Globally-centered autocovariances in MCMC

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

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

August 20, 2020

Abstract

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

1 Introduction

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

A key quantity of interest that drives MCMC output analysis is the autocovariance function (ACvF). From their use in autocorrelation plots to assessing Monte Carlo variability of estimates, to determining when to stop MCMC simulation, autocovariances drive many visual and quantitative inferences users make from MCMC output. However, tools on estimating ACvF are largely constructed for output from one MCMC chain. a little bit more here.

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

stationary Markov chain (see Meyn and Tweedie, 2009, for definitions) employed to learn characteristics about F. The process is covariance stationary so that the ACvF at lag k depends only on k and is defined as

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

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

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

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

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

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

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

ACvFs are required to estimate the long-run variance of Monte Carlo averages. Specifically, spectral variance estimators are used to estimate $\Sigma = \sum_{k=-\infty}^{\infty}\Gamma(k)$. We replace $\hat{\Gamma}(k)$ with G-ACvF in spectral variance estimators (SVE) and show strong consistency of this estimator under weak conditions. In the spirit of Andrews (1991), we also obtain large-sample bias and variance of the resulting estimator of Σ. Spectral variance estimators can be slow to prohibitively expensive to calculate (Liu and Flegal, 2018). To relieve the computational burden, we adapt the fast algorithm of Heberle and Sattarhoff (2017) for G-ACvFs that dramatically reduces computation time. The globally-centered spectral variance estimator (G-SVE) is employed in the computation of effective sample size through the estimator given by Vats et al. (2019b). We will show that using G-SVE

for estimating Σ safeguards us against early termination of the MCMC process and improves the reliability of estimators.

1.1 Demonstrative example

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

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

where $f(x; a, b)$ is the density of a normal distribution with mean a and variance b. Two Markov chains are started at each mode. Let n denote the number of samples for each chain. Figure 1 indicates that in the first 10^4 samples, the chains do not jump modes so that both Markov chains yield different estimates of the population mean, μ. At 10^5 sample size, both Markov chains have traversed the state space reasonably and yield similar estimates of μ. We make two points here: first, for small sample size, we observe that the locally-centered ACF gives misleading estimates whereas G-ACF accounts for the discrepancy in sample means between two chains. This is evident from the ACF plots in Figure 2. Second, for a large sample size, the estimates from G-ACrF as well as ACrF are equivalent. Therefore, G-ACF can be used in any situation with a promise of estimates that are at least as good as empirical ACF estimators.

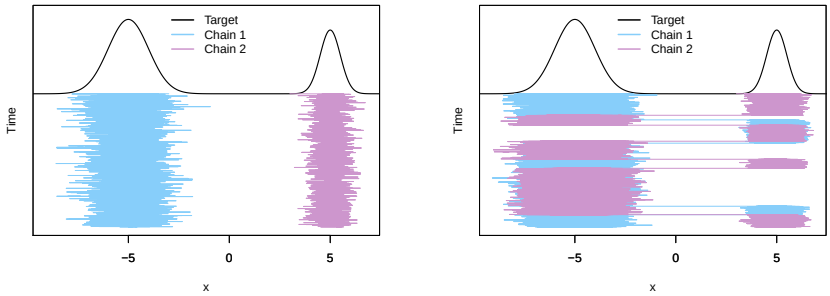

Figure 1: Target density and trace plots for two chains at $n = 10^4$ (left) and $n = 10^5$ (right)

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

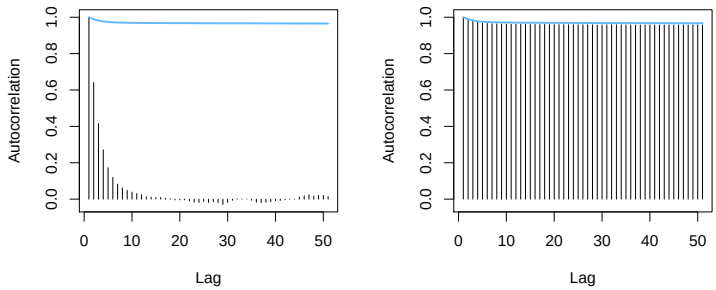

Figure 2: ACF plots using locally-centering (left) and global centering (right). Histogram estimates at $n = 10^4$ with the blue line plot being estimates from $n = 50000$ for chain 1.

the globally-centered autocovariance function (G-ACvF) estimator for the s^{th} Markov chain as

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

In the event that all m Markov chains have been run long enough that $\bar{X}_s \approx \bar{\bar{X}}$ are similar, then $\hat{\Gamma}_s(k) \approx \hat{\Gamma}_{G,s}(k)$. However, for shorter runs or for slow mixing chains, $\Gamma(k)$ is more appropriately estimated by $\hat{\Gamma}_{G,s}$ as it utilizes information from all chains and accounts for disparity between estimates of μ. We quantify the gains in bias for the G-ACvF estimator below. Let

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

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

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

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

and consequently,

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

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

Remark 1. Polynomial ergodicity and the moment conditions are required to ensure Φ and Σ are finite. The above result can be stated more generally for α-mixing processes, but we limit our

attention to Markov chains.

