CONSTRUCTING OPTIMAL TRANSITION MATRIX FOR MARKOV CHAIN MONTE CARLO VIA GLOBAL OPTIMIZATION

DRAFT AS OF November 30, 2014

SHENG-JHIH WU* AND MOODY T. CHU †

Abstract. The notion of asymptotic variance has been used as a means for performance evaluation of MCMC methods. An imperative task when constructing a Markov chain for the Monte Carlo simulation with prescribed stationary distribution is to optimize the asymptotic variance. Cast against an appropriately chosen coordinate system, the worst-case analysis of the asymptotic variance arising in MCMC can be formulated as a minimax problem. While the part of maximization is readily solvable as an eigenvalue problem, the part of minimization remains challenging. This work proposes employing global optimization techniques as a general means for numerical construction of the optimal transition matrix for MCMC methods. The framework of computation that eventual leads to an actual implementation is discussed in detail. Experimental results evidence the complexity of the underlying problem, including global solutions across the board for reversible problems and high occurrence of local solutions for nonreversible problems. In all, the approach via the global optimization techniques is a feasible and practical means for numerical construction of the optimal Markov chain.

Key words. MCMC, asymptotic variance, global optimization, minimax problem, inverse eigenvector problem, stochastic matrix

AMS subject classifications. 15B51, 49K35, 60J10, 65C05, 65F18, 90C26

1. Introduction.

For decades Markov chain Monte Carlo (MCMC) methods have been employed as a practical tool in a wide variety of applications such as Bayesian statistics, computational physics, genetics, and machine learning. See, for example, [3, 13, 23, 25, 26]. The methods become particularly useful when generating independent and identically distributed (i.i.d.) samples is not feasible or when the underlying distribution is not completely known. The basic idea underlying the MCMC is to construct a Markov chain with the desired distribution as its invariant distribution with the hope that, as the procedure runs long enough, the samples generated from the Markov chain serve as a good approximation to the would-be samples drawn from the unknown distribution. A key question to ask is how good an approximation is and the answer depends on the comparison criteria [2, 4, 18]. In the literature, one of the commonly employed measurements for gauging the performance of an MCMC algorithm is the so-called asymptotic variance which is the focus of this paper.

We briefly explain why the notion of asymptotic variance is a reasonable criterion for evaluating the performance of the MCMC methods. Let $S = \{1, 2, \cdots, n\}$ represent a finite state space and π be a probability distribution on S. It is often the case that we are interested in evaluating the expectation $\mathcal{E}(f) = \sum_{x\in S} f(x)\pi(x)$, where f is a real-valued function defined on S. When the closed-form calculation is not easy, we could appeal to the MCMC. Assume $X_0, X_1, \ldots$ is a discrete time Markov chain on S with some transition probability matrix P and invariant distribution π. We then use the time average $\frac{1}{n}\sum_{i=0}^{n-1} f(X_i)$ as an estimator to approximate the space average $\mathcal{E}(f)$, since by the strong law of large numbers we should have

$$\frac{\sum_{i=0}^{n-1} f(X_i)}{n} \xrightarrow{a.s.} \mathcal{E}(f). \tag{1.1}$$

In such a scenario, the asymptotic variance is defined by

$$\nu(f, P) := \lim_{n\to\infty} n\mathcal{E}_{\mu_0}\left[\frac{\sum_{i=0}^{n-1} f(X_i)}{n} - \pi(f)\right]^2, \tag{1.2}$$

where μ_0 denotes an arbitrary initial distribution. Obviously, smaller asymptotic variance indicates a better approximation. For this reason, various techniques have been proposed in the literature with the aim of reducing

*Center for Advanced Statistics and Econometrics Research, School of Mathematical Sciences, Soochow University, Suzhou, China (szwu@suda.edu.cn).

†Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205, USA. (chu@math.ncsu.edu.) This research was supported in part by the National Science Foundation under grants DMS-1014666 and DMS-1316779.

the asymptotic variance. Far from being complete, we refer to research endeavors reported in [5, 11, 14, 20, 22, 24]. To this date, the theory on the optimal asymptotic variance for reversible Markov chain is well established, but that for general case remains open [14, 19]. We propose to employ global optimization techniques to tackle this problem numerically. The goal of this paper is twofold — to reframe the problem under the context of linear algebra so as to facilitate the optimization procedure and to report some interesting observations.

We need not to repeat the tremendous importance of the stationary vector $\boldsymbol{\pi} \in \mathbb{R}^n$ satisfying $\boldsymbol{\pi}^\top P = \boldsymbol{\pi}^\top$ for a given row stochastic matrix P in the applications of Markov chains. On the other hand, in order to employ the MCMC process explained above, we need to have in hand an irreducible and aperiodic Markov chain such that the invariant distribution of the corresponding transition probability matrix P is the same as the prescribed distribution $\boldsymbol{\pi}$. Constructing a (row) stochastic matrix with a prescribed left eigenvector $\boldsymbol{\pi}$ associated with the Perron root 1 is a special type of the so called inverse eigenvalue problems [6]. More specifically, since the main concern in the construction is on maintaining the eigenvector $\boldsymbol{\pi}$, we refer to the current problem as an inverse eigenvector problem. In addition to the obvious structural constraint that P must be a row stochastic matrix, in the context of MCMC methods, we are also looking for a matrix P so that the corresponding asymptotic variance is minimized.

We should further clarify the meaning of optimal asymptotic variance before proceeding. Thus far, the definition $\nu(f, P)$ in (1.2) is case dependent, that is, it depends on the function f. So that an MCMC procedure is applicable to general cases, it is perhaps more reasonable to be concerned about the worst possible asymptotic variance among all possible f. In this paper, we investigate how such a worst-case analysis can be formulated for numerical calculation. Solving the inverse eigenvector problem with merely stochastic constraint is not difficult, but endeavors to minimize the worst-case asymptotic variance have been only partially successful. Though the theory is yet to be fully developed, numerical construction of such an optimal transition probability matrix P for the MCMC process could serve not only as a feasible alternative approach, but also have significant practical importance. In the following we explain how to use the global optimization techniques as a general means to construct such an optimal matrix P numerically.

The discoveries of our investigation can be summarized as follows:

1. For the reversible Markov chain, we find that, in addition to what is known as the classical solution [11, 14], the optimization problem may have many different optimizers. However, all of these solutions result in the same global optimal value, that is, all of them are global solutions and there are many.
2. Departing from the reversibility, we show numerically that the minimization of the worst-case problem for general stochastic matrices have multiple local solutions, suggesting that landscape of the objective function could be extremely complicated. Without resorting to the global optimization techniques, we demonstrate the high likelihood of producing a local optimizer and, hence, a less effective MCMC process.
3. We find that the optimal solution often occurs as the boundary of the feasible set. The Karush-Kuhn-Tucker (KKT) optimality conditions [10] will require us to discern active set from inactive set. The complexity involved in this process probably can only be resolved through numerical calculation. In contrast to the reversible case, it might be difficult, if possible at all, to obtain a closed-form solution for general irreversible Markov chains. Nonetheless, we offer to formulate the problem readily solvable through existent global optimization techniques.

