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 6, 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. 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 is the sample autocovariance matrix at lag k for the sth chain is

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

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 remarkably improvements. A demonstrative example is at the end of this Section.

ACvFs are required to estimate the log-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 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

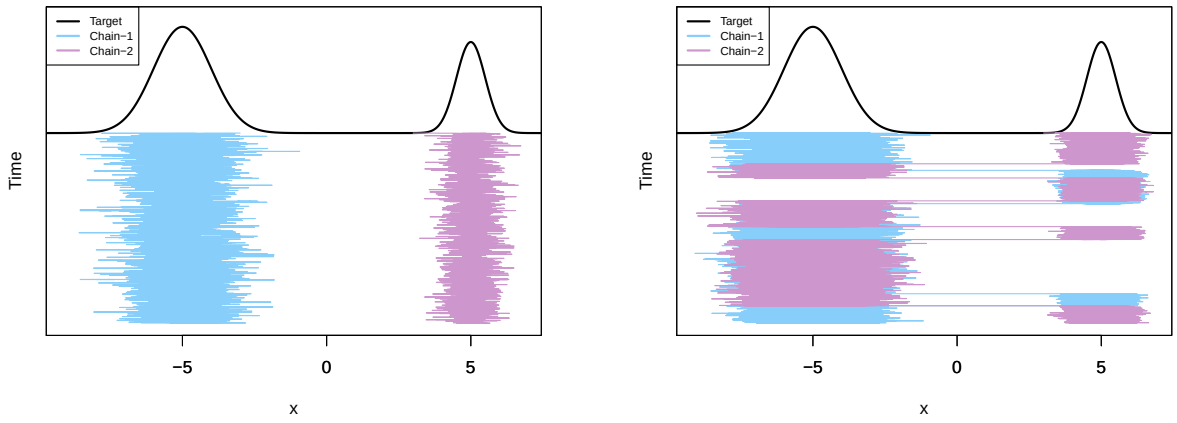
(Liu and Flegal, 2018). We relieve the computational burden, we adapt the fast algorithm of Heberle and Sattarhoff (2017) for G-ACvFs that dramatically reduces computation time. The globally-centered spectral variance estimators in employed in the computation of effective sample size and safeguards against early termination of the MCMC process and improved reliability of estimators.

1.1 Demonstrative example

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

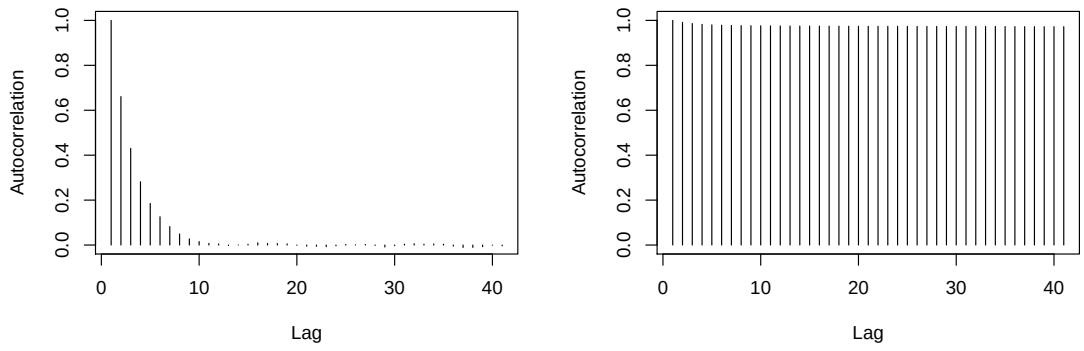
$$f(x) = 0.7\, f(x; -5, 1) + 0.3\, f(x; 5, 0.5)\,,$$

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. 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. This is evident from the ACF plots in Figure 2a. Second, for a large sample size, the estimates from G-ACF as well as ACF are equivalent in Figure 1b.

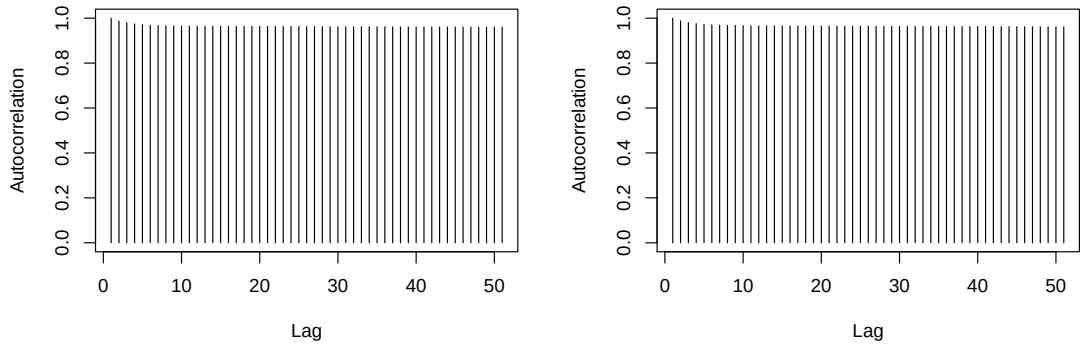


(a) $n = 10^4$ (b) $n = 10^5$

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



(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

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

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

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 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 A.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,

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

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 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 2.
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

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

Set $\{Y_{st}\}_{t\geq 1} = \{g(X_{st})\}_{t\geq 1}$ for all $s = 1, \ldots, m$ and set $\bar{Y}_s = n^{-1}\sum_{t=1}^{n} Y_{st}$ 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.
Reliability of $\bar{\bar{Y}}$ is assessed by its Monte Carlo variability via an asymptotic sampling distribution (Flegal et al., 2008; Roy, 2019; Vats et al., 2020). A Markov chain central limit theorem (CLT) holds if there exists a $p \times p$ positive-definite matrix Σ such that

$$\sqrt{mn}(\bar{\bar{Y}} - \mu) \xrightarrow{d} N(0, \Sigma) \tag{2}$$

where

$$\Sigma = \sum_{k=-\infty}^{\infty} \text{Cov}_F \left(Y_{11}, Y_{1(1+k)} \right) \tag{3}$$

The goal in output analysis for MCMC is to estimate Σin order to assess variability in $\bar{\bar{Y}}$. There is a rich literature of estimators of Σ available for single-chain MCMC implementations. The most common are spectral variance estimators (SVE) (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 multiple parallel chains. Batch means estimators are computationally more efficient than SVE, SVE 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.