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

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

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

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

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

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

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

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

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

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

3 Variance estimators for multiple Markov chains

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

(CLT) if there exists a $p \times p$ positive-definite matrix Σ such that

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

where

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

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

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

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

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

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

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

3.1 Globally-centered spectral variance estimators

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

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

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

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

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

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

3.1.1 Theoretical results

We will present three main results for the G-SV estimator. First, we provide conditions for strong consistency. Strong consistency is particularly important to ensure sequential stopping rules yield correct coverage at termination (Glynn and Whitt, 1992). A critical assumption is that of a strong invariance principle which the following theorem establishes. Let $B(n)$ denotes a standard p-dimensional Brownian motion.

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

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

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

Our next two results establish large-sample bias and variance for G-SV and mimic those of $\hat{\Sigma}_s$ which can be found in Hannan (1970). However, Hannan (1970) proved the results assuming μ_g

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

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

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

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

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

3.1.2 Fast implementation

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

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

Heberle and Sattarhoff (2017) compute the $n\times 1$ vector $T(w)A_s$ using an algorithm based on fast Fourier transforms. Let $w^*=(1\ w_1\ w_2\ \ldots,\ w_{n-1},\ 0,\ w_{n-1},\ldots,w_1)^T$ and set $C(w^*)$ to be a symmetric circulant matrix such that $C(w^*)_{1:n,1:n}=T(w)$. With inputs $C(w^*)$ and A_s, Algorithm 1 produces $\hat{\Sigma}_s$ exactly. For more details, see Heberle and Sattarhoff (2017).
We note a similar decomposition is possible for the G-SV estimator. Setting $B_s=(Y_{s1}-\bar{\bar{Y}}\ \ldots\ Y_{sn}-\bar{\bar{Y}})^T$ and calling Algorithm 1 with inputs $C(w^*)$ and B_s yields $\hat{\Sigma}_{G,s}$. This algorithm has complexity $\mathcal{O}(pn\log n)$ and is thus orders of magnitude faster. Of particular importance is the fact that the bandwidth b_n has close to no impact on the computation time.

4 Effective sample size

An important method of assessing the reliability of $\bar{\bar{Y}}$ in estimating μ_g is to calculate the effective sample size (ESS). ESS are number of independent and identically distributed samples that would yield the same Monte Carlo variability as this correlated sampler for estimating μ_g. Let $|\cdot|$ denote

Algorithm 1: Heberle and Sattarhoff (2017) Algorithm

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

Compute eigenvalues λ_i of $C(w^*)_{(1)}$ using a discrete Fourier transform, $i = 1, \ldots, 2n$

Construct $2n \times p$ matrix $A_s^* = (A^T \quad 0_{n\times q})^T$

for $j = 1, 2, \ldots, p$ **do**

  Calculate $V^* A_s^{*(j)}$ by DFT of $A^*_{s(j)}$.

  Multiply $V^* A^{*(i)}_{s(j)}$ with the eigenvalue λ_i for all $i = 1, ..., 2n$ to construct $\Lambda V^* A^*_{s(j)}$.

  Calculate $C(w^*)A^*_{s(j)} = V\Lambda V^* A^*_{s(j)}$ by inverse FFT of $\Lambda V^* A^*_{s(j)}$.

end

Select the first n rows of $C(w^*)A_s^*$ to form $T(w)A_s$.

Premultiply by A_s^T and divide by n.

Output: $\hat{\Sigma}_s$

determinant. A multiple chain version of the ESS as defined by Vats et al. (2019b) is

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

The simulation terminates when the ESS is greater than a pre-specified, theoretically motivated, lower-bound W_p. In our setting of m parallel chains of n samples each, define

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

and estimate ESS with

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

We use the locally-centered $\hat{\Upsilon}(0)$ to estimate $\Upsilon(0)$ instead of $\hat{\Upsilon}_G(0)$ when calculating ESS. Both $\hat{\Upsilon}(0)$ and $\hat{\Upsilon}_G(0)$ are consistent for $\Upsilon(0)$; however using $\hat{\Upsilon}(0)$ produces lower sample sizes when the parallel chains do not agree, indicating that more MCMC draws are needed.
For the sake of comparison, $\widehat{\text{ESS}}_A$ is constructed similarly using $\hat{\Sigma}_A$ instead of $\hat{\Sigma}_G$ to estimate Σ.

