Numerical Estimation of the Second Largest Eigenvalue of a Reversible Markov Transition Matrix

by

Kranthi Kumar Gade

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science

New York University

January, 2009

Jonathan Goodman

© Kranthi Kumar Gade

All Rights Reserved, 2009

To my family

ACKNOWLEDGMENTS

Grad school has been an arduous journey and crossing the chequered flag would have been an impossible task if not for the support and encouragement of many people. First and foremost among them is my advisor Jonathan Goodman – with his infinite wisdom and knowledge, his cheerful demeanor, his willingness to be accessible at all times, and his penchant for being tough when required all make him as close to a perfect advisor as there can be. I have learned something from each and every conversation with him, be it about mathematics, physics, politics or even literature – I wish to thank Jonathan for the treasure trove of memorable moments.

Secondly, I would like to thank Michael Overton for his constant support and guidance. The thing I admire most in Mike is his attention to detail and I learned from him that the style of writing is as important as the content of the writing itself – if I ever get any compliments about my technical writing skills, they should go to Mike. Mike is one of the nicest people I ever met – he has helped me with my personal problems and supported me financially for a good part of my graduate school – I would not have finished this thesis without his mentoring.

Over the years, I have learned a lot from my association with Yann LeCun, Margaret Wright, C. Sinan Güntürk, David Bindel, Eric Vanden-Eijnden, Adrian Lewis, and Persi Diaconis. A special note of thanks to all the administrative staff at Courant who made my stay there very smooth and easy-going: Rosemary Amico, Santiago Pizzini, Anina Karmen, and Naheed Afroz.

On a personal level, I would like to thank my family – my parents for their love and my sisters and brothers-in-law for their encouragement. They had a strong belief and confidence in me, though they could never understand why one would spend so many years of one's life on an obscure problem of computing the second eigenvalue of a Markov transition matrix. For that and everything else, I shall forever be grateful to them. I hope the next 100 pages will convince

them that the 6 long years and thousands of miles of separation from them was worth it.

My roommates have been a pleasant source of diversion from my work and I have had scores of them through the duration of my graduate study: Prashant Puniya, Sumit Chopra, Govind Jajoo, Arka Sengupta, Siddhartha Annapureddy, Abhishek Trivedi – they partook in my joys and helped with my hurdles. A note of thanks to all my friends at Courant who made my stay very memorable: Vikram Sharma, David Kandathil, Emre Mengi, Nelly Fazio, Antonio Nicolosi, Marc Millstone, and Tim Mitchell. Last, but definitely not the least, I am forever indebted to my best friends: Sreekar Bhaviripudi, Shashi Borade, Swapna Bhamidipati, Nagender Bandi, Rajesh Menon, Rajesh Vijayaraghavan, Bharani Kacham, Kranthi Mitra, and Aameek Singh.

ABSTRACT

We discuss the problem of finding the second largest eigenvalue of an operator that defines a reversible Markov chain. The second largest eigenvalue governs the rate at which the statistics of the Markov chain converge to equilibrium. Scientific applications include understanding the very slow dynamics of some models of dynamic glass. Applications in computing include estimating the rate of convergence of Markov chain Monte Carlo algorithms.

Most practical Markov chains have state spaces so large that direct or even iterative methods from linear algebra are inapplicable. The size of the state space, which is the dimension of the eigenvalue problem, grows exponentially with the system size. This makes it impossible to store a vector (for sparse methods), let alone a matrix (for dense methods). Instead, we seek a method that uses only time correlation from samples produced from the Markov chain itself.

In this thesis, we propose a novel Krylov subspace type method to estimate the second largest eigenvalue from the simulation data of the Markov chain using test functions which are known to have good overlap with the slowest mode. This method starts with the naive Rayleigh quotient estimate of the test function and refines it to obtain an improved estimate of the second largest eigenvalue. We apply the method to a few model problems and the estimate compares very favorably with the known answer. We also apply the estimator to some Markov chains occuring in practice, most notably in the study of glasses. We show experimentally that our estimator is more accurate and stable for these problems compared to the existing methods.

TABLE OF CONTENTS

Dedication ii

Acknowledgments iii

Abstract v

List of Figures viii

List of Tables xi

1 Introduction 1

2 Methodology of MCMC 4

2.1 Reversibility 6

3 Estimation of λ_* for reversible chains using simulation data 9

3.1 Prony's method 11

3.2 Krylov subspace algorithm for estimating spectral gap 13

3.2.1 Choice of the lag parameter 22

3.2.2 Estimating the LGEM of the pencil $\widehat{A}_{n,r} - \lambda\widehat{B}_{n,r}$ 25

3.3 Error bars for $\hat{\lambda}_*$ using batch means method 29

4 Model problems 32

4.1 AR(1) process 32

4.1.1 Observable $H_1 + H_2 + H_3 + H_4$ 35

4.2 Ehrenfest Urn Model 39

4.2.1 Identity observable 42

5 More accurate and stable estimates for λ_* 46

5.1 Series sum estimate 47

5.2 Least squares estimate 50

5.2.1 Obtaining a single estimate 55

6 Results 58

6.1 East model 58

6.2 Fredrickson-Andersen model 66

7 Multiple Observables 75

7.1 Results for Ising model with Glauber dynamics 78

Conclusion 83

Bibliography 84

LIST OF FIGURES

4.1 *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for AR(1) process with observable* $H_1 + H_2 + H_3 + H_4$ *using KSP singleton method and TLS Prony method for* $n = 3$ *and different values of* r*. Error bars are computed using* (4.1.5). 37

4.2 *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for AR(1) process with observable* $H_1 + H_2 + H_3 + H_4$ *using KSP singleton method and TLS Prony method for* $n = 4$ *and different values of* r. 38

4.3 *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for AR(1) process with observable* $H_1 + H_2 + H_3 + H_4$ *using KSP singleton method and TLS Prony method for* $n = 7$ *and different values of* r. 39

4.4 *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for Ehrenfest urn process (with* $N = 30, p = 0.4$*) with identity observable using KSP singleton method and TLS Prony method for* $n = 5$ *and different values of* r. 41

4.5 *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for Ehrenfest urn process (with* $N = 30, p = 0.4$*) with identity observable using KSP singleton method and TLS Prony method for* $n = 10$ *and different values of* r. 43

4.6 *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for Ehrenfest urn process (with* $N = 30, p = 0.4$*) with identity observable using KSP singleton method and TLS Prony method for* $n = 1$ *and different values of* r. 44

5.1 *The series sum estimate for the reduced spectral radius* $\hat{\lambda}_*^{SS}[n] \pm \hat{\sigma}[\lambda_*^{SS}[n]]$ *for AR(1) process with observable* $H_1 + H_2 + H_3 + H_4$ *using KSP and TLS Prony methods for* $n = 1, 2, \ldots, 10$. 48

5.2 *The series sum estimate for the reduced spectral radius $\hat{\lambda}_*^{SS}[n] \pm \hat{\sigma}[\lambda_*^{SS}[n]]$ for Ehrenfest urn process with identity observable using KSP and TLS Prony methods for $n = 1, 2, \ldots, 10$.* 49

5.3 *The LS, ML and series sum estimates with error bars, for the reduced spectral radius for AR(1) process with observable the sum of first four eigenfunctions for $n = 1, 2, \ldots, 10$.* 53

5.4 *The LS, ML and series sum estimates with error bars, for the reduced spectral radius for Ehrenfest urn process with identity observable for $n = 1, 2, \ldots, 10$. Even though the estimates may look all over the place, note the vertical scale which runs from 0.965 to 0.9683.* 54

6.1 *Plot depicting the overlap of the Aldous-Diaconis function with the eigenvectors of transition matrix corresponding to the East model with $\eta = 10, p = 0.1$. Only the 10 largest eigenvalues (apart from 1) have been plotted.* 64

6.2 *Plot of the KSP LS estimate, the Rayleigh quotient estimate of the spectral gap and actual spectral gap for East model ($\eta = 10, p = 0.1$) with Aldous-Diaconis function.* . 66

6.3 *Plot of the KSP LS estimate and the Rayleigh quotient estimate of the spectral gap for the East model ($\eta = 25, p = 1/25$) with Aldous-Diaconis function.* 67

6.4 *Plots of functions g_1, g_2 defined in (6.2.3).* 69

6.5 *Plot depicting the overlap of the function $f_{\text{FA-1f}}$ with the eigenvectors of transition matrix corresponding to the FA-1f model with $\eta = 3, p = 0.1$. Only the 10 largest eigenvalues (apart from 1) have been plotted.* 71

6.6 *Plot of the KSP LS estimate, the Rayleigh quotient estimate of the spectral gap and actual spectral gap for FA-1f model ($\eta = 3, p = 0.1$) with the function $f_{\text{FA-1f}}$ in (6.2.5).* 72

6.7 *Plot depicting the variation of the KSP estimate of spectral gap for FA1f process in two dimensions with varying up-spin probability p. The size of the grid is fixed to be 20×20 and p takes values $\frac{1}{20}, \frac{1}{50}, \frac{1}{100}, \frac{1}{150}, \frac{1}{200}, \frac{1}{250}, \frac{1}{300}, \frac{1}{350}, \frac{1}{400}$.* 74

7.1 *Comparison between the LS estimates corresponding to the multiple observable case $\{f_1, f_2, f_3\}$ and the single observable f_1 for the AR(1) process, where $f_1 = \frac{1}{2}H_1 + H_2 + H_3 + H_4, f_2 = H2 + H3, f_3 = H_4$ (H_i is the ith Hermite polynomial).* 78

7.2 *Comparison between the LS estimates corresponding to the multiple observable case $\{M, M^3, M^5\}$ and the single observable M for the Glauber dynamics process, where M is the total magnetization. The spin collection is a 2D grid of size 3×3 with periodic boundary conditions at temperature $T = T_c/4$, where T_c is the critical temperature for 2D Ising model.* 80

7.3 *Comparison between the LS estimates corresponding to the multiple observable case $\{M, M^3, M^5\}$ and the single observable M for the Glauber dynamics process, where M is the total magnetization. The spin collection is a 2D grid of size 10×10 with periodic boundary conditions at temperature $T = T_c/4$, where T_c is the critical temperature for 2D Ising model.* 81

LIST OF TABLES

5.1 ML, LS and series sum least squared error estimates for AR(1) process with $\lambda_* = 0.99$ and observable as sum of first four eigenfunctions. 56

5.2 ML, LS and series sum least squared error estimates for Ehrenfest urn process with $\lambda_* = 29/30$ and identity observable. 57

1
INTRODUCTION

In many fields, including statistics, computer science and statistical physics, Markov processes that satisfy a condition called reversibility are frequently encountered. For many such processes, most notably in physics, the slowest mode has a physical significance (we consider specifically the case of kinetically constrained spin models for glass transitions in section 6). The slowest mode for a reversible Markov process is the eigenfunction corresponding to the second largest eigenvalue of its transition matrix and its "decay rate" is given by the eigenvalue itself (the transition matrix, being stochastic, always has the first eigenvalue as 1 and all the eigenvalues in $[-1, 1]$).

Another important class of reversible Markov processes occur in Markov Chain Monte Carlo (MCMC), which is an important tool of scientific computing. The basic idea of MCMC is that for many distributions which are difficult to sample from directly, a Markov chain which converges to the distribution can be constructed more readily. The distribution is then sampled by simulating the Markov chain. Once the chain has been constructed, the other important issue is how long to perform its simulation. This simulation time of a Markov chain depends on the time it takes to converge to its equilibrium (invariant) distribution – in other words, its rate of convergence. For a reversible chain, a concrete bound for the rate of convergence in terms of the second largest eigenvalue modulus (SLEM) of the transition matrix has been established (Diaconis and Stroock, 1991, Prop. 3). Following Gade and Overton, 2007, we use the term *reduced spectral radius* for the SLEM. We also use the term *spectral gap* often – it is defined as the absolute value of the difference between 1 and the reduced spectral radius. The smaller the spectral gap, the slower the chain converges to stationarity.

To assess the rate of convergence of a Markov chain, it is clearly desirable to have an es-

timate of its spectral gap. Only for special chains can this quantity be computed exactly – the Ehrenfest urn model (section 4.2) is an example. A more frequently used approach is to obtain analytical bounds for the spectral gap – most notably using Cheeger's or Poincaré's inequalities (see Diaconis and Stroock, 1991, Lawler and Sokal, 1988, Sinclair and Jerrum, 1989).

Using an analytical approach to obtain tight bounds for the spectral gap is a hard task – and one that needs to be performed separately for each Markov chain. For chains for which it has been done using Cheeger's and Poincaré's inequalities Diaconis and Stroock, 1991, the bound depends on the choice of a so-called *canonical path*. Moreover, in most cases these analytical approaches yield bounds which are valid in an *asymptotic* sense. If one wishes to use them to determine the run length of the chain in question, the bounds may well have constants which cannot be ignored in practice.

The other approach to estimate the spectral gap is to use the simulation data generated by the process itself. There are precedents for a data-based approach, notably Garren and Smith, 2000 and Pritchard and Scott, 2001. The approach in Pritchard and Scott, 2001 is applicable in the situation where the transition matrix depends on a set of unknown parameters, which are estimated from the simulation data; the estimate of the spectral gap is then the spectral gap of the estimated transition matrix. For most chains which occur in practice (for example, the Metropolis chain for simulating the Ising model), the transition matrix is known exactly, but it is too large to store, let alone to perform linear algebra operations like eigenvalue computations.

The method outlined in Garren and Smith, 2000 is another data-based approach. It is similar to the well-known Prony's method in that it estimates the reduced spectral radius using least squares fitting of exponentials. It is shown to have good theoretical properties in an asymptotic sense, but there are no concrete results to show how it works when applied to Markov chains which occur in practice, the ones we are most interested in. Furthermore, one often has some intuition of how the *slowest mode*, the eigenfunction corresponding to the reduced spectral ra-

dius, should look like for the chain in question. There is no way of incorporating this important *a priori* knowledge with the method suggested in Garren and Smith, 2000.

In this paper, we propose a novel simulation data-based method which estimates the spectral gap (and hence the rate of convergence) of a Markov process. This method is a Krylov subspace type method and uses, as a starting observable, a function that is known to have a substantial overlap with the slowest mode. The naive Rayleigh quotient of the function is then refined to obtain a better estimate of the reduced spectral radius. We then apply the method to a range of model problems, including examples where the exact spectral decomposition is known (such as the AR(1) process and Ehrenfest urn process in sections 4.1 and 4.2 respectively). We show empirically that our estimate is more accurate than any existing methods, including Prony's method and the Rayleigh quotient estimate. The examples we consider include kinetically constrained spin models, like the east model and the Fredrickson-Andersen model, where the spectral gap is small, and hence difficult to estimate.

2
METHODOLOGY OF MCMC

In this section, we briefly review MCMC, the notion of reversibility and how, when the reversibility condition holds, the spectral gap quantifies the rate of convergence of the slowest mode.

Consider an irreducible, aperiodic Markov chain $\{X_n, n \geq 0\}$ on a finite discrete state space $\mathcal{X}$. If $|\mathcal{X}| = N$, this chain can be represented by an $N \times N$ transition matrix P. We use the terms "transition matrix" and "Markov kernel" interchangeably. For finite state spaces, an irreducible, aperiodic Markov chain is also called ergodic and the following fact about ergodic Markov chains is well-known (Bremaud, 1999, Chap. 3 Theorem 3.3, Chap. 4 Theorem 2.1).

Proposition 2.1. *Let P be an ergodic Markov kernel on a finite state space $\mathcal{X}$. Then P admits a unique steady state distribution π, that is,*

$$\forall x, y \in \mathcal{X}, \ \lim_{t \to \infty} P^t(x, y) = \pi(y).$$

This unique π is an invariant (or stationary) distribution for the chain, that is,

$$\sum_x \pi(x) P(x, y) = \pi(y).$$

The methodology of MCMC is typically as follows: given a measurable function (frequently called an *observable*) $f : \mathcal{X} \to \mathbb{R}$, one needs to estimate the quantity $E_\pi[f]$, the expectation of f in the distribution π. If π is difficult to sample from directly, an ergodic Markov chain $\{X_t, t \geq 0\}$ is constructed with π as its invariant distribution; the average of $f(X_t)$, $t \geq 0$, is then an estimate for $E_\pi[f]$. The basis of this methodology is the following Ergodic Theorem for Markov chains.

Proposition 2.2. *Ergodic Theorem for Markov chains: Let* $\{X_t, t \geq 0\}$ *be an ergodic Markov chain with state space* $\mathcal{X}$ *and stationary distribution* π*, and let* $f : \mathcal{X} \to \mathbb{R}$ *be such that* $E_\pi[|f(X)|] < \infty$*. Then for any initial distribution* ν*,*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} f(X_t) = E_\pi[f(X)] \quad \textit{a.s.}$$

From Proposition 2.2, we see that the estimator $\hat{f}_{\nu,T} = \dfrac{1}{T}\sum_{t=0}^{T} f(X_t)$ converges to $E_\pi[f(X)]$ with probability 1 and in fact can be shown to do so with fluctuations of size $T^{-1/2}$ (central limit theorem Maxwell and Woodroofe, 2000). Note that the 'ν' in $\hat{f}_{\nu,T}$ means that $X_0 \sim \nu$, for some arbitrary distribution ν. The estimator $\hat{f}_{\nu,T}$ of $E_\pi[f(X)]$ is biased unlike $\hat{f}_{\pi,T}$, but it turns out that the bias is only of the order $1/T$ and asymptotically much smaller than the statistical fluctuations that we have. Although in theory, it is acceptable if the sampling is started at $t = 0$, in practice a certain initial *burn-in* time is allowed to elapse before the sampling begins (see Sokal, 1989, Section 3 for more details).

How long should we simulate the chain to get a good estimate of $E_\pi[f]$? This run length depends on the variance of the estimator $\hat{f}_{\nu,T}$ – which in turn depends on the correlations among the samples $f(X_0), f(X_1), \ldots$. To measure these correlations, we define, for a fixed s, the *equilibrium autocovariance function* at lag s, $C_f(s)$ as:

$$C_f(s) \equiv \text{cov}_\pi[f(X_0), f(X_s)]. \tag{2.0.1}$$

From the ergodicity property of the Markov chain, $C_f(s)$ can also be written as:

$$C_f(s) = \lim_{t\to\infty} \text{cov}_\nu[f(X_t), f(X_{t+s})].$$

The quantity $C_f(s)$ measures the covariance between $f(X_t)$ and $f(X_{t+s})$ for very large t – sufficiently large that the distribution of X_t becomes independent of that of X_0. It is clear from

the definition that C_f is an even function, that is, $C_f(s) = C_f(-s)$. The *autocorrelation function* at lag s, $\rho_f(s)$ is defined as:

$$\rho_f(s) \equiv C_f(s)/C_f(0). \tag{2.0.2}$$

For a given observable f, we define the *integrated autocorrelation time*, $\tau_{\mathrm{int},f}$ as:

$$\tau_{\mathrm{int},f} \equiv \sum_{t=-\infty}^{t=\infty} C_f(t)/C_f(0) = \sum_{t=-\infty}^{t=\infty} \rho_f(t). \tag{2.0.3}$$

To simplify things, from now on, let us assume that the distribution of X_0 is π – this is equivalent to assuming that the sampling begins after the "initial transient" has disappeared. An expression for the variance of the estimator $\hat{f}_T = \frac{1}{T}\sum_{t=0}^{T} f(X_t)$ can now be given in terms of $\tau_{\mathrm{int},f}$ (Sokal, 1989, Equation (2.20)):

$$\mathrm{Var}[\hat{f}_T] = \frac{1}{T}\tau_{\mathrm{int},f} C_f(0) + O\left(\frac{1}{T^2}\right), \quad \text{for } T \gg \tau_{\mathrm{int},f}. \tag{2.0.4}$$

Note an additional factor of 2 in (Sokal, 1989, Equation(2.20)) – which can be attributed to a difference in the definitions of $\tau_{\mathrm{int},f}$ (see Sokal, 1989, Equation(2.16)).

2.1 Reversibility

An important notion in Markov chain theory is the notion of reversibility.

Reversible chain: P is said to be *reversible* with respect to a distribution π if

$$\forall x, y \in \mathcal{X},\ \pi(x)P(x,y) = \pi(y)P(y,x).$$

It is not hard to prove that if P is reversible with respect to π and P is ergodic, then π is the stationary distribution for P. The condition of reversibility is called the *detailed balance* condition in the statistical mechanics literature. In practice, we usually know π up to a normalizing

constant and we wish to construct an ergodic kernel P for which π is the invariant distribution. Given π, the kernel P is not uniquely determined and detailed balance is often imposed as an additional condition because it is more conveniently verified than the invariance of π.

The well-known *Metropolis algorithm* does this: it starts with a base chain on the relevant state space and modifies it to construct a chain which is reversible with respect to the target distribution π Diaconis and Saloff-Coste, 1995. In applications, the state space $\mathcal{X}$ is often a huge set and the stationary distribution π is given by $\pi(x) \propto e^{-H(x)}$ with $H(x)$ easy to calculate. The unspecified normalizing constant is usually impossible to compute. The Metropolis chain is constructed in such a way that this constant cancels out in the simulation of the local moves of the chain. See Diaconis and Saloff-Coste, 1995 for a detailed analysis of the Metropolis algorithm.

For an ergodic finite state space Markov chain, the invariant distribution π is strictly positive on the state space $\mathcal{X}$. Let $\ell^2(\Pi)$ denote the space of functions $f : \mathcal{X} \to \mathbb{R}$ for which

$$\|f\|_\pi^2 = \sum_{x \in \mathcal{X}} |f(x)|^2 \pi(x) < \infty.$$

Then $\ell^2(\Pi)$ is a Hilbert space with the inner product

$$\langle f, g \rangle_\pi = \sum_{x \in \mathcal{X}} f(x) g(x) \pi(x), \quad \forall f, g \in \ell^2(\Pi).$$

A necessary and sufficient condition for a transition matrix P to be reversible with respect to π is that P is self-adjoint in $\ell^2(\Pi)$ Bremaud, 1999, Chap. 6 Thoerem 2.1. An immediate conclusion is that all the eigenvalues of P are real and that the eigenvectors form an orthonormal basis in $\ell^2(\Pi)$. For any observable $f : \mathcal{X} \to \mathbb{R}$, the following result (Bremaud, 1999, Equation (7.18)) gives an upper bound for $\tau_{\mathrm{int},f}$ in terms of the eigenvalues of P.

Proposition 2.3. *Let (P, π) be a reversible Markov chain on a finite set $\mathcal{X}$. Let P be ergodic with eigenvalues $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_N > -1$ and let the corresponding eigenvectors be*

$v_1, v_2, \ldots, v_N$. *The integrated autocorrelation time for* f *satisfies the bound*

$$\tau_{\mathrm{int},f} \leq \frac{1+\lambda_*}{1-\lambda_*} = \frac{2}{\alpha} - 1, \tag{2.1.1}$$

where the reduced spectral radius $\lambda_* = \max(\lambda_2, -\lambda_N)$ *and the spectral gap* $\alpha = 1 - \lambda_*$. *In the bound above, equality occurs if and only if* $f = v_2$ *if* $\lambda_* = \lambda_2$ *and* $f = v_N$ *if* $\lambda_* = -\lambda_N$.

From the result above, it is clear that λ_* gives a tight upper bound on the convergence time of any observable if it is measured in terms of the autocorrelation time. From (2.0.4), for $T \gg \tau_{\mathrm{int},f}$, we have

$$\mathrm{Var}[\hat{f}_T] \approx \frac{1}{T}\tau_{\mathrm{int},f} C_f(0) \leq \frac{1}{T} C_f(0) \left(\frac{2}{\alpha} - 1\right).$$

For a particular Markov chain, if we require an estimate of the worst autocorrelation time for any observable, then estimation of λ_* is one approach. In the next section, we outline one such method for estimating λ_* which uses data from the simulation of the chain. In section 4 and 6, we apply it to estimate the spectral gap of various model problems.

3

ESTIMATION OF λ_* FOR REVERSIBLE CHAINS USING SIMULATION DATA

Before we describe our approach to estimate λ_* using simulation data, we need to answer the first obvious question that arises: why can't we use linear algebra techniques to compute λ_*? This is because the size of the Markov chains that occur in practice are usually exponentially large. For instance, in the Metropolis chain for the simulation of 2D Ising model, the number of (spin) states is 2^{n^2} if we start with an $n \times n$ grid (each spin can take two possible values: -1 or $+1$). The size of state space is still finite, but for all practical purposes one can consider it as infinite – we can evaluate each entry of the transition matrix P separately but it is impractical to store the matrix as a whole. Even storing a vector of size 2^{n^2} for moderate values of n is impossible.

Hence we *cannot* use standard eigenvalue-locating algorithms (like Lanczos, for instance) to compute λ_* – indeed, if there were a way of performing these operations, then we would directly take the inner product $\langle f, \pi \rangle$ to evaluate $\langle f \rangle_\pi = E_\pi[f]$ instead of going through the elaborate procedure of building a Markov chain to estimate $\langle f \rangle_\pi$!