SV estimators are constructed as weighted and truncated sums of estimated lag covariances. Let $w : \mathbb{R} \to [-z, z]$ be a lag window function and let $b_n \in \mathbb{N}$ be a truncation point. The SV estimator of Σ from a single chain is DV: limits should be $-n+1$ and $n-1$.

$$\hat{\Sigma}_{SV,s} = \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \hat{\Gamma}_s(k) \tag{4}$$

An overall estimate of Σ from all m chains is the average SV estimator (ASV):

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

3.1 Globally-centered spectral variance estimators

DV: There are some issues.

- $|w_n| < z$ or $|w_n| < c$?
- Need new notation for $\Gamma(k)$ since $\Gamma(k)$ is for the X and we need a new notation for Y. This change needs to be made for this whole section and the appendix!

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

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

For $\hat{\Sigma}_G$ to accurately estimate Σ, conditions need to be imposed on the weight function, w_n.
DV: Did we discuss that W_n should be define as an infinite sum and should be W?

Assumption 1. The lag window $w(x)$ is a continuous, even function with $w(0) = 1, |w(x)| < c$, $\int_{-\infty}^{\infty} w^2(x)dx < \infty$, and $W_n < \infty$ where

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

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

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 demonstrate for MCMC as it is required to ensure sequential stopping rules yield correct coverage at termination (Glynn and Whitt, 1992). A critical assumption is that of a strong invariance principle below. 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 $w \in \Omega$ 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 $\hat{\Sigma}_{SV,s} \xrightarrow{a.s.} \Sigma$ for all s, and $n^{-1}b_n \log\log n \to 0$ as $n \to \infty$, then $\hat{\Sigma}_G \to \Sigma$ with probability 1, as $n \to \infty$.*

Move these two sub-assumptions into the theorem 4 below.

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. $\sum_{-\infty}^{\infty} |k|^q \|\Gamma_s(k)\| < \infty$ DV: This assumption is satisfied by the polynomial ergodicity, so can remove it.

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

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

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

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

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

I have not looked at it after here.

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 4, 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
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

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

Algorithm 1: Herberle's Algorithm

Construct $C(w^*)$ and A^* from the MCMC samples.
Compute the DFT of the first column of $C(w^*)$. This gives the eigenvalues of $C(w^*)$.
for $j = 1, 2, ..., p$ **do**
    Calculate $V^*A_j^*$ by FFT of A_j
    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 $\Lambda V^*A_j^*$.
    Determine $C(w^*)A^* = V\Lambda V^*A_j^*$ by inverse FFT of $\Lambda V^*A_j^*$.
end
Select the first n rows of $C(w^*)A^*$ to form $T(w)A$.
Premultiply by A^T and divide by n.

.

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. (2019b) 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

In this section we consider three different target distributions and sample multiple Markov chains using MCMC to analyse the performance of our globally-centered estimators experimentally. We will make the following comparisons - (1) A-ACF and G-ACF; (2) A-SVE and G-SVE; (3) estimates of ESS/mn using A-SVE and G-SVE. In all the examples we consider the function $g(x) = x$, which

implies the Markov chain $\{Y_t\}$ and $\{X_t\}$ are the same. There are only a handful of MCMC chains where the true autocovariance and asymptotic variance are known. 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)). A distinct 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). Through this example, we will also show that in case of nicely mixing Markov chains, these estimators give almost equivalent results. Lastly, we consider a real-life example of finding unknown locations of sensors using noisy distance data. 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 single-chain estimators for autocovariance, asymptotic variance, 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. 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

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

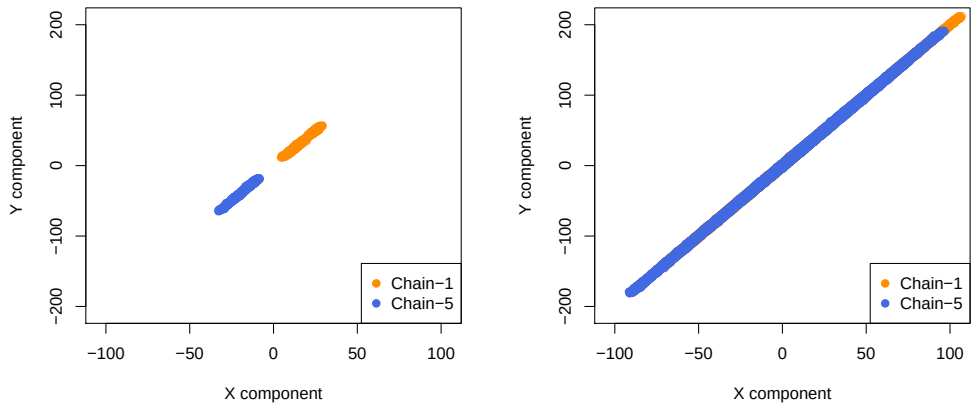
where $X_t \in \mathbb{R}^p$, Φ is a $p \times p$ matrix, $\epsilon_t \overset{i.i.d.}{\sim} N(0, \Omega)$, and Ω is a positive definite $p \times p$ matrix. The invariant distribution for this Markov chain is $N(0, \Psi)$ where $\vec{\Psi} = (I_{p^2} - \Phi \otimes \Phi)\vec{\Omega}$. The lag-k autocovariance 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. There is very high correlation between the two components in the target distribution 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. Let σ_{ij} denote



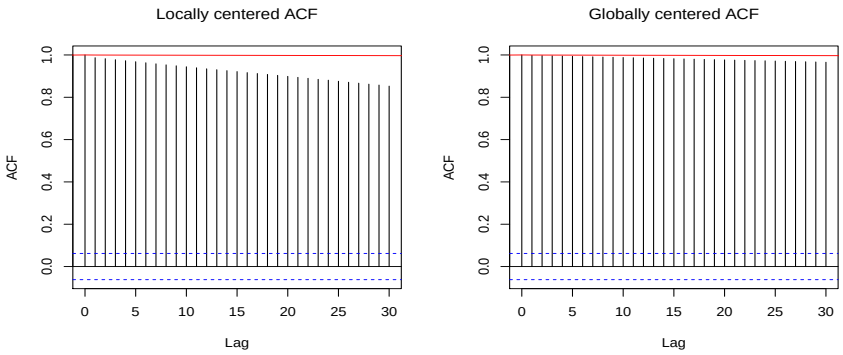
(a) $n = 1000$ (b) $n = 10000$

Figure 3: Scatter plot of two out of five slowly mixing Markov chains starting from opposite sides of known mean. Sample size of 1000 (left) and 10000 (right).