This paper is organized as follows. In Section 2, we propose using weighted inner product (by $\boldsymbol{\pi}$) to facilitate the description of the asymptotic variance (1.2). The advantage is that the worst-case problem is transformed into a minimax problem, where the maximization is readily resolved by taking the logarithmic norm of some suitable restricted map. So it only remains to solve the minimization of the logarithmic norm, which is the core of the challenge. In Section 3 we briefly review the case of reversible MCMC. As the reversibility is equivalent to self-adjoint with respect to the weighted inner product, we can take advantage of many properties available from symmetry. In particular, a constructive proof in the form of a recurrence relation can be derived, which leads to a closed-form solution. In Section 4 we discuss the feasible set and the optimality condition in general. In particular, we set up the logistics needed for global optimization which leads to some interesting discoveries.

2. Worst-case analysis as a minimax optimization. We begin this section with the introduction of the concept of $\boldsymbol{\pi}$-orthogonality. Then we provide background of the worst-case analysis for asymptotic variance and formulate the maximin problem.

2.1. $\boldsymbol{\pi}$-orthogonality. To explain the idea, we first introduce the weighted inner product between arbitrary $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ via

$$\langle \mathbf{a}, \mathbf{b} \rangle_{\boldsymbol{\pi}} := \mathbf{b}^\top \Pi \mathbf{a} = \langle \Pi^{\frac{1}{2}} \mathbf{a}, \Pi^{\frac{1}{2}} \mathbf{b} \rangle, \tag{2.1}$$

where $\Pi := \mathrm{diag}\{\pi_1, \ldots, \pi_n\}$. It follows that with respect to the induced $\boldsymbol{\pi}$-norm

$$\|\mathbf{a}\|_{\boldsymbol{\pi}} := \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle_\pi} = \|\Pi^{\frac{1}{2}} \mathbf{a}\|_2,$$

the vector $\mathbf{1} := [1, \ldots, 1]^\top$ has unit length. The notion of $\boldsymbol{\pi}$-orthogonality also naturally arises. By (2.1), we say that $\mathbf{b} \perp_{\boldsymbol{\pi}} \mathbf{a}$ if and only if $\Pi^{\frac{1}{2}} \mathbf{b} \perp \Pi^{\frac{1}{2}} \mathbf{a}$, or $\mathbf{b} \perp \Pi \mathbf{a}$, or $\Pi \mathbf{b} \perp \mathbf{a}$. In particular, we see that

$$\mathbf{1} \perp_{\boldsymbol{\pi}} \mathbf{a} \Longleftrightarrow \boldsymbol{\pi} \perp \mathbf{a}. \tag{2.2}$$

Similarly, we can extend to the $\boldsymbol{\pi}$-Frobenius inner product between two matrices $A, B \in \mathbb{R}^{n \times m}$ by

$$\langle A, B \rangle_{F_{\boldsymbol{\pi}}} := \mathrm{trace}(B^\top \Pi A) = \sum_{i=1}^{n} \sum_{j=1}^{m} \pi_i a_{ij} b_{ij}. \tag{2.3}$$

Built upon the notion of mutually π-orthogonal columns, the matrix $V \in \mathbb{R}^{n \times n}$ of the form

$$V = [\mathbf{1}, \underbrace{\mathbf{v}_2, \ldots, \mathbf{v}_n}_{V_2}] \tag{2.4}$$

is $\boldsymbol{\pi}$-orthogonal if and only if the matrix

$$Q := \Pi^{\frac{1}{2}} V = [\Pi^{\frac{1}{2}} \mathbf{1}, \underbrace{\Pi^{\frac{1}{2}} \mathbf{v}_2, \ldots, \Pi^{\frac{1}{2}} \mathbf{v}_n}_{Q_2}] \tag{2.5}$$

is orthogonal. Since the first column $\Pi^{\frac{1}{2}} \mathbf{1}$ of Q is fixed, the columns of Q_2 can be taken to be the orthonormal basis of its orthogonal complement subspace. Once Q_2 is formed, then

$$V_2 = \Pi^{-\frac{1}{2}} Q_2. \tag{2.6}$$

Though it is not unique, note that V_2 is known once the desirable stationary distribution $\boldsymbol{\pi}$ is specified. For instance, suppose that $\boldsymbol{\pi} > 0$. Then it can readily be checked that, with $\mathbf{v}_1 = \mathbf{1}$, the vectors

$$\mathbf{v}_k = \frac{1}{\sqrt{\pi_{k-1}(\sum_{j=k}^{n} \pi_j)(\sum_{j=k-1}^{n} \pi_j)}} \left[\underbrace{0, \ldots, 0}_{k-2 \text{ times}}, \sum_{j=k}^{n} \pi_j, \underbrace{-\pi_{k-1}, \ldots, -\pi_{k-1}}_{n-k+1 \text{ times}} \right]^\top, \; k = 2, \ldots, n \tag{2.7}$$

form a $\boldsymbol{\pi}$-orthonormal basis for $\mathbb{R}^{n \times n}$. If so desire, we may use this basis in all the following discussion.

LEMMA 2.1. *The $\boldsymbol{\pi}$-orthogonal complement of the unit vector $\mathbf{1}$, denoted by $\mathbf{1}^{\perp_{\boldsymbol{\pi}}}$, is invariant under the transformation by any stochastic matrix P with stationary distribution $\boldsymbol{\pi}$.*

Proof. Suppose $\mathbf{w} \in \mathbf{1}^{\perp_{\boldsymbol{\pi}}}$. Then, by (2.2), $\langle \mathbf{w}, \boldsymbol{\pi} \rangle = 0$. On the other hand, we have $\langle P\mathbf{w}, \mathbf{1} \rangle_{\boldsymbol{\pi}} = \mathbf{1}^\top \Pi P \mathbf{w} = \langle \boldsymbol{\pi}, P\mathbf{w} \rangle = 0$, because $\boldsymbol{\pi}^\top P = \boldsymbol{\pi}^\top$. ∎

Therefore, it makes sense to consider the restricted map

$$P\big|_{\mathbf{1}^{\perp_{\boldsymbol{\pi}}}} : \mathbf{1}^{\perp_{\boldsymbol{\pi}}} \to \mathbf{1}^{\perp_{\boldsymbol{\pi}}}.$$

Corresponding to every element $\mathbf{w} \in \mathbf{1}^{\perp_{\boldsymbol{\pi}}}$, there is a unique $\mathbf{z} \in \mathbb{R}^{n-1}$ such that $\mathbf{w} = \Pi^{-\frac{1}{2}} Q_2 \mathbf{z}$. Using columns of V_2 as a $\boldsymbol{\pi}$=orthonormal basis for $\mathbf{1}^{\perp_{\boldsymbol{\pi}}}$, we can identify $P\big|_{\mathbf{1}^{\perp_{\boldsymbol{\pi}}}}$ as a linear map from $\mathbb{R}^{n-1}$ to $\mathbb{R}^{n-1}$ which has a matrix representation

$$P\big|_{\mathbf{1}^{\perp_{\boldsymbol{\pi}}}} = Q_2^\top \Pi^{\frac{1}{2}} P \Pi^{-\frac{1}{2}} Q_2 = V_2^\natural P V_2, \tag{2.8}$$

where the product

$$V_2^\natural := Q_2^\top \Pi^{\frac{1}{2}} \tag{2.9}$$

is only the so called $\{1,2,4\}$-inverse of V_2 and is not the Moore-Penrose pseudo-inverse $V_2^\dagger$ in general[1].