5 Examples

In this section we consider three different target distributions and sample multiple Markov chains using MCMC methods to experimentally analyse the performance of our globally-centered estimators. We will make the following three comparisons - (1) A-ACF vs G-ACF through ACrF plots; (2) A-SVE vs G-SVE; (3) $\widehat{\text{ESS}}_G$ vs $\widehat{\text{ESS}}_A$. The quality of estimation of Σ is studied by coverage probabilities for a 95% confidence interval in cases where the true mean μ is known. The conver-

gence of local and global estimators as n increases is studied through two types of running plots (1) logarithm of Frobenius norm of estimated Σ denoted by $\|M\|_F$ for a matrix M, and (2) logarithm of estimated ESS/mn. In all the examples we consider the function $g(x) = x$, which implies the Markov chain $\{Y_t\}$ and $\{X_t\}$ are the same.

5.1 Vector Autoregressive Process

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

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

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

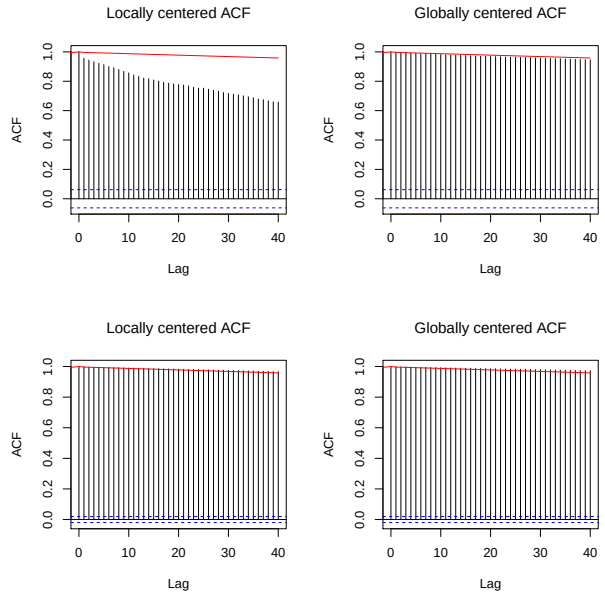

Figure 3: VAR(1). ACF (left) and G-ACF (right) for the second chain for $m = 5$. (Top) $n = 10^3$ and (bottom) $n = 10^4$. The red line is the true ACF.

We compare the the locally and globally-centered autocorrelations against the truth. In Figure 3 are the estimated ACF plots for the first component of the second chain against the truth in red. For a run length of 10^3 (top row), the commonly used locally centered ACF underestimates the true correlation structure giving a false sense of security about the mixing of the chain. The G-ACF, on the other hand, is much closer to the truth. This difference is negligible at the larger run length of $n = 10^4$ when each of the 5 chains have sufficiently explored the state space.

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

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

Table 1: VAR(1). Coverage probabilities at 95% nominal level using A-SVE and G-SVE.

The quality of estimation of Σ and ESS is assessed by running plots from 50 replications of run length 50000. In Figure 4, we present running plots of $\log(\|\hat{\Sigma}\|_F)$ for both $\hat{\Sigma}_G$ and $\hat{\Sigma}_A$ and the running plots of $\log(\widehat{\text{ESS}})$ for both $\widehat{\text{ESS}}_G$ and $\widehat{\text{ESS}}_A$. It is evident that the locally centered estimator of Σ severely underestimates the truth, leading to an overestimation of ESS. The GSV estimator is able to estimate Σ more accurately early on, safeguarding against early termination using ESS.

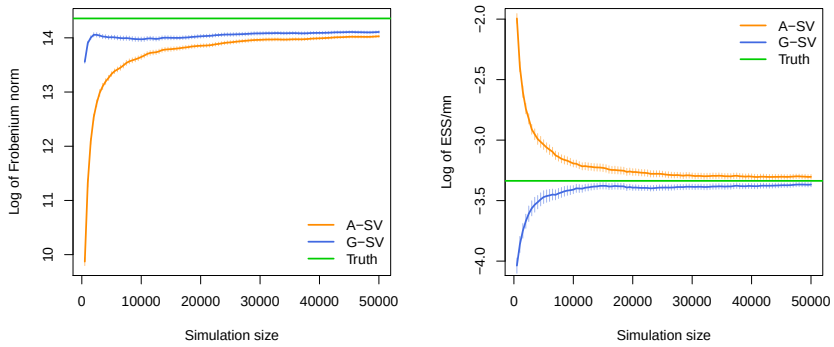

Figure 4: VAR(1). (Left) Running plot for logarithm of Frobenius norm of A-SV and G-SV.(Right) Running plot for logarithm of $\widehat{\text{ESS}}/mn$ using A-SV and G-SV.