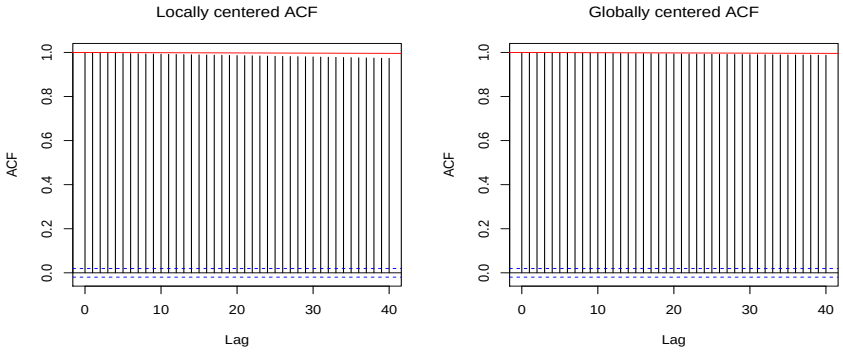
the ij^{th} element of Σ. We sample five parallel chains and plot the first and fifth chain in Figure **??** for two different simulation sizes. The first chain starts at $(\sigma_{11}/2, \sigma_{22}/2)$ and the fifth chains starts from $(-\sigma_{11}/2, -\sigma_{22}/2)$. For a sample size of 1000, the chains have not moved enough from their starting points. Whereas, for a $n = 10,000$, the chains seem to have explored the sample space fairly well.

The locally and globally-centered autocorrelations averaged over all the chains have been shown in Figure **??** for two different simulation sizes where the true ACrF is shown in red. When the individual chains have not explored well, the local sample mean is closer to the samples and possibly far away from the true mean. As a consequence, the empirical ACrF estimator severely underestimate the known true autocorrelations. When the starting values are nicely dispersed, the global mean provides a more reliable estimate of actual mean and hence, G-ACrF is closer to reality. In subfigure 4b, the convergence property of sample means has kicked in for each chain. Therefore, both the estimators provide equivalent results.

The quality of estimation of Σ and ESS/mn has been studied through two running plots - (a) convergence of $\log(\|\hat{\Sigma}_A\|_F)$ and $\log(\|\hat{\Sigma}_G\|_F)$ as n increases and (b) convergence of $\log(E\hat{S}S_A/mn)$ and $\log(E\hat{S}S_G/mn)$ as n increases. In practice, it is preferred to slightly overestimate both the values than to underestimate. In Figure 5, 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 the 95% confidence interval. We run the simulations for a maximum



(a) $n = 10^3$

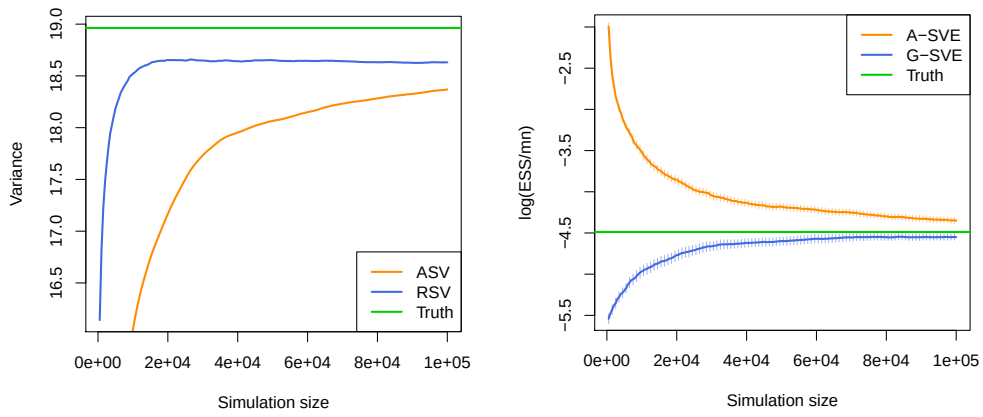


(b) $n = 10^4$

Figure 4: 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): $n = 10^3$, not converged yet; (b): $n = 10^4$, converged

sample size of 10^5 until when, it can be seen that $\|\hat{\Sigma}_G\|_F$ and $\|\hat{\Sigma}_A\|_F$ have not converged. However, the globally-centered estimator is always closer to reality. Before actual convergence, it can be observed that $E\hat{S}S_G/mn$ overestimates the effective sample size but it is preferred to sample more as compared to premature stopping of the sampling process.

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 SV estimator for variance gives better coverage than A-SVE. At $n = 10^5$, the convergence nature starts appearing and both the estimators give good coverage probability.



(a) log-Frobenius norm

(b) $\log(\hat{ESS}/mn)$

Figure 5: (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 $\hat{ESS}/mn$ calculated using A-SVE and G-SVE from $n = 500$ to $n = 10^5$.