We use the word "overlap" many times, so it is important to give a precise definition. The overlap between an observable f and an eigenfuntion v is defined as:

$$\text{overlap}(f, v) = \frac{\langle f, v \rangle_\pi}{\langle f, f \rangle_\pi}. \tag{3.0.1}$$

Suppose by physical intuition or otherwise, we have an observable f which has a substantial overlap with the slowest mode. Test functions which are used to prove "good" upper bounds on the spectral gap are examples of such observables. To motivate our method of estimating λ_*, let us assume, for simplicity, that we know the expansion of f in the (orthonormal) eigenvector

basis of P, that is,

$$f = a_1 v_{k_1} + a_2 v_{k_2} + \cdots + a_m v_{k_m}, \tag{3.0.2}$$

where $m \ll N$ and $a_i \neq 0$, $1 \leq k_i \leq N$ for each $i \leq m$. The a_i are normalized such that $\langle f \rangle_\pi = 1$. The idea is that f has a nonzero component along only a handful of eigenvectors even though the eigenvector basis is exponentially large. Let v_{k_i} be ordered such that $|\lambda_{k_1}| \geq |\lambda_{k_2}| \geq \cdots \geq |\lambda_{k_m}|$. Let the slowest mode be denoted by v_* – it is either v_2 or v_N depending on whether the reduced spectral radius is λ_2 or $-\lambda_N$ respectively. Since f is assumed to have an overlap with v_*, the reduced spectral radius $\lambda_* = |\lambda_{k_1}|$. Without loss of generality, we can assume that $\langle f \rangle_\pi = E_\pi[f] = 0$ (otherwise consider the observable $f - \langle f \rangle_\pi$). The autocovariance function for f can then be written as:

$$C_f(s) = \langle f, P^s f \rangle_\pi = a_1^2 \lambda_{k_1}^s + a_2^2 \lambda_{k_2}^s + \cdots + a_m^2 \lambda_{k_m}^s. \tag{3.0.3}$$

Given a simulation run $f_t = f(X_t)$ for $t = 0, \ldots, T$ of the Markov chain using f as the observable, the problem is to estimate $|\lambda_{k_1}|$. An estimate for $C_f(s)$ is given by the following expression (Anderson, 1971, Chapter 8, (8)):

$$\widehat{C}_f(s) = \frac{1}{T - |s|} \sum_{t=0}^{T-|s|-1} (f_t - \hat{f}_T)(f_{t+s} - \hat{f}_T). \tag{3.0.4}$$

Analogous to (3.0.3), we can write the following set of equations for the estimates of λ_{k_i} and a_i in terms of the estimates for $C_f(s)$:

$$\widehat{C}_f(s) = \hat{a}_1^2 \hat{\lambda}_{k_1}^s + \hat{a}_2^2 \hat{\lambda}_{k_2}^s + \cdots + \hat{a}_m^2 \hat{\lambda}_{k_m}^s, \tag{3.0.5}$$

where for $1 \leq i \leq m$, $\hat{a}_i$, $\hat{\lambda}_{k_i}$ denote estimates for a_i, λ_{k_i} respectively. If we have the estimates $\widehat{C}_f(s)$ for $1 \leq s \leq 2m$, we have $2m$ equations in $2m$ unknowns. This problem is exactly identical to the so-called *shape-from-moments* problem in which the vertices of a planar polygon need to be recovered from its measured complex moments. The moments correspond to the

autocovariance estimates $\widehat{C}_f(s)$ and the vertices of the planar polygon to the eigenvalues λ_{k_i}. The shape-from-moments problem has connections to applications in array processing, system identification, and signal processing and has been well-studied – see Golub et al., 2000 and the literature cited therein for analysis of the case when the moments are exactly known. Schuermans et al., 2006 and Elad et al., 2004 consider the case when the given moments are noisy. Broadly, the methods proposed for the solution of the shape-from-moments problem can be classified into two categories: Prony-based and pencil-based.

3.1 Prony's method

We first describe Prony's method for the shape-from-moments problem, that is, to estimate λ_{k_i} given the autocovariance estimates $\widehat{C}_f(s)$ as in equation (3.0.5). Suppose for a moment that we know the exact autocovariance values as in equation (3.0.3). Then consider the mth degree polynomial

$$\begin{aligned} p(\lambda) &= (\lambda - \lambda_{k_1})(\lambda - \lambda_{k_2}) \dots (\lambda - \lambda_{k_m}) \\ &= b_m \lambda^m + b_{m-1}\lambda^{m-1} + \dots + b_0, \end{aligned}$$

where the coefficients b_i depend on λ_{k_i} for $i = 1, 2, \dots, m$ and $b_m = 1$. Then clearly,

$$\sum_{i=1}^{m} a_i^2 p(\lambda_{k_i}) = \sum_{s=0}^{m} b_s \left(\sum_{i=1}^{m} a_i^2 \lambda_{k_i}^s \right) = \sum_{s=0}^{m} b_s C_f(s) = 0.$$

The last equality follows because λ_{k_i} are the roots of the polynomial $p(\lambda)$, that is, $p(\lambda_{k_i}) = 0$ for $i = 1, 2, \dots, m$. If we have autocovariance estimates $C_f(s)$ for $s = 0, 1, \dots, T$ for $T > m$, then by an argument similar to the one above, we can show that

$$\sum_{s=0}^{m} b_s C_f(s + l) = 0,$$

for $l = 0, 1, \ldots, T - m$. The equation above can be written in matrix form as (recall that $b_m = 1$):

$$\begin{bmatrix} b_0 & \ldots & b_{m-1} & 1 & & 0 \\ & \ddots & & \ddots & \ddots & \\ 0 & & b_0 & \ldots & b_{m-1} & 1 \end{bmatrix} \begin{bmatrix} C_f(0) \\ C_f(1) \\ \vdots \\ C_f(T) \end{bmatrix} = 0. \tag{3.1.1}$$

Reordering the equations, we obtain

$$-\begin{bmatrix} C_f(m) \\ C_f(m+1) \\ \vdots \\ C_f(T) \end{bmatrix} = \begin{bmatrix} C_f(0) & C_f(1) & \ldots & C_f(m-1) \\ C_f(1) & C_f(2) & \ldots & C_f(m) \\ \vdots & \vdots & \ddots & \vdots \\ C_f(T-m) & C_f(T-m+1) & \ldots & C_f(T-1) \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{m-1} \end{bmatrix}.$$

Since we have $T - m + 1$ equations in m unknowns, requiring $T \geq 2m - 1$ leads to an overdetermined but consistent system of equations. The system of equations can be written as

$$-w = Wb. \tag{3.1.2}$$

From the equation above, b can be computed as

$$b = -W^{+}w,$$

where W^{+} is the Moore-Penrose pseudo-inverse of W. Once b is obtained, the eigenvalues λ_{k_i} can be found by computing the roots of the polynomial

$$p(\lambda) = \prod_{i=1}^{m}(\lambda - \lambda_{k_i}) = \lambda^m + \sum_{i=0}^{m-1} b_i \lambda^i.$$

The root-finding problem can be converted to an eigenvalue problem by using the companion matrix method.

Since we do not have the exact autocovariance values, but only their estimates, an alternate method called the total-least-squares (TLS) Prony has been suggested in Elad et al., 2004. The basic idea is that in the presence of noise, equation (3.1.2) does not hold exactly but the matrix $[W \quad w]$ (the matrix W padded by an additional column w) is expected to be nearly singular. Hence the TLS problem is solved using the singular value decomposition (SVD) – basically the estimate $\hat{b}$ of the vector b is the right singular vector corresponding to the smallest singular value. This is normalized so that the last entry $\hat{b}_m = 1$.

The main obstacle to the application of the TLS Prony method is that it requires a knowledge of m, the number of exponentials, λ_{k_i}, that represent the autocovariance numbers. It is usually not known in practice, since the eigenvectors v_{k_i} in the representation of f are not known. Since we are interested in only $|\lambda_{k_1}|$, the largest modulus eigenvalue, we can use the following heuristic: we *assume* a particular value for m, apply the TLS Prony method and return the estimate $\hat{\lambda}_{k_i}$ with the largest modulus. But in practice, Prony-based methods are very ill-conditioned; pencil-based methods, the most notable being the Generalized Pencil of Function (GPOF) method Hua and Sarkar, 1990, are considered better numerical methods.

The GPOF method also requires a knowledge of m, but we now show that there is a connection between the GPOF method and a *Krylov subspace* method, which indicates that this heuristic can work well for reasonable *assumed* values of m with a judicious choice of f – in essence, that the choice of m may not matter that much if we are interested in only estimating $|\lambda_{k_1}|$, the largest modulus eigenvalue.

3.2 Krylov subspace algorithm for estimating spectral gap

Consider the Krylov subspace of dimension n generated by the matrix P and the vector f:

$$\mathcal{K}_n[f] = \text{span}\{f, Pf, P^2 f, \ldots, P^{n-1} f\}. \tag{3.2.1}$$

If f has a substantial overlap with the slowest mode v_*, then presumably v_* can be well-approximated by a vector in $\mathcal{K}_n[f]$ for $n \ll N$ even though m is unknown. For any $u \in \mathcal{K}_n[f]$, we can write

$$u = \sum_{j=1}^{n} \xi_j P^{j-1} f,$$

for some $\xi_1, \ldots, \xi_n \in \mathbb{R}$. Then the *Rayleigh quotient* for u is:

$$q(u) = \frac{\langle u, Pu \rangle_\pi}{\langle u, u \rangle_\pi}. \tag{3.2.2}$$

The denominator of this expression is

$$\begin{aligned}
\langle u, u \rangle_\pi &= \sum_{i,j=1}^{n} \xi_i \xi_j \langle P^{i-1} f, P^{j-1} f \rangle_\pi \\
&= \sum_{i,j=1}^{n} \xi_i \xi_j \langle f, P^{i+j-2} f \rangle_\pi \\
&= \sum_{i,j=1}^{n} \xi_i \xi_j E_\pi[f(X_0) f(X_{i+j-2})] \\
&= \sum_{i,j=1}^{n} \xi_i \xi_j \mathrm{cov}_\pi[f(X_0), f(X_{i+j-2})] \\
&= \sum_{i,j=1}^{n} \xi_i \xi_j C_f(i+j-2),
\end{aligned} \tag{3.2.3}$$

where $C_f(i+j-2)$ is the autocovariance function for the observable f at lag $i+j-2$. Similar to equation (3.2.3), we can write $\langle u, Pu \rangle_\pi$ as:

$$\langle u, Pu \rangle_\pi = \sum_{i,j=1}^{n} \xi_i \xi_j C_f(i+j-1). \tag{3.2.4}$$

If we form $n \times n$ matrices A and B with entries $A(i,j) = C_f(i+j-1)$ and $B(i,j) = C_f(i+j-2)$, then substituting (3.2.4) and (3.2.3) into (3.2.2) yields

$$q(u) = \frac{\langle \xi, A\xi \rangle}{\langle \xi, B\xi \rangle}. \tag{3.2.5}$$

If a small perturbation of v_* lies in the subspace $\mathcal{K}_n[f]$, then

$$\lambda_* \approx \max_{\xi \neq 0} \left| \frac{\langle \xi, A\xi \rangle}{\langle \xi, B\xi \rangle} \right|. \tag{3.2.6}$$

The right hand side of (3.2.6) is the largest generalized eigenvalue modulus for the generalized eigenvalue problem $A\xi = \lambda B\xi$. In other words, the problem of estimating λ_* has been reduced to a generalized eigenvalue problem involving matrices of much smaller size with a judicious choice of the observable f.

Now we are ready to describe the first version of the *Krylov Subspace Pencil (KSP)* algorithm for estimating λ_* for a reversible ergodic Markov chain:

1. Choose an observable f with a sizable overlap with the slowest mode of the Markov chain.

2. Start with a random initial state X_0 and simulate the chain for a "long time" T with f as the observable, that is, collect samples $f(X_0), f(X_1), \ldots, f(X_T)$.

3. Choose a small number n, say around 10, and estimate the autocovariance function for f at lags $s = 0, 1, \ldots, 2n-1$ using the expression (3.0.4).

4. Form the matrices

$$\widehat{A} = \begin{bmatrix} \widehat{C}_f(1) & \widehat{C}_f(2) & \dots & \widehat{C}_f(n) \\ \widehat{C}_f(2) & \widehat{C}_f(3) & \dots & \widehat{C}_f(n+1) \\ \widehat{C}_f(3) & \widehat{C}_f(4) & \dots & \widehat{C}_f(n+2) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{C}_f(n) & \widehat{C}_f(n+1) & \dots & \widehat{C}_f(2n-1) \end{bmatrix},$$

$$\widehat{B} = \begin{bmatrix} \widehat{C}_f(0) & \widehat{C}_f(1) & \dots & \widehat{C}_f(n-1) \\ \widehat{C}_f(1) & \widehat{C}_f(2) & \dots & \widehat{C}_f(n) \\ \widehat{C}_f(2) & \widehat{C}_f(3) & \dots & \widehat{C}_f(n+1) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{C}_f(n-1) & \widehat{C}_f(n) & \dots & \widehat{C}_f(2n-2) \end{bmatrix}. \tag{3.2.7}$$

5. Return the largest generalized eigenvalue modulus (LGEM) of the pencil $\widehat{A} - \lambda\widehat{B}$ as the estimate for λ_*.

We have intentionally left the description of the algorithm vague – we have more to say about how to choose the observable f, the run length T, and the dimension of the Krylov subspace n.

Given an observable $f : \mathcal{X} \to \mathbb{R}$, for matrices A, B given by

$$
\begin{aligned}
A &= \begin{bmatrix} C_f(1) & C_f(2) & \dots & C_f(n) \\ C_f(2) & C_f(3) & \dots & C_f(n+1) \\ C_f(3) & C_f(4) & \dots & C_f(n+2) \\ \vdots & \vdots & \ddots & \vdots \\ C_f(n) & C_f(n+1) & \dots & C_f(2n-1) \end{bmatrix}, \\
B &= \begin{bmatrix} C_f(0) & C_f(1) & \dots & C_f(n-1) \\ C_f(1) & C_f(2) & \dots & C_f(n) \\ C_f(2) & C_f(3) & \dots & C_f(n+1) \\ \vdots & \vdots & \ddots & \vdots \\ C_f(n-1) & C_f(n) & \dots & C_f(2n-2) \end{bmatrix},
\end{aligned}
\tag{3.2.8}
$$

equation (3.2.5) shows that any generalized eigenvalue of the pencil $A - \lambda B$ is the Rayleigh quotient of a vector $u \in \mathcal{K}_n[f]$. Since the eigenvalues of the transition matrix P all lie in the interval $(-1, 1)$, this implies that all the generalized eigenvalues of the pencil $A - \lambda B$ should lie in the interval $(-1, 1)$.

The matrices A and B above have the so-called Hankel structure; it is a well-known fact that real Hankel matrices in general can be severely ill-conditioned Tyrtyshnikov, 1994. In fact, if we know the representation (3.0.2) of f in the eigenvector basis of P, we choose the pencil size $n = m$ and can write the following representations for the $n \times n$ matrices A and B (Zamarashkin and Tyrtyshnikov, 2001):

$$
\begin{aligned}
A &= V_n \, \mathrm{Diag}(a_1^2 \lambda_{k_1}, a_2^2 \lambda_{k_2}, \dots, a_n^2 \lambda_{k_n}) \, V_n^t, \\
B &= V_n \, \mathrm{Diag}(a_1^2, a_2^2, \dots, a_n^2) \, V_n^t,
\end{aligned}
\tag{3.2.9}
$$

where V_n is the Vandermonde matrix of λ_{k_i}:

$$V_n = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \lambda_{k_1} & \lambda_{k_2} & \dots & \lambda_{k_n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{k_1}^{n-1} & \lambda_{k_2}^{n-1} & \dots & \lambda_{k_n}^{n-1} \end{bmatrix}. \tag{3.2.10}$$

Equation (3.2.9) can be easily derived as follows:

$$\begin{aligned} V_n \ \mathrm{Diag}(a_1^2, a_2^2, \dots, a_n^2) \ V_n^t &= \begin{bmatrix} a_1^2 & a_2^2 & \dots & a_n^2 \\ a_1^2\lambda_{k_1} & a_2^2\lambda_{k_2} & \dots & a_n^2\lambda_{k_n} \\ \vdots & \vdots & \ddots & \vdots \\ a_1^2\lambda_{k_1}^{n-1} & a_2^2\lambda_{k_2}^{n-1} & \dots & a_n^2\lambda_{k_n}^{n-1} \end{bmatrix} \begin{bmatrix} 1 & \lambda_{k_1} & \dots & \lambda_{k_1}^{n-1} \\ 1 & \lambda_{k_2} & \dots & \lambda_{k_2}^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_{k_n} & \dots & \lambda_{k_n}^{n-1} \end{bmatrix} \\ &= [X_{i,j}]_{n\times n}, \end{aligned}$$

where for $i, j = 1, 2, \dots, n$,

$$\begin{aligned} X_{i,j} &= \begin{bmatrix} a_1^2\lambda_{k_1}^{i-1} & a_2^2\lambda_{k_2}^{i-1} \dots & a_n^2\lambda_{k_n}^{i-1} \end{bmatrix} \begin{bmatrix} \lambda_{k_1}^{j-1} \\ \lambda_{k_2}^{j-1} \\ \vdots \\ \lambda_{k_n}^{j-1} \end{bmatrix} \\ &= a_1^2\lambda_{k_1}^{i+j-1} + a_2^2\lambda_{k_2}^{i+j-1} + \dots + a_n^2\lambda_{k_n}^{i+j-1} \\ &= C_f(i+j-1), \end{aligned}$$

from equation (3.0.3). Thus the matrix $X = B$. We can similarly prove the equation for matrix A in (3.2.9).

A lower bound on the condition number of B has been proved in Zamarashkin and Tyrtyshnikov, 2001 using (3.2.9). The idea is to consider the matrix $K_n = V_n \ \mathrm{Diag}(a_1, a_2, \dots, a_n)$; the matrix B can then be written as $B = K_n K_n^t$. The following relation has been proved in

Zamarashkin and Tyrtyshnikov, 2001:

$$\mathrm{cond}_2(K_n) \geq 2^{n-2}\left(\frac{2}{d-c}\right)^{n-1},$$

where $c = \min(\lambda_{k_1}, \lambda_{k_2}, \ldots, \lambda_{k_n})$ and $d = \max(\lambda_{k_1}, \lambda_{k_2}, \ldots, \lambda_{k_n})$. The lower bound on the condition number of B is then

$$\mathrm{cond}_2(B) \geq 2^{2n-4}\left(\frac{2}{d-c}\right)^{2n-2}. \tag{3.2.11}$$

Since $\lambda_{k_i}, i = 1, \ldots, n$ are all between -1 and 1, $d - c \leq 2$, in which case $\mathrm{cond}_2(B) \geq 2^{2n-4}$ for any distribution of λ_{k_i} – the condition number of B is much worse when they are all clustered.

Since B is most likely very ill-conditioned, how does this affect the generalized eigenvalues of the pencil $A - \lambda B$? In particular, are the generalized eigenvalues of $\widehat{A} - \lambda\widehat{B}$ "close enough" to the generalized eigenvalues of the actual pencil $A - \lambda B$? This question is important because the actual matrices A and B are unknown – we only have their estimates $\widehat{A}, \widehat{B}$ from the simulation run. Also, in step 5 of the KSP algorithm that we outlined in the last section, we return the LGEM of $\widehat{A} - \lambda\widehat{B}$ as the estimate for λ_*. We now show that indeed the generalized eigenvalues of $A - \lambda B$ are in most cases very ill-conditioned and naively returning the LGEM of $\widehat{A} - \lambda\widehat{B}$ might well result in a value greater than 1, which is impossible as an estimate for λ_*. A more sophisticated procedure should hence be designed to estimate λ_*.

To show that the generalized eigenvalues of $A - \lambda B$ may be ill-conditioned, we first make the observation that the matrix B is always positive semidefinite, that is, $B \succeq 0$. This is easy to see from (3.2.3), since for any $\xi \in \mathbb{R}^n$, we can write

$$\langle \xi, B\xi \rangle = \sum_{i,j}^{n} \xi_i \xi_j C_f(i + j - 2) = \langle u, u \rangle_\pi, \tag{3.2.12}$$

where $u = \sum_{i=1}^{n} \xi_i P^{i-1} f$. Since $\langle u, u \rangle_\pi \geq 0$ for any $u \in \mathbb{R}^N$, it follows that $B \succeq 0$.

Also, note that B is singular if $m < n$ (recall that n denotes the size of square matrices A, B and m denotes the minimal number of eigenvectors of P as a linear combination of which

one can represent f; see equation (3.0.2)). This is easy to see because then the dimension of the Krylov subspace $\mathcal{K}_n[f]$ is $m < n$, that is, each $P^{i-1}f$ can be represented as a linear combination of m vectors; we can therefore find a non-zero $\xi \in \mathbb{R}^n$ such that $\sum_{i=1}^{n} \xi_i P^{i-1} f = 0$. Equation (3.2.12) then asserts that B is singular. If $m < n$, A is singular as well and in fact has a common $n - m$ dimensional null space with B. The estimates $\widehat{A}$ and $\widehat{B}$ are noisy and hence not exactly singular but have an $n - m$ dimensional subspace with "tiny" eigenvalues (if $\widehat{A}$ and $\widehat{B}$ are accurate enough). If ξ is in this subspace, then the ratio $\langle \xi, \widehat{A}\xi \rangle / \langle \xi, \widehat{B}\xi \rangle$ is pure noise. So we have to extract out this subspace somehow before returning the LGEM of $\widehat{A} - \lambda \widehat{B}$.

If $n \geq m$ and all the λ_{k_i} are distinct, then B and A are nonsingular. Let us now examine the conditioning of the generalized eigenvalues of A and B. To simplify the analysis, let us assume that $n = m$. Let $\xi_1, \xi_2, \ldots, \xi_n$ be the generalized eigenvectors of A and B corresponding to generalized eigenvalues $\mu_1, \mu_2, \ldots, \mu_n$. If we perturb A and B slightly, we can write the first-order perturbation equation for μ_j – first-order perturbation theory for eigenvalues is well-known and we adopt the following notation from Beckermann et al., 2007:

$$(A + \epsilon \tilde{A})\xi_j(\epsilon) = \mu_j(\epsilon)(B + \epsilon \tilde{B})\xi_j(\epsilon),$$

where $\tilde{A}, \tilde{B}$ are normalized such that $\|\tilde{A}\| \leq 1$, $\|\tilde{B}\| \leq 1$ and $\epsilon > 0$ is small. In a small neighborhood around a simple eigenvalue $\mu_j(0) = \mu_j$ with eigenvector $\xi_j(0) = \xi_j$, the function $\epsilon \mapsto \mu_j(\epsilon)$ is differentiable and has the following derivative at $\epsilon = 0$:

$$\frac{d\mu_j}{d\epsilon}(0) = \frac{\langle \xi_j, (\tilde{A} - \mu_j \tilde{B})\xi_j \rangle}{\langle \xi_j, B\xi_j \rangle}. \tag{3.2.13}$$

From the representation of A and B given by (3.2.9), it is easy to see that $\mu_j = \lambda_{k_j}$ and $\xi_j = V_n^{-t} e_j$ for $1 \leq j \leq n$. Substituting these into the equation above yields

$$\frac{d\mu_j}{d\epsilon}(0) = \frac{\langle V_n^{-t} e_j, (\tilde{A} - \mu_j \tilde{B}) V_n^{-t} e_j \rangle}{\langle V_n^{-t} e_j, B V_n^{-t} e_j \rangle} = \frac{\langle V_n^{-t} e_j, (\tilde{A} - \mu_j \tilde{B}) V_n^{-t} e_j \rangle}{a_j}. \tag{3.2.14}$$

This equation is identical to Beckermann et al., 2007, Equation (8). The conditioning of the eigenvalue μ_j hence depends on a_j and the norm $\|V_n^{-t}e_j\|$. We are interested in only μ_1 (the one with the largest modulus among μ_j) and since the observable f is assumed to be chosen with a substantial overlap with the slowest mode v_*, a_1 can be assumed to be "large". Let's now consider the term $V_n^{-t}e_j$. If V_n is as shown in (3.2.10), the matrix V_n^{-t} has the form $V_n^{-t} = UL$, where the matrices U and L are given by (Turner, 1966):

$$U = \begin{bmatrix} 1 & -\lambda_{k_1} & \lambda_{k_1}\lambda_{k_2} & -\lambda_{k_1}\lambda_{k_2}\lambda_{k_3} & \cdots \\ 0 & 1 & -(\lambda_{k_1}+\lambda_{k_2}) & \lambda_{k_1}\lambda_{k_2}+\lambda_{k_2}\lambda_{k_3}+\lambda_{k_3}\lambda_{k_1} & \cdots \\ 0 & 0 & 1 & -(\lambda_{k_1}+\lambda_{k_2}+\lambda_{k_3}) & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

$$L = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ \frac{1}{\lambda_{k_1}-\lambda_{k_2}} & \frac{1}{\lambda_{k_2}-\lambda_{k_1}} & 0 & \cdots \\ \frac{1}{(\lambda_{k_1}-\lambda_{k_2})(\lambda_{k_1}-\lambda_{k_3})} & \frac{1}{(\lambda_{k_2}-\lambda_{k_1})(\lambda_{k_2}-\lambda_{k_3})} & \frac{1}{(\lambda_{k_3}-\lambda_{k_1})(\lambda_{k_3}-\lambda_{k_2})} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (3.2.15)$$

It has been observed experimentally that when the λ_{k_i} are clustered together, the estimation of λ_{k_1} is a very ill-conditioned problem. The structure of V_n^{-t} from (3.2.15) gives an intuition of why this happens. If indeed λ_{k_i} are clustered together, the product terms $(\lambda_{k_1}-\lambda_{k_2})(\lambda_{k_1}-\lambda_{k_3})$ etc. are small, making $\|Le_1\|$ large – this could potentially lead to $\|V_n^{-t}e_1\|$ being large. From (3.2.14), it then follows that the estimation of $\mu_1 = \lambda_{k_1}$ is an ill-conditioned problem. We do not claim that this is a rigorous proof – the question of giving the exact conditions on the distribution of λ_{k_i} under which we get an ill-conditioned problem of estimating λ_{k_1} is not easy to answer; all we intend to give here is an intuitive explanation for an experimentally observed fact.

