Journal of Multivariate Analysis **66**, 22–37 (1998)
Article No. MV971735

On the Geometrical Convergence of Gibbs Sampler in R^d*

Chii-Ruey Hwang and Shuenn-Jyi Sheu

Institute of Mathematics, Academia Sinica, Taipei, Taiwan

Received August 23, 1995; revised December 4, 1997

The geometrical convergence of the Gibbs sampler for simulating a probability distribution in R^d is proved. The distribution has a density which is a bounded perturbation of a log-concave function and satisfies some growth conditions. The analysis is based on a representation of the Gibbs sampler and some powerful results from the theory of Harris recurrent Markov chains. © 1998 Academic Press

AMS 1991 subject classifications: 62E25, 65C05, 82B31.

Key words and phrases: Stochastic relaxation, Gibbs sampler, Markov chain, geometrical convergence, Harris recurrence, Monte Carlo Markov chain, Metropolis algorithm, nonlinear autoregression.

1. INTRODUCTION

Stochastic relaxation is a powerful idea used extensively for sampling probability distributions in high-dimensional spaces. The convergence of the stochastic relaxation to the equilibrium and the rate of convergence are important issues remained to be settled in general.

The Gibbs sampler and the Metropolis algorithm are among the best known examples of stochastic relaxations commonly used in applications [9]. Their behaviors are far from well understood [2, 5, 6, 7, 12, 15]. In this paper we shall consider the Gibbs sampler in the Euclidean space.

For the Gaussian case the Gibbs sample has a specific representation [3]. We show that a similar representation holds under our conditions. And this result demonstrates the intrinsic difference between these two types of algorithms. Note that the Metropolis algorithm in the Euclidean space is a perturbation of the gradient dynamics [8, 14]. Another main result is the exponential convergence of the Gibbs sampler. Related works for the Metropolis algorithm can be found in [12, 15].

First we mention some known results on the convergence for the Gibbs sampler. Exponential convergence in variational norm for the deterministic

* Research supported by NSC 83-0208-M-001-078.

0047-259X/98 $25.00

and random updating strategies in the Gaussian case are obtained in [3]. The rates of convergence for these two strategies are also compared. In [1], the exponential convergence of the Gibbs sampler for bounded perturbations of Gaussian densities is proved.

For simplicity we assume that the given distribution has a log-concave density which is proportional to $\exp(-V(x))$. The main results continue to hold under slightly general conditions, see Section 2 and the remark after Theorem 3.5. We apply some powerful results from the theory of Harris recurrent Markov chain [13] to establish the convergence of the Gibbs sampler and obtain the exponential rate of convergence.

It is known that the Markov chain generated by the Gibbs sampler is Harris recurrent if the probability density is continuous and positive. Therefore,

$$\|P^n(x,\cdot)-\mu(\cdot)\|_{\mathrm{var}}\to 0, \qquad n\to\infty,$$

where P^n is the n-step transition probabilities, $\mu(\cdot)$ is the given distribution and $\|\cdot\|_{\mathrm{var}}$ is the variational norm of measures. Moreover, if the Markov chain is geometrically recurrent, then $P^n(x,\cdot)$ converges to $\mu(\cdot)$ exponentially for almost all x. These results and some related notions are presented in the Appendix.

Under suitable conditions, we show that the Markov chain generated is geometrically recurrent. Furthermore we deduce that it converges exponentially to the equilibrium for an initial point.

The main ingredient of our approach comes from the following observation. Denote the Markov chain in R^d generated by the Gibbs sampler as $X^{(0)}$, $X^{(1)}$, $X^{(2)}$, ..., and

$$X^{(n)}=(x_1^{(n)}, Z^{(n)}), \qquad x_1^{(n)}\in R, \qquad Z^{(n)}\in R^{d-1},$$

then there are maps Ψ, ϕ,

$$\Psi\colon R^{d-1}\to R^{d-1}$$
$$\phi\colon R^{d-1}\to R$$

such that

$$X^{(n+1)}=\Phi(Z^{(n)})+\eta^{(n)},$$

$$\Phi=\begin{pmatrix}\phi\\ \Psi\end{pmatrix}.$$

Ψ is shown to satisfy the property that Ψ^m, the mth iteration of Ψ, is a contraction map. Note that for the Gibbs sampler the transition from $X^{(n)}$

to $X^{(n+1)}$ is done by updating the coordinates one by one. Now compare this updating scheme with a deterministic one. ϕ here is the argument in the first coordinate which minimizes V with other coordinates fixed. Φ is defined similarity by performing the same minimization procedure throughout the other coordinates. $\eta^{(n)}$ is the resulting error. Details of this approach will be spelled out later.