5.2 Boomerang Distribution

Our estimators perform exceptionally better than the classical 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

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

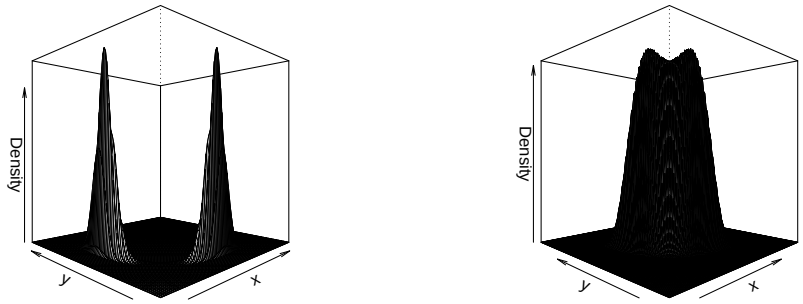
The conditional distribution of x given y and vice versa is a normal distribution given by:

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

We report results for two different bi-modal settings (parameterizations) in this example. Firstly, we use a carefully chosen parameterization of $A = 1,\ B = 3,\ C = 8$ wherein the two modes are well-separated with a very low connection (Setting-1, Figure 6a). 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 model are closer and well-connected (Setting-2, Figure 6b). We sample five parallel

n	A-SVE	G-SVE
1000	0.526	0.975
5000	0.622	0.963
10,000	0.694	0.953
50,000	0.864	0.944
100,000	0.901	0.943

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

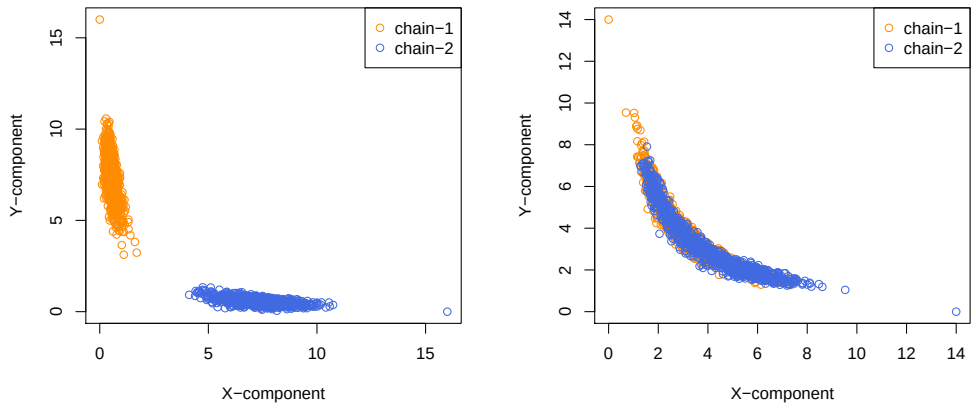


(a) $A = 1, B = 3, C = 8$ (b) $A = 1, B = 10, C = 7$

Figure 6: Perspective plots of the target distribution for two different parameterizations.

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 in Figure 7. Figure 7 shows the scatter plot for both settings for $n = 1000$. 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 even for $n = 1000$.

In Figure 8a, the autocorrelations are severely underestimated by A-ACF because the chains have not jumped modes. Whereas, in Figure 8b, both ACF and G-ACF give almost indistinguishable results indicating that A-ACF and G-ACF have converged. We can observe in Figure 9 that A-ACF and G-ACF give similar results for simulation setting-2. The Markov chains explore both the modes in 1000 iterations only. Therefore, the global and local means do not differ significantly.



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

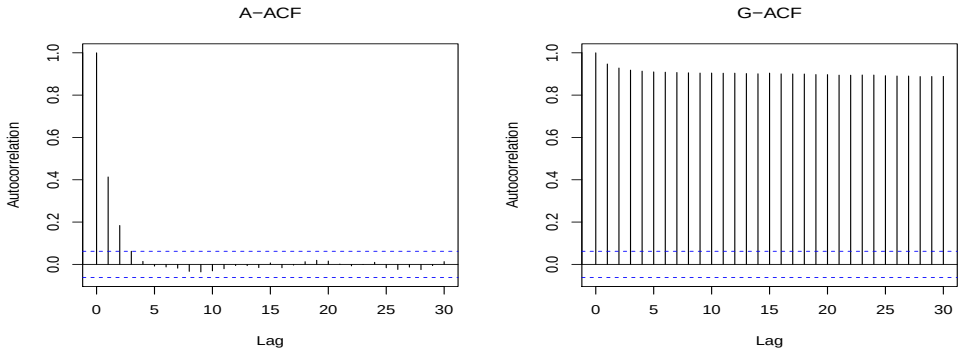
Figure 7: Scatter plots for Setting-1 (left) and Setting-2 (right) at a sample size $n = 1000$.

A good estimate of ESS is crucial to determine when to stop the simulations. We can see in Figure 10 that for the first few thousand samples, 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.

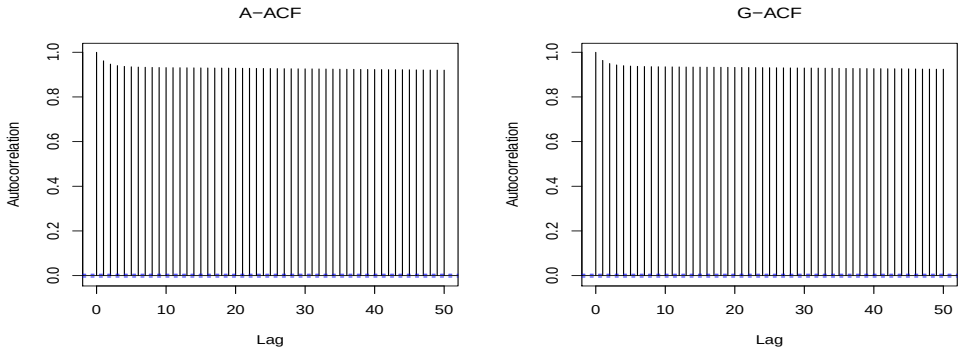
Setting-1					Setting-2				
n	$m = 2$		$m = 5$		n	$m = 2$		$m = 5$	
	A-SVE	G-SVE	A-SVE	G-SVE		A-SVE	G-SVE	A-SVE	G-SVE
1000	0.800	0.877	0.145	0.268	1000	0.873	0.884	0.875	0.898
5000	0.633	0.727	0.436	0.650	2000	0.881	0.888	0.897	0.914
10000	0.585	0.690	0.552	0.750	5000	0.902	0.908	0.919	0.929
50000	0.778	0.815	0.805	0.872	10000	0.923	0.927	0.925	0.926
100000	0.851	0.867	0.863	0.897	50000	0.951	0.952	0.939	0.939

Table 2: Boomerang. Coverage probabilities for $m = 2$ and $m = 5$ for both the simulation settings.