If the estimation of λ_{k_1} is indeed ill-posed, the alternative is to use a value of n smaller than m and expect that the LGEM of the $n \times n$ pencil $\widehat{A}_n - \lambda\widehat{B}_n$ is close to the LGEM of the

actual $m \times m$ pencil. In fact, in the degenerate case, choosing $n = 1$ yields $\widehat{A}_1 = \widehat{C_f}(1) = \hat{a}_1^2\hat{\lambda}_{k_1} + \hat{a}_2^2\hat{\lambda}_{k_2} + \cdots + \hat{a}_m^2\hat{\lambda}_{k_m}$, $\widehat{B}_1 = \widehat{C_f}(0) = \hat{a}_1^2 + \hat{a}_2^2 + \cdots + \hat{a}_m^2$. If the eigenvalues λ_{k_i} are all clustered together, then the LGEM of the pencil $\widehat{A} - \lambda\widehat{B}$,

$$|\hat{\mu}_1| = \frac{|\widehat{C_f}(1)|}{\widehat{C_f}(0)} = \frac{|\hat{a}_1^2\hat{\lambda}_{k_1} + \hat{a}_2^2\hat{\lambda}_{k_2} + \cdots + \hat{a}_m^2\hat{\lambda}_{k_m}|}{\hat{a}_1^2 + \hat{a}_2^2 + \cdots + \hat{a}_m^2}$$

is a good approximation of $|\lambda_{k_1}|$. Except in this degenerate case, it is not immediately clear how good an approximation $|\hat{\mu}_1|$ is of $|\lambda_{k_1}|$ if $n < m$. A simple symbolic manipulation experiment has been performed in `Maple` to see how the LGEM of the 2×2 pencil $\widehat{A}_2 - \lambda\widehat{B}_2$ compares with $|\lambda_{k_1}|$ when $m = 3$, that is, when

$$\begin{aligned}
\widehat{A}_2 &= \begin{bmatrix} \hat{a}_1^2\hat{\lambda}_{k_1} + \hat{a}_2^2\hat{\lambda}_{k_2} + \hat{a}_3^2\hat{\lambda}_{k_3} & \hat{a}_1^2\hat{\lambda}_{k_1}^2 + \hat{a}_2^2\hat{\lambda}_{k_2}^2 + \hat{a}_3^2\hat{\lambda}_{k_3}^2 \\ \hat{a}_1^2\hat{\lambda}_{k_1}^2 + \hat{a}_2^2\hat{\lambda}_{k_2}^2 + \hat{a}_3^2\hat{\lambda}_{k_3}^2 & \hat{a}_1^2\hat{\lambda}_{k_1}^3 + \hat{a}_2^2\hat{\lambda}_{k_2}^3 + \hat{a}_3^2\hat{\lambda}_{k_3}^3 \end{bmatrix} \\
\widehat{B}_2 &= \begin{bmatrix} \hat{a}_1^2 + \hat{a}_2^2 + \hat{a}_3^2 & \hat{a}_1^2\hat{\lambda}_{k_1} + \hat{a}_2^2\hat{\lambda}_{k_2} + \hat{a}_3^2\hat{\lambda}_{k_3} \\ \hat{a}_1^2\hat{\lambda}_{k_1} + \hat{a}_2^2\hat{\lambda}_{k_2} + \hat{a}_3^2\hat{\lambda}_{k_3} & \hat{a}_1^2\hat{\lambda}_{k_1}^2 + \hat{a}_2^2\hat{\lambda}_{k_2}^2 + \hat{a}_3^2\hat{\lambda}_{k_3}^3 \end{bmatrix}.
\end{aligned}$$

Even in this simple case, the symbolic expressions for the generalized eigenvalues of the pencil $\widehat{A}_2 - \lambda\widehat{B}_2$ are very complicated – it is difficult to find conditions on $\hat{a}_i$ and $\hat{\lambda}_{k_i}$ under which using $n < m$ leads to a good approximation for $|\lambda_{k_1}|$. Moreover, the value of m is never known in practice – all we have are the autocovariance estimates $\widehat{C}_f(s)$ for $s = 0, 1, \ldots$. Fortunately, it turns out the choice of n is less important than the choice of another parameter called the *lag parameter*. We now define what we mean by the lag parameter and give a justification of why it is more important than the choice of n, which is henceforth termed the *pencil size parameter*.

3.2.1 Choice of the lag parameter

From equations (3.2.7) and (3.2.9), it is clear that the ill-conditioning of the Hankel matrices prevents us from using a large value for the pencil size parameter n; the value $n = 10$ is one of the largest we can manage. A look at (3.2.7) reveals that the largest lag autocovariance estimate

that we then consider is $\widehat{C}_f(2n-1)$. If we use $n = 10$, then we use only the autocovariance estimates for lags $s = 0, 1, \ldots, 19$ to estimate the spectral gap, leaving the remaining estimates unused. The other and more important drawback of this approach is this: if the spectral gap of the chain is very small and the observable f has a substantial overlap with the slowest mode, then it is highly likely that Pf is not very different from f; the basis corresponding to the Krylov subspace $\mathcal{K}_n[f]$ for small n is close to being degenerate and the problem of estimating λ_{k_1} from this basis is very ill-conditioned.

To get around this problem, what we can instead do is to consider an alternate Markov chain with transition matrix P^r for some $r > 1$ – this parameter r is the so-called lag parameter. The reduced spectral radius for this chain is λ_*^r; if $\hat{\mu}$ is an estimate of it, then an estimate of λ_*, the reduced spectral radius of the original chain, is $(\hat{\mu})^{1/r}$. In estimating λ_*^r, instead of using the Krylov subspace $\mathcal{K}_n[f]$ from (3.2.1), we use the subspace:

$$\mathcal{K}_{n,r}[f] = \text{span}\{f, P^r f, P^{2r} f, \ldots, P^{(n-1)r} f\}. \tag{3.2.16}$$

The matrices $\widehat{A}$ and $\widehat{B}$ then take the form:

$$\widehat{A}_{n,r} = \begin{bmatrix} \widehat{C}_f(r) & \widehat{C}_f(2r) & \ldots & \widehat{C}_f(nr) \\ \widehat{C}_f(2r) & \widehat{C}_f(3r) & \ldots & \widehat{C}_f((n+1)r) \\ \widehat{C}_f(3r) & \widehat{C}_f(4r) & \ldots & \widehat{C}_f((n+2)r) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{C}_f(nr) & \widehat{C}_f((n+1)r) & \ldots & \widehat{C}_f((2n-1)r) \end{bmatrix},$$

$$\widehat{B}_{n,r} = \begin{bmatrix} \widehat{C}_f(0) & \widehat{C}_f(r) & \ldots & \widehat{C}_f((n-1)r) \\ \widehat{C}_f(r) & \widehat{C}_f(2r) & \ldots & \widehat{C}_f(nr) \\ \widehat{C}_f(2r) & \widehat{C}_f(3r) & \ldots & \widehat{C}_f((n+1)r) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{C}_f((n-1)r) & \widehat{C}_f(nr) & \ldots & \widehat{C}_f((2n-2)r) \end{bmatrix}. \tag{3.2.17}$$

The basic idea is that in the representation of f as in (3.0.2), even if the eigenvalues λ_{k_i} are clustered, their powers $\lambda_{k_i}^r$, for an appropriate choice of r are well-separated, and hence the generalized eigenvalue problem for the pencil $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$ is likely to be better conditioned for $r > 1$ than for $r = 1$.

What is an appropriate value for the lag parameter r? Too small a value does not make a significant difference from $r = 1$, while if r is chosen to be large, the entries in $\widehat{A}_{n,r}$ and $\widehat{B}_{n,r}$ are likely to be very noisy. The lag parameter needs to be chosen with care depending on how slow the observable f is. One measure of the slowness of f is its integrated autocorrelation time $\tau_{\mathrm{int},f}$ as given by (2.0.3). From (2.0.3), it seems that the natural estimator for $\tau_{\mathrm{int},f}$ is:

$$\hat{\tau}_{\mathrm{int},f} = 1 + 2\sum_{t=1}^{T} \widehat{C}_f(t)/\widehat{C}_f(0).$$

But as pointed out in Sokal, 1989, the variance of this estimator does not go to zero as the sample time $T \to \infty$. This is because for large t, $\widehat{C}_f(t)$ has much noise and little signal; if we add several of these terms, Var$[\hat{\tau}_{\mathrm{int},f}]$ does not go to zero as $T \to \infty$ – it's an inconsistent estimator in the terminology of statistics Anderson, 1971. One way of dealing with this problem is to "cut off" at some t so as to retain only the signal, while disregarding the noise. An automatic windowing algorithm has been proposed in Sokal, 1989, which we describe here.

procedure $TauEstimate(\widehat{C}_f(0..T), c)$

1: $M \leftarrow 1$

2: $\tau \leftarrow 1$

3: **repeat**

4: $\quad \tau \leftarrow \tau + \frac{2\widehat{C}_f(M)}{\widehat{C}_f(0)}$

5: $\quad M \leftarrow M + 1$

6: **until** $M > c\tau$

7: return τ

In this procedure, the cut-off window M is chosen in a self-consistent fashion; more precisely, M is chosen to be the smallest integer such that $M \geq c\hat{\tau}_{\mathrm{int},f}$. There is no rigorous analysis of this self-consistent procedure, but it is known to work well in practice if we have lots of data, like for instance, $T \gtrsim 1000\hat{\tau}_{\mathrm{int},f}$ (Sokal, 1989). The value of the consistency parameter c is typically chosen to be around 8; if $\hat{\rho}_f(t) = \widehat{C}_f(t)/\widehat{C}_f(0)$ were roughly a pure exponential, then $c = 4$ should suffice, since the relative error is then $e^{-c} = e^{-4} \approx 2\%$. But since $\rho_f(t)$ is expected to have asymptotic or pre-asymptotic decay slower than pure exponential, a slightly higher value of c is necessary Sokal, 1989.

To estimate $\hat{\tau}_{\mathrm{int},f}$ using the self-consistent procedure described above, we use the autocovariance estimates $\widehat{C}_f(t)$ for lags $t = 0, 1, \ldots, c\hat{\tau}_{\mathrm{int},f}$. Given the pencil size parameter n and the consistency parameter c, our heuristic for choosing the lag parameter r is this: choose r to be the maximum integer such that no estimate of $\widehat{C}_f(t)$ is used beyond lag $t = c\hat{\tau}_{\mathrm{int},f}$. A look at (3.2.17) reveals that the maximum autocovariance estimate we use to form the matrices $\widehat{A}_{n,r}$ and $\widehat{B}_{n,r}$, in terms of n and r, is $\widehat{C}_f((2n-1)r)$. If we equate $(2n-1)r$ to $c\hat{\tau}_{\mathrm{int},f}$, we get the following equation for the lag parameter:

$$r = \left\lfloor \frac{c\hat{\tau}_{\mathrm{int},f}}{2n-1} \right\rfloor . \tag{3.2.18}$$

3.2.2 Estimating the LGEM of the pencil $\widehat{A}_{n,r} - \lambda\widehat{B}_{n,r}$

Once the pencil size parameter n and the lag parameter r are determined, the next step is to determine the LGEM $\hat{\mu}_{n,r}$ of the pencil $\widehat{A}_{n,r} - \lambda\widehat{B}_{n,r}$. An estimate of the reduced spectral radius of the original chain is then $\hat{\lambda}_* = (\hat{\mu}_{n,r})^{1/r}$. We mentioned previously that the choice of the pencil size parameter n is less important than the choice of the lag parameter r. If n is chosen to be a fixed number, say 10, what happens in the case $m < n$ (recall that m is the number of eigenvectors of the transition matrix P in terms of which the observable f can be represented)?

In that case, as we noted in the last section, the actual matrices $A_{n,r}$ and $B_{n,r}$ have a null space of dimension $n-m$. The noisy estimates $\widehat{A}_{n,r}$ and $\widehat{B}_{n,r}$ may not have a common null space but only eigenvalues which are highly ill-conditioned.

In practice, m is not known, so we need a way of rejecting the "bad" set of generalized eigenvalues and look for the LGEM only among the "good" set. If $A_{n,r}$ and $B_{n,r}$ were known exactly, then the singular structure of the pencil $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$ can be figured out by computing its Kronecker Canonical Form (KCF). For given matrices A and B, there exist matrices U and V such that

$$U^{-1}(A - \lambda B)V = S - \lambda T, \tag{3.2.19}$$

where $S = \mathrm{Diag}(S_1, S_2, \ldots, S_b)$ and $T = \mathrm{Diag}(T_1, T_2, \ldots, T_b)$ are block diagonal. Each block $S_i - \lambda T_i$ must be one of the following forms: $J_j(\alpha), N_j, L_j$, or L_j^T, where

$$J_j(\alpha) = \begin{bmatrix} \alpha - \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \alpha - \lambda \end{bmatrix}, \text{ and } N_j(\alpha) = \begin{bmatrix} 1 & -\lambda & & \\ & \ddots & \ddots & \\ & & \ddots & -\lambda \\ & & & 1 \end{bmatrix},$$

that is, $J_j(\alpha)$ is a Jordan block of size $j \times j$ corresponding to the finite generalized eigenvalue α, while N_j corresponds to an infinite generalized eigenvalue of multiplicity j. The $J_j(\alpha)$ and $N_j(\alpha)$ constitute the *regular structure* of the pencil. The other two types of diagonal blocks are

$$L_j(\alpha) = \begin{bmatrix} -\lambda & 1 & & \\ & \ddots & \ddots & \\ & & -\lambda & 1 \end{bmatrix}, \text{ and } L_j^T(\alpha) = \begin{bmatrix} -\lambda & & \\ 1 & \ddots & \\ & \ddots & -\lambda \\ & & 1 \end{bmatrix}.$$

The block $j \times (j+1)$ block L_j is called a *singular block of right minimal index* j. It has a one-dimensional right null space, $[1, \lambda, \lambda^2, \ldots, \lambda^j]^T$ for any λ. Similarly, the block L_j^T has

a one-dimensional left null space and is called a *singular block of left minimal index* j. The blocks L_j and L_j^T constitute the singular structure of the pencil $A - \lambda B$. The regular and singular structures constitute the Kronecker structure of the pencil. The Kronecker Canonical Form given by (3.2.19) is the generalization to a matrix pencil of the Jordan Canonical Form (JCF) of a square matrix. See Demmel and Kågström, 1993a for more details about the KCF of a matrix pencil.

Computing the KCF of a pencil is a hard problem in general because the matrices U and V reducing $A - \lambda B$ to $S - \lambda T$ may be arbitrarily ill-conditioned. We instead compute the Generalized Upper Triangular (GUPTRI) form of the pencil, which is a generalization of the Schur Canonical Form of a square matrix. The GUPTRI form of a pencil $A - \lambda B$ is given by

$$U^*(A - \lambda B)V = \begin{bmatrix} A_{\rm r} - \lambda B_{\rm r} & * & * \\ 0 & A_{\rm reg} - \lambda B_{\rm reg} & * \\ 0 & 0 & A_{\rm l} - \lambda B_{\rm l} \end{bmatrix},$$

where the matrices U and V are unitary and $*$ denote arbitrary conforming matrices. The block $A_{\rm r} - \lambda B_{\rm r}$ contains only right singular blocks in its KCF; indeed, the same L_j blocks as in the KCF of $A - \lambda B$. Similarly, the KCF of $A_{\rm l} - \lambda B_{\rm l}$ contains only left singular blocks and the same L_j^T blocks as in the KCF of $A - \lambda B$. The pencil $A_{\rm reg} - \lambda B_{\rm reg}$ is upper-triangular and regular and has the same regular structure in its KCF as that of $A - \lambda B$.

The computation of the GUPTRI form of a pencil is stable because it involves only unitary transformations. There exist efficient algorithms and software to compute the GUPTRI form of a general pencil; most notable of them is Demmel and Kågström, 1993a and Demmel and Kågström, 1993b. Since the pencil $A_{n,r} - \lambda B_{n,r}$ is symmetric and has only finite generalized eigenvalues, its KCF contains only J_j and L_j blocks.

Now, given the noisy estimates $\widehat{A}_{n,r}$ and $\widehat{B}_{n,r}$ of $A_{n,r}$ and $B_{n,r}$ respectively, the LGEM of the pencil $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$ is the LGEM of the regular structure in its GUPTRI form. In general, any

singular structure that is present in the pencil $A_{n,r} - \lambda B_{n,r}$ might be absent from the GUPTRI form of $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$ because of the noise in the estimates. In other words, even if $A_{n,r} - \lambda B_{n,r}$ is singular, it might so happen that $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$ is regular, but with some highly ill-conditioned eigenvalues. So we need to ignore eigenvalues with large condition numbers ($> 10^{12}$) and return as the LGEM, the eigenvalue with the largest modulus among the remaining ones. For a generalized eigenvalue μ of the pencil $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$ with the corresponding generalized eigenvector ξ, from equation (3.2.13), it is clear that one measure of the conditioning of μ is given by $\frac{\langle \xi, \xi \rangle}{\langle \xi, \widehat{B}_{n,r} \xi \rangle}$.

Our heuristic for computing the LGEM of a given pencil $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$ can be summarized as follows:

1. Compute the GUPTRI form of the pencil $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$.

2. Extract the regular structure of the pencil from its GUPTRI form. Let its size be n_1 and let the regular eigenvalues be $\mu_1, \mu_2, \ldots, \mu_{n_1}$ with their generalized eigenvectors being $\xi_1, \xi_2, \ldots, \xi_{n_1}$.

3. The LGEM of the pencil is given by $\max_i |\mu_i|$ over all $i = 1, 2, \ldots, n_1$ such that μ_i is real and $\frac{|\langle \xi_i, \xi_i \rangle|}{|\langle \xi_i, \widehat{B}_{n,r} \xi_i \rangle|} < C$, where C is typically chosen to be a large number, such as 10^{12}.

Before we describe other methods to estimate λ_*, let us backtrack a bit and summarize the modified KSP algorithm to estimate λ_* given a sample run $f(X_0), f(X_1), \ldots, f(X_T)$.

1. Estimate the autocovariance function for f at lags $s = 0, 1, \ldots, T-1$ using the expression (3.0.4).

2. Estimate the autocorrelation time $\hat{\tau}_{\mathrm{int},f}$ using the procedure *TauEstimate* described in section 3.2.1. The estimate is self-consistent only if the run length T is sufficiently large compared to the estimate $\hat{\tau}_{\mathrm{int},f}$.

3. Fix the pencil size parameter n to say, 10 and determine the lag parameter r from equation (3.2.18).

4. Form the matrices $\widehat{A}_{n,r}$ and $\widehat{B}_{n,r}$ from equation (3.2.17).

5. Find the LGEM, $\hat{\mu}$, of the pencil $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$; the estimate for λ_* is then $(\hat{\mu})^{1/r}$.

This method of estimating λ_* by considering a single value of n, r and taking the rth root of the LGEM of $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$ is henceforth termed the *KSP singleton method*. The drawback of this method is that it uses very little of the data that is available; even though we have estimates $\widehat{C}_f(s)$ for $s = 0, 1, \ldots, T-1$, the singleton method estimates λ_* using only $\widehat{C}_f(rk)$ for $k = 0, 1, \ldots, 2n-1$. In section 5, we devise more stable and accurate estimates for λ_* by not limiting the data used as in the singleton method.

Any statistical estimate is meaningless without some kind of error bars – we now give a simple method of computing error bars for the estimate $\hat{\lambda}_*$, called the method of batch means.

3.3 Error bars for $\hat{\lambda}_*$ using batch means method

The method of batch means is a simple and effective procedure to compute error bars for an estimated quantity in MCMC. Suppose we have samples $\{f_1, f_2, \ldots, f_T\}$ from the simulation of a Markov chain. If we are trying to estimate the quantity $\langle f \rangle_\pi$, we know from the ergodic theorem that the time average $\hat{f} = \frac{\sum_{i=1}^{T} f_i}{T}$ converges to $\langle f \rangle_\pi$ as $T \to \infty$. One method of computing error bars for $\hat{f}$ is to estimate the integrated autocorrelation time $\hat{\tau}_{\mathrm{int},f}$ from the procedure *TauEstimate* given in section 3.2.1; one can then compute error bars for $\hat{f}$ from (2.0.4). An alternative method is to use the method of batch means, which we describe now.

The basic idea is to divide the simulation run into a number of contiguous batches and then use the sample means from the batches (batch means) to estimate the overall mean and its vari-

ance. To be more precise, let the run length $T = BK$ – we can then divide the whole run into B batches of size K each. The bth batch consists of the samples $f_{(b-1)K+1}, f_{(b-1)K+2}, \ldots, f_{bK}$ for $b = 1, 2, \ldots, B$, with its sample mean given by

$$\hat{f}^{(b)} = \frac{1}{K} \sum_{i=1}^{K} f_{(b-1)K+i}.$$

The overall mean $\hat{f}$ can then be computed as the mean of these batch means:

$$\hat{f} = \sum_{b=1}^{B} \frac{1}{B} \hat{f}^{(b)}.$$

Let us assume that the original process $\{f_i\}$ is weakly stationary, that is, $E[f_i] = \langle f \rangle_\pi$ and $\text{Var}[f_i] = \text{Var}_\pi[f]$ for all i and $\text{Cov}(f_i, f_j)$ depends only on $|j - i|$. Then the batch means process $\hat{f}^{(1)}, \ldots, \hat{f}^{(B)}$ is also weakly stationary and we can write the following expression for the variance of $\hat{f}$ (Alexopoulos et al., 1997):

$$\text{Var}[\hat{f}] = \frac{\text{Var}[\hat{f}^{(b)}]}{B} + \frac{1}{B^2} \sum_{i \neq j} \text{Cov}[\hat{f}^{(i)}, \hat{f}^{(j)}].$$

As the batch size $K \to \infty$, the covariance between the batch means $\text{Cov}[\hat{f}^{(i)}, \hat{f}^{(j)}] \to 0$. If K is very large, it is a reasonable approximation that the batch means are *independent* of each other. In that case, $\text{Var}[\hat{f}] \approx \frac{\text{Var}[\hat{f}^{(b)}]}{B}$. The quantity $\text{Var}[\hat{f}^{(b)}]$ can be estimated by the standard estimator

$$\hat{\sigma}_b^2[f] = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{f}^{(b)} - \hat{f})^2.$$

If the batch size K is large enough that the distribution of $\hat{f}^{(b)}$ is approximately normal, then we can write the $(1 - \alpha)$ confidence interval for $\langle f \rangle_\pi$:

$$\hat{f} \pm t_{B-1,\alpha/2} \frac{\hat{\sigma}_b[f]}{\sqrt{B}},$$

where $t_{B-1,\alpha/2}$ is the upper $\alpha/2$ critical value of Student's t-distribution, although the t-distribution is very close to normal even for moderate values of B, such as 100.

We extend the method of batch means to compute the error bars for *any* function of the samples and not just the sample means. To illustrate this for the estimate of λ_*, let us assume as before that we have a large simulation run of length T split into B batches of size K each. Let $\hat{\lambda}_*^{(b)}$ be the estimate of reduced spectral radius computed from batch b using, say, the KSP singleton method described in section 3.2.2. The overall estimate $\hat{\lambda}_*$ is then computed as the mean of these batch estimates. As we mentioned before, if the batch size K is large, then it is a reasonable approximation that the batch estimates $\hat{\lambda}_*^{(b)}$ are independent of each other. As shown above, we can then write the $(1-\alpha)$ confidence interval for the estimate of λ_*:

$$\hat{\lambda}_* \pm t_{B-1,\alpha/2} \frac{\hat{\sigma}_b[\lambda_*]}{\sqrt{B}},$$

where $\hat{\sigma}_b[\lambda_*]$ is given by the expression

$$\hat{\sigma}_b^2[\lambda_*] = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\lambda}_*^{(b)} - \hat{\lambda}_*)^2. \tag{3.3.1}$$

How do we choose the batch size K and the number of batches B? It is observed in practice that K should be chosen large relative to the autocorrelation time $\hat{\tau}_{\mathrm{int},f}$ – for instance, $K \approx 1000\hat{\tau}_{\mathrm{int},f}$; a value of $B = 100$ usually suffices.