Note that the above formulation is very similar to that in the nonlinear autoregressive (NLAR) time series [4, 16, 17] where the innovations $\eta^{(n)}$ are i.i.d. However, in the current setup the innovations are not i.i.d. in general. Instead we establish that there are $\beta > 0$, $c > 0$ such that

$$E[\exp(\beta\,|\eta^{(n)}|^2) \mid X^{(n)}] \leqslant c,$$

for all n. This will be sufficient to deduce the geometrical convergence of the dynamics.

It is very tempting to use the NLAR formulation, i.e., by applying a deterministic dynamic to the current state plus an innovation as the updating scheme to generate the Monte Carlo Markov chain. The difficulty lies in finding feasible dynamics and innovations in practice. But if we consider a similar situation in the continuous time setup, $-\nabla V$ and Brownian motion may be used as the deterministic dynamics and innovations respectively [10]. However, the discrete time approximation of this diffusion has a different limiting distribution.

The paper is arranged as follows. The main results are presented in Section 3. Assumptions, definitions, and some elementary properties are presented in Section 2. Some relevant results of Harris recurrent Markov chains are in the Appendix, e.g., the Gibbs sampler for a positive continuous density is Harris recurrent. This very result was announced in [11] without a complete proof.

2. ASSUMPTIONS AND PRELIMINARY RESULTS

Let μ be a probability distribution on R^d with density (still denoted by μ) given by

$$\mu(X) = \frac{1}{M} \exp(-V(X)),$$

where M is the normalizing constant and $V(\cdot)$ is assumed to satisfy the following condition.

(A) $V(\cdot)$ is smooth, strictly convex and there are $\alpha_1, \alpha_2 > 0$ such that

$$\alpha_1 \leqslant \left(\frac{\partial^2 V}{\partial x_i \, \partial x_j}(X)\right) \leqslant \alpha_2 \qquad \forall X \in R^d.$$

The inequalities are meant for nonnegative definite matrices. Our main results still hold under conditions slightly more general than (A). (See the remark after Theorem 3.5.) Under (A), $V(\cdot)$ has a unique minimum and, without loss of generality, we assume

$$V(0) = \min V.$$

The following lemmas are easily obtained.

LEMMA 2.1. $\frac{1}{2}\alpha_1 |X|^2 \leqslant V(X) \leqslant \frac{1}{2}\alpha_2 |X|^2$.

LEMMA 2.2. $\alpha_1 |X| \leqslant |\nabla V(X)| \leqslant \alpha_2 |X|$. *Here ∇V is the gradient of V.*

Now we introduce the Gibbs sampler. Let $X = (x_1, ..., x_d)$, $Y = (y_1, ..., y_d)$ be in R^d. Denote

$$\begin{aligned} W^{(k)} &= (y_1, ..., y_{k-1}, x_{k+1}, ..., x_d), \\ Y^{(k)} &= (y_1, ..., y_k, x_{k+1}, ..., x_d), \end{aligned} \tag{2.1}$$

and

$$P_k(y_k \mid W^{(k)}) = \frac{1}{Z_k(W^{(k)})} \exp(-V(Y^{(k)}))$$

with

$$Z_k(W^{(k)}) = \int \exp(-V(Y^{(k)})) \, dy_k.$$

The Gibbs sampler we consider is a Markov chain with transition density $P(X, Y)$ given by

$$P(X, Y) = P_1(y_1 \mid W^{(1)}) \, P_2(y_2 \mid W^{(2)}) \cdots P_d(y_d \mid W^{(d)}). \tag{2.2}$$

The following property is easily verified,

$$\int \mu(X) \, P(X, Y) \, dX = \mu(Y).$$

Namely, the Gibbs sampler has $\mu(\cdot)$ as its invariant measure. Moreover, by the ergodic property established in the Appendix, $\mu(\cdot)$ is the unique invariant measure.