It is easy to see that with respect to the $\boldsymbol{\pi}$-inner product (2.1), the corresponding adjoint of P, denoted by $P^{\top_{\boldsymbol{\pi}}}$, is given by

$$P^{\top_{\boldsymbol{\pi}}} = \Pi^{-1} P^\top \Pi. \tag{2.10}$$

If P is a row stochastic matrix satisfying $\boldsymbol{\pi}^\top P = \boldsymbol{\pi}^\top$, then so is $P^{\top_{\boldsymbol{\pi}}}$.

2.2. Reformulation. Assuming the generic condition that the Perron root of P is unique, it is known that the asymptotic variance $\nu(f,P)$ can be expressed as

$$\nu(f,P) = 2\left\langle \left((I-P-\mathbf{1}\boldsymbol{\pi}^\top)^{-1} - \mathbf{1}\boldsymbol{\pi}^\top\right)(f-(\boldsymbol{\pi}^\top f)\mathbf{1}), f-(\boldsymbol{\pi}^\top f)\mathbf{1}\right\rangle_{\boldsymbol{\pi}} - \langle f-(\boldsymbol{\pi}^\top f)\mathbf{1}, f-(\boldsymbol{\pi}^\top f)\mathbf{1}\rangle_{\boldsymbol{\pi}}. \tag{2.11}$$

The readers are referred to [1, 16] for this expression. By re-centering f if necessary, we may assume without loss of generality henceforth that $\boldsymbol{\pi}^\top f = 0$. Then, for $f \in \mathbf{1}^{\perp_{\boldsymbol{\pi}}}$, the asymptotic variance is reduced to

$$\nu(f,P) = 2\left\langle (I-P)^{-1} f, f\right\rangle_{\boldsymbol{\pi}} - \langle f, f\rangle_{\boldsymbol{\pi}} \tag{2.12}$$

where the inverse $(I-P)^{-1}$ should be interpreted as the inverse of the restricted map $(I-P)\big|_{\mathbf{1}^{\perp_{\boldsymbol{\pi}}}}$. That is,

$$(I-P)^{-1}\big|_{\mathbf{1}^{\perp_{\boldsymbol{\pi}}}} = \left(I_{n-1} - V_2^\natural P V_2\right)^{-1}, \tag{2.13}$$

where we stress that I_{n-1} is the identity matrix of size $n-1$. Since $\nu(f,P)$ in (2.12) is homogeneous in f, it makes sense to assume further that $\|f\|_{\boldsymbol{\pi}} = 1$. Noticed that, for a fixed transition probability matrix P, asymptotic variance depends on the choice of the function f. In the literature, there are various ways to analyze the asymptotic variance over all functions f within classes of interest without exploiting any possible prior knowledge on these functions. The aspect of averaging the asymptotic variance over functions is taken in [4]. On the other hand, the worst-case analysis is employed in [11]. More precisely, the authors there consider the maximization of asymptotic variance over functions of interest as the criterion for evaluating the performance of MCMC algorithms. In the present paper, we adopt the worst-case analysis. Therefore, the worst scenario of the asymptotic variance for each fixed P amounts to finding

$$\sup_{f\in\mathbf{1}^{\perp_{\boldsymbol{\pi}}}, \|f\|_{\boldsymbol{\pi}}=1} \left\langle (I-P)^{-1} f, f\right\rangle_{\boldsymbol{\pi}}, \tag{2.14}$$

which we translate as the maximization problem

$$\max_{\mathbf{z}\in\mathbb{R}^{n-1}, \|\mathbf{z}\|_2=1} \left\langle \left(I_{n-1} - V_2^\natural P V_2\right)^{-1} \mathbf{z}, \mathbf{z}\right\rangle. \tag{2.15}$$

[1] A matrix $X \in \mathbb{R}^{n\times m}$ satisfying the properties that $AXA = A$, $XAX = X$, and $(XA)^\top = XA$ is called a $\{1,2,4\}$-inverse of the matrix $A \in \mathbb{R}^{m\times n}$. If, in addition, X also satisfies $(AX)^\top = AX$, then it is the unique Moore-Penrose pseudo-inverse of A. In the current context, $V_2 V_2^\natural$ is not symmetric, while $V_2^\dagger = \left(Q_2^\top \Pi^{-1} Q_2\right)^{-1} Q_2^\top \Pi^{-\frac{1}{2}}$ would be much more complicated than $V_2^\natural$.

Note that we have changed $\sup$ to $\max$ because the optimal value is attainable over the unit sphere. We note that throughout the rest of the present paper, the asymptotic variance is referred to the worst case of the asymptotic variance for simplicity of presentation. For convenience, introduce the abbreviation

$$R(P;\boldsymbol{\pi}) := \left(I_{n-1} - V_2^\natural P V_2\right)^{-1}. \tag{2.16}$$

At first glance, the notation $R(P;\boldsymbol{\pi})$ might seem redundant because once P is given, then its stationary distribution $\boldsymbol{\pi}$ is determined. In our acceleration problem, however, our task is to construct a matrix P with a specified $\boldsymbol{\pi}$. In this context, the matrix P is not known a priori and the expression $R(P;\boldsymbol{\pi})$ does depend on the parameter $\boldsymbol{\pi}$ in the way that V_2 is defined in (2.6).

It is clear from linear algebra that one way to characterization the maximization of (2.15) is via the symmetric eigenvalue formulation.