4
MODEL PROBLEMS

In this section, we describe some processes which serve as model problems for testing our methods of estimating spectral gap. Chief among these occurring in practice are the East model, Fredrickson-Andersen (FA) model and Ising model with Glauber dynamics. The East model and the FA model belong to the class of kinetically constrained spin models for glass transitions. The reason these are chosen as model problems is that each has a spectral gap that is very small and hence in general more difficult to estimate. For these Markov chains, we show how our method is more accurate than the existing methods. Also for each of these two models, we have an observable that has a good overlap with the slowest mode. Before we go on to describe these, we consider simpler problems like the AR(1) process and the urn model for which the complete spectral decomposition, that is, the eigenvalues and eigenvectors, are known.

4.1 AR(1) process

The main motivation for choosing a process for which the eigenvalues and eigenfunctions are known is that we can better test the heuristics that we proposed in the last section for the choice of the lag and pencil size parameters. The AR(1) process is given by the following:

$$X_t = aX_{t-1} + bZ_t, \quad t = 1, 2, \ldots, \tag{4.1.1}$$

where Z_t is a Gaussian noise term with mean 0 and variance 1 and a and b are the parameters of the model. If $|a| < 1$, the process is weakly stationary and $E[X_t] = 0$ and Var$[X_t] = b^2/(1-a^2)$. In fact, if $|a| < 1$, we can also show that the distribution of X_t is Gaussian for large t. Writing $X_{t-1} = aX_{t-2} + bZ_{t-1}$ in the defining equation (4.1.1), we get $X_t = a^2X_{t-2} + abZ_{t-1} + bZ_t$.

Continuing this n times yields

$$X_t = a^n X_{t-n} + b\sum_{k=0}^{n-1} a^k Z_{t-k}.$$

For large n, $a^n \to 0$ and since X_t is the sum of n independent Gaussian random variables, its distribution is Gaussian as well. Also, it is easy to check that the Gaussian distribution is the invariant distribution for this process. To simplify things, we choose $b = \sqrt{1-a^2}$, in which case the invariant distribution is standard normal.

Though we described the KSP method of estimating the spectral gap only for discrete state space chains in section 3.2, it also applies to a particular class of general state space reversible chains called the Hilbert-Schmidt class (Dunford and Schwartz, 1988). Consider a Markov chain on a general state space $(\mathcal{X}, \mathcal{F}, \Pi)$ with stationary distribution Π whose density is π with respect to a dominating measure ν on $(\mathcal{X}, \mathcal{F})$. Let the Markov chain be given by a transition probability density $P(x, y)$ for $x, y \in \mathcal{X}$ with respect to ν. Let $\ell^2(\pi)$ be the space of functions $f : \mathcal{X} \to \mathbb{R}$ for which

$$\|f\|_\pi^2 = \int_{\mathcal{X}} |f(x)|^2 \Pi(dx) = \int_{\mathcal{X}} |f(x)|^2 \pi(x)\nu(dx) < \infty.$$

Then $\ell^2(\pi)$ is a Hilbert space with the inner product

$$\langle f, g\rangle_\pi = \int_{\mathcal{X}} f(x)g(x)\Pi(dx), \quad \forall f, g \in \ell^2(\pi).$$

Assume that the function $y \mapsto P(x, y)/\pi(y)$ is in $\ell^2(\pi)$ for all x. This defines an operator P on $\ell^2(\pi)$ given by

$$\mathsf{P}f(x) = \int_{\mathcal{X}} P(x, y)f(y)\nu(dy) = \int_{\mathcal{X}} [P(x, y)/\pi(y)]f(y)\pi(dy).$$

Then P is self-adjoint if and only if the Markov chain is reversible. It belongs to the Hilbert-Schmidt class if (Garren and Smith, 2000)

$$\int_{\mathcal{X}}\int_{\mathcal{X}} (P(x, y)/\pi(y))^2 \Pi(dx)\Pi(dy) = \int_{\mathcal{X}}\int_{\mathcal{X}} P(x, y)^2[\pi(x)/\pi(y)]\nu(dx)\nu(dy) < \infty. \tag{4.1.2}$$

A self-adjoint Hilbert-Schmidt operator P on $\ell^2(\pi)$ has a discrete spectrum of eigenvalues $\{\lambda_k, k = 1, 2, \ldots\}$ with the corresponding eigenfunctions $\{e_k, k = 1, 2, \ldots\}$ forming an orthonormal basis of $\ell^2(\pi)$ Dunford and Schwartz, 1988, pp. 1009–1034. The reason our KSP method applies to Hilbert-Schmidt class of Markov chains is precisely because they have a discrete spectrum of eigenvalues.

For the AR(1) process in equation (4.1.1), $\mathcal{X} = \mathbb{R}$ and the dominating measure ν is the Lebesgue measure on $\mathbb{R}$. The transition probability density with respect to the Lebesgue measure is given by $P(x, y) = \frac{1}{\sqrt{2\pi(1-a^2)}} e^{\frac{-(y-ax)^2}{2(1-a^2)}}$. The invariant distribution $\pi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$. We can verify the condition (4.1.2) for AR(1) process as follows:

$$\begin{aligned}
\int_{y=-\infty}^{\infty}\int_{x=-\infty}^{\infty} P(x,y)^2[\pi(x)/\pi(y)]dxdy &= \int_{y=-\infty}^{\infty}\int_{x=-\infty}^{\infty} \frac{1}{2\pi(1-a^2)} e^{-\frac{(y-ax)^2}{(1-a^2)}} e^{-\frac{x^2}{2}} e^{\frac{y^2}{2}} dxdy \\
&= \int_{y=-\infty}^{\infty}\int_{x=-\infty}^{\infty} \frac{1}{2\pi(1-a^2)} e^{-\frac{(y-ax)^2}{1-a^2} - \frac{x^2}{2} - \frac{y^2}{2}} dxdy \\
&= \int_{y=-\infty}^{\infty}\int_{x=-\infty}^{\infty} \frac{1}{2\pi(1-a^2)} e^{\frac{-y^2(1+a^2)-x^2(1+a^2)+4axy}{2(1-a^2)}} dxdy \\
&= \int_{\mathbb{R}^2} \frac{1}{2\pi(1-a^2)} e^{-\frac{1}{2}z^T C^{-1} z} dz, \qquad (4.1.3)
\end{aligned}$$

where $z = [x, y]^T$ and $C^{-1} = \frac{1}{1-a^2}\begin{bmatrix} 1+a^2 & 2a \\ 2a & 1+a^2 \end{bmatrix}$. The determinant of C is 1. The following is a Gaussian integral in two dimensions and hence equals 1.

$$\int_{\mathbb{R}^2} \frac{1}{2\pi} e^{-\frac{1}{2}z^T C^{-1} z} dz = \int_{\mathbb{R}^2} \frac{1}{2\pi \ \det(C)} e^{-\frac{1}{2}z^T C^{-1} z} dz = 1.$$

Therefore, the integral in (4.1.3) equals $1/(1-a^2)$, which is finite. The AR(1) process given by (4.1.1) hence belongs to the Hilbert-Schmidt class.

We can show that the AR(1) process has eigenvalues $\{1, a, a^2, \ldots\}$ with the eigenfunction for the kth eigenvalue a^{k-1} being the Hermite polynomial of degree k which is given by:

$$H_k(x) = e^{x^2/2} c_k \partial_x^k e^{-x^2/2}, \qquad (4.1.4)$$

where c_k depends only on k. The Hermite polynomials form an orthonormal basis for $\ell^2(\pi)$, that is, for $j \neq k$,

$$\int_{x=-\infty}^{\infty} H_k(x) H_j(x) e^{-x^2/2} dx = 0,$$

and $\{H_k(x), k = 0, 1, 2, \ldots\}$ span the space $\ell^2(\pi)$. For $k \geq 1$, the number c_k is chosen so as to make the norm of H_k one, that is, to ensure that

$$\frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{\infty} H_k^2(x) e^{-x^2/2} dx = 1.$$

The reduced spectral radius of the AR(1) process is a with the corresponding slowest mode being $H_1(x) = x$. Let us start with an observable which has an overlap with $H_1(x)$ and then use the KSP method to estimate the parameter a. By applying our method to a process like this, for which we know the eigenvalues and eigenfunctions, we can

- evaluate the heuristics that we proposed for choosing the pencil size parameter n and the lag parameter r, and
- analyze the error in the estimate.

4.1.1 Observable $H_1 + H_2 + H_3 + H_4$

In this section, we consider the problem of estimating the spectral gap of the AR(1) process with the observable as the sum of the first four eigenfunctions, namely, the observable $f = H_1 + H_2 + H_3 + H_4$. As we mentioned before, the kth Hermite polynomial H_k is an eigenfunction of the AR(1) process with eigenvalue a^k. The reason this particular observable is chosen for our experiment is because of the following factors:

- Since we know the spectral decomposition of the process, we can compare the estimates of spectral gap using different values of the pencil size and lag parameters with the exact result. We can thus evaluate the heuristics for the choice of n and r.

- A careful choice of n, r is essential for this problem because the overlap with the slowest mode H_1 is not any more substantial than with other modes. With the naive choice of $n = r = 1$, the KSP method, which is then identical to returning the Rayleigh quotient of f as the estimate of λ_*, gives the answer $(a + a^2 + a^3 + a^4)/4$.

The autocorrelation time for the observable f is $\tau_{\mathrm{int},f} = (\tau_{\mathrm{int},H_1} + \tau_{\mathrm{int},H_2} + \tau_{\mathrm{int},H_3} + \tau_{\mathrm{int},H_4})/4$, where τ_{int,H_i} is the autocorrelation time corresponding to H_i and is given by $\tau_{\mathrm{int},H_i} = (1 + a^i)/(1 - a^i)$. The AR(1) process is simulated with f as the observable and the parameter a chosen to be 0.99. The total run length of the chain is 10^9 divided into 100 batches.

To motivate the heuristics for the choice of the pencil size parameter n and the lag parameter r, for each batch (of size 10^7) let us use the KSP singleton method described in section 3.2 to estimate λ_* for $n = 1, 2, \ldots, 10$ and for each n, $r = 1, 2, \ldots, \lfloor \frac{c\hat{\tau}_{\mathrm{int},f}}{2n-1} \rfloor$. Basically, we do not use autocovariance estimates for lags beyond $c\hat{\tau}_{\mathrm{int},f}$, where $\hat{\tau}_{\mathrm{int},f}$ is the estimate of $\tau_{\mathrm{int},f}$ obtained from the *TauEstimate* procedure and c is the consistency parameter therein.

For the AR(1) process, we also compare the KSP singleton method with the TLS Prony method for different values of n and r. Although in section 3.1, we described Prony's method for the case $r = 1$, it can be extended to $r > 1$ by simply replacing $C_f(k)$ with $C_f(kr)$ for $k = 0, 1, \ldots, T$ in equation (3.1.1). The estimate obtained in that case is an estimate of λ_*^r.

Let $\hat{\lambda}_{n,r}^{(b)}$ be the estimate of λ_* for AR(1) process obtained using data from batch b (and parameters n,r) either by using the KSP singleton method or the TLS Prony method and let $\hat{\lambda}_{n,r}$ be the mean of these batch estimates. Also, let $\hat{\sigma}_b[\lambda_{n,r}]$ be the sample standard deviation estimate of $\hat{\lambda}_{n,r}^{(b)}$ which analogous to equation (3.3.1), is given by

$$\hat{\sigma}_b^2[\lambda_{n,r}] = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\lambda}_{n,r}^{(b)} - \hat{\lambda}_{n,r}\right)^2. \tag{4.1.5}$$

An estimate for the standard deviation of $\hat{\lambda}_{n,r}$ is then given by $\hat{\sigma}[\lambda_{n,r}] = \frac{\hat{\sigma}_b[\lambda_{n,r}]}{\sqrt{B}}$.

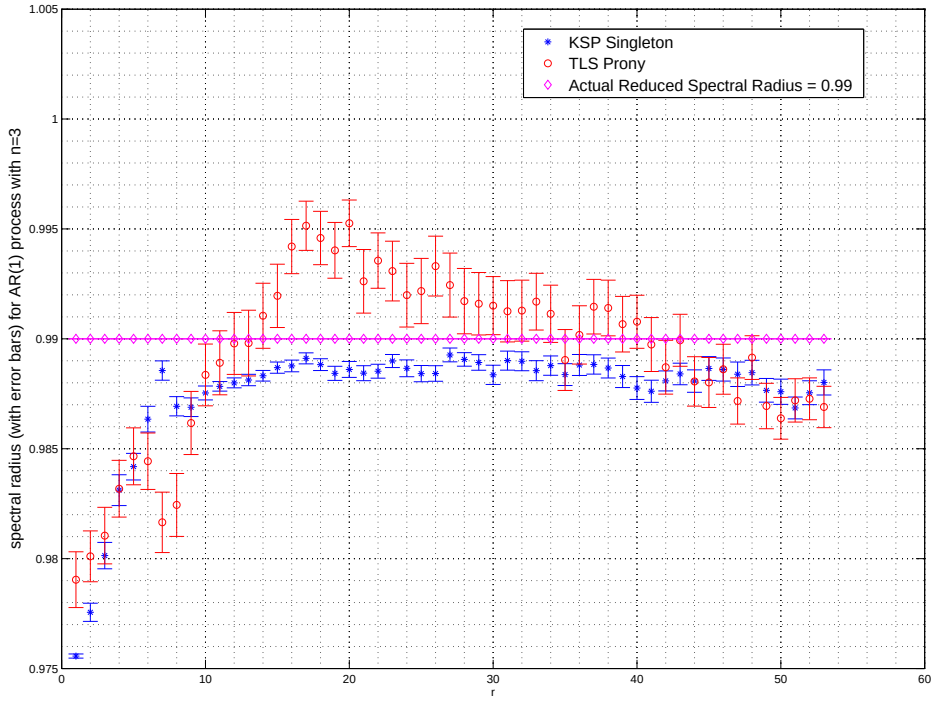

Figure 4.1: *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for AR(1) process with observable* $H_1 + H_2 + H_3 + H_4$ *using KSP singleton method and TLS Prony method for* $n = 3$ *and different values of* r*. Error bars are computed using* (4.1.5).

Figure 4.1 plots the estimate $\hat{\lambda}_{n,r}$ obtained by applying the KSP singleton and the TLS Prony methods and one standard deviation error bar, namely $\hat{\sigma}[\lambda_{n,r}]$, for $n = 3$ and different values of r. Figures 4.2 and 4.3 give similar plots for $n = 4$ and $n = 7$ respectively. A couple of distinctions between the two estimation methods are immediately obvious:

- The TLS Prony estimate has the larger error bars, especially for $n = 3$ and $n = 7$.
- The bias for the TLS Prony estimate becomes large and positive for larger values of r,

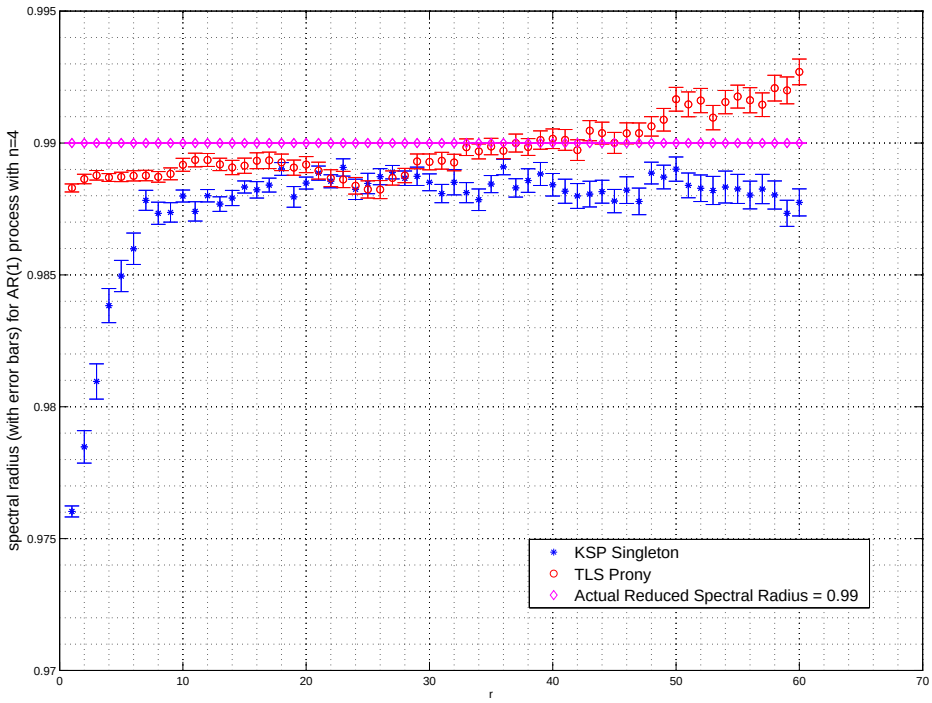

Figure 4.2: *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for AR(1) process with observable* $H_1 + H_2 + H_3 + H_4$ *using KSP singleton method and TLS Prony method for* $n = 4$ *and different values of* r.

especially for $n = 4$ and $n = 7$. For the KSP estimate on the other hand, for increasing values of r, the bias does not become noticeably larger, but the error bars get bigger.

- The KSP estimate has a negative bias for small r, but it stabilizes for increasing values of r instead of fluctuating as in the TLS Prony case.

The figures above reinforce the claim we made at the beginning of section 3.2 that the choice of the pencil size parameter n is not as important as that of the lag parameter r. Even though the

actual value of n is 4 (the observable being the sum of four eigenfunctions), the choice $n = 3$ or $n = 7$ work equally well with an appropriate choice of r (basically ignoring very small values).

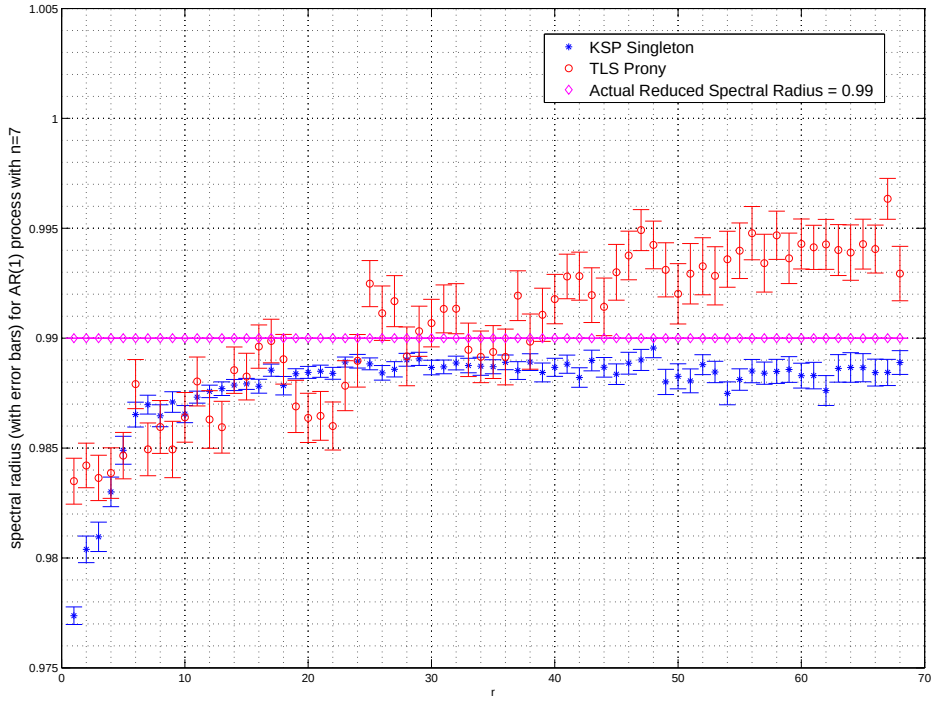

Figure 4.3: *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for AR(1) process with observable* $H_1 + H_2 + H_3 + H_4$ *using KSP singleton method and TLS Prony method for* $n = 7$ *and different values of* r.

4.2 Ehrenfest Urn Model

The next model problem that we consider is the Ehrenfest urn process, which is another example of a reversible Markov chain whose spectral decomposition is completely known. Our descrip-

tion follows the one given in Karlin and McGregor, 1965. In this process, we consider two urns and N balls distributed in the urns. The chain is said to be in state i if there are i balls in urn I and $N - i$ balls in urn II. At each instant, a ball is drawn at random from among all the balls and placed in urn I with probability p and in urn II with probability $q = 1 - p$. The probability transition matrix is given by the $(N + 1) \times (N + 1)$ matrix $P = (P_{ij})$, where

$$P_{ij} = \begin{cases} \kappa_i & \text{if } j = i + 1, \\ \zeta_i & \text{if } j = i - 1, \\ 1 - (\kappa_i + \zeta_i) & \text{if } j = i, \\ 0 & \text{if } |j - i| > 1 \end{cases}$$

and $\kappa_i = (N - i)p/N$, $\zeta_i = iq/N$ for $i, j = 0, 1, \ldots, N - 1$. This matrix is reversible with respect to the positive weights

$$\rho_i = \frac{\kappa_0 \kappa_1 \ldots \kappa_{i-1}}{\zeta_1 \zeta_2 \ldots \zeta_i} = \binom{N}{i} \left(\frac{p}{q}\right)^i,$$

that is, $\rho_i P_{ij} = \rho_j P_{ji}$ for $i, j = 0, 1, \ldots, N$. The invariant distribution for state i, therefore, is given by $\pi_i = \rho_i / \sum_{i=0}^{N} \rho_i$. All the eigenvalues of P are real and there are $N + 1$ linearly independent eigenvectors. Let v be an eigenvector of P corresponding to eigenvalue λ. Writing the equation $Pv = \lambda v$ in expanded form, we obtain

$$\begin{aligned} \lambda v_0 &= -\kappa_0 v_0 + \kappa_0 v_1 \\ \lambda v_i &= \zeta_i v_{i-1} + (1 - (\zeta_i + \kappa_i)) v_i + \kappa_i v_{i+1}, \ 1 \le i \le N - 1, \\ \lambda v_N &= -\zeta_N v_{N-1} + \zeta_N v_N. \end{aligned}$$

If v_0 is known we can solve these equations recursively for $v_1, v_2, \ldots, v_N$. The solution is of the form $v_i = K_i(\lambda) v_0$ where $K_i(\lambda)$ is a real polynomial of degree i. We can assume without loss

of generality that $v_0 = 1$. The eigenvalue λ is then the root of the $(N+1)$ degree polynomial

$$R(\lambda) = \lambda K_N(\lambda) + \zeta_N K_{N-1}(\lambda) - \zeta_N K_N(\lambda).$$

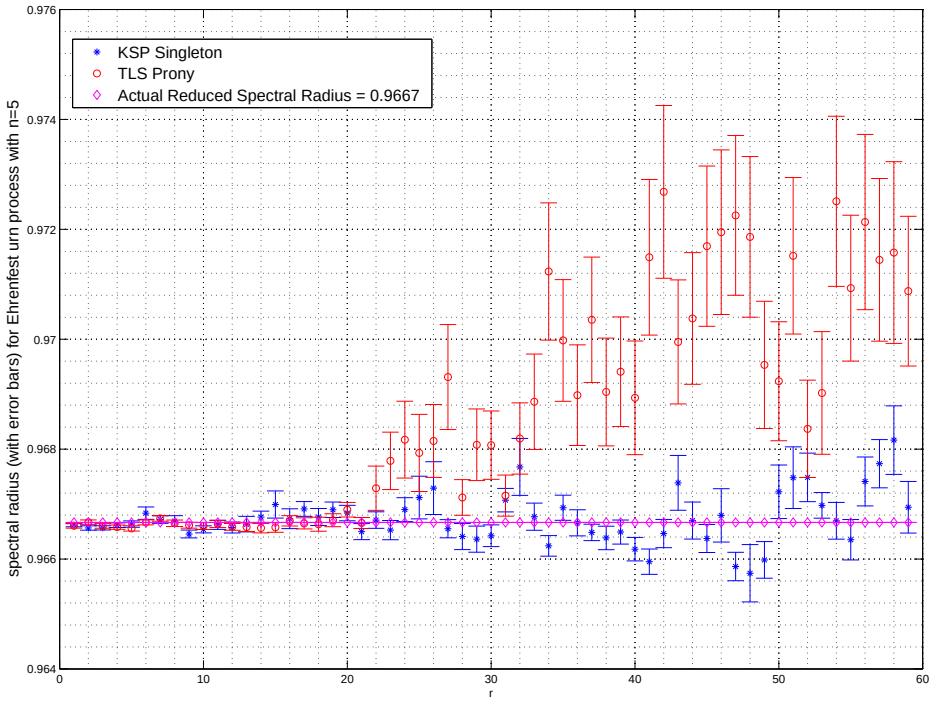

Figure 4.4: *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for Ehrenfest urn process (with* $N = 30, p = 0.4$*) with identity observable using KSP singleton method and TLS Prony method for* $n = 5$ *and different values of* r.

It is known that the eigenvalues of the transition matrix P are $0, 1/N, 2/N, \ldots, 1 - 1/N, 1$ (irrespective of the values of p, q) with the corresponding eigenvectors being the so-called *Krawtchouk* polynomials (Karlin and McGregor, 1965).

Consider the binomial distribution with weights $\binom{N}{j} p^j q^{N-j}$ on the $N+1$ points $j = 0, 1, \ldots, N$.