Since a numerical approximation of the true mean is available, we construct 95% confidence interval using A-SVE and G-SVE. We report the coverage probabilities for a 1000 replications at different sample sizes. For both the simulation settings, we provide coverage probabilities for $m = 2$ and



(a) $n = 1000$



(b) $n = 100000$

Figure 8: A-ACF and G-ACF of component-1 at two different sample size for simulation setting-1.

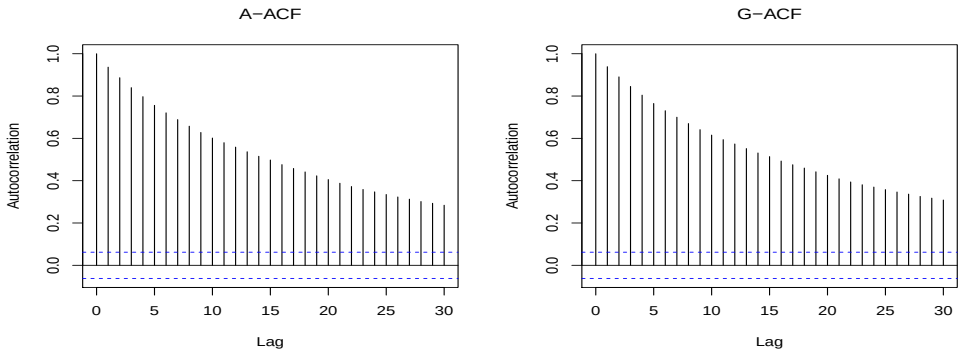
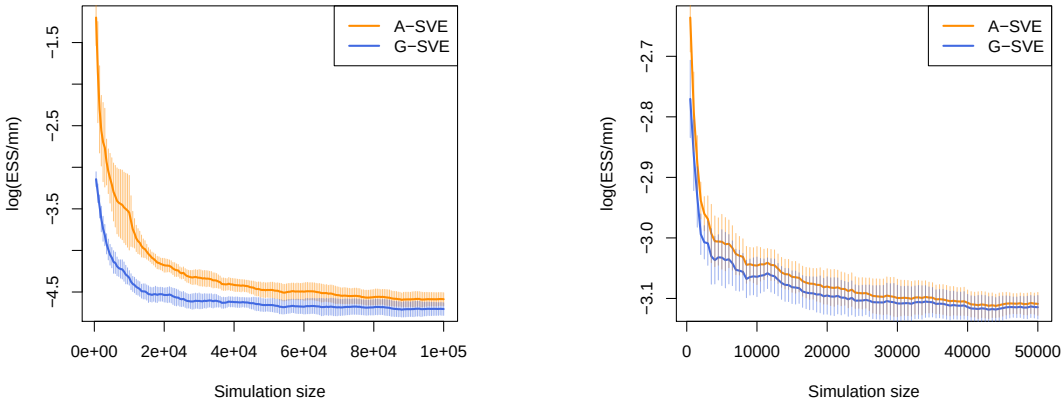


Figure 9: A-ACF and G-ACF of component-1 for simulation setting-2



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

Figure 10: Running plot for $\hat{\text{ESS}}/mn$ using ASV and G-SVE for $m = 2$.

$m = 5$. Table 2 reports all the results. In setting-1, it can be observed that G-SVE gives consistently higher coverage probability than A-SVE. Whereas for setting-2, the results are almost similar.

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 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,

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

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

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,

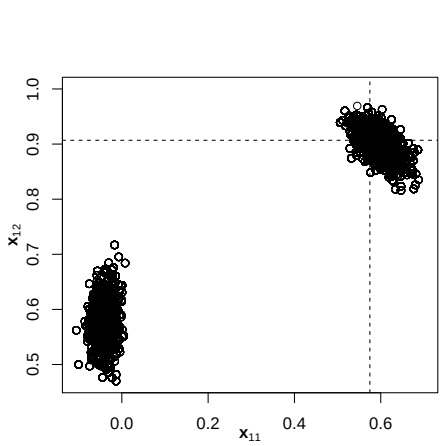
$$\pi(x_1, x_2, x_3, x_4 | y, w) \propto L(x_1, x_2, x_3, x_4) \times \exp\left(-\frac{\sum_{i=1}^{4} x_i^T x_i}{2 \times 0.02^2} \right) \tag{5}$$

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 Attractive 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.

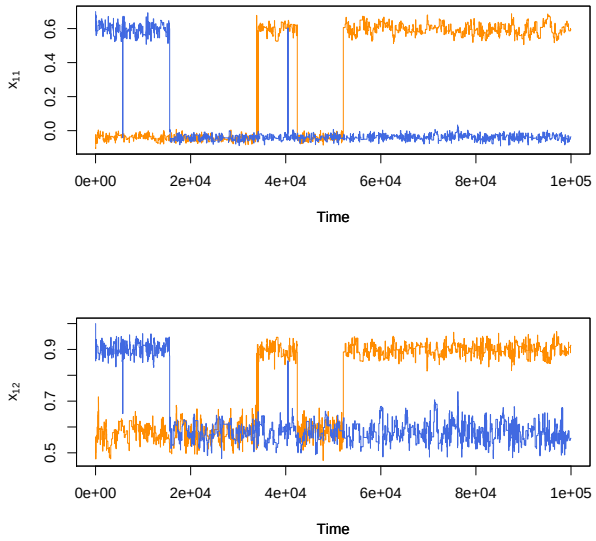
We will use the RAM algorithm with a jumping scale equal to 0.5 to sample five parallel Markov chains with well-separated starting points. The total simulation size for each chain is fixed at $100,000$. We sample five chains starting from dispersed starting values. Since the truth about actual mean and asymptotic variance is not known in this case, we are not interested in coverage probabilities. Instead, we will focus on a comparative study of local and global estimators in terms of - (1) autocorrelations, (2) log Frobenius norm of estimated Σ, and (3) log of estimated ESS/mn.