LEMMA 2.2. *The maximizer of (2.15) is attained at the unit eigenvector $\mathbf{z}_{max}$ associated with the largest eigenvalue λ_{max} of the symmetric matrix $R(P;\boldsymbol{\pi}) + R(P;\boldsymbol{\pi})^\top$ and the optimal objective value is the logarithmic norm of $R(P;\boldsymbol{\pi})$, that is, $\frac{\lambda_{max}}{2}$.*

The inverse problem of reconstructing a Markov chain P to accelerate asymptotic variance, therefore, can be formulated as

$$\begin{aligned} \min_{P} \quad & \tfrac{1}{2}\lambda_{max}\left(R(P;\boldsymbol{\pi}) + R(P;\boldsymbol{\pi})^\top\right), \\ \text{subject to} \quad & P \in \mathcal{N}(\boldsymbol{\pi}), \end{aligned} \tag{2.17}$$

where $\mathcal{N}(\boldsymbol{\pi})$ is the set

$$\mathcal{N}(\boldsymbol{\pi}) := \{P \in \mathbb{R}^{n\times n} \,\big|\, \boldsymbol{\pi}^\top P = \boldsymbol{\pi}^\top,\ P \text{ is row stochastic}\}. \tag{2.18}$$

Example 1. Consider the case $n = 2$. Write the row stochastic matrix P as

$$P = \begin{bmatrix} x & 1-x \\ y & 1-y \end{bmatrix},$$

where x and y are to be determined. With respect to a prescribed a stationary distribution $\boldsymbol{\pi} = [\pi_1, \pi_2]^\top$ satisfying $\pi_1 + \pi_2 = 1$, $\pi_1, \pi_2 > 0$, the feasible set $\mathcal{N}(\boldsymbol{\pi})$ can be identified as the segment

$$\mathcal{N}(\boldsymbol{\pi}) \equiv \left\{(x,y) \in \mathbb{R}^2 \big| \pi_1 x + \pi_2 y = \pi_1, 0 \le x, y \le 1\right\}.$$

See Figure 2.1. By (2.5), we can construct

$$V_2 = \begin{bmatrix} -\sqrt{\frac{\pi_2}{\pi_1}} \\ \sqrt{\frac{\pi_1}{\pi_2}} \end{bmatrix}$$

and obtain

$$V_2^\natural = \left[-\sqrt{\pi_1\pi_2}, \sqrt{\pi_1\pi_2}\right].$$

With $P \in \mathcal{N}(\boldsymbol{\pi})$, it follows that

$$R(P;\boldsymbol{\pi}) = \begin{cases} \frac{\pi_2}{1-x}, & \text{if} \quad \pi_2 > \pi_1; \\ \frac{\pi_1}{y}, & \text{if} \quad \pi_2 \le \pi_1; \end{cases}$$

achieves its minimum value π_2 at $(0, \frac{\pi_1}{\pi_2})$, if $\pi_2 > \pi_1$; or the minimum value π_1 at $(1 - \frac{\pi_2}{\pi_1}, 1)$, if $\pi_2 \le \pi_1$. It is interesting to note that the minimum is attained at the boundary of $\mathcal{N}(\boldsymbol{\pi})$.

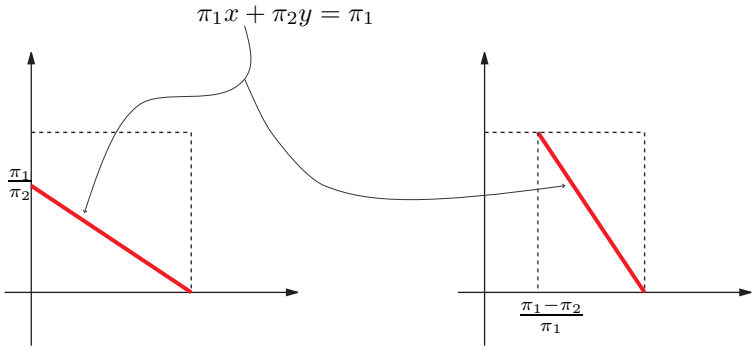

FIGURE 2.1. *Feasible set $\mathcal{N}(\boldsymbol{\pi})$ for the case $n = 2$. Left: $\pi_2 > \pi_1$; Right: $\pi_2 \leq \pi_1$.*

The above discussion indicates that the task of minimizing the asymptotic variance in a MCMC method is a minimax problem. In one formulation, the part of maximization is readily solved for every given P by means of the symmetric eigenvalue computation. It remains to solve the constrained minimization problem (2.17). In this approach, we have to deal with the intricate change of $\lambda_{max}\left(R(P;\boldsymbol{\pi}) + R(P;\boldsymbol{\pi})^\top\right)$ relative to the change in P. Alternatively, we could also work on the saddle-point problem

$$\min_{P\in\mathcal{N}(\boldsymbol{\pi})} \max_{\mathbf{z}\in\mathcal{S}^{n-2}} \left\langle \left(I - V_2^\natural P V_2\right)^{-1} \mathbf{z}, \mathbf{z} \right\rangle, \tag{2.19}$$

where $\mathcal{S}^{n-2}$ stands for the unit sphere in $\mathbb{R}^{n-1}$ and we treat $\mathbf{z}$ and P independently with the hope of avoiding stagnation at a local solution prematurely. Though techniques are available for the minimax problem (2.19), our emphasis in this paper is on using the global optimization techniques. Since the global solution is to be found, it no longer matters whether the maximization problem is solved first or not. It suffices to tackle the constrained optimization problem (2.17).

By now the reference to the given stationary distribution $\boldsymbol{\pi}$ is clear. We abbreviate $\mathcal{N}(\boldsymbol{\pi})$ as $\mathcal{N}$ and $R(P;\boldsymbol{\pi})$ as $R(P)$ hereafter.

3. Reversible MCMC. Let a matrix P be denoted in terms of its entries by writing $P = [p_{ij}]$. A row stochastic matrix P is said to be reversible relative to $\boldsymbol{\pi}$ if and only if it satisfies the so called detailed balance condition,

$$\Pi P = P^\top \Pi.$$

Equivalently, P is reversible if and only if $P = P^{\top_{\boldsymbol{\pi}}}$, that is, P is self-adjoint respect to the $\boldsymbol{\pi}$-inner product (2.1). Similar to symmetric matrices, it is readily verifiable that all eigenvalues of a reversible stochastic matrix are real-valued, and that there exists a basis consisting of mutually $\boldsymbol{\pi}$-orthogonal eigenvectors of P. Exploiting the relatively simpler eigenstructure, research efforts on other topics such as mixing time or mean recurrence time for reversible Markov chains have been extensive and fruitful. For some general comparison between reversible and non-reversible Markov chains in MCMC, see the discussions in [5, 8, 20].

Many popular MCMC methods also assume reversibility. The reason is that if we limit our construction of a Markov chain to the class of reversible matrices, then the minimum asymptotic variance is achievable via a closed-form solution. We can deal with the formulation (2.15) directly without resorting to the formulation (2.17). Furthermore, since eigenvectors that are $\boldsymbol{\pi}$-orthogonal to $\mathbf{1}$ are in the same space spanned by V_2, finding the reversible P that minimizes the asymptotic variance is equivalent to finding the reversible P whose second largest eigenvalue λ_2 is minimized. In fact, the rate of convergence of a reversible Markov chain to its equilibrium is also related to the second eigenvalue λ_2 [7, 9, 12, 15, 28].