5.2 Boomerang Distribution

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

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

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

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

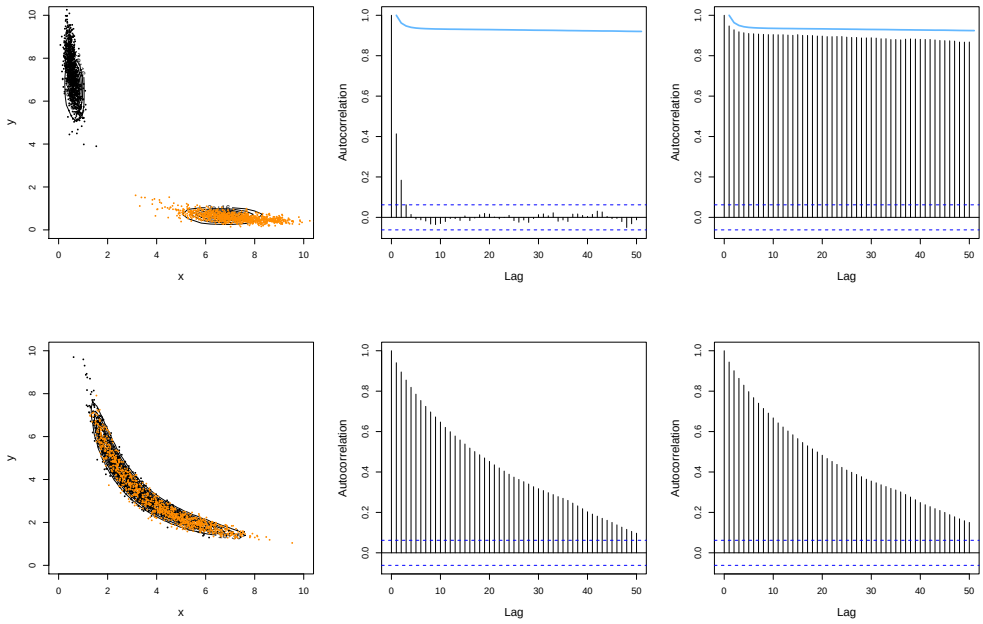

Figure 5: Boomerang. Top row is Setting 1 and bottom row is Setting 2. Contour plots of the target distributions overlaid with scatter plots for two chains for both settings. Locally-centered ACF (middle) and G-ACF (right) plots for one chain for $n = 1000$ with the blue line being the respectie ACF at $n = 10^4$.

We consider two settings; in setting 1 we let $A = 1,\ B = 3,\ C = 8$ which results in well-separated models. In setting 2, we let $A = 1,\ B = 10,\ C = 7$ which yields in a boomerang shape for the contours. The contour plots for these two settings are in the left in Figure 5, overlaid with scatter plots of two parallel draws of the Gibbs sampler. Setting 2 is chosen specifically to illustrate that

the locally and globally centered ACvFs perform similarly when the Markov chain moves freely across the state space. what is m here?

For Setting 1, when the Markov chains have not been able to jump modes, locally-centered ACF severely underestimates autocorrelation. This is seen in the middle column in Figure 5, where the locally-centered autocorrelations at $n = 1000$ (histogram) are drastically different from the locally-centered autocorrelations at $n = 10^4$. Somewhere between $n = 1000$ and $n = 10^4$, the Markov chains jump modes and it is only then that the locally-centered ACFs provide better estimates. The G-ACFs, on the other hand, produce similar ACF estimates at $n = 1000$ and $n = 10^4$ by measuring deviations about the global mean. For setting 2, both methods yield similar ACF estimates reinforcing our claim that there is much to gain by using G-ACvF and nothing to lose.

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

Table 2: Setting-1

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

Table 3: Setting-2

The true mean of the target distribution can be obtained using numerical approximation. Using this and both the A-SV and G-SV estimators, we construct 95% confidence regions and report coverage probabilities for 1000 replications for both $m = 2$ and $m = 5$. Table 2, 3 reports all results. In setting 1, systematically, the G-SV estimator yields far superior coverage than the A-SV estimator for all values of n. Whereas for setting 2, the results are almost similar indicating the equivalence of A-SV and G-SV for fast mixing Markov chains.

In Figure 6 we present running plots of estimates of $\log \mathrm{ESS}/mn$ for both setting 1 and setting 2. As the sample size increases, $\widehat{\mathrm{ESS}}_G$ and $\widehat{\mathrm{ESS}}_A$ become closer, but early on for setting 1, $\widehat{\mathrm{ESS}}_G$ estimates are much smaller giving more adequate estimates of ESS.

5.3 Sensor Network Localization

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

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

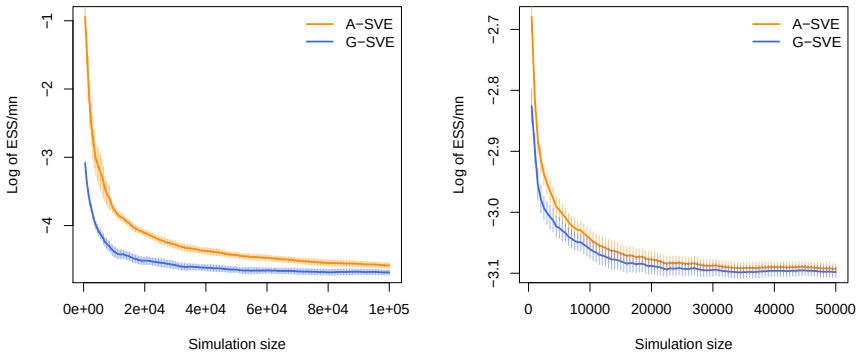

Figure 6: Boom: Running plot of $\log(\widehat{\text{ESS}}_A)/mn$ and $\log(\widehat{\text{ESS}}_G)/mn$ with $m = 5$ for Setting 1 (left) and Setting 2 (right).