Figure 11a shows the scatter plot of location-1 for a single Markov chain. Similarly, the other three locations also have a bimodal marginal distribution in each component. To illustrate the *sticky* nature of the Markov chains obtained from RAM sampling method, we plot the evolution of two chains (starting near different modes) with time. Observe that each chain explores one particular mode for a long time before jumping to another mode.

Figure 12 plots the locally and globally centered ACrF estimators for individual chains as well as the average over all the chains. The thick lines represent the averaged estimator and the thin lines correspond to single chains. Figure 13 presents the running plot of log Frobenius norm of A-ACF



(a) Location-1; $n = 10^5$



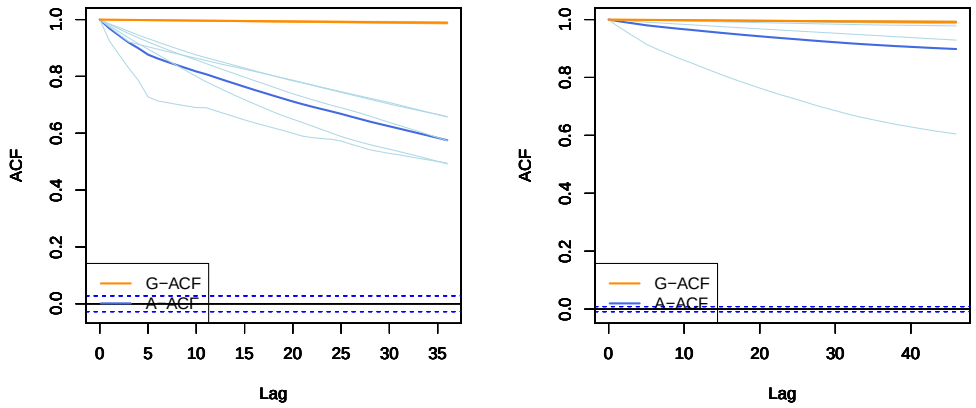
(b) x_{11} and x_{12}; $n = 10^5$

Figure 11: Sensor Network. (a) Scatter plot (x_{12} vs x_{11}) for a Markov chain of length $n = 10^5$; (b) Trace plots of x_{11} and x_{12} for two chains started near the two modes.

and G-ACF along with 95% confidence interval calculated using 50 replications for each value of n (x-axis). In both the plots, G-SVE and $E\hat{S}S_G/mn$ prove to be drastically better than the classical single-chain estimators.

6 Discussion

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.



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

Figure 12: A-ACF and G-ACF for component-1 at two different sample size. The locally (light blue) and globally (light orange) centered ACF for individual chain has been plotted.

A Appendix

A.1 Preliminaries

Lemma 1. *(Csörgo and Révész (2014)). Suppose Assumption* **??** *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* $\forall i = 1, ..., p$

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

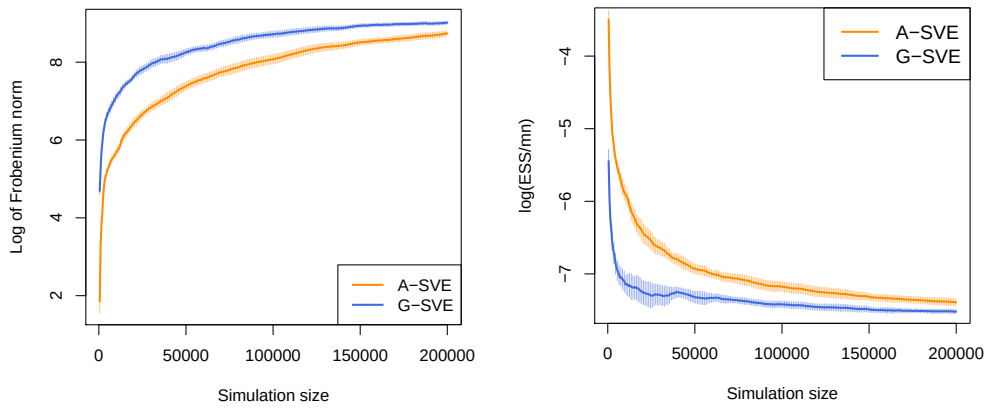
$$\sup_{0 \leq s \leq b_n} \left| B^{(i)}(n) - B^{(i)}(n-s) \right| < (1+\epsilon) \left(2b_n \left(\log \frac{n}{b_n} + \log \log n \right) \right)^{1/2}, \text{ and}$$

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

A.2 Proof of Theorem 1

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

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



(a) log Frobenius norm (b) $\log(E\hat{S}S/mn)$

Figure 13: Sensor Network. Running plot for log of Frobenius norm of estimated Σ (left) and log of estimated ESS/mn (right) calculated using A-SVE and G-SVE.

$$
\begin{aligned}
&= \left[\frac{1}{n}\sum_{t=1}^{n-|k|}\left(X_{st}-\bar{X}_s\right)\left(X_{s(t+k)}-\bar{X}_s\right)^T\right] + \left[\frac{1}{n}\sum_{t=1}^{|k|}\left(\bar{X}_s-\bar{\bar{X}}\right)\left(\bar{X}_s-X_{st}\right)^T\right] \\
&\quad + \left[\frac{1}{n}\sum_{t=n-|k|+1}^{n}\left(\bar{X}_s-X_{st}\right)\left(\bar{X}_s-\bar{\bar{X}}\right)^T\right] + \left[\frac{n-|k|}{n}\left(\bar{X}_s-\bar{\bar{X}}\right)\left(\bar{X}_s-\bar{\bar{X}}\right)^T\right] \\
&= \hat{\Gamma}_s(k) - \frac{1}{n}\sum_{t=1}^{|k|}A_{st}^T - \frac{1}{n}\sum_{t=n-|k|+1}^{n}A_{st} + \frac{n-|k|}{n}\left(\bar{X}_s-\bar{\bar{X}}\right)\left(\bar{X}_s-\bar{\bar{X}}\right)^T, \qquad (6)
\end{aligned}
$$