Renaming the state variable if necessary, we may assume without loss of generality that the given stationary distribution $\boldsymbol{\pi}$ has been arranged in the order $0 < \pi_1 \leq \pi_2 \leq \ldots \leq \pi_n$. Then it has been established that one such a reversible solution matrix can be characterized as follows[2]. Define

$$P^{(1)} := \left[\begin{array}{c|ccc} 0 & \frac{\pi_2^{(1)}}{\beta_2} & \frac{\pi_3^{(1)}}{\beta_2} & \frac{\pi_n^{(1)}}{\beta_2} \\ \hline 1-\alpha_2 & & & \\ \vdots & & \alpha_2 P^{(2)} & \\ 1-\alpha_2 & & & \end{array}\right] \tag{3.1}$$

and recursively

$$P^{(k)} := \left[\begin{array}{c|ccc} 0 & \frac{\pi_{k+1}^{(k)}}{\beta_{k+1}} & \cdots & \frac{\pi_n^{(k)}}{\beta_2} \\ \hline 1-\alpha_{k+1} & & & \\ \vdots & & \alpha_{k+1} P^{(k+1)} & \\ 1-\alpha_{k+1} & & & \end{array}\right], \quad k = 2, \ldots, n-1 \tag{3.2}$$

where

$$\begin{aligned} \boldsymbol{\pi}^{(1)} &:= \boldsymbol{\pi} \\ \boldsymbol{\pi}^{(k+1)} &:= \frac{\boldsymbol{\pi}^{(k)}}{\beta_{k+1}}, \quad k = 1, \ldots n-2, \end{aligned}$$

and

$$\begin{aligned} \alpha_{k+1} &:= 1 - \frac{\pi_k^{(k)}}{\beta_{k+1}} \\ \beta_{k+1} &:= 1 - \pi_k^{(k)}, \quad k = 1, \ldots, n-2. \end{aligned}$$

Then the optimizer P whose second largest eigenvalue λ_2 is minimal among all reversible matrices is given by

$$P = P^{(1)} + \left(1 - \sum_{j=1}^{n-1} P_{nj}^{(1)}\right) \mathbf{e}_n \mathbf{e}_n^\top,$$

where $\mathbf{e}_n$ is the n-th standard unit vector [11, Theorem 1]. Besides the Perron root of unity, the other eigenvalues of P are precisely $\alpha_k - 1$, $k = 2, \ldots n$, which are all negative and λ_2 is the largest among these negative values. Similar to Example 1, it is interesting to note that the minimal asymptotic variance for reversible Markov chains is attained at the boundary of $\mathcal{N}$.

4. Numerical approach. For general stochastic matrices, the construction of a Markov chain in closed form with minimal asymptotic variance has been an open question. In this section, we approach this problem (2.17) by conventional numerical optimization techniques which enable us to gain some interesting insight into the problem. In particular, we offer numerical evidence that

1. Without the reversibility, there are many local solutions to the problem (2.17). Consequently, finding the absolute minimal asymptotic variance is challenging in general.
2. The reversible problem might have different optimizers, but all of them are global.
3. Existent global optimization techniques can help to construct the optimal Markov chain numerically, even if the theory of a closed-form solution is not available.

[2]A Markov chain with prescribed stationary distribution and minimal asymptotic variance is not unique, as we shall see from Example 4 in Section 4.3.

4.1. Feasible set. We first observe a few topological properties of the feasible $\mathcal{N}$. For most optimization software, it suffices to realize that the constraints imposed upon (2.17) are either linear equality or box constraints. However, the feasible set $\mathcal{N}$ has more innate structures. It is a monoid, that is, it is closed under matrix multiplication, has the identity, but it is not a group since not all its elements are invertible. It is compact and convex. Also, $\mathcal{N}$ has boundaries which turn out to be where the difficulty resides. The tangent space $\mathscr{T}_P\mathcal{N}$ of $\mathcal{N}$ at an interior point $P \in \mathcal{N}$ can be characterized as follows, which shows that $\mathscr{T}_P\mathcal{N}$ is independent of P and, hence, the set $\mathcal{N}$ is an affine subset.

LEMMA 4.1. *Suppose P is an interior point to $\mathcal{N}$. Then the tangent space $\mathscr{T}_P\mathcal{N}$ of $\mathcal{N}$ at P can be parameterized by $(n-1)^2$ free parameters. Specifically, any $W \in \mathscr{T}_P\mathcal{N}$ is of the form*

$$W = V_2 Z \Xi \tag{4.1}$$

where $Z \in \mathbb{R}^{(n-1)\times(n-1)}$ is arbitrary and $\Xi := [I_{n-1}, -\mathbf{1}_{n-1}] \in \mathbb{R}^{(n-1)\times n}$, where the identity matrix I_{n-1} and the vector $\mathbf{1}_{n-1}$ of all 1's are of size $n-1$.

Proof. The matrix $W \in \mathscr{T}_P\mathcal{N}$ if and only if $\boldsymbol{\pi}^\top W = 0$ and $W\mathbf{1} = 0$. The first equation implies that columns of W are in $\mathbf{1}^{\perp_{\boldsymbol{\pi}}}$. The second equation implies that the last column of W is the negative sum of the preceding $n-1$ columns. ∎

We also characterize how an arbitrary vector can be projected onto $\mathscr{T}_P\mathcal{N}$. Note that the projection is taken with respect to the $\boldsymbol{\pi}$-Frobenius inner product for the sake of convenience.

LEMMA 4.2. *Given any $X \in \mathbb{R}^{n\times n}$, its projection onto $\mathscr{T}_P\mathcal{N}$ with respect to the $\boldsymbol{\pi}$-Frobenius inner product is given by*

$$\mathscr{P}_{\mathscr{T}_P\mathcal{N}} X = \Pi^{-\frac{1}{2}} Q_2 Q_2^\top \Pi^{\frac{1}{2}} X \Omega = V_2 V_2^\natural X \Omega, \tag{4.2}$$

with

$$\Omega := \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & & \\ \vdots & & \ddots & \vdots \\ -\frac{1}{n} & & & 1-\frac{1}{n} \end{bmatrix}$$

Proof. By Lemma 4.1, the projection of X onto the tangent space $\mathscr{T}_P\mathcal{N}$ with respect to the $\boldsymbol{\pi}$-Frobenius inner product is equivalent to minimizing $\|X - V_2 Z\Xi\|_{F_{\boldsymbol{\pi}}}$ by suitable $Z \in \mathbb{R}^{(n-1)\times(n-1)}$, which we rewrite as

$$\|X - V_2 Z\Xi\|_{F_{\boldsymbol{\pi}}} = \|\Pi^{\frac{1}{2}} X - \Pi^{\frac{1}{2}} V_2 Z\Xi\|_F = \|\Pi^{\frac{1}{2}} X - Q_2 Z\Xi\|_F = \|Q_2^\top \Pi^{\frac{1}{2}} X - Z\Xi\|_F.$$

The least squares solution to the last objective function is given by $Z = Q_2^\top \Pi^{\frac{1}{2}} X \Xi^\dagger$. It is easy to check that $\Omega = \Xi^\dagger \Xi$. ∎