$\pi(x_i, x_j) = \exp \|x_i - x_j\|^2/2R^2$ 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 $w_{ij} = 1$. We follow the Markov chain structure as described by Tak et al. (2018) and sample from the four bivariate conditionals for each sensor location using a Gibbs sampler. In their paper on Repelling Attractive Metropolis (RAM) algorithms, Tak et al. (2018) compare the performance of different sampling techniques and show that RAM improves the acceptance rate by a factor of at least 5.5 over Metropolis using the same jumping scale. RAM algorithm supplies Markov chains with higher jumping frequency between the modes.

This is an 8-dimensional estimation problem where the components corresponds to x and y coordinates of four unknown sensor locations. We will use the RAM algorithm with a jumping scale equal to 0.5 to sample five parallel Markov chains with well-separated starting points. The total simulation size for each chain is fixed at $100,000$. Since the truth about actual mean and asymptotic variance is not known in this case, we are not interested in coverage probabilities.

To illustrate the *sticky* nature of the Markov chains, we plot the evolution of two chains (starting near different modes) with time. Figure 7 shows the trace plot of x and y coordinate of sensor location-1. It is evident that the marginal posterior for both the components is bimodal. Similarly,

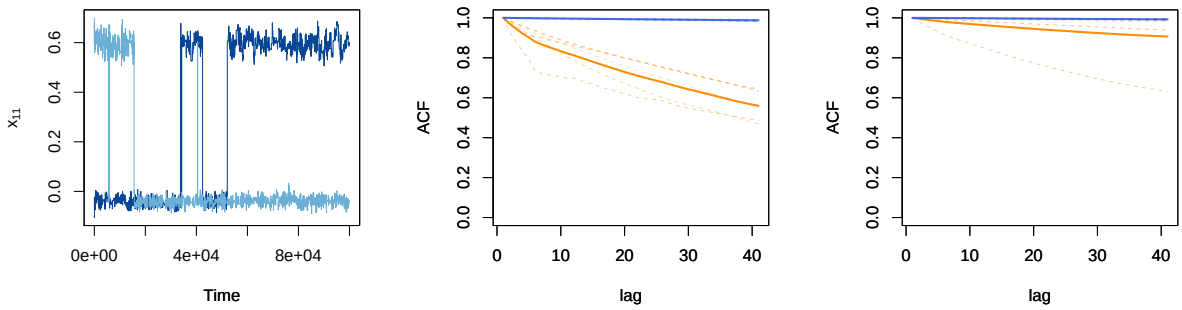

Figure 7: Sensor: Trace plots for x (left) and y (right) coordinate of sensor location-1.

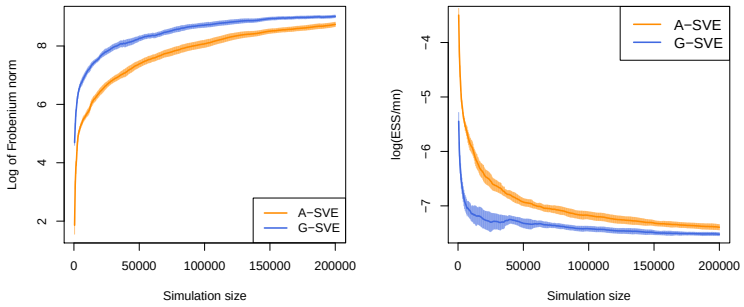

Figure 8: Sensor: Running plot of $\log(\|\Sigma\|_F)$ (left) and $\log(\text{ESS})/mn$ (right) estimated using A-SV and G-SV along with standard errors for 10 replications.

the other three locations also have a bimodal marginal distribution in each component. Observe that each chain explores one particular mode for a long time before jumping to another mode. Figure **??** plots the locally and globally centered ACF estimators for individual chains as well as the average over all the chains. The thick solid line plots the averaged estimator and the thin dotted lines correspond to ACF calculated using single chains. Figure 8 presents the running plot of $\log \|\Sigma\|_F$ and $\log \text{ESS}/mn$ estimated using A-SVE and G-SVE along with standard errors from 10 replications for each value of n. In both the plots, G-SVE and $\widehat{\text{ESS}}_G/mn$ reach the convergence value faster.

6 Discussion

For slowly mixing Markov chains, the traditional single chain empirical estimator for ACvF can dramatically underestimate the truth. This has a severe impact on ACF plots, Monte Carlo variance, and stopping time of MCMC algorithms. We provide a globally-centered estimate of the

ACvF that leads to improvements in all three aspects of MCMC output analysis.