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},

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

$$= \frac{m-1}{mn}\left(\sum_{k=0}^{n-1}\Gamma(k) - n\text{Var}\left[\bar{X}_1\right]\right). \tag{7}$$

Similarly,

$$\mathbb{E}\left[A_{11}^T\right] = \mathbb{E}\left[A_{11}\right]^T = \frac{m-1}{mn}\left(\sum_{k=0}^{n-1}\Gamma(k)^T - n\text{Var}\left[\bar{X}_1\right]\right). \tag{8}$$

Further,

$$\begin{aligned}\mathbb{E}\left[\left(\bar{X}_1 - \bar{\bar{X}}\right)\left(\bar{X}_1 - \bar{\bar{X}}\right)^T\right] &= \mathbb{E}\left[\bar{X}_1\bar{X}_1^T - \bar{X}_1\bar{\bar{X}}^T - \bar{\bar{X}}\bar{X}_1^T + \bar{\bar{X}}\bar{\bar{X}}^T\right] \\ &= \left(\text{Var}(\bar{X}_1) + \mu\mu^T - \text{Var}(\bar{\bar{X}}) - \mu\mu^T\right) \\ &= \frac{m-1}{m}\text{Var}(\bar{X}_1)\,. \end{aligned} \tag{9}$$

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

$$\mathbb{E}[\hat{\Gamma}(k)] = \left(1 - \frac{|k|}{n}\right)\left(\Gamma(k) - \text{Var}(\bar{X})\right). \tag{10}$$

As a consequence, if $\text{Var}(\bar{X})$ is finite, then $\text{Var}(\bar{X}) \to 0$ as $n \to \infty$. Using (10), (7), (8), and (9) in (6),

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

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

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

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

then broken down as following,

$$\mathbb{E}\left[\hat{\Gamma}_{G,s}(k)\right] = \left(1 - \frac{|k|}{n}\right)\Gamma(k) + O_1 + O_2\,. \tag{11}$$

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,

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

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,

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

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,

$$\begin{aligned}
\|\bar{X}_s - \mu\|_\infty \leq \|\bar{X}_s - \mu\| &= \frac{1}{mn}\left\|\sum_{t=1}^{n} X_{st} - n\mu\right\| \\
&\leq \frac{1}{n}\left\|\sum_{t=1}^{n} X_{st} - n\mu - LB(n)\right\| + \frac{\|LB(n)\|}{n} \\
&< \frac{D\psi(n)}{n} + \frac{\|LB(n)\|}{n} \\
&< \frac{D\psi(n)}{n} + \frac{1}{n}\|L\|\left(\sum_{i=1}^{p}|B^{(i)}(n)|^2\right)^{1/2} \\
&\leq \frac{D\psi(n)}{n} + \frac{1}{n}\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}\,. \qquad (12)
\end{aligned}$$

Similarly,

$$\|\bar{\bar{X}} - \mu\|_\infty \leq \frac{D\psi(n)}{n} + \frac{1}{n}\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}\,. \qquad (13)$$

Now consider,

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

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

$$
\begin{aligned}
|M_1| &< 2\|(\bar{\bar{X}} - \mu)\|_\infty \left[\sum_{k=-b_n+1}^{b_n-1} \left[\frac{D\psi(k)}{n} + \frac{\|L\| p^{1/2}(1+\epsilon)\sqrt{2k \log\log k}}{n} \right] \right] + 2(2b_n - 1)\|\bar{X}_1 - \mu\|_\infty \\
&\leq 2(2b_n - 1)\|(\bar{\bar{X}} - \mu)\|_\infty \left[\frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2}(1+\epsilon)\sqrt{n \log\log n}}{n} + \|\bar{X}_1 - \mu\|_\infty \right] \\
&\leq 4(2b_n - 1) \left[\frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2}(1+\epsilon)\sqrt{n \log\log n}}{n} \right]^2 \quad \text{(by (12) and (13))}\,. \qquad (14)
\end{aligned}
$$

For M_2,

$$
\begin{aligned}
|M_2| &= \left| \frac{1}{m} \sum_{s=1}^{m} \left\{ \left(\mu - \bar{\bar{X}}\right)_i \left(\mu - \bar{\bar{X}}\right)_j \sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n}\right) w\left(\frac{k}{b_n}\right) \right\} \right| \\
&\leq \|\bar{\bar{X}} - \mu\|_\infty^2 \left[\sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n}\right) w\left(\frac{k}{b_n}\right) \right] < \|\bar{\bar{X}} - \mu\|_\infty^2 \left[\sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \right] \\
&\leq b_n \|\bar{\bar{X}} - \mu\|_\infty^2 \int_{-\infty}^{\infty} |w(x)| dx \\
&\leq C b_n \left[\frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2}(1+\epsilon)\sqrt{n \log\log n}}{n} \right]^2 \quad \text{(by (13))}\,. \qquad (15)
\end{aligned}
$$

Using (14) and (15),

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

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 $n \to \infty$.

□

Proof of theorem 3. We have the following decomposition,

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

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,

$$\left|\hat{\Sigma}^{ij}_G-\Sigma^{ij}\right| = \left|\hat{\Sigma}^{ij}_{SV}-\Sigma^{ij}+N_1+N_2+M_1+M_2\right| \le \left|\hat{\Sigma}^{ij}_{SV}-\Sigma^{ij}\right| + |N_1| + |N_2| + |M_1+M_2| \quad (16)$$