The following maps are essential. For each k, $k=1, 2, ..., d$, and $W \in R^{d-1}$, there uniquely exists a point $\phi_k(w)$ in R satisfying

$$V(w_1, ..., w_{k-1}, \phi_k(W), w_k, ..., w_{d-1}) = \min_y V(w_1, ..., w_{k-1}, y, w_k, ..., w_{d-1}). \tag{2.3}$$

Then, the following mapping from R^{d-1} to R is well defined.

$$\hat{V}_k(W) = V(w_1, ..., w_{k-1}, \phi_k(W), w_k, ..., w_{d-1}). \tag{2.4}$$

LEMMA 2.3. $\exists c_1, c_2 > 0$ *such that*

$$\begin{aligned} c_1 \exp(-\hat{V}_k(W^{(k)})) &\leqslant Z_k(W^{(k)}) \\ &\leqslant c_2 \exp(-\hat{V}_k(W^{(k)})). \end{aligned}$$

Proof. It is easy to see

$$\begin{aligned} \tfrac{1}{2}\alpha_1(y_k - \phi_k(W^{(k)}))^2 &\leqslant V(Y^{(k)}) - \hat{V}_k(W^{(k)}) \\ &\leqslant \tfrac{1}{2}\alpha_2(y_k - \phi_k(W^{(k)}))^2. \end{aligned}$$

The result follows by taking

$$c_1 = \sqrt{\frac{2\pi}{\alpha_2}} \quad c_2 = \sqrt{\frac{2\pi}{\alpha_1}}.$$

LEMMA 2.4. *There is $c > 0$ such that*

$$|\nabla \phi_k| \leqslant c \qquad \forall k.$$

Proof. Since

$$\frac{\partial V}{\partial x_k}(w_1, ..., w_{k-1}, \phi_k(W), w_k, ..., w_{d-1}) = 0,$$

we differentiate this with respect to w_j to get

$$\frac{\partial^2 V}{\partial x_j x_k} + \frac{\partial^2 V}{\partial {x_k}^2}\frac{\partial \phi_k}{\partial w_j} = 0 \qquad \text{if} \quad j \leqslant k-1,$$

and

$$\frac{\partial^2 V}{\partial x_{j+1}\, \partial x_k} + \frac{\partial^2 V}{\partial x_k{}^2} \frac{\partial \phi_k}{\partial w_j} = 0 \qquad \text{if} \quad j \geqslant k.$$

Here the functions are evaluated at the point $(w_1, ..., w_{k-1}, \phi_k(W), w_k, ..., w_{d-1})$. It is easy to deduce the result from these by using condition (A).

Let $X^{(0)}$, $X^{(1)}$, $X^{(2)}$, ... be the Markov chain in R^d generated by the Gibbs sampler. For each n, let $W^{(n,k)}$ be the random vectors in R^{d-1} defined by (2.1) with $X = X^{(n)}$, $Y = X^{(n+1)}$.

LEMMA 2.5. *The random variables $\xi_k^{(n)}$ defined by*

$$x_k^{(n+1)} = \phi_k(W^{(n,k)}) + \xi_k^{(n)}$$

have the following property: There are β, $c > 0$ such that

$$E_X[e^{\beta |\xi_k^{(n)}|^2} \mid \mathscr{F}_n] \leqslant c$$

for all X and n. $\mathscr{F}_n$ is the σ-algebra generated by $X^{(0)}, ..., X^{(n)}$.