The polynomials orthogonal with respect to this distribution are the Krawtchouk polynomials given explicitly by the following expression (Karlin and McGregor, 1965):

$$K_n(j) = \sum_{i=0}^{n} (-1)^i \frac{\binom{n}{i}\binom{j}{i}}{\binom{N}{i}} \frac{1}{p^i}, \quad n, j = 0, 1, \ldots, N. \tag{4.2.1}$$

We can visualize the Krawtchouk polynomials as an $(N+1) \times (N+1)$ matrix $K = (K_{n,j})$ for $n, j = 0, 1, \ldots, N$. Note that $K_n(j)$ is symmetric with respect to n and j, that is, $K_n(j) = K_j(n)$, which implies that K is a symmetric matrix. The nth column/row of K, that is, K_n is the eigenvector corresponding to the eigenvalue $1 - n/N$.

4.2.1 Identity observable

In this section, we consider the problem of estimating the reduced spectral radius of the urn process with the simple identity observable, namely, $f(j) = j$, where j is the state of the Markov chain, that is, the number of balls in urn I. As we mentioned before, the eigenvector K_1 corresponds to the second largest eigenvalue $1 - 1/N$. From equation (4.2.1), it is clear that the first Krawtchouk polynomial is given by $K_1(j) = 1 - \frac{j}{Np}$. It turns out that among all Krawtchouk polynomial vectors K_n, f has an overlap only with K_1 (apart from K_0, which is irrelevant for the estimation of spectral gap because we consider only the autocovariance numbers).

So in effect, we are trying to estimate the reduced spectral radius of the urn process with the observable as the eigenfunction corresponding to the reduced spectral radius itself! This might sound like an easy proposition because in that case, the Rayleigh quotient of f is an adequate estimate of reduced spectral radius. But we would like to observe how our KSP method behaves for an observable which is the exact eigenfunction itself. Especially, if we use pencil size parameter $n > 1$, the actual matrices $A_{n,r}, B_{n,r}$ corresponding to the estimates in equation (3.2.17) have an $n-1$ dimensional common null space and only one regular generalized eigenvalue. To extract that particular generalized eigenvalue from the estimates $\hat{A}_{n,r}$ and $\hat{B}_{n,r}$ might be a bit of

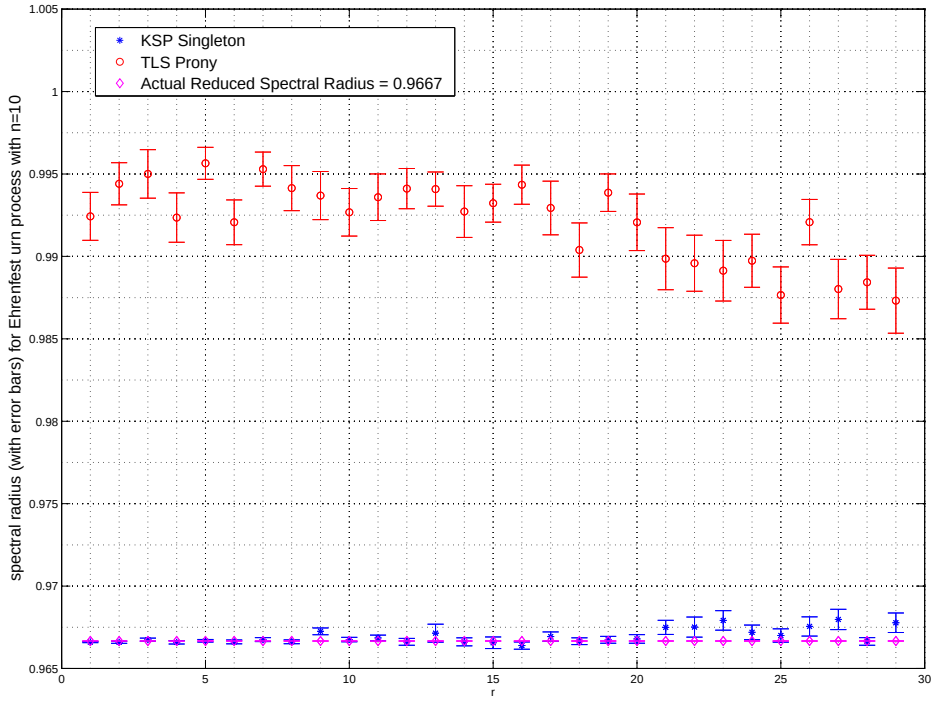

Figure 4.5: *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for Ehrenfest urn process (with* $N = 30, p = 0.4$*) with identity observable using KSP singleton method and TLS Prony method for* $n = 10$ *and different values of* r.

a challenge. But it turns out that the procedure that we outlined in section (3.2.2) for estimating the LGEM of $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$ by using the GUPTRI form is adequate for filtering out the singular generalized eigenvalues of $\widehat{A}_{n,r} - \lambda \widehat{B}_{n,r}$.

We performed the simulation of the urn process with the values $N = 30, p = 0.4$ – the run length is 10^8 split into 100 batches. The reduced spectral radius is then $\lambda_* = 1 - 1/30 = 0.9667$. We then applied the KSP singleton method and TLS Prony method with different values of pencil size parameter n and lag parameter r. Figures 4.4, 4.5 and 4.6 plot the estimates $\hat{\lambda}_{n,r}$ with their

error bars for $n = 5, 10, 1$ respectively. The figures indicate that TLS Prony misses the exact λ_* completely, especially for large values of n and r.

As for the KSP method, it is clear that large n do not pose much of a problem and definitely not as much of a problem as for TLS Prony. Even when n is as large as 10, as clear from Figure 4.5, the heuristic that we outlined in section 3.2.2 accurately picks out the only regular generalized eigenvalue (which is 0.9667), while ignoring the other singular ones.

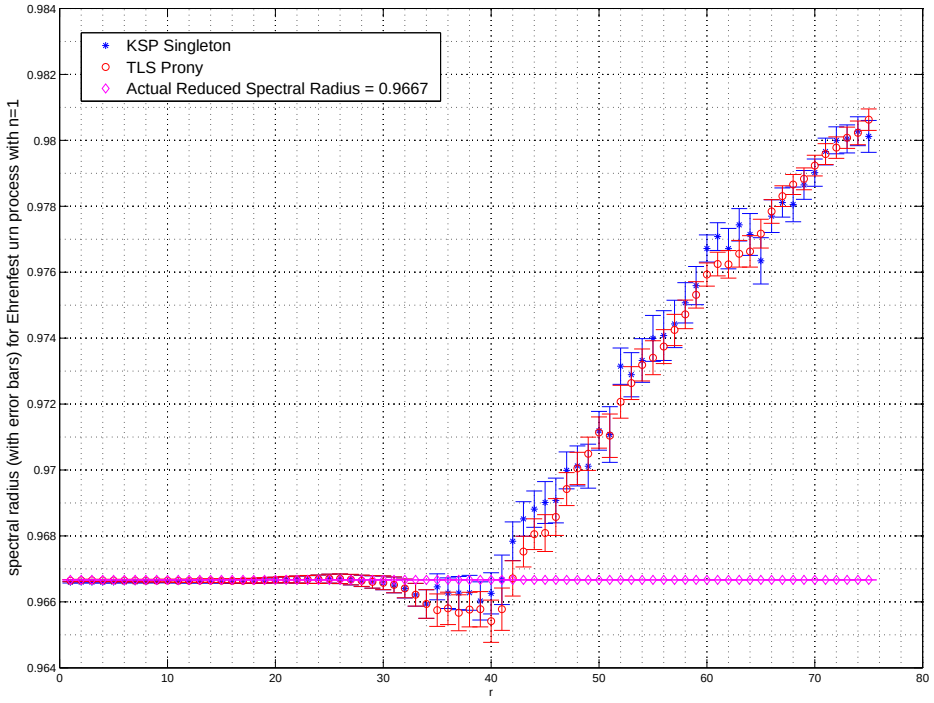

Figure 4.6: *The reduced spectral radius estimate* $\hat{\lambda}_{n,r} \pm \hat{\sigma}[\lambda_{n,r}]$ *for Ehrenfest urn process (with* $N = 30, p = 0.4$*) with identity observable using KSP singleton method and TLS Prony method for* $n = 1$ *and different values of* r.

Even if large values of n do not pose a problem, large values of r may. Especially from

Figure 4.6, it is clear that for $n = 1$ and for approximately $r > 40$, there is a large positive bias in the estimates $\hat{\lambda}_{n,r}$. This can be attributed to the fact that for large r, $C_f(nr)$ is all noise and little "signal". This in turn means that $\hat{\lambda}_{n,r}$ is noisy and $(\hat{\lambda}_{n,r})^{(1/r)} \to 1$ as $r \to \infty$. This tendency gets magnified for small n especially, as is obvious from Figure 4.6, and shows up even for moderately large values of r. This is precisely the reason we restrict the autocovariance estimates we use to lag $t < c\hat{\tau}_{\mathrm{int},f}$, which automatically restricts the lag parameter r. Even with that restriction, we may be using too many r values for some cases such as, for instance, the urn process with $n = 1$ (Figure 4.6).

That brings us to an important issue with KSP singleton method: if we are to use a single value of n and r, what values should we use? It is clear that no values of n and r, fixed heuristically *a priori*, work well for *all* processes. Can we devise estimates which use the estimates $\hat{\lambda}_{n,r}$ for a range of values of n and r? The answer is yes, and it is precisely the topic of next section.

5

MORE ACCURATE AND STABLE ESTIMATES FOR λ_*

We have seen from the results for the AR(1) process and the Ehrenfest urn process that the estimates returned by the KSP singleton method vary widely depending on the choice of n and r. We instead intend to use a set of pencil size parameter values $\mathcal{N} = \{n_1, n_2, \ldots, n_k\}$ and for each $i = 1, \ldots, k$, a set of lag parameter values $\mathcal{R}[n_i] = \{r_1^{(i)}, r_2^{(i)}, \ldots, r_{l_i}^{(i)}\}$. We then wish to return a single estimate $\hat{\lambda}_*$ that "fits" all the estimates $\hat{\lambda}_{n,r}$ from this range of values. The idea is that using a single value of n, r results in a noisy estimate of λ_*, but the use of a range of values diminishes the impact of noise by "averaging" it out.

We have seen that 10 is one of the largest values we can consider for the pencil size parameter, so we fix the pencil size parameter set to be

$$\mathcal{N} = \{1, 2, \ldots, 10\}. \tag{5.0.1}$$

For each n, the lag parameter set $\mathcal{R}[n_i]$ is chosen to include all values not exceeding the one given by (3.2.18), that is,

$$\mathcal{R}[n_i] = \left\{1, 2, \ldots, \left\lfloor \frac{c\ \hat{\tau}_{\mathrm{int},f}}{2n_i - 1} \right\rfloor \right\} \text{ for } n_i = 1, \ldots, 10. \tag{5.0.2}$$

If the estimated $\hat{\tau}_{\mathrm{int},f}$ is large, the $\mathcal{R}[n_i]$ above is a large set, thus pushing up the computation time to estimate $\hat{\lambda}_{n,r}$ for each value of n, r. Even though this time is small compared to that required for obtaining the simulation data, in practice it suffices to pick around 100 values at equal intervals from $\left\{1, 2, \ldots, \left\lfloor \frac{c\ \hat{\tau}_{\mathrm{int},f}}{2n_i - 1} \right\rfloor \right\}$ as the set $\mathcal{R}[n_i]$.

5.1 Series sum estimate

For each $n_i \in \mathcal{N}$ and $r_j^{(i)} \in \mathcal{R}[n_i]$ ($i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, l_i$), let $\hat{\mu}_{i,j}$ be the LGEM of $\widehat{A}_{n_i, r_j^{(i)}} - \lambda \widehat{B}_{n_i, r_j^{(i)}}$, that is, $\hat{\mu}_{i,j}$ is an estimate of $\lambda_*^{r_j^{(i)}}$. Consider the quantity

$$\hat{\mu}_i = \sum_{j=1}^{l_i} \hat{\mu}_{i,j}.$$

Then for the particular pencil size parameter value $n = n_i$, an estimate for the reduced spectral radius is a solution of the following polynomial equation:

$$\sum_{j=1}^{l_i} \lambda^{r_j^{(i)}} = \hat{\mu}_i. \tag{5.1.1}$$

The roots of the polynomial can be obtained by using the companion matrix method and the estimate for λ_* is the one that lies between 0 and 1. But if l_i is very large (that is, if the estimated autocorrelation time $\hat{\tau}_{\mathrm{int},f}$ is large), then the companion matrix could be huge. Instead, let us assume that the set of values $\mathcal{R}[n_i]$ is chosen to be an arithmetic series, that is, $r_j^{(i)} = r_1^{(i)} + (j-1)\Delta r^{(i)}$ for $1 \le j \le l_i$ where $\Delta r^{(i)} = r_2^{(i)} - r_1^{(i)}$. Then for each $1 \le i \le k$, we can write

$$\sum_{j=1}^{l_i} \hat{\mu}_{i,j} \approx \sum_{j=1}^{l_i} \lambda_*^{r_j^{(i)}} = \sum_{j=1}^{l_i} \lambda_*^{r_1^{(i)} + (j-1)\Delta r^{(i)}} = \frac{\lambda_*^{r_1^{(i)}} (1 - \lambda_*^{l_i \Delta r^{(i)}})}{1 - \lambda_*^{\Delta r^{(i)}}}.$$

If l_i is large enough that we can make the approximation $\lambda_*^{l_i \Delta r^{(i)}} \approx 0$, then the equation above can be written as

$$\hat{\mu}_i = \sum_{j=1}^{l_i} \hat{\mu}_{i,j} \approx \frac{\lambda_*^{r_1^{(i)}}}{1 - \lambda_*^{\Delta r^{(i)}}}.$$

The series sum estimate for λ_* corresponding to the pencil size parameter $n = n_i$, $\hat{\lambda}_*^{SS}[n_i]$, is obtained by solving the following polynomial equation for λ:

$$\lambda^{\Delta r^{(i)}} \hat{\mu}_i + \lambda^{r_1^{(i)}} - \hat{\mu}_i = 0. \tag{5.1.2}$$

The quantities $\Delta r^{(i)}$ and $r_1^{(i)}$ are much smaller than $r_{l_i}^{(i)}$ and hence the polynomial equation can be solved quite easily using the companion matrix method. If l_i is not so large that $\lambda_*^{l_i \Delta r^{(i)}}$ can be ignored, then we still need to solve equation (5.1.1) to obtain an estimate for λ_*.

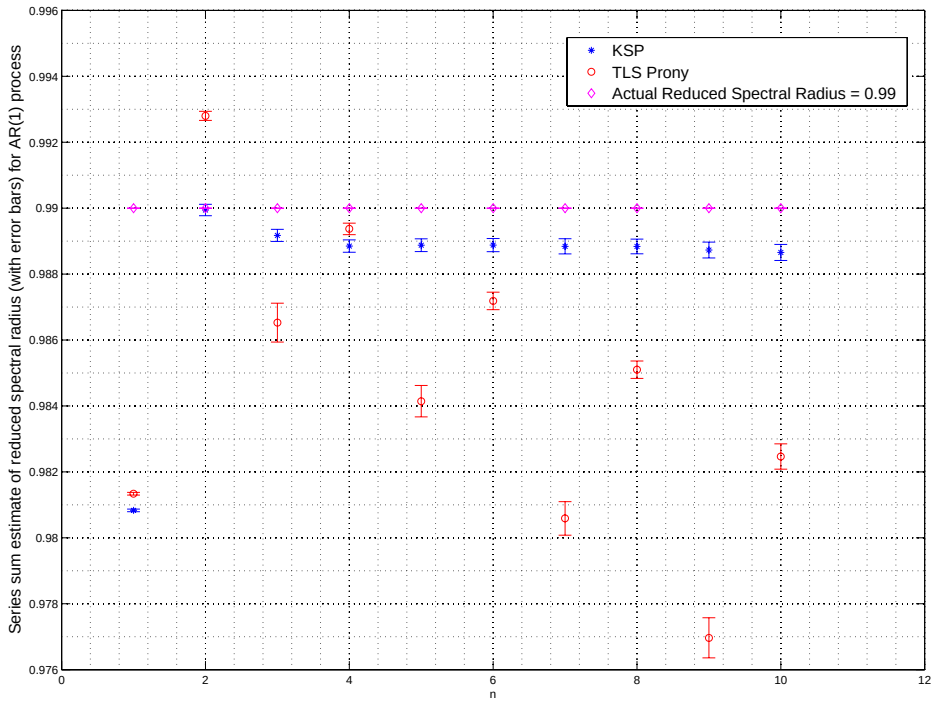

Figure 5.1: *The series sum estimate for the reduced spectral radius* $\hat{\lambda}_*^{SS}[n] \pm \hat{\sigma}[\lambda_*^{SS}[n]]$ *for AR(1) process with observable* $H_1 + H_2 + H_3 + H_4$ *using KSP and TLS Prony methods for* $n = 1, 2, \ldots, 10$.

In Figure 5.1, we plot $\hat{\lambda}_*^{SS}[n]$ with one standard deviation error bar $\hat{\sigma}[\lambda_*^{SS}[n]]$ for AR(1) process of section 4.1 for $n = 1, 2, \ldots, 10$ using both KSP and TLS Prony methods. As expected, the series sum estimates for TLS Prony method are all over the place for almost all values of n. In contrast, for the KSP method, the series sum estimate is pretty close to the actual answer 0.99 except for the case $n = 1$.

We can explain the anomaly for $n = 1$ by noting that for AR(1) process, the estimates $\hat{\lambda}_{1,r}$ have a negative bias for all $r < c\hat{\tau}_{\text{int},f}$. The observable f is the sum of four eigenfunctions, so using Rayleigh quotient (in effect, using pencil size parameter $n = 1$) yields a value less than the actual λ_*. The estimates $\hat{\lambda}_{1,r}$ do increase with increasing r but $c\hat{\tau}_{\text{int},f}$ is not sufficient for the bias to turn positive from negative. Contrast that with the urn process from Figure 4.6 where the limit of $c\hat{\tau}_{\text{int},f}$ for r may have been "too large". The essence basically is that for $n = 1$, the behavior of the estimates depends heavily on what r values we choose. But for $n > 1$, the estimates $\hat{\lambda}_{n,r}$ are more robust with respect to the choice of r values.

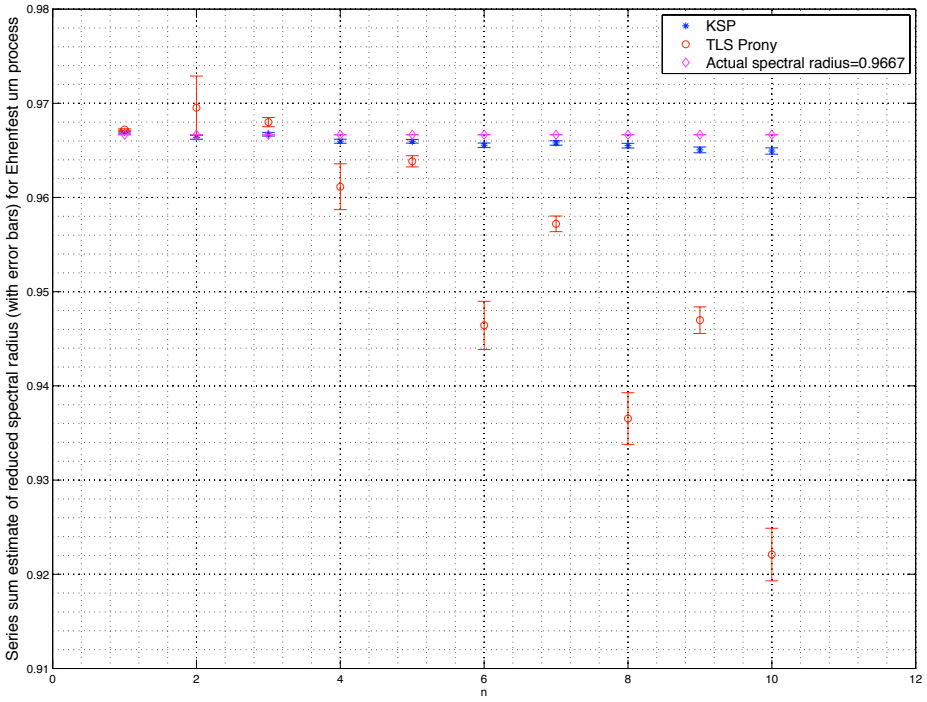

Figure 5.2: *The series sum estimate for the reduced spectral radius* $\hat{\lambda}_*^{SS}[n] \pm \hat{\sigma}[\lambda_*^{SS}[n]]$ *for Ehrenfest urn process with identity observable using KSP and TLS Prony methods for* $n = 1, 2, \ldots, 10$.

Figure 5.2 plots the series sum estimate for the urn process of section 4.2 with identity observable again for $n = 1, 2, \ldots, 10$. It is clear that the series sum estimate using KSP is very close to the exact answer 0.9667 for all values of n.

5.2 Least squares estimate

A more natural method than the series sum estimate and which works better in practice is the one obtained by using the *least squares* technique which we describe now. The least squares method is a modification of the maximum likelihood estimation method applied under suitable assumptions.

As before for each $n_i \in \mathcal{N}$ and $r_j^{(i)} \in \mathcal{R}[n_i]$, let $\hat{\mu}_{i,j}$ be the LGEM of $\widehat{A}_{n_i,r_j^{(i)}} - \lambda \widehat{B}_{n_i,r_j^{(i)}}$, that is, $\hat{\mu}_{i,j}$ is an estimate of $\lambda_*^{r_j^{(i)}}$. We can write

$$\hat{\mu}_{i,j} = \lambda_*^{r_j^{(i)}} + \eta_{i,j}, \tag{5.2.1}$$

where $\eta_{i,j}$ is the noise term. What is the distribution of $\hat{\mu}_{i,j}$? From the central limit theorem for Markov chains, we can conclude that for a sufficiently large run length of the Markov chain, the autocovariance estimates $\widehat{C}_f(s)$ are normally distributed. In addition, the following is a well-known fact:

Proposition 5.1. *Suppose X is approximately a multivariate normal with mean $\mathbf{x}_0$. If h is a differentiable function at $\mathbf{x}_0$, then $Y = h(X)$ is approximately normal with mean and variance given by the linearization of h about $\mathbf{x}_0$.*

Let us take X to be the autocovariance estimate vector with components $\widehat{C}_f(s)$, $s = 1, 2, \ldots, (2n_i - 1)r_j^{(i)}$, and the function h to be the LGEM of $\widehat{A}_{n_i,r_j^{(i)}} - \lambda \widehat{B}_{n_i,r_j^{(i)}}$; clearly h is a differentiable function if $\widehat{B}_{n_i,r_j^{(i)}}$ is positive definite (even if $\widehat{B}_{n_i,r_j^{(i)}}$ is singular, the largest *regular* generalized eigenvalue modulus is a differentiable function).

From Proposition 5.1, we can conclude that $\hat{\mu}_{i,j}$ is normally distributed if the run length of the Markov chain is sufficiently large. This then implies from (5.2.1) that $\eta_{i,j}$ has a Gaussian distribution, that is, $\eta_{i,j} \sim N(b_{i,j}, \sigma_{i,j})$. We now describe the least squares method of estimating λ_* for a fixed i, that is, for a fixed pencil size parameter value n_i.

Henceforth, the notation v_i denotes the column vector with components $v_{i,j}$ for $j = 1, 2, \ldots, l_i$, where l_i is the cardinality of $\mathcal{R}[n_i]$. Is it a reasonable assumption that the $\eta_{i,j}$ are independent for different j? Clearly not: since $\hat{\mu}_{i,j}$ are estimated from the same set of autocovariance estimates $\widehat{C}_f(s)$, they are indeed correlated. It is therefore logical to model η_i as a correlated Gaussian random variable vector, that is,

$$\eta_i \sim N(b_i, \Sigma_i), \tag{5.2.2}$$

where b_i represents the bias and Σ_i the covariance matrix of the estimates $\hat{\mu}_{i,j}$.

The covariance matrix Σ_i can be estimated from the $\hat{\mu}_{i,j}$ data (estimated using the KSP singleton method). The bias b_i cannot be estimated without first estimating λ_*. To get around this problem, we make the approximation that the noise η_i has mean zero, that is, that the estimates $\hat{\mu}_{i,j}$ are unbiased. We know that this is not true from the results for AR(1) and urn processes in sections 4.1 and 4.2 respectively (see Figures 4.1-4.6).

We can treat the $\hat{\mu}_{i,j}$ data from different batches as independent samples; then we can estimate the covariance matrix of singleton estimates Σ_i from equation (5.2.2) for a particular n_i. There are standard methods to estimate the covariance matrix of a random variable vector from a set of independent samples. Once we have the estimate of covariance matrix $\hat{\Sigma}_i$, we can write the maximum likelihood (ML) estimator of λ_* for pencil size parameter n_i as:

$$\hat{\lambda}_*^{ML}[n_i] = \arg\min_{\lambda} (\hat{\mu}_i - \lambda^{r^{(i)}})^T \hat{\Sigma}_i^{-1} (\hat{\mu}_i - \lambda^{r^{(i)}}), \tag{5.2.3}$$

where $\lambda^{r^{(i)}}$ is the column vector $[\lambda^{r_j^{(i)}}]$ for $j = 1, \ldots, l_i$.