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

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

The choice of starting vector is crucial for multiple chain sampling. In case the chains are started very near to each other, all the chains will explore nearly the same neighborhood in the state space. As a consequence, the G-ACvF will be equivalent to A-ACvF, and no significant improvement will be observed in estimation quality. We have used a deterministic initialization method in our examples using our prior knowledge about the location of modes in the target distribution. Comparing existing initialization methods MA: maybe cite a few and developing new ones could be an interesting direction for future research.

7 Appendix

7.1 Preliminaries

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

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

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

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

7.2 Proof of Theorem 1

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

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

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

$$\begin{aligned}
&\mathbb{E}\left[A_{11}\right] \\
&= \mathbb{E}\left[\left(X_{11}-\bar{X}_1\right)\left(\bar{X}_1-\bar{\bar{X}}\right)^T\right] \\
&= \mathbb{E}\left[X_{11}\bar{X}_1^T\right] - \frac{1}{m}\mathbb{E}\left[X_{11}\bar{X}_1^T\right] - \frac{m-1}{m}\mathbb{E}\left[X_{11}\bar{X}_2^T\right] + \frac{1}{m}\mathbb{E}\left[\bar{X}_1\bar{X}_1^T\right] + \frac{m-1}{m}\mathbb{E}\left[\bar{X}_1\bar{X}_2^T\right] - \mathbb{E}\left[\bar{X}_1\bar{X}_1^T\right] \\
&= \frac{m-1}{m}\left(\mathbb{E}\left[X_{11}\bar{X}_1^T\right] - \mathbb{E}\left[X_{11}\bar{X}_2^T\right] + \mathbb{E}\left[\bar{X}_1\bar{X}_2^T\right] - \mathbb{E}\left[\bar{X}_1\bar{X}_1^T\right]\right) \\
&= \frac{m-1}{m}\left(\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[X_{11}X_{1t}^T\right] - \mathbb{E}\left[X_{11}\right]\mathbb{E}\left[\bar{X}_2^T\right] + \mathbb{E}\left[\bar{X}_1\right]\mathbb{E}\left[\bar{X}_2^T\right] - \mathrm{Var}\left[\bar{X}_1\right] - \mathbb{E}\left[\bar{X}_1\right]\mathbb{E}\left[\bar{X}_1^T\right]\right) \\
&= \frac{m-1}{mn}\left(\sum_{k=0}^{n-1}\Gamma(k) - n\mathrm{Var}\left[\bar{X}_1\right]\right)\ . \qquad (8)
\end{aligned}$$

Similarly,

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

Further,

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

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

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

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

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

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

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

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

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

where,

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

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

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

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

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

7.3 Strong consistency argument

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

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

The average pseudo spectral variance estimator is

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

Further, let

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

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

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

where

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

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

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

Consequently

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

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

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

$$\leq \frac{D\psi(n)}{n} + \frac{1}{n}\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}\,. \tag{13}$$

Similarly,

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

Now consider,

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

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

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

For M_2,

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

Using (15) and (16),

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

where

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

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

□

Proof of theorem 3. We have the following decomposition,

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

$$= \hat{\Sigma}^{ij}_{SV} + N_1 + N_2\,,$$

where

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

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

Using the above and Lemma 2,

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

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

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

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

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

$$\leq \frac{1}{m}\sum_{s=1}^{m}\sum_{k=-b_n+1}^{b_n-1}\frac{\|\bar{Y}_s-\mu_g\|_\infty}{n}\left\|\sum_{t=1}^{|k|}\left(Y_{st}-\mu_g\right)\right\| + \frac{1}{m}\sum_{s=1}^{m}\frac{b_n(b_n-1)}{n}\left\|\bar{Y}_s-\mu_g\right\|_\infty^2 .$$

Using SIP on the summation of k terms,

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

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

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

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

7.4 Proof of Theorem 4

By Equation 12,

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

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

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

where

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

Similar to Hannan (2009), we break P_1 into three parts. Note that notation $A = o(z)$ for matrix A implies $A^{ij} = o(z)$ for every (i,j)th element of the matrix A. Consider,

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

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

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

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

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

For $q \geq 1$,

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

For $q < 1$,

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

So,

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

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

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

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

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

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

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

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

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

which completes the result.