4.2. Optimality conditions. For easy reference, let $S : \mathbb{R}^{n\times n} \to \mathbb{R}^{(n-1)\times(n-1)}$ be the map

$$S(P) := \frac{1}{2}\left(R(P) + R(P)^\top\right).$$

Note that $S(P)$ is symmetric. Consequently, the maximal eigenvalue of $S(P)$, denoted as $\lambda_{max} : \mathbb{R}^{n\times n} \to \mathbb{R}$, is differentiable in P [17, 21]. We know that the Fréchet derivative of λ_{max} of $S(P)$ at P acting on a general matrix $H \in \mathbb{R}^{n\times n}$ satisfies the relationship

$$\frac{\partial \lambda_{max}(S(P))}{\partial P}.H = \left\langle \mathbf{x}(P), \left(\frac{\partial S(P)}{\partial P}.H\right)\mathbf{x}(P)\right\rangle, \tag{4.3}$$

where $\mathbf{x}(P)$ is the unit eigenvector of $S(P)$ associated with the eigenvalue $\lambda_{max}(S(P))$. We can calculate from (2.16) that

$$\frac{\partial S(P)}{\partial P}.H = \frac{1}{2}\left(R(P)V_2^\natural HV_2R(P) + \left(R(P)V_2^\natural HV_2R(P)\right)^\top\right). \tag{4.4}$$

Upon substitution into (4.3), we see from the Riesz representation theorem that the gradient of λ_{max} with respect to the Frobenius inner product (2.3) can be expressed as

$$\nabla\lambda_{max}(S(P)) = {V_2^\natural}^\top R(P)^\top \mathbf{x}(P)\mathbf{x}(P)^\top R(P)^\top V_2^\top. \tag{4.5}$$

Equivalently, if the $\boldsymbol{\pi}$-Frobenius inner product is used to represent the action of the Fréchet derivative, then we write the gradient as $\nabla_{\boldsymbol{\pi}}\lambda_{max} = \Pi^{-1}\nabla\lambda_{max}$.

When P is in the interior of $\mathcal{N}$, the projection of $-\nabla\lambda_{max}$ onto the tangent space $\mathscr{T}_P\mathcal{N}$ points to a steepest descent direction. When P is on the boundary of $\mathcal{N}$ and $-\nabla\lambda_{max}$ points away from $\mathcal{N}$, we need to take the projection of $-\nabla\lambda_{max}$ onto the tangent space of the boundary of $\mathcal{N}$. This requirement further complicates the optimality condition. For instance, by Lemma 4.2, we have

$$\mathscr{P}_{\mathscr{T}_P\mathcal{N}}\left(-\nabla\lambda_{max}\right) = -V_2R(P)^\top\mathbf{x}(P)\mathbf{x}(P)^\top R(P)^\top V_2^\top\Omega \tag{4.6}$$

If there is a matrix $P \in \mathcal{N}$ that zeros out the right-hand side of (4.6), then definitely it is a local critical point for the objective function $\lambda_{max}(S(P))$. However, it is possible, and seemingly always the case for our problem, that a minimizer for λ_{max} occurs at the boundary where (4.6) is not zero.

Example 2. We already know the solution to (2.17) for Example 1 is obvious. In particular, we know that the boundary of the feasible set is active. Suppose that we carry out the steps outlined above. The 2×2 projected gradient corresponding to (4.6) becomes

$$\mathscr{P}_{\mathscr{T}_P\mathcal{N}}\left(-\nabla\lambda_{max}\right) = -\frac{1}{2}\begin{bmatrix} \frac{\pi_2^2}{\pi_1(1-x)^2} & -\frac{\pi_2^2}{\pi_1(1-x)^2} \\ \frac{-\pi_2}{(1-x)^2} & \frac{\pi_2}{(1-x)^2} \end{bmatrix},$$

which clearly shows that at a boundary point the projected gradient is never zero. So merely checking the derivative information is not enough.

We can continue to explore the projection of the gradient onto the tangent space of the boundary of $\mathcal{N}$ and further refine the optimality under the KKT conditions. Such a theory has been well established in the field of constrained optimization. There is no need to reproduce the results here. Above all, we think that the KKT conditions for the general problem (2.17) would be so involved that a closed-form solution similar to that for the reversible problem might not be immediately available. Nonetheless, having an optimal transition probability matrix will effectuate the performance of an MCMC process. The demand for such a construction is indispensable. For practical purpose, we move ahead to finish the construction numerically by utilizing the many available optimization software packages in which the various techniques such as checking the active set and so on have already been built. Many of these codes can be more efficient in calculation if the analytic form of $\nabla\lambda_{max}$ in (4.5) is given.

4.3. Multiple Solutions. We use the routine fmincon in MATLAB as the optimization tool to experiment with the problem (2.17). For demonstration, we invoke the sequential quadratic programming (sqp) algorithm which computes a quasi-Newton approximation to the Hessian of the Lagrangian, where the gradient (4.5) of the objective function is supplied. Termination tolerance on the function value (TolFun), tolerance on the constraint violation (TolCon), and termination tolerance on P (TolX) are all set at 10^{-12} for high precision calculation. We call for the multi-start global optimization solver MultiStart with 100 starting points each of which is required to respect the box constraint initially and then is corrected to meet the feasibility within the prescribed tolerance. MultiStart runs the local solver sqp, starting at the points uniformly but randomly distributed within bounds [27]. A positive exit flag indicates a successful return from the local solver.

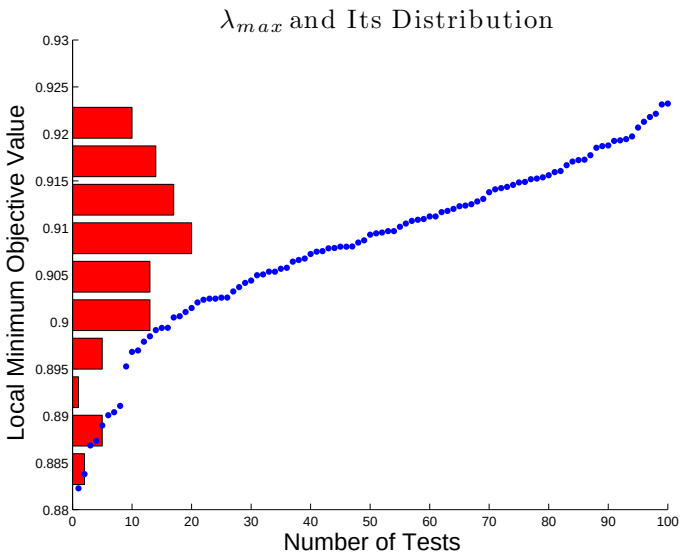

FIGURE 4.1. *Values (blue dots) and histogram of local minimal* λ_{max}.