The main drawback of this approach is that the covariance matrix $\hat{\Sigma}_i$ may not be estimated

accurately. The number of batches we consider is a relatively small number (such as $B = 100$), so if $l_i = |\mathcal{R}[n_i]| > B$, the estimated covariance matrix is not even positive definite. In such a situation, we can use one of the so-called shrinkage covariance estimators. An extensive review of the Bayes shrinkage estimators can be found in Daniels and Kass, 2001; Schäfer and Strimmer, 2005 proposes a set of improved shrinkage estimators with applications to functional genomics. Even if $l_i < B$, the estimated $\hat{\Sigma}_i$ may be noisy and since equation (5.2.3) involves its inverse $\hat{\Sigma}_i^{-1}$, the effect of noise is accentuated.

We wish to compare the ML estimate with the least squares (LS) estimate, which corresponding to $n = n_i$ is given by

$$\hat{\lambda}_*^{LS}[n_i] = \arg\min_{\lambda} \sum_{j=1}^{l_i} \left(\hat{\mu}_{i,j} - \lambda^{r_j^{(i)}} \right)^2. \tag{5.2.4}$$

Basically, the least squares loss function is the same as the ML loss function with the covariance matrix replaced by the identity matrix. Figure 5.3 plots the ML, LS and series sum estimates (with error bars) for the AR(1) process of section 4.1. As an aside, we note that since we have demonstrated in previous sections that KSP is a more accurate method than TLS Prony, we do not consider the estimates from TLS Prony any further and concentrate only on KSP.

A couple of conclusions from Figure 5.3 are immediately obvious: the least squares estimate is the best of the three, slightly better than the series sum estimate. The ML estimate is by far the worst of the three, with a large negative bias. Let us try and understand why this is so.

First of all, note that the LS loss function puts equal weight on each of the loss terms, measured by $\hat{\mu}_{i,j} - \lambda^{r_j^{(i)}\,2}$, for each j, which is not the case with ML loss function. If $\hat{\Sigma}_i$ is diagonal, then the ML loss function takes the form

$$\sum_{j=1}^{l_i} \frac{1}{\hat{\Sigma}_i[j,j]} \left(\hat{\mu}_{i,j} - \lambda^{r_j^{(i)}} \right)^2. \tag{5.2.5}$$

It is clear from Figures 4.1-4.3 that for small j, the bias is large (and negative) but the variance is smaller – with increasing j, the bias decreases, but the variance gets bigger. The jth diagonal

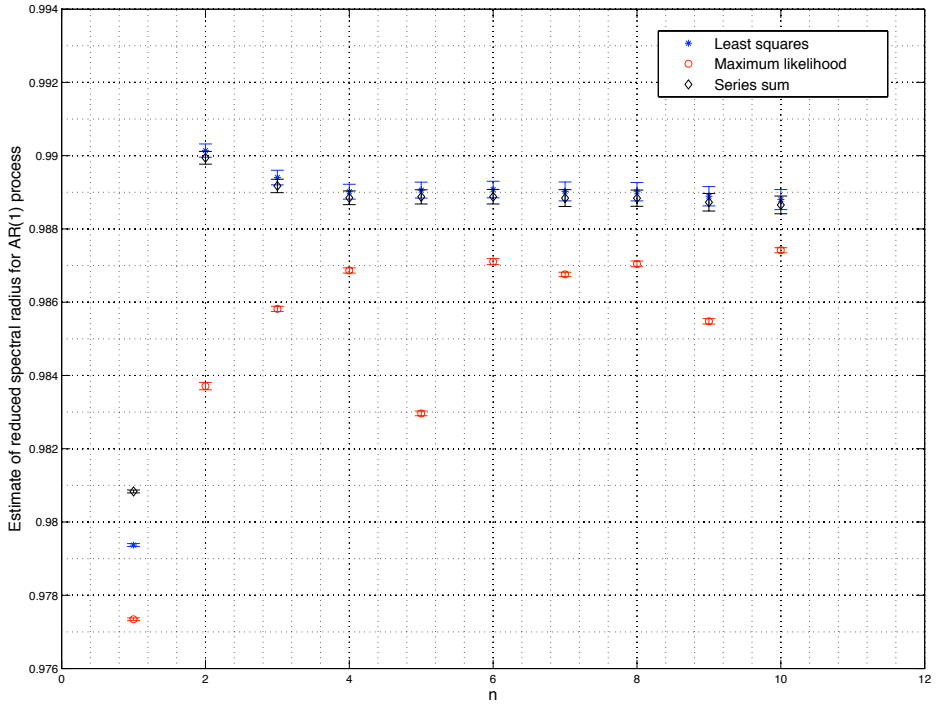

Figure 5.3: *The LS, ML and series sum estimates with error bars, for the reduced spectral radius for AR(1) process with observable the sum of first four eigenfunctions for* $n = 1, 2, \ldots, 10$.

entry $\hat{\Sigma}_i[j, j]$ represents the variance of $\hat{\mu}_{i,j}$ and it increases with increasing j. As a result, it is clear from (5.2.5) that in the ML loss function, the loss terms for small values of j have more weight attached to them compared to the ones for large values of j. But we have seen that for AR(1) process, the estimates $\hat{\mu}_{i,j}$ have large bias for small values of j and because of the additional weight placed on these by the ML loss function as compared to the LS loss function, the whole ML estimate (the solution of the optimization problem (5.2.3)) gets pulled down vis-a-vis the LS estimate. This explains why the ML estimate always is smaller than the simple LS estimate in the case of the AR(1) process.

Although we have explained the phenomenon of why the ML estimate is always worse than the LS estimate using the simplifying assumption that the covariance matrix $\hat{\Sigma}_i$ is diagonal, we can nevertheless draw the following conclusion: that ignoring the bias of $\hat{\mu}_{i,j}$ *worsens* the ML estimate as compared to the LS estimate. In other words, we cannot assume that the noise term η_i in Equation (5.2.1) has mean zero, proceed to minimize the associated loss function and expect better results than with the simple LS loss function.

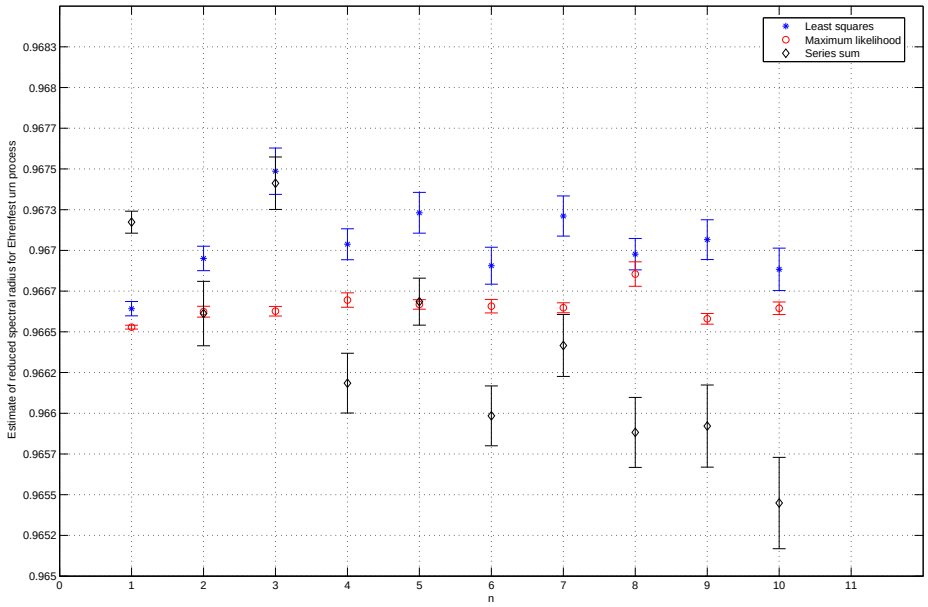

Figure 5.4: *The LS, ML and series sum estimates with error bars, for the reduced spectral radius for Ehrenfest urn process with identity observable for* $n = 1, 2, \ldots, 10$*. Even though the estimates may look all over the place, note the vertical scale which runs from 0.965 to 0.9683.*

This point brings us to the next question: if for a particular example, the $\hat{\mu}_{i,j}$ estimates do not have large bias for small j, can the ML estimates perform better than the LS estimates? In other words, we pose the question: if it is valid to assume that η_i from (5.2.1) have mean zero,

then how does the ML estimate perform in comparison to the LS estimate?

The Ehrenfest urn process with the identity observable is a classic example where $\hat{\mu}_{i,j}$ do not have a large bias for small j (see Figures 4.4-4.6). Figure 5.4 plots the LS, ML and series sum estimates with error bars for the urn process with identity observable. From Figure 5.4, it is clear that the ML estimate is superior to the LS estimate and the series sum estimates for this particular example.

The basic point is that if the observable is sufficiently close to the slowest mode that its Rayleigh quotient is a good approximation to the reduced spectral radius (which is the case with the identity observable for the urn process), then the estimates $\hat{\mu}_{i,j}$ do not have large bias for small values of j. For an example like this, the ML estimate is superior to the simple LS estimate. Also from Figures 5.3 and 5.4, it is clear for both AR(1) and Ehrenfest urn processes that the LS and ML estimates are respectively more accurate (less biased) than the series sum estimate. For AR(1) process, the explanation of the anomaly for the ML and LS estimates for $n = 1$ is similar to that for the series sum estimate (which is basically that the estimates $\hat{\lambda}_{1,r}$ have a negative bias for all values of $r < c\hat{\tau}_{\mathrm{int},f}$).

5.2.1 Obtaining a single estimate

We now have the estimates of λ_*, $\hat{\lambda}_*[n]$, for pencil size parameter values $n = 1, 2, \ldots, 10$. Out of these estimates, how should we pick the one "best" estimate, in some sense? The obvious idea is to pick the estimate with the least squared error. The squared error of an estimate, as is well-known, is the sum of the variance and the square of the bias. The variance of the estimates $\hat{\lambda}_*[n]$ is estimated using the batch means method, but what about the bias? Clearly, we are faced with the same problem as when we considered the maximum likelihood estimate, namely, how can we estimate the bias without knowing the true value of λ_*?

It turns out that we can approximate the bias for $\hat{\lambda}_*^{LS}[n]$ by the expression $\hat{\lambda}_*^{LS}[n] - \bar{\lambda}_*^{LS}$

where $\bar{\lambda}_*^{LS}$ is simply the sample mean of $\hat{\lambda}_*^{LS}[n]$ for $n = 1, 2, \ldots, 10$. The estimate with the least square error is simply $\hat{\lambda}_*^{LS}[n_{\text{lse}}]$, where n_{lse} is defined as

$$n_{\text{lse}} = \arg\min_n \left(\hat{\lambda}_*^{LS}[n] - \bar{\lambda}_*^{LS}\right)^2 + \hat{\sigma}[\lambda_*^{LS}]^2, \text{ and } \bar{\lambda}_*^{LS} = \frac{1}{10}\sum_{n=1}^{10} \hat{\lambda}_*^{LS}[n].$$

We have seen that the estimate $\hat{\lambda}_*^{LS}[1]$ for pencil size parameter $n = 1$ depends heavily on the lag parameter set $\mathcal{R}[1]$ that is chosen. Since there is no way of knowing whether $\mathcal{R}[1]$ is chosen properly beforehand, we can instead look at the LS and ML estimate plots. From Figure 5.3, it is clear that for AR(1) process, the estimate for $n = 1$ seems like an outlier compared to other values of n. So for cases like this, we can exclude the $n = 1$ estimate when computing $\bar{\lambda}_*^{LS}$, that is, $\bar{\lambda}_*^{LS}$ is instead computed as $\bar{\lambda}_*^{LS} = \frac{1}{9}\sum_{n=2}^{10} \hat{\lambda}_*^{LS}$.

Tables 5.1 and 5.2 list the least squared error estimates of reduced spectral radius for AR(1) and Ehrenfest urn processes using the various estimation methods. They just emphasize the point that we made before: that the LS estimate is the most accurate for the AR(1) process, while the ML one works the best for the urn process.

Estimation method	Estimate	Standard deviation	n_{lse}	Squared error
Maximum likelihood	0.9860	5.5605×10^{-5}	2	1.5747×10^{-5}
Least squares	0.9891	2.1779×10^{-4}	5	9.3169×10^{-7}
Series sum	0.9888	1.9368×10^{-4}	5	1.3019×10^{-6}

Table 5.1: ML, LS and series sum least squared error estimates for AR(1) process with $\lambda_* = 0.99$ and observable as sum of first four eigenfunctions.

Estimation method	Estimate	Standard deviation	n_{lse}	Squared error
Maximum likelihood	0.9667	2.9182×10^{-5}	5	8.5425×10^{-10}
Least squares	0.9670	9.522×10^{-5}	4	1.4674×10^{-7}
Series sum	0.9662	1.8355×10^{-4}	4	2.6585×10^{-7}

Table 5.2: ML, LS and series sum least squared error estimates for Ehrenfest urn process with $\lambda_* = 29/30$ and identity observable.

6
RESULTS

In this section, we describe how the KSP method that we described in the previous section estimates spectral gap for Markov chains that occur in practice. In particular, we consider the East model and Fredrickson-Andersen (FA) models as our model problems. These models belong to the class of kinetically constrained spin models for glass transitions. These models have been proposed to explain the dynamics of glass transitions, which, even after all these years, remains poorly understood. The slowest modes for these models are physically significant because they help explain the slow relaxation times near glass transition temperatures – see the next section for more details.

The other reason the East model and Fredrickson-Andersen models are chosen as model problems is that each has an observable that has a good overlap with the slowest mode. Aldous and Diaconis, 2002 and Cancrini et al., 2007 report observables for the east model and the FA models respectively, and prove them to be the slowest in an asymptotic sense. Using these observables, we obtain estimates of specral gap for these two processes that are more accurate than those obtained by existing methods.

6.1 East model

Before we describe the East model, let us give a synopsis of glass transitions in general. Glasses are typically prepared by rapid cooling of a molten liquid, that is, if the cooling rate is large enough, crystallization is avoided and the system becomes metastable (supercooled liquid). Good glass formers are those substances that, for some reason, either possess no crystalline configurations or have great difficulty in accessing these configurations (Fredrickson, 1988). If this supercooled liquid is further cooled, at some temperature T_g (which depends on the material),

its viscosity and various structural relaxation times increase dramatically, exceeding the observation time. The temperature T_g is said to be a glass transition temperature and depends on the rate of cooling.

The phenomenon of glass transition has been widely studied and it is now an accepted fact that it is a dynamical phenomenon and not a thermodynamic phase transition as was thought previously (Fredrickson and Andersen, 1985). Long relaxation times for a thermodynamic phase transition correspond to diverging correlation lengths which have not been observed near a glass transition. If indeed glass transition is a dynamical phenomenon, how do we model it? Once we design a model process, an inkling of the slowest mode for this process helps us understand, qualitatively, the nature of glass transitions. In order to quantify this understanding, we need to estimate the relaxation time (the autocorrelation time, in other words) of this slowest mode.

The long relaxation times in glasses typically show a pronounced divergence as temperature is lowered and are fitted experimentally by the Vogel-Fulcher law

$$\tau = \tau_0 e^{-A/(T-T0)}.$$

The exponential inverse temperature squared (EITS) form

$$\tau = \tau_0 e^{B/T^2} \tag{6.1.1}$$

has been proposed as an alternative (Sollich and Evans, 2003). Various approaches have been developed to understand the phenomena of glass transitions, most notable of them being mode coupling theories (Götze, 1989), the random first order scenario (Kirkpatrick et al., 1989) and kinetically constrained spin models (Fredrickson and Andersen, 1984).

Before we describe kinetically constrained spin models, we take a small detour and give a brief synopsis of the Ising model. It is a widely known mathematical model in statistical mechanics and was first proposed to explain the phenomenon of ferromagnetism. The Ising

model is defined on a discrete collection of spins (typically a two-dimensional grid) s_i which take values -1 and 1. The spins s_i interact in pairs with the total energy (Hamiltonian) given by the following expression:

$$E = -J \sum_{i,j} s_i s_j,$$

where J is the energy of interaction and the sum is over neighboring spins. The model is a statistical model, so the energy is the logarithm of the probability, that is, the probability of each configuration of spins is the Boltzmann distribution with inverse temperature β:

$$P(S) \propto e^{-\beta E}.$$

To generate configurations from this probability distribution, various reversible Markovian dynamics have been proposed, the most notable of them being the Metropolis algorithm and the Glauber dynamics.

Kinetically constrained spin models were first proposed in Fredrickson and Andersen, 1984 where the system is an Ising spin system with the usual Hamiltonian and the same equilibrium distribution and properties; the only difference being the transition probabilities by which the system evolves. Since the detailed balance condition does not uniquely determine the transition probabilities, there are multiple ways of defining transition probabilities satisfying detailed balance condition with respect to a particular equilibrium distribution.

In the facilitated spin models defined in Fredrickson and Andersen, 1984, a spin flips only if at least a certain fraction of its neighboring spins are up; for the one-spin facilitated model (FA1f) for instance, a spin flips only if at least one of its neighboring spins is up. A possible physical interpretation of the facilitated spin models in terms of real fluids is to identify *down* spins with lower than average compressibility and *up* spins with higher than average compressibility (Fredrickson and Andersen, 1984). With this interpretation, a highly compressible region can force a neighboring region to relax while a low compressibility region leads to "jamming" of the

dynamics of the system.

East model belongs to the class of kinetically constrained spin models, where the spin system is a one-dimensional spin system and a spin flips only if the spin to its left is up. Kinetically constrained spin models in general have been fairly successful in modeling the dynamics of glass transition. For the East model for instance, at low temperatures, an EITS form for the relaxation time has been reported in Sollich and Evans, 2003. We adopt the following definition of East model process in continuous time from Aldous and Diaconis, 2002.

Consider a one-dimensional chain of $\eta + 1$ spins σ_i for $i = 0, \ldots, \eta$, which can take values 0 and 1, that is, $\sigma_i \in \{0, 1\} \ \forall \ i$ with the leftmost spin σ_0 always 1 (we consider only finite systems, while in Aldous and Diaconis, 2002, the chain is infinite). The transition rate matrix P_{cont} in continuous time is given by (for $i > 0$):

$$P_{\text{cont}}(\sigma \to \sigma^{(i)}) = \begin{cases} p & \text{if } \sigma_{i-1} = 1 \text{ and } \sigma_i = 0 \\ 1 - p & \text{if } \sigma_{i-1} = 1 \text{ and } \sigma_i = 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $\sigma^{(i)}$ is the spin state obtained from σ by flipping σ_i. We can change the continuous time Markov process above to a discrete time one by the following: at each time, a spin $i > 0$ is selected at random and if the spin to its left is up, then σ_i, if previously 0, is set to 1 with probability p and if previously 1, is set to 0 with probability $1 - p$. Clearly, the discrete time probability matrix P_{disc} is related to the continuous time *rate* matrix P_{cont} as:

$$P_{\text{disc}} = I - \frac{1}{\eta} P_{\text{cont}}, \tag{6.1.2}$$

where I is the $2^\eta \times 2^\eta$ identity matrix. This matrix P_{disc} is also called the *embedded discrete time Markov chain*. The implication of equation (6.1.2) is that P_{disc} and P_{cont} have the same set of eigenvectors and the eigenvalues are related by the simple relation $\lambda_{\text{disc},i} = 1 - \frac{1}{\eta}\lambda_{\text{cont},i}$ for

$i = 1, \ldots, 2^\eta$. The reason we look at the relation between P_{disc} and P_{cont} is that we wish to consider the observable, henceforth termed the Aldous-Diaconis function, described in Aldous and Diaconis, 2002 as the starting observable f for the KSP method to estimate the spectral gap of P_{disc}. Even though the Aldous-Diaconis function has been used to prove lower bounds on relaxation time for P_{cont} in (Aldous and Diaconis, 2002), we can use that as the observable for our KSP method since the relation (6.1.2) implies that P_{disc} and P_{cont} have the same set of eigenvectors.

We chose the East model as a test case to demonstrate our method of estimating the spectral gap because of the following reasons:

- Kinetically constrained models are notoriously slow processes, that is, their spectral gaps are usually very small. For the east model in particular, the relaxation time, $\tau(p)$ defined as 1/(spectral gap) satisfies $\log \tau(p) \sim \frac{\log^2(1/p)}{\log 2}$ as $p \downarrow 0$. The task of estimating the spectral gap for these processes is very difficult from a numerical point of view.

- There is a known "good" starting observable for the East model. It has been proved in Aldous and Diaconis, 2002 that the Aldous-Diaconis observable has a relaxation time within a factor 2 of the upper bound and hence could have substantial overlap with the slowest mode, especially in the regime $p \downarrow 0$.

To start with, let us describe the Aldous-Diaconis function from Aldous and Diaconis, 2002 – note that the observable is defined in terms of a continuous time *coalescing random jumps* (CRJ) process which we describe now. For a particular configuration of spins σ, let $S_0(\sigma) \subseteq \{0, 1, 2, \ldots, \eta\}$ be the set of up spin sites in σ at time 0,

$$S_0(\sigma) = \{0 \leq i \leq \eta : \ \sigma_i = 1\}.$$

The CRJ process shrinks the set of up spins by the following rule: at time t, the up spin at site i goes down at rate $p^{D(S_t(\sigma),i)}$, where $D(S_t(\sigma), i)$ is the distance to the next up spin to the left

and is given by $D(S_t(\sigma), i) = \min\{|j - i| : j < i,\ j \in S_t(\sigma)\}$. Once the process is started with a spin configuration σ such that $S_0(\sigma) \neq \{0\}$, eventually there is a time when there is only one site (apart from site 0) that contains an up spin – let that random site be $L(\sigma)$. The value of the Aldous-Diaconis function is then

$$f_{AD}(\sigma) = f_{AD}(S_0(\sigma)) = \begin{cases} \text{Prob}(L(\sigma) > \eta/2), & \text{if } S_0(\sigma) \neq \{0\}, \\ 0 & \text{if } S_0(\sigma) = \{0\}. \end{cases}$$

Note that we use the notation $f_{AD}(\sigma)$ and $f_{AD}(S_0(\sigma))$ interchangeably since there is a one-to-one mapping between a spin configuration σ and its set of up spin sites $S_0(\sigma)$. Though the observable looks complicated, it turns out that $f_{AD}(S_0)$ can be written in terms of $f_{AD}(S_0 \setminus \{i\})$ for $i \in S_0$. The key idea is that given m independent exponential random variables with rates $\lambda_i, i = 1, \ldots, m$, the probability that the ith random variable is the minimum of these is $\lambda_i / \sum_{i=1}^{m} \lambda_i$. Given the set of up spin sites S_0, the time at which the spin i flips is an exponential random variable with rate $p^{D(S_0,i)}$ and hence the probability that it flips before all others is $p^{D(S_0,i)} / \sum_{i \in S_0} p^{D(S_0,i)}$. Hence we can write the following recurrence relation for $f_{AD}(S_0)$.

$$f_{AD}(S_0) = \frac{1}{\sum_{i \in S_0} p^{D(S_0,i)}} \sum_{i \in S_0} p^{D(S_0,i)} f_{AD}(S_0 \setminus \{i\}). \tag{6.1.3}$$

Equation (6.1.3) coupled with the relation $f_{AD}(\{0\}) = 0$ helps us evaluate $f_{AD}(S_0)$. A naive application of this procedure yields an $O(m!)$ procedure for evaluating f_{AD}, where $m = |S_0|$. A more efficient version of this procedure can be easily designed by storing the values of $f_{AD}(S)$ once they have been computed, for intermediate values of S. The complexity of this memoized procedure is then $O(2^m)$.

Now we consider using East model as a test case for the KSP method using the Aldous-Diaconis function as the observable. For our first experiment, the number of spins is chosen to be $\eta = 10$ and the probability of an up spin $p = 0.1$. For $\eta = 10$, the transition matrix of the

East model process is of size $2^{10} \times 2^{10}$, for which we can compute the eigenvalues and compare the spectral gap estimate we obtain with its exact value.

As we described for AR(1) process, we obtain least squares estimate $\hat{\lambda}_*^{LS}[n]$ for different pencil size parameter values n. For the East model with Aldous-Diaconis observable, like the AR(1) process we have seen earlier, the least squares estimate works better than the ML estimate. This is again because the bias for singleton estimates of $\lambda_*^{r_j^{(i)}}$, $\hat{\mu}_{i,j}$, is large for small $r_j^{(i)}$.