7.5 Proof of Theorem 5

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

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

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

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

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

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

We have shown from the proof of strong consistency that,

$$\begin{aligned}
&\left|\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right| \\
&\leq \frac{1}{m}\sum_{s=1}^{m}\left|\sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right)\sum_{t=1}^{n-|k|}\left[\left(\frac{(Y_{st}-\mu_g)_i(\mu_g-\bar{\bar{Y}})_j}{n}\right) + \left(\frac{(\mu_g-\bar{\bar{Y}})_i(Y_{s(t+k)}-\mu_g)_j}{n}\right)\right]\right. \\
&\qquad \left. +(\mu_g-\bar{\bar{Y}})(\mu_g-\bar{\bar{Y}})^T\sum_{k=-b_n+1}^{b_n-1}\left(\frac{n-|k|}{n}\right)w\left(\frac{k}{n}\right)\right| < D^2 g_1(n) + Dg_2(n) + g_3(n)\,.
\end{aligned}$$

By (22), there exists an N_0 such that

$$\begin{aligned}
\left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right)^2 &= \left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right)^2 I(0 \leq n \leq N_0) + \left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right)^2 I(n > N_0) \\
&\leq \left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right)^2 I(0 \leq n \leq N_0) + \left(D^2 g_1(n) + Dg_2(n) + g_3(n)\right)^2 I(n > N_0) \\
&:= g_n^*(X_{11}, \ldots, Y_{1n}, \ldots, Y_{m1}, \ldots, Y_{mn})\,.
\end{aligned}$$

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

$$\mathbb{E}\left|g_n^*\right| \leq \mathbb{E}\left[\left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}_A^{ij}\right)^2\right] + \mathbb{E}\left[\left(D^2 g_1(n) + Dg_2(n) + g_3(n)\right)^2\right] < \infty\,.$$

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

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

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

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

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

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

$$\leq \mathbb{E}\left[\left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right)^2\right] + 2\left|\left(\mathbb{E}\left[\left(\hat{\Sigma}_G^{ij} - \tilde{\Sigma}^{ij}\right)^2\right] \mathrm{Var}\left(\tilde{\Sigma}^{ij}\right)\right)^{1/2}\right|$$

$$= o(1) + 2\left(o(1)\left(O\left(\frac{b_n}{n}\right) + o\left(\frac{b_n}{n}\right)\right)\right) = o(1)\,.$$

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

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

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

8 Additional Examples

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

8.1 Bayesian Poisson Change Point Model

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

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

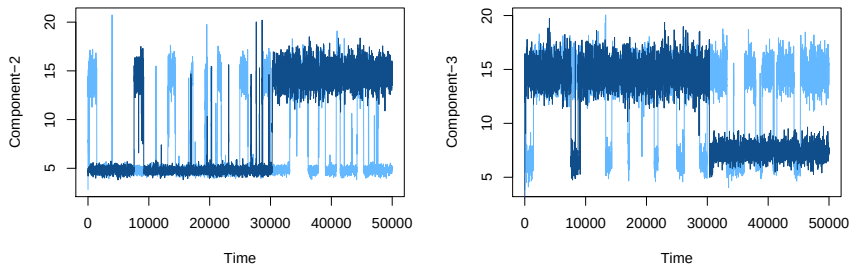

Figure 9: Trace plot for second (left) and third (right) component of two chains started from random points.The two colors denote the two different chains.

This is a 7-dimensional estimation problem wherein the marginal distribution of majority of components display a multimodal nature. Figure 9 shows the evolution of two chains started from random points with time for the second component. We will report the ACF plots for component-2. Similar behavior is observed for ACF plots of other components as well
Figure 10 demonstrates a striking advantage of G-ACF in estimating the true autocorrelations. G-ACF gives a more realistic and accurate estimation of autocorrelations in almost ten times lesser chain length.

8.2 Network crawling

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

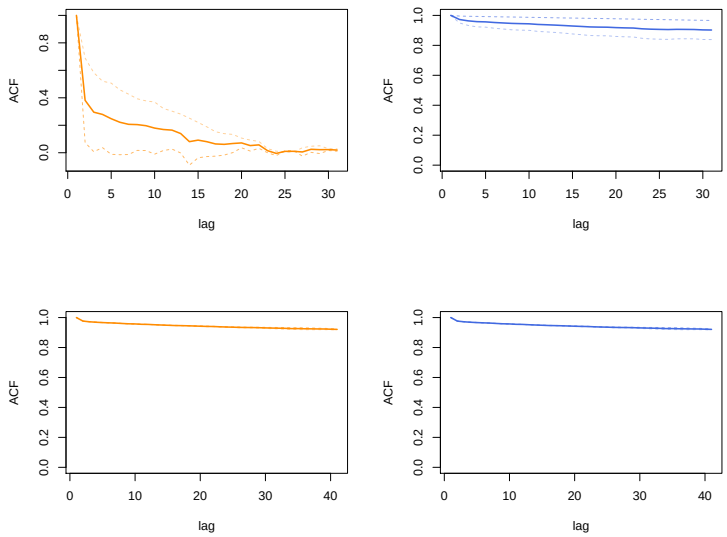

Figure 10: Change point: ACF (left) and G-ACF (right) for individual chains and average over m chains for $n = 1000$ (left) and $n = 10000$ (right).