Example 3. Corresponding to the prescribed stationary distribution[3]

$$\boldsymbol{\pi} = [0.3200, 0.0765, 0.2007, 0.1446, 0.2581]^\top,$$

plotted in Figure 4.1 are the 100 local minimal values of λ_{max} found by fmincon. We rearrange these values in increasing order for ease of comparison. The potentially global minimum value is estimated to be 0.8823 attained at[4]

$$P_1 = \begin{bmatrix} 0.0221 & 0.0000 & 0.3543 & 0.1206 & 0.5031 \\ 0.8339 & 0.0000 & 0.0932 & 0.0075 & 0.0654 \\ 0.5619 & 0.0935 & 0.0020 & 0.3319 & 0.0107 \\ 0.0237 & 0.1740 & 0.1790 & 0.0018 & 0.6215 \\ 0.5152 & 0.1262 & 0.2090 & 0.1496 & 0.0000 \end{bmatrix}.$$

The horizontal bars on the left edge of the plot are the histogram of these calculated objective values which are distributed over the indicated bins. We check the exit flag to ascertain that each run of the local solver has converged. Therefore, though they are clustered, each of the dots in Figure 4.1 represents a true local minimal value. What is most puzzling is that out of 100 start points, we find 100 distinct local minima, a strong indication that the topology associated with the objective function in (2.17) is fairly complicated [4]. Without resorting to the global optimization technique with sufficiently many tests, the optimal asymptotic variance problem probably cannot be easily found. Most likely, a causal starting point will lead convergence to a local solution only.

Example 4. On the other hand, corresponding to the same prescribed stationary distribution $\boldsymbol{\pi}$, suppose we impose the additional requirement of reversibility. Plotted in Figure 4.2 are the 25 entries for each of the optimizers found by fmincon. Details of these drawings are immaterial, but it is quite obvious from a glance that not all the minimizers are the same. Thus, there are multiple solutions even for the reversible problem. The downward spikes at locations 1,7, 13, 19, and 25 are diagonal entries which are small, but are not identically

[3]For the ease of running text, all numbers are displayed in 4 digits only.

[4]It is potentially global because the 100 multi-starts cannot constitute an exhaustive search. It is rare, but possible, that a true global optimizer has not been found yet.

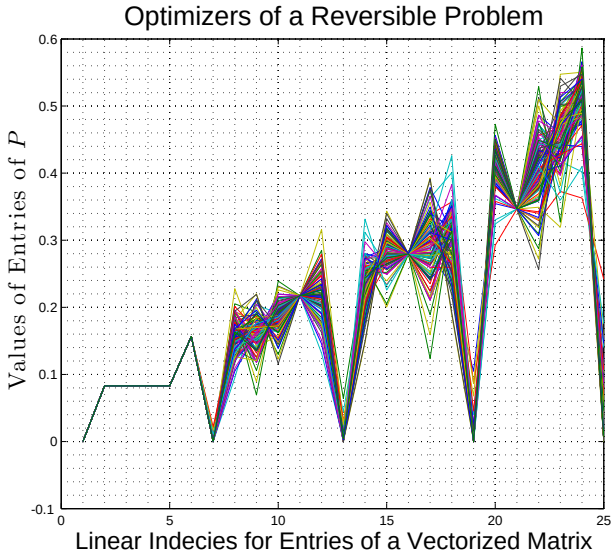

FIGURE 4.2. *Entry values of local reversible minimizer.*

zero. For instance,

$$P_2 = \begin{bmatrix} 0 & 0.1566 & 0.2174 & 0.2794 & 0.3465 \\ 0.0828 & 0.0000 & 0.2302 & 0.2820 & 0.4049 \\ 0.0828 & 0.1659 & 0.0289 & 0.2313 & 0.4911 \\ 0.0828 & 0.1581 & 0.1799 & 0.1048 & 0.4744 \\ 0.0828 & 0.1830 & 0.3081 & 0.3825 & 0.0435 \end{bmatrix}$$

is one such a reversible solution. In contrast, the optimizer obtained from the receipt in Section 3 is

$$P_3 = \begin{bmatrix} 0 & 0.1566 & 0.2174 & 0.2794 & 0.3465 \\ 0.0828 & 0 & 0.2364 & 0.3039 & 0.3769 \\ 0.0828 & 0.1703 & 0 & 0.3334 & 0.4134 \\ 0.0828 & 0.1703 & 0.2593 & 0 & 0.4875 \\ 0.0828 & 0.1703 & 0.2593 & 0.3931 & 0.0944 \end{bmatrix}$$

which always has zero diagonal entries except the last one. What is most interesting is that for the reversible problem we have the same optimal value 0.9235 universally at all local solutions which, therefore, are also global solutions.

Example 5. To catch a glimpse into a possible landscape associated with the objective function λ_{max}, consider the subclass of 3×3 row stochastic matrices

$$P = \begin{bmatrix} \widehat{p}_{11} & \widehat{p}_{12} & \widehat{p}_{13} \\ x & p_{22} & p_{23} \\ p_{31} & y & p_{33} \end{bmatrix},$$

where the first row $[\widehat{p}_{11}, \widehat{p}_{12}, \widehat{p}_{13}]$ of P is fixed. We use x and y as the variables which, once chosen, determine the remaining entries p_{22}, p_{23}, p_{31} and p_{33} from the equations $\boldsymbol{\pi}^\top P = \boldsymbol{\pi}^\top$ and $P\mathbf{1} = \mathbf{1}$. Obviously, since all entries of P must be nonnegative, x and y are subject to the constraints

$$\pi_1 \widehat{p}_{12} - \pi_2 \leq \pi_2 x - \pi_3 y \leq \pi_1 \widehat{p}_{11},$$
$$\pi_1 - \pi_3 - \pi_1 \widehat{p}_{11} \leq \pi_2 x - \pi_3 y \leq \pi_1 - \pi_1 \widehat{p}_{11}.$$

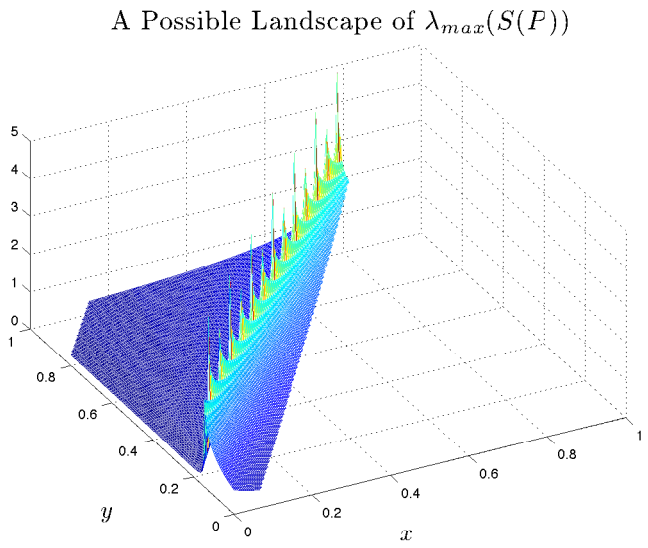

FIGURE 4.3. *Logarithm of $\lambda_{max}(R(P))$ over feasible domain.*

We are interested in seeing the "cross-section" of $\lambda_{max}(S(P))$ when the first row of P is fixed. Consider an example with randomly generated $\boldsymbol{\pi} = [0.1099, 0.5097, 0.3804]^\top$ and $[\widehat{p}_{11}, \widehat{p}_{12}, \widehat{p}_{13}] = [0.3133, 0.4746, 0.2121]$. Because the objective values λ_{max} with P satisfying the corresponding constraints vary significantly, we plot in Figure 4.3 the values of $\log \lambda_{max}(S(P))$ over the feasible domain to demonstrate the possible complexity of the objective function. We mention in passing that with the first row and $\boldsymbol{\pi}$ as are given, the minimum of $\lambda_{max}(S(P))$ is 1.2881 attainable at $(x, y) = (0.0281, 0.8287)$, whereas without such a restriction on the first row the global minimum of $\lambda_{max}(S(P))$ is $.8391$ attainable at

$$P = \begin{bmatrix} 0.0000 & 0.9752 & 0.0248 \\ 0.0261 & 0.2984 & 0.6755 \\ 0.2538 & 0.6583 & 0.0879 \end{bmatrix}.$$