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

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

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

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

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

Using SIP on the summation of k terms,

$$
\begin{aligned}
&\left|\frac{1}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1}w\left(\frac{k}{b_n}\right)\frac{1}{n}\sum_{t=1}^{n-|k|}\left[\left(X_{st}-\bar{X}_s\right)_i\left(\bar{X}_s-\mu\right)_j\right]\right|\\
&<\frac{1}{m}\sum_{s=1}^{m}\|\bar{X}_s-\mu\|_\infty\sum_{k=-b_n+1}^{b_n-1}\left[\frac{D\psi(k)}{n}+\frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2k\log\log k}}{n}\right]+\frac{1}{m}\sum_{s=1}^{m}\frac{b_n(b_n-1)}{n}\|\bar{X}_s-\mu\|_\infty^2\\
&<\frac{(2b_n-1)}{m}\sum_{s=1}^{m}\|\bar{X}_s-\mu\|_\infty\left[\frac{D\psi(n)}{n}+\frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}}{n}\right]+\frac{1}{m}\sum_{s=1}^{m}\frac{b_n(b_n-1)}{n}\|\bar{X}_s-\mu\|_\infty^2\\
&\leq\left(2b_n-1+\frac{b_n^2}{n}-\frac{b_n}{n}\right)\left[\frac{D\psi(n)}{n}+\frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}}{n}\right]^2\to 0\,. \quad \text{(by (12))}
\end{aligned}
$$

Similarly, the second part of $N_1\to 0$ with probability 1. Following the steps in (15),

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

Thus, in (16), every term goes to 0 and $\hat{\Sigma}_G^{ij}\to\Sigma^{ij}$ with probability 1 as $n\to\infty$. □

A.4 Proof of Theorem 4

By Equation 11,

$$
\mathbb{E}\left[\hat{\Gamma}_G(k)\right]=\left(1-\frac{|k|}{n}\right)\Gamma(k)+\mathcal{O}\left(n^{-1}\right). \tag{17}
$$

$$
\mathbb{E}\left[\hat{\Gamma}_{G-}(k)\right]=\left(1-\frac{|k|}{n}\right)\Gamma(k)+O_1+O_2\,. \tag{18}
$$

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

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

where

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

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,

$$P_1 = -\sum_{|k|\geq n}\Gamma(k) - \sum_{k=-n+1}^{n-1} w\left(\frac{|k|}{n}\right)\frac{|k|}{n}\Gamma(k) - \sum_{k=-n+1}^{n-1}\left(1-w\left(\frac{|k|}{n}\right)\right)\Gamma(k)\,. \tag{19}$$

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

$$-\sum_{|k|\geq n}\Gamma(k) \leq \sum_{|k|\geq n}\left|\frac{k}{n}\right|^q\Gamma(k) = \frac{1}{b_n^q}\left|\frac{b_n}{n}\right|^q\sum_{|k|\geq n}|k|^q\Gamma(k) = o\left(\frac{1}{b_n^q}\right), \tag{20}$$

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,

$$\sum_{k=-n+1}^{n-1} w\left(\frac{|k|}{n}\right)\frac{|k|}{n}\Gamma(k) = o\left(\frac{1}{b_n^q}\right) \tag{21}$$

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,

$$\begin{aligned}\sum_{k=-n+1}^{n-1}\left(1-w\left(\frac{k}{b_n}\right)\right)\Gamma(k) &= -\frac{1}{b_n^q}\sum_{k=-n+1}^{n-1}\left(\frac{|k|}{b_n}\right)^{-q}\left(1-w\left(\frac{|k|}{b_n}\right)\right)|k|^{q}\Gamma(k)\\ &= -\frac{1}{b_n^q}\sum_{k=-n+1}^{n-1}\left[k_q+o(1)\right]|k|^{q}\Gamma(k)\\ &= -\frac{k_q\Phi^{(q)}}{b_n^q} + o\left(\frac{1}{b_n^q}\right).\end{aligned} \tag{22}$$

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 (20), (21), and (22) in (19), 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.

A.5 Proof of Theorem 5

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

$$\left|\hat{\Sigma}_G - \tilde{\Sigma}\right| \to 0 \text{ with probability } 1\,. \tag{23}$$

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 (23), 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 $n \to \infty$,

$$\mathbb{E}\left[\left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right)^2\right] \to 0\,. \tag{24}$$

We will use (24) 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 (24),

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

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

$$\frac{n}{b_n}\text{Var}(\tilde{\Sigma}^{ij}) = \left[\Sigma^{ii}\Sigma^{jj} + \left(\Sigma^{ij}\right)\right]\int_{-\infty}^{\infty} w^2(x)dx + o(1) \tag{25}$$

Plugging (25) 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 14. 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

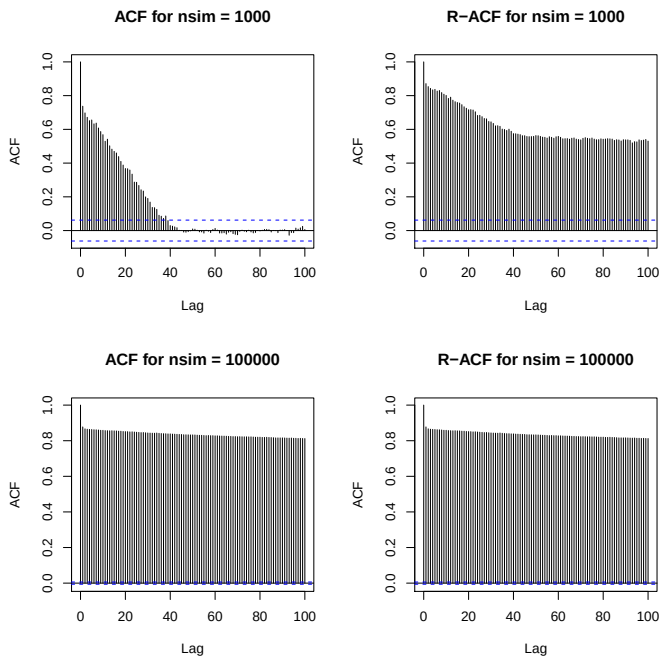


Figure 14: ACF and G-ACF for the first component of the first chain.

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.
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.

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.

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

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.

Damerdji, H. (1991). Strong consistency and other properties of the spectral variance estimator. *Management Science*, 37:1424–1440.

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.

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

Liu, Y. and Flegal, J. M. (2018). Weighted batch means estimators in Markov chain Monte Carlo. *Electronic Journal of Statistics*, 12:3397–3442.

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.

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.

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.

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., 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.