}|k|^q\Gamma(k)=o\left(\frac{1}{b_n^q}\right), \tag{17}$$

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

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

For $q\geq 1$,

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

For $q<1$,

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

So,

$$\sum_{k=-n+1}^{n-1}w\left(\frac{|k|}{n}\right)\frac{|k|}{n}\Gamma(k)=o\left(\frac{1}{b_n^q}\right) \tag{18}$$

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,

$$\begin{aligned}\sum_{k=-n+1}^{n-1}\left(1-w\left(\frac{k}{b_n}\right)\right)\Gamma(k)&=-\frac{1}{b_n^q}\sum_{k=-n+1}^{n-1}\left(\frac{|k|}{b_n}\right)^{-q}\left(1-w\left(\frac{|k|}{b_n}\right)\right)|k|^q\Gamma(k)\\&=-\frac{1}{b_n^q}\sum_{k=-n+1}^{n-1}\left[k_q+o(1)\right]|k|^q\Gamma(k)\\&=-\frac{k_q\Phi^{(q)}}{b_n^q}+o\left(\frac{1}{b_n^q}\right).\end{aligned} \tag{19}$$

Using (17), (18), and (19) in (16), we get

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

which completes the result. □

A.5 Proof of Theorem 4

Due to the strong consistency proof from theorem 2, as $n \to \infty$,

$$\left|\hat{\Sigma}_G - \tilde{\Sigma}\right| \to 0 \text{ with probability } 1\,. \tag{20}$$

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,

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

By (20), there exists an N_0 such that

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

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 $n \to \infty$,

$$\mathbb{E}\left[\left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right)^2\right] \to 0\,. \tag{21}$$

We will use (21) 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 (21),

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

Finally,

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

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

$$\frac{n}{b_n}\text{Var}(\tilde{\Sigma}^{ij}) = \left[\Sigma^{ii}\Sigma^{jj} + \left(\Sigma^{ij}\right)\right]\int_{-\infty}^{\infty} w^2(x)dx + o(1) \tag{22}$$

Plugging (22) into variance of $\hat{\Sigma}_G$ gives the result of the theorem.

B Additional Examples

We present two additional examples illustrating the difference between the ACF and G-ACF plots.

B.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 the following Bayesian Poisson change-point model:

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

Following Martin et al. (2011), we will use `MCMCpoissonChange` from `MCMCpack` to fit the model with $k = 7$ which samples the latent states based on the algorithm in Chib (1998).
We run two parallel Markov chains from randomly chosen starting points and present the resulting ACF and G-ACF in Figure 8. The early G-ACF estimates are far closer to the G-ACF and ACF estimates at 10^5.

B.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 each node 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). The stationary distribution for this algorithm is a uniform distribution over the nodes.

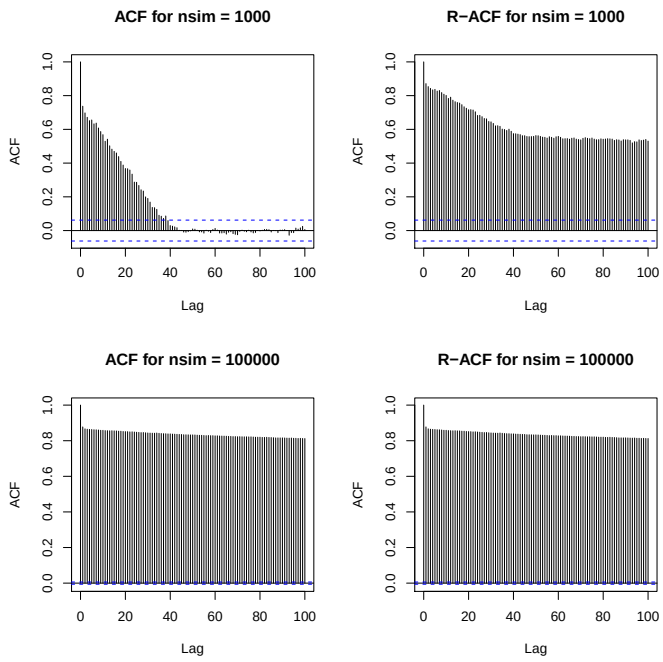


Figure 8: ACF and G-ACF for the first component of the first chain.

We start two parallel Markov chains from two students belonging to different G-es ??? and study its impacts on the average features of their immediate social group.

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.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

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.

Flegal, J. M. and Gong, L. (2015). Relative fixed-width stopping rules for markov chain monte carlo simulations. *Statistica Sinica*, pages 655–675.

Gelman, A. and Meng, X.-L. (1991). A note on bivariate distributions that are conditionally normal. *The American Statistician*, 45(2):125–126.

Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

Geyer, C. (2011). Introduction to markov chain monte carlo. *Handbook of markov chain monte carlo*, 20116022:45.

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., Whitt, W., et al. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability*, 2(1):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(3):684–700.

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.

Jones, G. L. et al. (2004). On the markov chain central limit theorem. *Probability surveys*, 1(299-320):5–1.

Kuelbs, J. and Philipp, W. (1980). Almost sure invariance principles for partial sums of mixing b-valued random variables. *The Annals of Probability*, pages 1003–1036.

Lan, S., Streets, J., and Shahbaba, B. (2014). Wormhole hamiltonian monte carlo. In *AAAI*, pages 1953–1959.

Ma, Y. and Genton, M. G. (2000). Highly robust estimation of the autocovariance function. *Journal of time series analysis*, 21(6):663–684.

Martin, A. D., Quinn, K. M., and Park, J. H. (2011). Mcmcpack: Markov chain monte carlo in r.

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.

Quenouille, M. H. (1947). Notes on the calculation of autocorrelations of linear autoregressive schemes. *Biometrika*, 34(3/4):365–367.

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.

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(2):321–337.

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.