5. Conclusions. Constructing a Markov chain with a prescribed stationary distribution vector is the first step needed in a MCMC method. To optimize the performance of the MCMC process, it is desirable that the worst-case asymptotic variance of the constructed Markov chain is minimal. In the present paper we revisit this important problem with the tool of global optimization in hand. As extensive search is possible through efficient numerical computation, it is found that the reversible problems enjoy having multiple global solutions which are different from the one known as classical in the literature. This multiple choice of optimal reversible Markov chains opens the door to factoring in other considerations when constructing a MCMC process. On the other hand, it is found that the irreversible problems will have multiple local solutions in general and that the basin of attraction for some of the local solutions might be larger than that for the global solution. Unless a global optimization approach is adopted, it might be difficult to construct the ultimate Markov chain for a MCMC process. In all, this paper suggests a framework of applying existent global optimization techniques to tackle the problem of Markov chains construction with prescribed stationary distribution vector and minimal worst-case asymptotic variance.

REFERENCES

[1] D. ALDOUS AND J. A. FILL, *Reversible markov chains and random walks on graphs*, 2002. Unfinished monograph, recompiled 2014, available at `http://www.stat.berkeley.edu/$\sim$aldous/RWG/book.html`.

[2] C. ANDRIEU AND J. THOMS, *A tutorial on adaptive MCMC*, Stat. Comput., 18 (2008), pp. 343–373.

[3] S. BROOKS, A. GELMAN, G. L. JONES, AND X.-L. MENG, eds., *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, CRC Press, Boca Raton, FL, 2011.

[4] T.-L. CHEN, W.-K. CHEN, C.-R. HWANG, AND H.-M. PAI, *On the optimal transition matrix for Markov chain Monte Carlo sampling*, SIAM J. Control Optim., 50 (2012), pp. 2743–2762.

[5] T.-L. CHEN AND C.-R. HWANG, *Accelerating reversible Markov chains*, Statist. Probab. Lett., 83 (2013), pp. 1956–1962.

[6] M. T. CHU AND G. H. GOLUB, *Inverse eigenvalue problems: theory, algorithms, and applications*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 2005.

[7] M. P. DESAI AND V. B. RAO, *On the convergence of reversible Markov chains*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 950–966.

[8] P. DIACONIS, S. HOLMES, AND R. M. NEAL, *Analysis of a nonreversible Markov chain sampler*, Ann. Appl. Probab., 10 (2000), pp. 726–752.

[9] P. DIACONIS AND D. STROOCK, *Geometric bounds for eigenvalues of Markov chains*, Ann. Appl. Probab., 1 (1991), pp. 36–61.

[10] R. FLETCHER, *Practical methods of optimization*, A Wiley-Interscience Publication, John Wiley & Sons, Ltd., Chichester, second ed., 1987.

[11] A. FRIGESSI, C.-R. HWANG, AND L. YOUNES, *Optimal spectral structure of reversible stochastic matrices, Monte Carlo methods and the simulation of Markov random fields*, Ann. Appl. Probab., 2 (1992), pp. 610–628.

[12] J. FULMAN AND E. L. WILMER, *Comparing eigenvalue bounds for Markov chains: when does Poincaré beat Cheeger?*, Ann. Appl. Probab., 9 (1999), pp. 1–13.

[13] W. R. GILKS, S. RICHARDSON, AND D. J. SPIEGELHALTER, eds., *Markov chain Monte Carlo in practice*, Interdisciplinary Statistics, Chapman & Hall, London, 1996.

[14] C.-R. HWANG, *Accelerating Monte Carlo Markov processes*, Cosmos, 1 (2005), pp. 87–94.

[15] S. INGRASSIA, *On the rate of convergence of the Metropolis algorithm and Gibbs sampler by geometric bounds*, Ann. Appl. Probab., 4 (1994), pp. 347–389.

[16] M. IOSIFESCU, *Finite Markov processes and their applications*, John Wiley & Sons, Ltd., Chichester; Editura Tehnică, Bucharest, 1980. Wiley Series in Probability and Mathematical Statistics.

[17] A. S. LEWIS, *Derivatives of spectral functions*, Math. Oper. Res., 21 (1996), pp. 576–588.

[18] A. MIRA, *Ordering and improving the performance of Monte Carlo Markov chains*, Statist. Sci., 16 (2001), pp. 340–350.

[19] A. MIRA AND C. J. GEYER, *On non-reversible Markov chains*, in Monte Carlo methods (Toronto, ON, 1998), vol. 26 of Fields Inst. Commun., Amer. Math. Soc., Providence, RI, 2000, pp. 95–110.

[20] R. NEAL, *Improving asymptotic variance of MCMC estimators: non-reversible chains are better*, Tech. Rep. 0406, Department of Statistics, University of Toronto.

[21] M. L. OVERTON AND R. S. WOMERSLEY, *Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices*, Math. Programming, 62 (1993), pp. 321–357.

[22] P. H. PESKUN, *Optimum Monte-Carlo sampling using Markov chains*, Biometrika, 60 (1973), pp. 607–612.

[23] C. P. ROBERT AND G. CASELLA, *Monte Carlo statistical methods*, Springer Texts in Statistics, Springer-Verlag, New York, second ed., 2004.

[24] G. O. ROBERTS AND J. S. ROSENTHAL, *Variance bounding Markov chains*, Ann. Appl. Probab., 18 (2008), pp. 1201–1214.

[25] ———, *Minimising MCMC variance via diffusion limits, with an application to simulated tempering*, Ann. Appl. Probab., 24 (2014), pp. 131–149.

[26] D. SORENSEN AND D. GIANOLA, *Likelihood, Bayesian, and MCMC methods in quantitative genetics*, Statistics for Biology and Health, Springer-Verlag, New York, 2002.

[27] Z. UGRAY, L. LASDON, J. PLUMMER, F. GLOVER, J. KELLY, AND R. MARTÍ, *Scatter search and local NLP solvers: a multistart framework for global optimization*, INFORMS J. Comput., 19 (2007), pp. 328–340.

[28] X.-D. ZHANG, *The smallest eigenvalue for reversible Markov chains*, Linear Algebra Appl., 383 (2004), pp. 175–186.