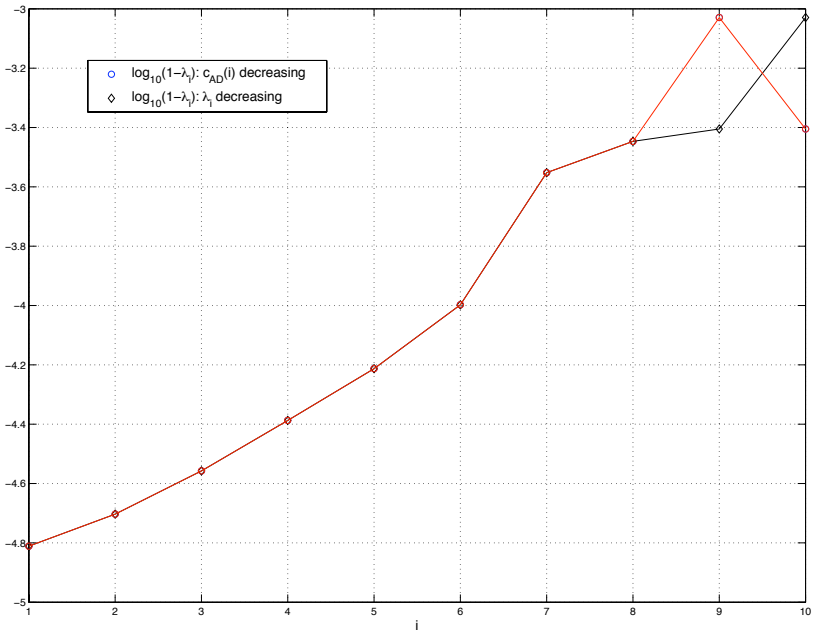

Figure 6.1: *Plot depicting the overlap of the Aldous-Diaconis function with the eigenvectors of transition matrix corresponding to the East model with* $\eta = 10, p = 0.1$. *Only the 10 largest eigenvalues (apart from 1) have been plotted.*

Before we investigate further how our estimate compares with the actual spectral gap, let us determine how good the Aldous-Diaconis function is as an observable for the KSP method.

In particular, does it have a good overlap with the slowest mode? Let the eigenvalues of the transition matrix P be λ_i with corresponding eigenvector v_i for $i = 1, 2, \ldots, 2^\eta$. The overlap of the Aldous-Diaconis function along the eigenvector v_i is given by the inner product $c_{AD}(i) = |\langle f_{AD}, v_i \rangle_\pi| / |\langle f_{AD}, f_{AD} \rangle_\pi|$. Figure 6.1 gives the log plots of the smallest values of gap(i); the circles depict $\log_{10}(\text{gap}(i))$ in decreasing order of $c_{AD}(i)$, while the diamonds depict $\log_{10}(\text{gap}(i))$ in decreasing order of λ_i.

It is clear from Figure 6.1 that the plots for $\log_{10}(\text{gap}(i))$ with $c_{AD}(i)$ decreasing and for $\log_{10}(\text{gap}(i))$ with λ_i decreasing coincide up to $i = 8$ – implying that the Aldous-Diaconis function has the maximum overlap with the eigenvectors corresponding to the largest eight eigenvalues. This means that the Aldous-Diaconis function is a good observable for the KSP method at least in the present case $\eta = 10, p = 0.1$.

Figure 6.2 plots the gaps of least squares estimates, $1 - \hat{\lambda}_*^{LS}[n]$, of the East model for different values of n for the particular choice $\eta = 10, p = 0.1$. It also plots the actual spectral gap (computed using linear algebra routines in MATLAB) and the gap of the Rayleigh quotient estimate, namely, $1 - \widehat{C}_f(1)/\widehat{C}_f(0)$). Note that as far as LS estimates are concerned, only the one for $n = 1$ is an outlier compared to other values of n. The estimate of spectral gap $1 - \lambda_*$ with the least squared error is the one with $n_{\text{lse}} = 3$ and is $1.6839 \times 10^{-5} \pm 3.1338 \times 10^{-7}$. The actual spectral gap computed using MATLAB is 1.5448×10^{-5}. Note that the KSP estimate of the spectral gap is 3 times smaller than the naive Rayleigh quotient estimate and is much closer to the actual answer.

The real power of this new method is apparent when we apply it to problems for which we *do not* know the actual answer – for which it is impossible to compute the spectral gap using linear algebra routines. As an example, we apply it to the East model for the configuration $\eta = 25, p = 1/25$. The transition matrix in this case is of size $2^{25} \times 2^{25}$. Figure 6.3 plots the estimated spectral gap using the KSP least squares method and the Rayleigh quotient method. It is apparent that

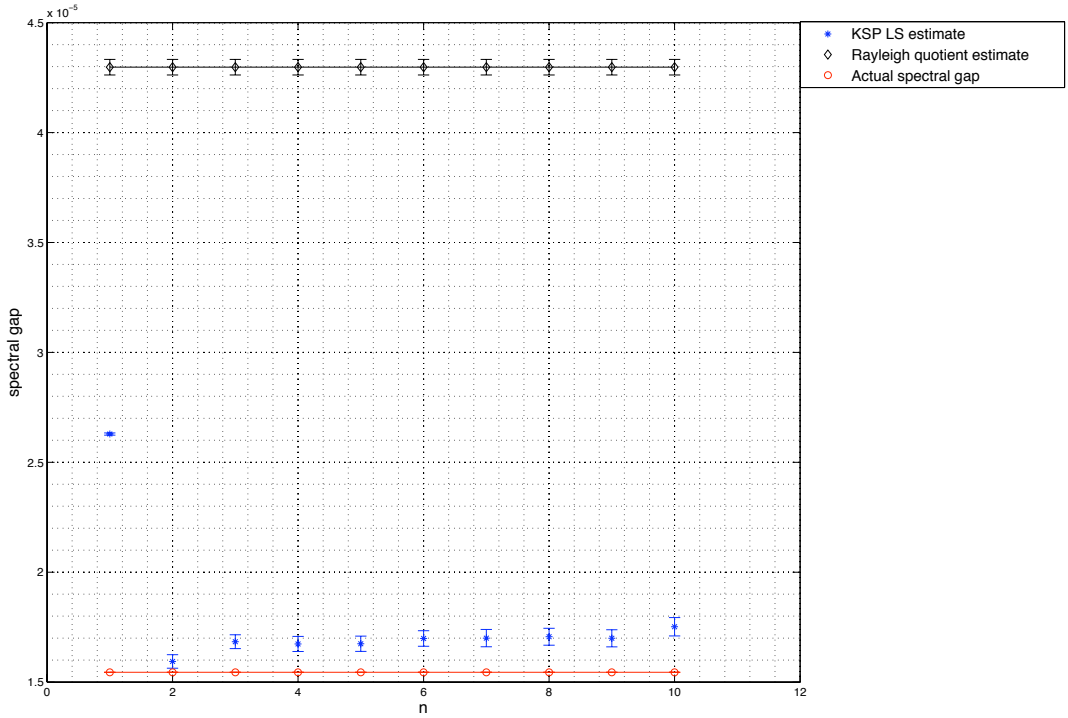

Figure 6.2: *Plot of the KSP LS estimate, the Rayleigh quotient estimate of the spectral gap and actual spectral gap for East model ($\eta = 10, p = 0.1$) with Aldous-Diaconis function.*

the Rayleigh quotient method gets progressively worse compared to the KSP method for smaller and smaller spectral gaps – in this case, the Rayleigh quotient estimate (6.858×10^{-7}) is almost an order of magnitude larger than the KSP estimate (2.1828×10^{-8}).

6.2 Fredrickson-Andersen model

As we mentioned in the last section, like the East model, Fredrickson-Andersen models belong to the class of kinetically constrained spin models; in the k-spin facilitated model (FA-kf) model

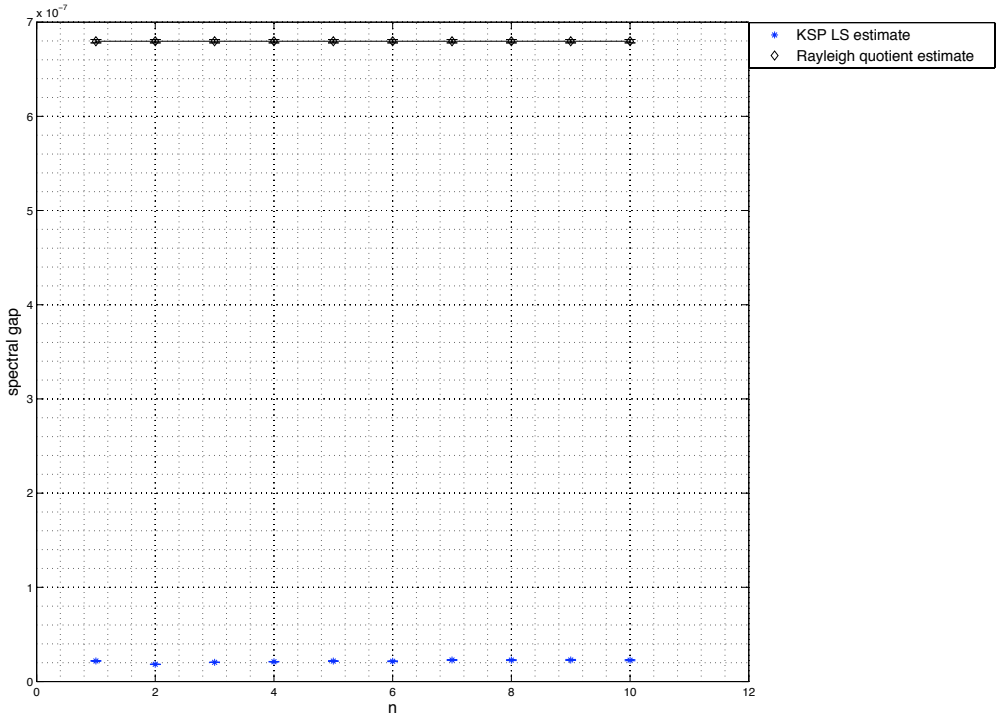

Figure 6.3: *Plot of the KSP LS estimate and the Rayleigh quotient estimate of the spectral gap for the East model ($\eta = 25, p = 1/25$) with Aldous-Diaconis function.*

for instance, a spin flips only if k of its neighbors are up. In this section, we apply the KSP method to estimate the spectral gap of the simplest model of this class, the FA-1f model in two dimensions.

Consider a set of η^d spins in d dimensions, σ_x for $x \in \{1, \ldots, \eta\}^d$, which can take values 0 and 1, that is, $\sigma_x \in \{0, 1\}$. A general kinetically constrained spin model can be described by a

transition matrix as follows:

$$P_{\text{cont}}(\sigma \to \sigma^{(x)}) = \begin{cases} p & \text{if } c_x(\sigma) \text{ and } \sigma_x = 0 \\ 1-p & \text{if } c_x(\sigma) \text{ and } \sigma_x = 1 \\ 0 & \text{otherwise,} \end{cases}$$

where $c_x(\sigma)$ is the constraint that needs to be satisfied for the spin at x to flip and $\sigma^{(x)}$ is the spin state obtained by flipping the spin at site x, σ_x. The constraint $c_x(\sigma)$ for FA-kf model is satisfied (true) if the number of up spin neighbors of x is at least k.

In Cancrini et al., 2007, asymptotically tight bounds on the spectral gap have been proved for the FA-1f model in two dimensions by substituting a carefully chosen test function into the variational characterization of the spectral gap. We describe the function now and use it as the starting observable for the KSP method and estimate the spectral gap. The description of this function has been adopted from Cancrini et al., 2007, Theorem 6.4. Although we state it for any dimension $d \geq 1$, our experiments are specifically for the case $d = 2$.

The boundary condition that we consider for the FA1f model in two dimensions is the zero boundary condition with the exception that the spin on the boundary adjacent to $\sigma_{(1,1)}$ is up. This ensures that the spin at location (1,1) is always "mobile", meaning ready to flip. This adjustment is similar to the East model, where we fix the spin to the left of σ_1 to always be up. If we had completely zero boundary conditions, then the spin configuration consisting of all down spins is an absorbing state, thus ensuring that the Markov chain is not ergodic on the set of all spin configurations.

Coming back to the observable for FA1f model, let p_0 be a large value of up-spin probability (chosen sufficiently close to 1). For the exact technical definition of p_0, please see Cancrini et al., 2007, Theorem 4.2. Let also

$$\ell_q = \left(\frac{\log(1-p_0)}{\log(1-p)}\right)^{1/d}. \tag{6.2.1}$$

Let g be a smooth function on $[0,1]$ with support in $[1/4, 3/4]$ and such that

$$\int_0^1 \alpha^{d-1} e^{-\alpha^d} g(\alpha) d\alpha = 0 \text{ and } \int_0^1 \alpha^{d-1} e^{-\alpha^d} g^2(\alpha) d\alpha = 1. \tag{6.2.2}$$

The function g has been so chosen for the purposes of the proof of Cancrini et al., 2007, Theorem

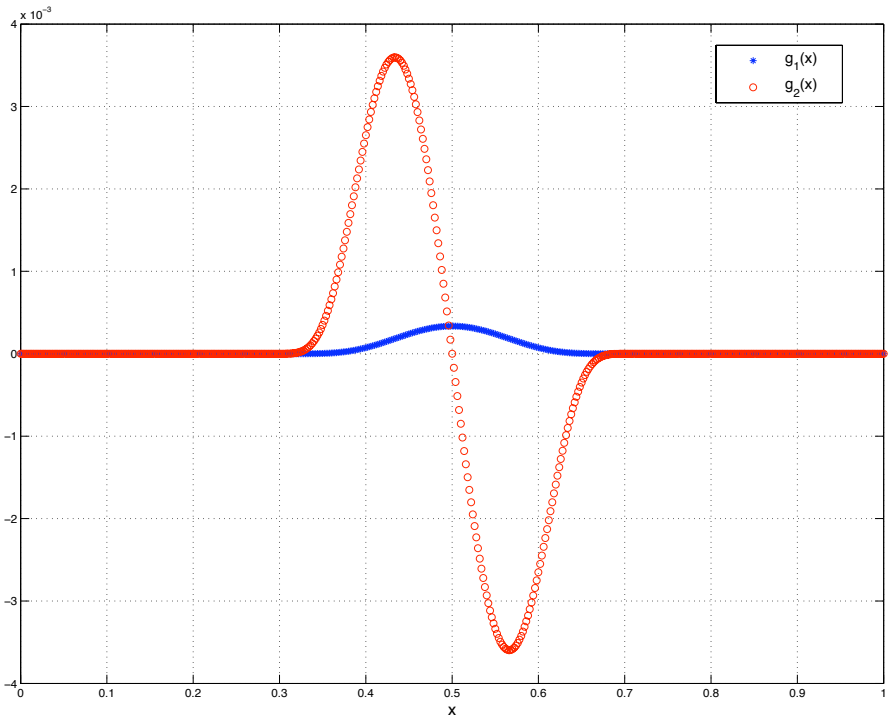

Figure 6.4: *Plots of functions g_1, g_2 defined in (6.2.3).*

4.2.

How to choose g to satisfy conditions in (6.2.2) has not been explained in Cancrini et al., 2007, which we do here. Two C^∞-continuous functions and with support in $[a, b]$ where $0 \leq$

$a, b \leq 1$ are given by

$$
\begin{aligned}
g_1(x) &= \begin{cases} 0 & \text{if } x \leq a \text{ or } x \geq b, \\ e^{-\frac{1}{x-a}-\frac{1}{b-x}} & \text{if } a < x < b. \end{cases} \\
g_2(x) &= \begin{cases} 0 & \text{if } x \leq a \text{ or } x \geq b, \\ e^{-\frac{1}{x-a}-\frac{1}{b-x}} \left(\frac{1}{(x-a)^2} - \frac{1}{(b-x)^2} \right) & \text{if } a < x < b. \end{cases}
\end{aligned} \tag{6.2.3}
$$

The functions g_1 and g_2 are plotted for $a = 1/4, b = 3/4$ in Figure 6.4. A function g satisfying the conditions in equation (6.2.2) can be defined as the linear combination of g_1 and g_2, namely,

$$g = (g_1 + \lambda g_2)/\mu, \tag{6.2.4}$$

where the parameters λ, μ are chosen to satisfy (6.2.2). They are evaluated numerically using a quadrature method.

Set

$$\xi(\sigma) = \sup\{\ell : \sigma(x) = 0 \text{ for all } x \text{ such that } \|x\|_\infty < \ell\}.$$

Note the difference in the definition of ξ from that given in Cancrini et al., 2007 – it is because the notation we use is slightly different. Whereas in Cancrini et al., 2007, a *down spin* induces mobility in a neighboring spin and p is the probability of a down spin, we use the more standard notation that an up spin induces mobility and p is the probability of an up spin.

Having defined ξ and g, the observable f is then simply

$$f_{\text{FA-1f}}(\sigma) = g(\xi(\sigma)/\ell_q), \tag{6.2.5}$$

where ℓ_q is as defined in (6.2.1).

For the FA-1f model with $\eta = 3, p = 0.1$, Figure 6.5 plots the overlap of the function $f_{\text{FA-1f}}$ with the eigenvectors of the transition matrix. To compute this overlap, the value $\eta = 3$ is one of the largest we can consider because the size of the matrix in the 2D model is $2^{\eta^2} \times 2^{\eta^2}$. It

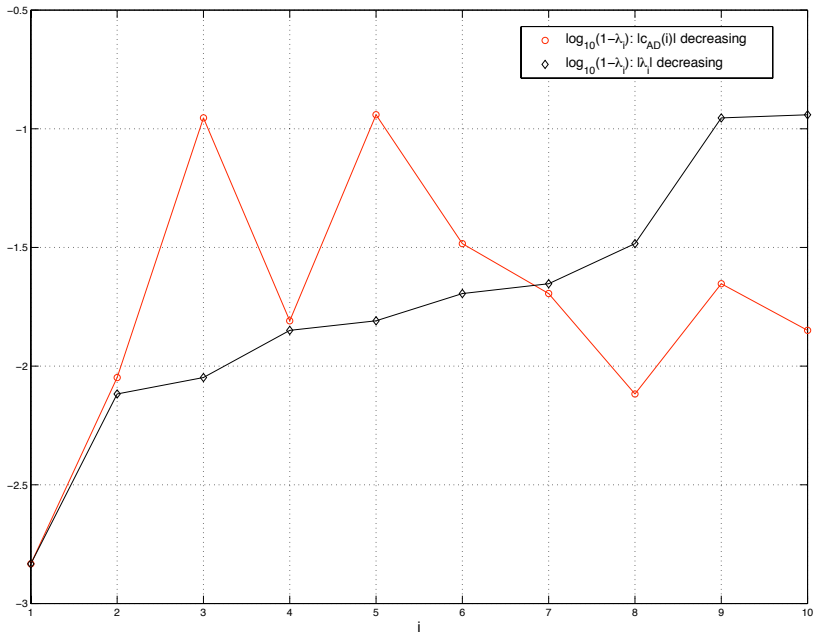

Figure 6.5: *Plot depicting the overlap of the function $f_{FA\text{-}1f}$ with the eigenvectors of transition matrix corresponding to the FA-1f model with $\eta = 3, p = 0.1$. Only the 10 largest eigenvalues (apart from 1) have been plotted.*

is clear from Figure 6.5 that the plot for $\log_{10}(1 - \lambda_i)$ with $c_{AD}(i)$ decreasing and the plot for $\log_{10}(1 - \lambda_i)$ with λ_i decreasing coincide only at $i = 1$ (in the range $i = 1, 2, \ldots, 10$).

This example is a good one to distinguish the KSP method a naive method, such as the Rayleigh quotient method, for estimating spectral gap. The Rayleigh quotient estimate, just being the weighted average of eigenvalues with the weights equal to the components of their eigenvectors along the observable, works fine if the eigenvalues with the maximum components are "close" to the largest eigenvalue (excluding 1). Intuitively speaking, the KSP method, on the other hand, requires that the observable has a significant overlap with the eigenvector cor-

responding to the largest eigenvalue, but does not depend on the "nearby" eigenvalues. Since it uses a Krylov subspace algorithm to find the maximum eigenvalue, it does a much better job of separating the largest eigenvalue from others.

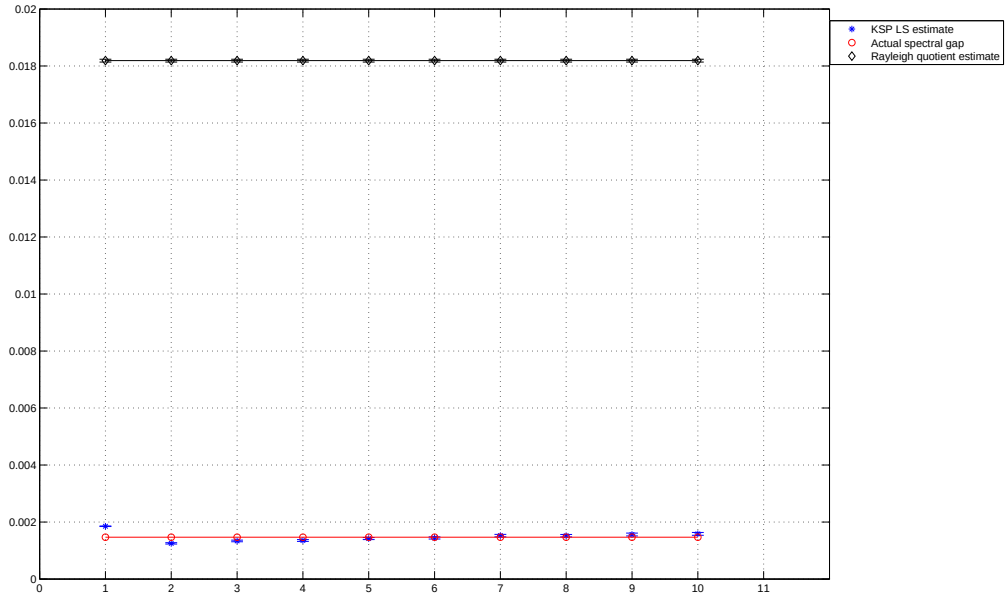

Figure 6.6: *Plot of the KSP LS estimate, the Rayleigh quotient estimate of the spectral gap and actual spectral gap for FA-1f model ($\eta = 3, p = 0.1$) with the function $f_{\text{FA-1f}}$ in (6.2.5).*

This is clearly demonstrated in Figure 6.6, where the KSP LS estimate of spectral gap has been plotted alongside the Rayleigh quotient estimate and the actual spectral gap (computed using MATLAB) for the FA-1f model with $\eta = 3, p = 0.1$. The KSP LS estimate is $1.4436 \times 10^{-3} \pm 3.8234 \times 10^{-5}$, with the actual spectral gap being 1.4685×10^{-3}. The Rayleigh quotient, on the other hand, is $1.8188 \times 10^{-2} \pm 4.9456 \times 10^{-5}$, which is almost ten times larger than the LS estimate.

We now fix the size of the two dimensional spin grid and vary the up-spin probability p; we wish to see how the spectral gap estimates vary with decreasing p. We then compare this variation with the asymptotic relation given in Cancrini et al., 2007, Theorem 6.4. The size of the grid is fixed as 20×20 and the up-spin probability takes values $p = \frac{1}{20}, \frac{1}{50}, \frac{1}{100}, \frac{1}{150}, \frac{1}{200}, \frac{1}{250}, \frac{1}{300}, \frac{1}{350}, \frac{1}{400}$. Figure 6.7 plots the log values of spectral gap vs. the log values of p for a 20×20 grid. We then fit a straight line to the resulting plot.

According to Cancrini et al., 2007, Theorem 6.4, the straight line fitting the log plot should have a slope of roughly 2. The observed slope is 2.255. The reason for this divergence from the true value is possibly because the KSP estimate of spectral gap for small values of p is not very accurate. Note that the gap estimate for $p = 1/400$ is *higher* than that for $p = 1/350$, which does not make physical sense. This is because for grid size fixed, decreasing p results in a slowing of the Markov chain, that is, a decreasing spectral gap.

One way to make the spectral gap estimate more accurate is to make the autocovariance estimates more accurate, in effect making the generalized eigenvalue estimation problem more accurate. One way to make the autocovariance estimates more accurate is to increase the batch size – using larger batch sizes can lead to more accurate estimates of spectral gap in theory, but manipulating such large data sets require enormous computing power and we do not undertake such an endeavor in this work.

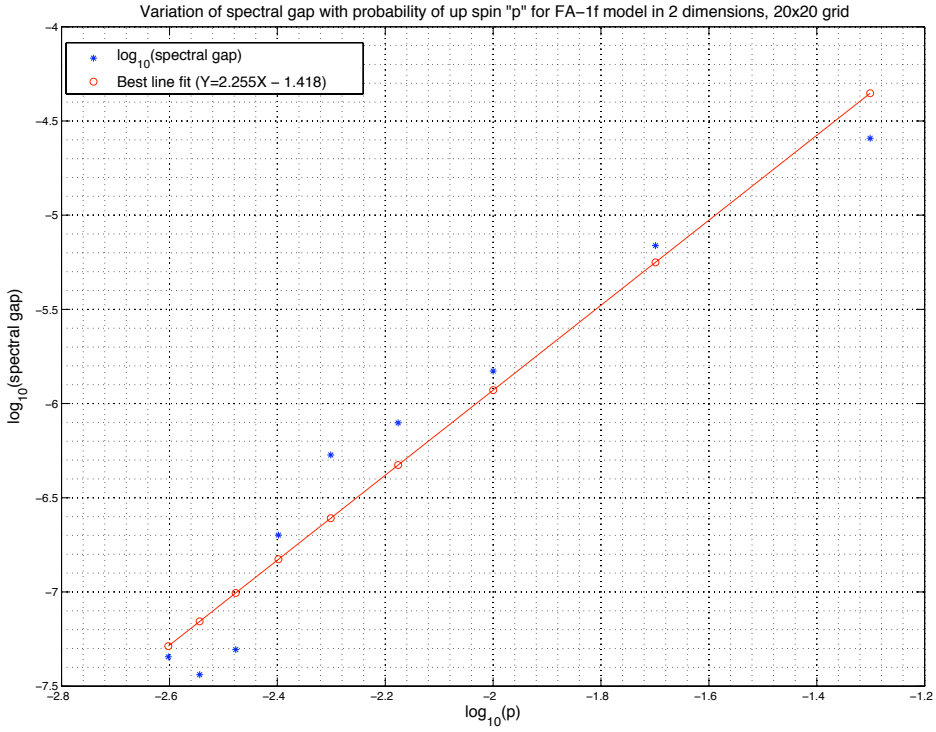

Figure 6.7: *Plot depicting the variation of the KSP estimate of spectral gap for FA1f process in two dimensions with varying up-spin probability p. The size of the grid is fixed to be 20×20 and p takes values* $\frac{1}{20}, \frac{1}{50}, \frac{1}{100}, \frac{1}{150}, \frac{1}{200}, \frac{1}{250}, \frac{1}{300}, \frac{1}{350}, \frac{1}{400}$.