Proof. We only consider $k = 1$. The rest are similar.

$$E_X[e^{\beta |\xi_1^{(n)}|^2} \mid \mathscr{F}_n] = \frac{1}{Z_1(W^{(n,1)})} \int e^{\beta |x - \phi_1(W^{(n,1)})|^2} e^{-V(x, W^{(n,1)})}\, dx.$$

This integral is smaller than

$$\begin{aligned} &\frac{1}{c_1} \int e^{\beta |x - \phi_1(W^{(n,1)})|^2} e^{-V(x, W^{(n,1)}) + V(\phi_1(W^{(n,1)}), W^{(n,1)}))}\, dx \\ &\leqslant \frac{1}{c_1} \int e^{\beta |x - \phi_1(W^{(n,1)})|^2} e^{-(1/2)\alpha_1 |x - \phi_1(W^{(n,1)})|^2}\, dx \\ &\leqslant c \end{aligned}$$

for some $c > 0$ if $\beta < \frac{1}{2}\alpha_1$, as asserted.

3. GEOMETRICAL CONVERGENCE OF THE GIBBS SAMPLER

We shall show the geometrical convergence of the Gibbs sampler starting from any initial point. Let $X^{(0)}$, $X^{(1)}$, $X^{(2)}$, ... be the Markov chain in R^d with transition (2.2). Denote $Z^{(0)}$, $Z^{(1)}$, $Z^{(2)}$, to be the random vectors

$$Z^{(n)} = (x_2^{(n)}, ..., x_d^{(n)}) \in R^{d-1},$$

where $X^{(n)} = (x_1^{(n)}, ..., x_d^{(n)})$. And define the following mappings, for $Z = (z_2, ..., z_d) \in R^{d-1}$,

$$\begin{aligned}
\psi_2(Z) &= \phi_2(\phi_1(Z), z_3, ..., z_d),\\
\psi_3(Z) &= \phi_3(\phi_1(Z), \psi_2(Z), z_4, ..., z_d),\\
&\vdots\\
\psi_d(Z) &= \phi_d(\phi_1(Z), \psi_2(Z), ..., \psi_{d-1}(z)),\\
\Psi &= (\psi_2, ..., \psi_d).
\end{aligned}$$

For $\phi_1, ..., \phi_d$, see (2.3).

THEOREM 3.1. *We have*

$$\begin{aligned}
x_1^{(n+1)} &= \phi_1(Z^{(n)}) + \xi_1^{(n)}\\
Z^{(n+1)} &= \Psi(Z^{(n)}) + \eta^{(n)}.
\end{aligned}$$

Here, $\eta^{(0)}, \eta^{(1)}, ...$ are random vectors in R^{d-1} satisfying

$$E[e^{\beta |\eta^{(n)}|^2} \mid \mathscr{F}_n^Z] \leqslant c$$

for all n for some β and $c > 0$, $\mathscr{F}_n^Z$ is the σ-algebra generated by $Z^{(0)}, Z^{(1)}, Z^{(2)}, ..., Z^{(n)}$.

This is a direct consequence of Lemmas 2.4 and 2.5. We omit its proof. In the following, we denote $\hat{V}$ on R^{d-1} by $\hat{V}(\cdot) = \hat{V}_1(\cdot)$. That is,

$$\begin{aligned}
\hat{V}(Z) &= V(\phi_1(Z), Z)\\
&= \min_x V(x, Z).
\end{aligned}$$

LEMMA 3.2. $\frac{1}{2}\alpha_2 |Z|^2 \leqslant \hat{V}(Z) \leqslant \frac{1}{2}\alpha_2 |Z|^2$.

Proof. By Lemma 2.1,

$$\begin{aligned}
\hat{V}(Z) &= V(\phi_1(Z), Z)\\
&\geqslant \tfrac{1}{2}\alpha_1(\phi_1(Z)^2 + |Z|^2)\\
&\geqslant \tfrac{1}{2}\alpha_1 |Z|^2.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\hat{V}(Z) &\leqslant V(0, Z)\\
&\leqslant \tfrac{1}{2}\alpha_2 |Z|^2,
\end{aligned}$$

as asserted.

LEMMA 3.3. $\sup_{Z\neq 0}(\hat{V}(\Psi(Z))/\hat{V}(Z))=\gamma<1$.

The proof will be given at the end of this section.