The goal is to draw nodes uniformly from the network by using a network crawler. Nilakanta et al. (2019) modified the data by removing 1,022 out of 1,461 nodes to obtain a well-connected graph. This resulting social network has 439 nodes and 573 edges. We use a Metropolis-Hastings algorithm with a simple random-walk proposal suggested by Gjoka et al. (2011). Each node is associated with five features namely - degree of connection, cluster coefficient, grade, binary sex indicator (1 for female, 0 for male), and binary race indicator (1 for white, 0 for others)
We believe that the students from different races engage in a "selective networking" which might cause formation of clusters in the network wherein students within a cluster engage more with each other than with the students outside it. We sample two parallel Markov chains starting from two students belonging to different races and study its impacts on the average features of their immediate social group. Figure 11 shows the ACF and G-ACF plots for the third feature at two different simulation sizes. As evident, G-ACF displays a clear advantage over ACF in just $n = 100$ samples.

References

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

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.

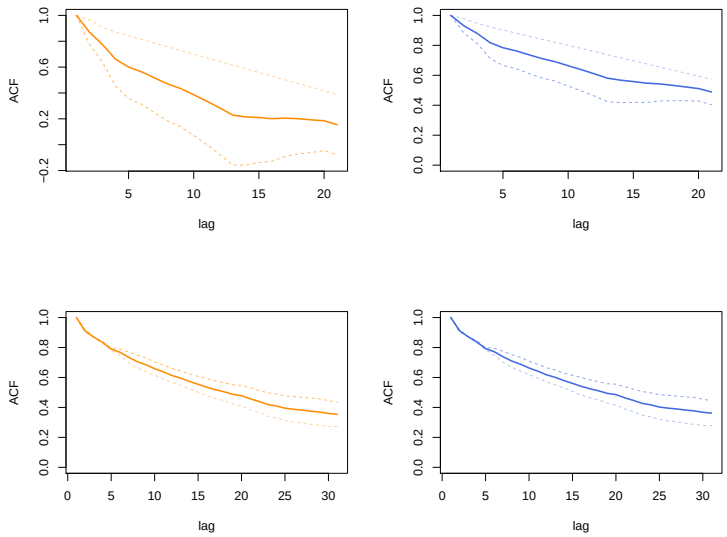

Figure 11: Crawling: ACF (left) and G-ACF (right) for individual chains and average over m chains for $n = 1000$ (left) and $n = 10000$ (right).

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

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

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

Dai, N. and Jones, G. L. (2017). Multivariate initial sequence estimators in Markov chain Monte Carlo. *Journal of Multivariate Analysis*, 159:184–199.

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

Flegal, J. M., Haran, M., and Jones, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, 23:250–260.

Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38:1034–1070.

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

Gjoka, M., Kurant, M., Butts, C. T., and Markopoulou, A. (2011). Practical recommendations on crawling online social networks. *IEEE Journal on Selected Areas in Communications*, 29(9):1872–1892.

Glynn, P. W. and Whitt, W. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability*, 2:180–198.

Gong, L. and Flegal, J. M. (2016). A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 25:684–700.

Gupta, K. and Vats, D. (2020). Estimating Monte Carlo variance from multiple Markov chains. *arXiv preprint arXiv:2007.04229*.

Hannan, E. J. (1970). Multiple time series: Wiley series in probability and mathematical statistics.

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

Heberle, J. and Sattarhoff, C. (2017). A fast algorithm for the computation of hac covariance matrix estimators. *Econometrics*, 5(1):9.

Ihler, A. T., Fisher, J. W., Moses, R. L., and Willsky, A. S. (2005). Nonparametric belief propagation for self-localization of sensor networks. *IEEE Journal on Selected Areas in Communications*, 23(4):809–819.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998). Markov chain Monte Carlo in practice: a roundtable discussion. *The American Statistician*, 52:93–100.

Kuelbs, J. (1976). A strong convergence theorem for Banach space valued random variables. *The Annals of Probability*, 4:744–771.

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

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

Meyn, S. P. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press.

Nilakanta, H., Almquist, Z. W., and Jones, G. L. (2019). Ensuring reliable monte carlo estimates of network properties. *arXiv preprint arXiv:1911.08682*.

Priestley, M. B. (1981). *Spectral analysis and time series: probability and mathematical statistics*. Number 04; QA280, P7.

Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., Tabor, J., Beuhring, T., Sieving, R. E., Shew, M., et al. (1997). Protecting adolescents from harm: findings from the national longitudinal study on adolescent health. *Jama*, 278(10):823–832.

Roy, V. (2019). Convergence diagnostics for Markov chain Monte Carlo. *Annual Review of Statistics and Its Application*, 7.

Song, W. T. and Schmeiser, B. W. (1995). Optimal mean-squared-error batch sizes. *Management Science*, 41(1):110–123.

Tak, H., Meng, X.-L., and van Dyk, D. A. (2018). A repelling–attracting metropolis algorithm for multimodality. *Journal of Computational and Graphical Statistics*, 27(3):479–490.

Tjstheim, D. (1990). Non-linear time series and markov chains. *Advances in Applied Probability*, 22(3):587–611.

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

Vats, D., 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.

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