7

MULTIPLE OBSERVABLES

In the previous sections, we considered the problem of estimating the reduced spectral radius of a reversible Markov chain using a single function as the observable. But suppose we have more than one observable with good overlap with the slowest mode – can we use all of these to improve the estimate obtained by using only one of these? It turns out that we can and we discuss that approach in this section.

Suppose we have k observables $f_1, f_2, \ldots, f_k$ each of which has overlap with the slowest mode. As before, consider the Krylov subspace

$$\mathcal{K}_n[\mathbf{f}] = \{P^{j-1} f_l : j = 1, \ldots, n \text{ and } l = 1, \ldots, k\}, \tag{7.0.1}$$

where $\mathbf{f} = \{f_1, \ldots, f_k\}$ is the set of observables. Any $u \in \mathcal{K}_n[\mathbf{f}]$ can be written as:

$$u = \sum_{l=1}^{k} \sum_{j=1}^{n} \xi_j^{(l)} P^{j-1} f_l,$$

for some $\xi_j^{(l)} \in \mathbb{R}$. The Rayleigh quotient of u is given by:

$$q(u) = \frac{\langle u, Pu \rangle_\pi}{\langle u, u \rangle_\pi}.$$

The denominator of this expression is

$$\begin{aligned} \langle u, u \rangle_\pi &= \sum_{l,m=1}^{k} \sum_{i,j=1}^{n} \xi_i^{(l)} \xi_j^{(m)} \langle P^{i-1} f_l, P^{j-1} f_m \rangle_\pi \\ &= \sum_{l,m=1}^{k} \sum_{i,j=1}^{n} \xi_i^{(l)} \xi_j^{(m)} \langle f_l, P^{i+j-2} f_m \rangle_\pi \\ &= \sum_{l,m=1}^{k} \sum_{i,j=1}^{n} \xi_i^{(l)} \xi_j^{(m)} C_{f_l, f_m}(i + j - 2), \end{aligned} \tag{7.0.2}$$

where C_{f_l,f_m} is the equilibrium *generalized autocovariance function* and is given by

$$C_{f_l,f_m}(s) = \text{cov}_\pi[f_l(X_0), f_m(X_s)]. \tag{7.0.3}$$

Equivalently, the expression $\langle u, Pu\rangle_\pi$ is given by

$$\langle u, Pu\rangle_\pi = \sum_{l,m=1}^{k} \sum_{i,j=1}^{n} \xi_i^{(l)} \xi_j^{(m)} C_{f_l,f_m}(i+j-1). \tag{7.0.4}$$

The Rayleigh quotient $q(u)$ for $u \in \mathcal{K}_n[\mathbf{f}]$ can be written as the generalized Rayleigh quotient of $nk \times nk$ matrices A and B where $A(ik+l, jk+m) = C_{f_l,f_m}(i+j-1)$ and $B(ik+l, jk+m) = C_{f_l,f_m}(i+j-2)$, that is, A and B have the form:

$$\begin{aligned} A &= \begin{bmatrix} C_{\mathbf{f}}(1) & C_{\mathbf{f}}(2) & \dots & C_{\mathbf{f}}(n) \\ C_{\mathbf{f}}(2) & C_{\mathbf{f}}(3) & \dots & C_{\mathbf{f}}(n+1) \\ C_{\mathbf{f}}(3) & C_{\mathbf{f}}(4) & \dots & C_{\mathbf{f}}(n+2) \\ \vdots & \vdots & \ddots & \vdots \\ C_{\mathbf{f}}(n) & C_{\mathbf{f}}(n+1) & \dots & C_{\mathbf{f}}(2n-1) \end{bmatrix}, \\ B &= \begin{bmatrix} C_{\mathbf{f}}(0) & C_{\mathbf{f}}(1) & \dots & C_{\mathbf{f}}(n-1) \\ C_{\mathbf{f}}(1) & C_{\mathbf{f}}(2) & \dots & C_{\mathbf{f}}(n) \\ C_{\mathbf{f}}(2) & C_{\mathbf{f}}(3) & \dots & C_{\mathbf{f}}(n+1) \\ \vdots & \vdots & \ddots & \vdots \\ C_{\mathbf{f}}(n-1) & C_{\mathbf{f}}(n) & \dots & C_{\mathbf{f}}(2n-2) \end{bmatrix}. \end{aligned} \tag{7.0.5}$$

Equation (7.0.5) is analogous to equation (3.2.8) except that here, $C_{\mathbf{f}}(i)$ is a block matrix of size $k \times k$.

All the estimators that we described for the case of a single observable (sections 3.2.2 and 5) extend directly to the multiple observable scenario. The only issue to be resolved is what values to use for the pencil size parameter set and the lag parameter set in the least squares estimator.

If the pencil size parameter is fixed to be n and the number of observables is k, then the size of matrices in the generalized eigenvalue problem is nk. If we stipulate that we do not use matrices larger than 10×10 in the generalized eigenvalue problem because of ill-conditioning, then (5.0.1) generalizes for the multiple observable case to:

$$\mathcal{N} = \left\{ 1, 2, \ldots, \left\lfloor \frac{10}{k} \right\rfloor \right\}. \tag{7.0.6}$$

As for the lag parameter set, the choice for the single observable case in (5.0.2) has the estimated autocorrelation time for that particular observable f. We extend (5.0.2) to the multiple observable scenario with $\hat{\tau}_{\mathrm{int},f}$ replaced by the minimum estimated autocorrelation time among all observables:

$$\mathcal{R}[n_i] = \left\{ 1, 2, \ldots, \left\lfloor \frac{c \left(\min\limits_{j} \hat{\tau}_{\mathrm{int},f_j} \right)}{2n_i - 1} \right\rfloor \right\} \text{ for } n_i = 1, 2, \ldots, \left\lfloor \frac{10}{k} \right\rfloor. \tag{7.0.7}$$

Using multiple observables improves the estimator in situations where none of the observables by itself gives an accurate estimate. A simple model problem that illustrates this is again the AR(1) process from section 4.1 with the following observables: $f_1 = \frac{1}{2}H_1 + H_2 + H_3 + H_4$, $f_2 = H_2 + H_3$, $f_3 = H_4$. Clearly, the only observable with an overlap with H_1, the slowest mode, is f_1, but which is less significant than the overlap that f_1 has with the other three observables. It is clear that none of f_1, f_2, f_3 is a good observable by itself for the KSP method, but using all three observables will definitely yield a better estimate because a simple linear combination of f_1, f_2, f_3 gives the slowest mode H_1: $2(f_1 - f_2 - f_3) = H_1$.

Figure 7.1 plots the least squares estimate for two scenarios: one, using multiple observables f_1, f_2, f_3 and the other using the single observable f_1. It is clear that the multiple observable estimate is more accurate, especially for $n = 1$, as expected.

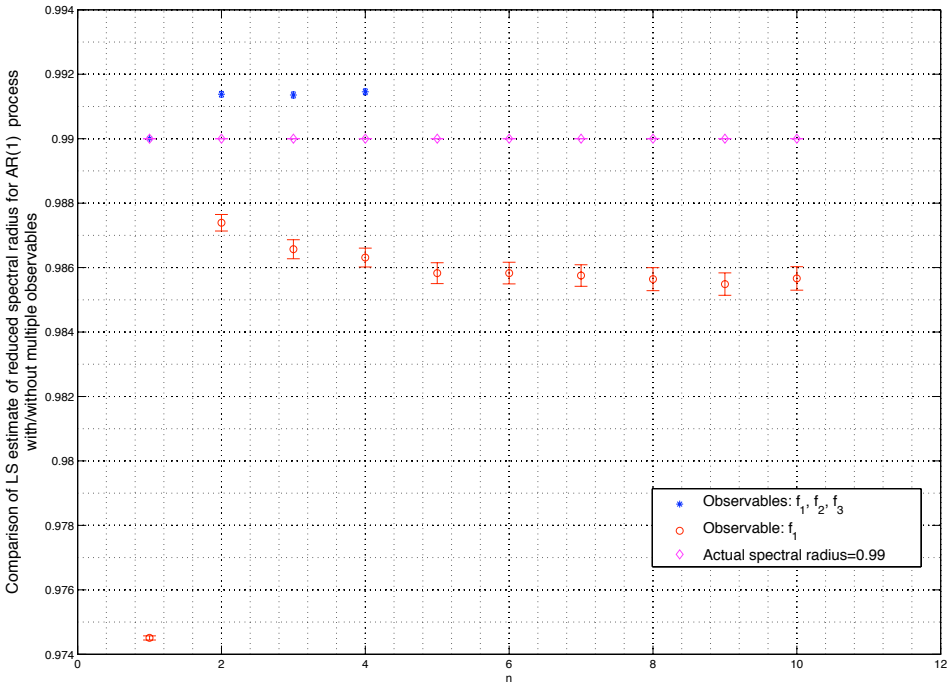

Figure 7.1: *Comparison between the LS estimates corresponding to the multiple observable case* $\{f_1, f_2, f_3\}$ *and the single observable* f_1 *for the AR(1) process, where* $f_1 = \frac{1}{2}H_1 + H_2 + H_3 + H_4$, $f_2 = H2 + H3$, $f_3 = H_4$ *(*H_i *is the* i*th Hermite polynomial).*

7.1 Results for Ising model with Glauber dynamics

The AR(1) model problem above has been designed specifically to show that using multiple observables is advantageous in some situations. But in real applications, we haven't found an example where this happens. We tried the specific example of Ising model with Glauber dynamics. Like the Metropolis algorithm, Glauber dynamics is a reversible Markov chain to generate samples from the Boltzmann distribution (see refer to section 6.1 for a brief introduction to the

Ising model). In Glauber dynamics, a spin is generated at random from the collection of all spins and is flipped with probability $\frac{1}{1+e^{\Delta H}}$, where ΔH is the energy difference due to the spin flip.

It is a known fact that for Glauber dynamics, total magnetization is a good "slow" mode – using it as the starting observable for the KSP method gives accurate estimate for the spectral gap, at least for the cases where the spectral gap is known exactly. Now that we have the framework for the multiple observable scenario, an interesting question to ask is whether using higher powers of magnetization as observables improve the estimator?

We investigate this point for Glauber dynamics on two-dimensional grids with periodic boundary conditions. Figures 7.2 and 7.3 report the results for 3×3 and 10×10 grids at temperature $T = T_c/4$, where T_c is the critical temperature for 2D Ising model. Figure 7.2 supports our earlier assertion that M is indeed a good observable for the KSP method (since we know the exact spectral gap for a 3×3 grid). It is clear from both of these figures that using higher powers of M does not improve the estimator of the spectral gap.

Even though we haven't found any real example where using multiple observables is better than using a single observable, the framework that we have developed is indeed very useful. Suppose that for a specific Markov chain, we know a set of observables with good overlap with the slowest mode – using the multiple observable framework, we can quantify this overlap precisely.

In this thesis, we describe a novel estimation method for the second largest eigenvalue of a reversible Markov chain using the simulation data from the chain. The method is empirically shown to be far superior to existing methods, both for model problems and some real problems which occur in statistical mechanics and glass transitions. In fact for a particular Markov chain with a tiny spectral gap, our estimate is more than 10 times smaller and more accurate than a naive Rayleigh quotient estimate (Figure 6.3).

Despite the overwhelming evidence that we present that testifies to the promise of this new

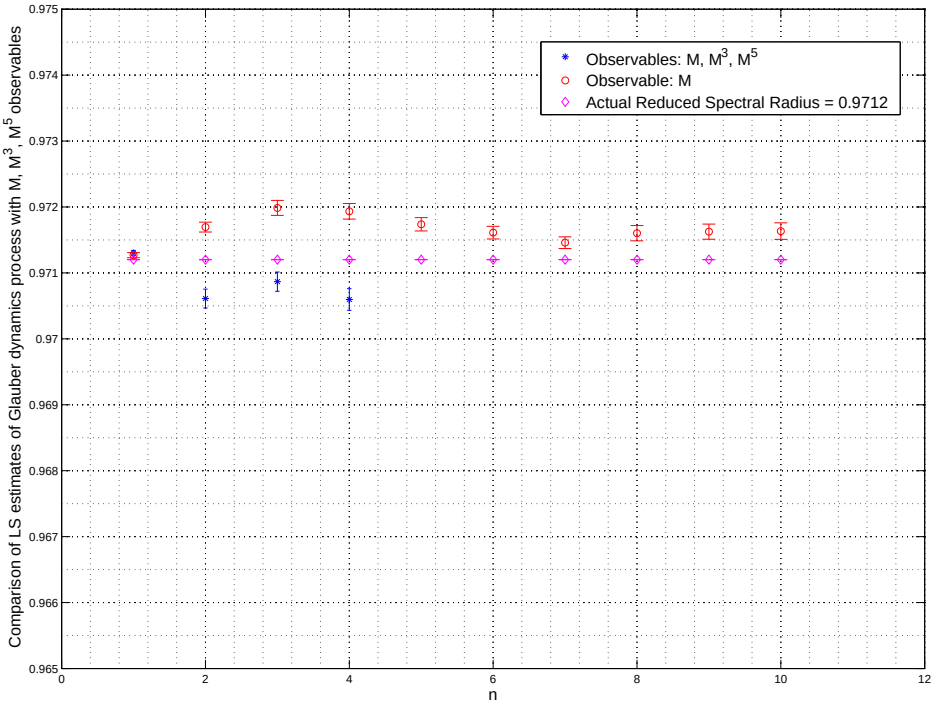

Figure 7.2: *Comparison between the LS estimates corresponding to the multiple observable case* $\{M, M^3, M^5\}$ *and the single observable* M *for the Glauber dynamics process, where* M *is the total magnetization. The spin collection is a 2D grid of size* 3×3 *with periodic boundary conditions at temperature* $T = T_c/4$*, where* T_c *is the critical temperature for 2D Ising model.*

method, we have to accept that the evidence is ultimately only empirical. We mostly give only intuitive explanations for why a particular heuristic works well in practice. Any thorough analysis gets extremely complicated. To give a perspective, to determine the distributional properties of the autocovariance estimates for a general Markov chain is a highly non-trivial problem (Anderson, 1971) – one has to make many simplifying assumptions (such as focusing only on specific simple processes like the AR(1) process) just to analyze the autocovariance estimates.

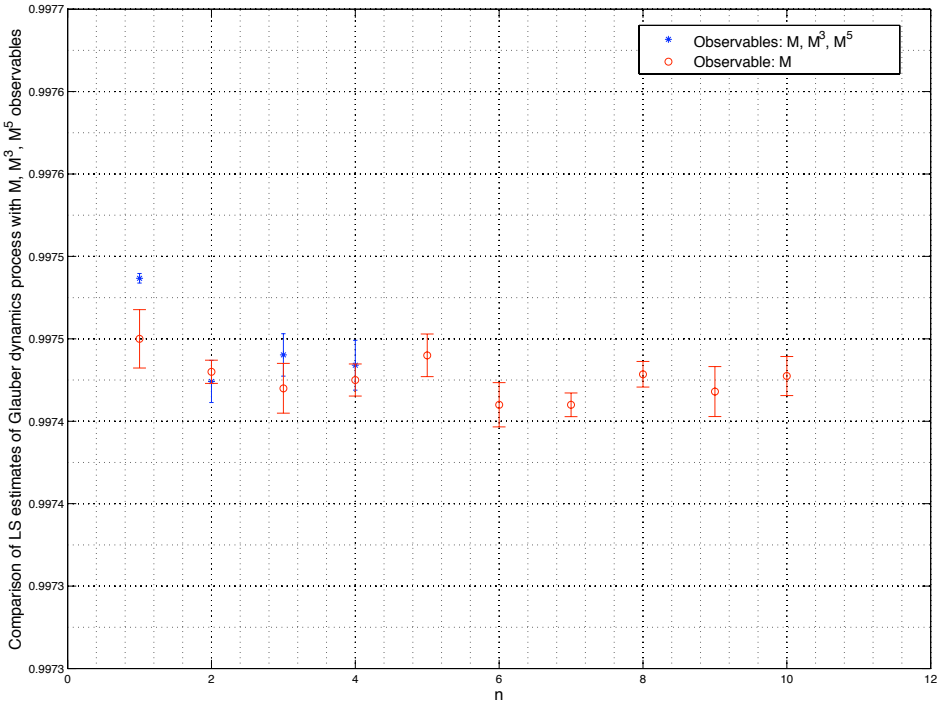

Figure 7.3: *Comparison between the LS estimates corresponding to the multiple observable case* $\{M, M^3, M^5\}$ *and the single observable* M *for the Glauber dynamics process, where* M *is the total magnetization. The spin collection is a 2D grid of size* 10×10 *with periodic boundary conditions at temperature* $T = T_c/4$*, where* T_c *is the critical temperature for 2D Ising model.*

On the other hand, what we have here is a highly ill-conditioned generalized eigenvalue problem (in most cases) with the autocovariance estimates as the entries in the matrices, so it is not surprising that the analysis gets very tricky.

Apart from the need for a more rigorous analysis, there are also many ways in which the method itself can be improved – such as, for instance, more informed choices for the lag parameter and the pencil size parameter. Although one particular choice may not work for all chains,

we need a more streamlined approach to tailor the choice of these parameters to the particular chain under consideration.

Although there is a huge literature regarding the second largest eigenvalue for Markov chains, it is mostly theoretical. The problem of numerically estimating it from the simulation data of the Markov chain has received scant attention. Our hope is that the apparent success of this new method will inspire more work in this direction.

CONCLUSION

In this thesis, we describe a novel estimation method for the second largest eigenvalue of a reversible Markov chain using the simulation data from the chain. The method is empirically shown to be far superior to existing methods, both for model problems and some real problems which occur in statistical mechanics and glass transitions. In fact for a particular Markov chain with a tiny spectral gap, our estimate is more than 10 times smaller and more accurate than a naive Rayleigh quotient estimate (Figure 6.3).

Despite the overwhelming evidence that we present that testifies to the promise of this new method, we have to accept that the evidence is ultimately only empirical. We mostly give only intuitive explanations for why a particular heuristic works well in practice. Any thorough analysis gets extremely complicated. To give a perspective, to determine the distributional properties of the autocovariance estimates for a general Markov chain is a highly non-trivial problem (Anderson, 1971) – one has to make many simplifying assumptions (such as focusing only on specific simple processes like the AR(1) process) just to analyze the autocovariance estimates. On the other hand, what we have here is a highly ill-conditioned generalized eigenvalue problem (in most cases) with the autocovariance estimates as the entries in the matrices, so it is not surprising that the analysis gets very tricky.

Apart from the need for a more rigorous analysis, there are also many ways in which the method itself can be improved – such as, for instance, more informed choices for the lag parameter and the pencil size parameter. Although one particular choice may not work for all chains, we need a more streamlined approach to tailor the choice of these parameters to the particular chain under consideration.

Although there is a huge literature regarding the second largest eigenvalue for Markov chains,

it is mostly theoretical. The problem of numerically estimating it from the simulation data of the Markov chain has received scant attention. Our hope is that the apparent success of this new method will inspire more work in this direction.

BIBLIOGRAPHY

Aldous, D. and Diaconis, P. (2002). The asymmetric one-dimensional constrained Ising model: rigorous results. *Journal of Statistical Physics*, 107(5):945–975.

Alexopoulos, C., Fishman, G. S., and Seila, A. F. (1997). Computational experience with the batch means method. In *Winter Simulation Conference*, pages 194–201.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

Beckermann, B., Golub, G., and Labahn, G. (2007). On the numerical condition of a generalized Hankel eigenvalue problem. *Numerische Mathematik*, 106:41–68.

Bremaud, P. (1999). *Markov chains, Gibbs fields, Monte Carlo simulation, and queues*. Springer-Verlag, New York, NY.

Cancrini, N., Martinelli, F., Roberto, C., and Toninelli, C. (2007). Kinetically constrained spin models. *Probability Theory and Related Fields*, 140:459–504.

Daniels, M. J. and Kass, R. E. (2001). Shrinkage estimators for covariance matrices. *Biometrics*, 57:1173–1184.

Demmel, J. and Kågström, B. (1993a). The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: robust software with error bounds and applications. Part I: theory and algorithms. *ACM Transactions on Mathematical Software*, 19:160–174.

Demmel, J. and Kågström, B. (1993b). The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: robust software with error bounds and applications. Part II: software and applications. *ACM Transactions on Mathematical Software*, 19:175–201.

Diaconis, P. and Saloff-Coste, L. (1995). What do we know about the Metropolis algorithm? In *STOC '95: Proceedings of the twenty-seventh annual ACM symposium on Theory of Computing*, pages 112–129, New York, NY, USA. ACM Press.

Diaconis, P. and Stroock, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1:36–61.

Dunford, N. and Schwartz, J. T. (1988). *Linear Operators Part II*. Wiley-Interscience, New York, NY.

Elad, M., Milanfar, P., and Golub, G. H. (2004). Shape from moments – an estimation theory perspective. *IEEE Transactions on Signal Processing*, 52(7):1814–1829.

Fredrickson, G. H. (1988). Recent developments in dynamical theories of the liquid-glass transition. *Annual Review of Physical Chemistry*, 39:149–180.

Fredrickson, G. H. and Andersen, H. C. (1984). Kinetic Ising model of the glass transition. *Physical Review Letters*, 53:1244–1247.

Fredrickson, G. H. and Andersen, H. C. (1985). Facilitated kinetic Ising models and the glass transition. *The Journal of Chemical Physics*, 83(11):5822–5831.

Gade, K. K. and Overton, M. L. (2007). Optimizing the asymptotic convergence rate of the Diaconis-Holmes-Neal sampler. *Advances in Applied Mathematics*, 38:382–403.

Garren, S. T. and Smith, R. L. (2000). Estimating the second largest eigenvalue of a Markov transition matrix. *Bernoulli*, 6(2):215–242.

Golub, G. H., Milanfar, P., and Varah, J. (2000). A stable numerical method for inverting shape from moments. *SIAM Journal on Scientific Computing*, 21(4):1222–1243.

Götze, W. (1989). Liquids, freezing and glass transition. In Hansen, J. P., Levesque, D., and Zinn-Justin, J., editors, *Proceedings of 51st Summer School in Theoretical Physics, Les Houches*, page 287.

Hua, Y. and Sarkar, T. K. (1990). Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(5):814–824.

Karlin, S. and McGregor, J. (1965). Ehrenfest urn models. *Journal of Applied Probability*, 2(2):352–376.

Kirkpatrick, T. R., Thirumalai, D., and Wolynes, P. G. (1989). Scaling and droplet notions for the dynamics of viscous liquids near an ideal glassy state. *Physical Review A*, 40:1045–1054.

Lawler, G. F. and Sokal, A. D. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. *Transactions of the American Mathematical Monthly*, 309(2):557–580.

Maxwell, M. and Woodroofe, M. (2000). Central limit theorems for additive functionals of Markov chains. *Annals of Probability*, 28(2):713–724.

Pritchard, G. and Scott, D. J. (2001). Empirical convergence rates for continuous-time Markov chains. *J. Appl. Probab.*, 38(1):262–269.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1):1–32.

Schuermans, M., Lemmerling, P., Lathauwer, L. D., and Huffel, S. V. (2006). The use of

total least squares data fitting in the shape-from-moments problem. *Signal Processing*, 86(5):1109–1115.

Sinclair, A. and Jerrum, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.*, 82:93–133.

Sokal, A. D. (1989). Monte Carlo methods in statistical mechanics: foundations and new algorithms. Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms, Cours de Troisieme Cycle de la Physique en Suisse Romande.

Sollich, P. and Evans, M. R. (2003). Glassy dynamics in the asymmetrically constrained kinetic Ising chain. *Physical Review E*, 68.

Turner, L. R. (1966). Inverse of the Vandermonde matrix with applications. Technical Report NASA TN D-3547, NASA.

Tyrtyshnikov, E. E. (1994). How bad are Hankel matrices? *Numerische Mathematik*, 67:261–269.

Zamarashkin, N. L. and Tyrtyshnikov, E. E. (2001). Eigenvalue estimates for Hankel matrices. *Sbornik : Mathematics*, 192(4):537–550.