COROLLARY 3.4. *$\exists m>0$ integer such that*

$$|\Psi^{(m)}(Z)|\leqslant \tfrac{1}{2}\,|Z|.$$

Proof. $\hat{V}(\Psi^{(m)}(Z))\leqslant \gamma^m \hat{V}(Z)$. This implies

$$\tfrac{1}{2}\alpha_1\,|\Psi^{(m)}(Z)|^2\leqslant \gamma^m\,\tfrac{1}{2}\alpha_2\,|Z|^2,$$

i.e.,

$$|\Psi^{(m)}(Z)|^2\leqslant \frac{\alpha_2}{\alpha_1}\,\gamma^m\,|Z|^2.$$

From this, $|\Psi^{(m)}(Z)|\leqslant \frac{1}{2}\,|Z|$ if m is large. This completes the proof.

We now state the main result.

THEOREM 3.5. *Let q be large enough*

$$S_q=\inf\{n\geqslant 1;\ |Z^{(n)}|\leqslant q\}.$$

Then there is $\beta>0$ such that

$$\sup_{|Z|\leqslant q}\ E_Z[e^{\beta S_q}]<\infty.$$

This implies that the Markov chain is of geometrical recurrence. Hence it converges to the equilibrium exponentially.

Proof. We shall show that, for λ small $e^{\lambda \hat{V}(Z^{(n)})+n}$ is supermartingale before S_q. Assume $|Z^{(n)}|>q$. By

$$\begin{aligned}
\hat{V}(Z^{(n+1)}) &= V(\phi_1(Z^{(n+1)}), Z^{(n+1)})\\
&= V(\phi_1(\Psi(Z^{(n)})), \Psi(Z^{(n)}))+\sum_{j=2}^{d}\frac{\partial V}{\partial x_j}(\phi_1(\Psi(Z^{(n)})), \Psi(Z^{(n)}))\,\eta_j^{(n)}\\
&\quad +O(|\eta^{(n)}|^2)\\
&= \hat{V}(\Psi(Z^{(n)}))+\sum_{j=2}^{d}\frac{\partial V}{\partial x_j}(\phi_1(\Psi(Z^{(n)})), \Psi(Z^{(n)}))\,\eta_j^{(n)}+O(|\eta^{(n)}|^2),
\end{aligned}$$

we have

$$E[e^{\lambda\hat{V}(Z^{(n+1)})} \mid \mathscr{F}_n^Z] \leqslant ce^{\lambda\hat{V}(\Psi(Z^{(n)}))} \int \exp\left(-\alpha\,|\eta|^2 + \lambda \sum_{j=2}^{d} \frac{\partial V}{\partial x_j}(X)\,\eta_j\right) d\eta_2 \cdots d\eta_n,$$

where $X = (\phi_1(\Psi(Z^{(n)})), \Psi(Z^{(n)})$ and α is some positive number. From this, by simple calculus, we have

$$E[e^{\lambda\hat{V}(Z^{(n+1)})} \mid \mathscr{F}_n^Z] \leqslant c_1 e^{\lambda\hat{V}(\Psi(Z^{(n)}))} e^{(\lambda^2/4\alpha^2)\,|\nabla V(X)|^2}.$$

But

$$\begin{aligned}\lambda\hat{V}(\Psi(Z^{(n)})) + \frac{\lambda^2}{4\alpha^2}\,|\nabla V(X)|^2 &\leqslant \lambda(1+c\lambda)\,\hat{V}(\Psi(Z^{(n)}))\\ &\leqslant \lambda(1+c\lambda)\,\gamma\hat{V}(Z^{(n)})\\ &= \lambda\hat{V}(Z^{(n)}) - \lambda(1-(1+c\lambda)\,\gamma)\,\hat{V}(Z^{(n)}).\end{aligned}$$

Choose λ such that

$$(1+c\lambda)\,\gamma < 1.$$

Then choose q large enough such that

$$\lambda(1-(1+c\lambda)\,\gamma)\,\hat{V}(Z^{(n)}) > 1 + \ell n c_1$$

if $|Z^{(n)}| > q$. This implies

$$E[e^{\lambda\hat{V}(Z^{(n+1)}) + (n+1)} \mid \mathscr{F}_n^Z] \leqslant e^{\lambda\hat{V}(Z^{(n)}) + n},$$

as asserted. From the last result, we deduce

$$E[e^{\lambda\hat{V}(Z^{(S_q)}) + S_q} \mid \mathscr{F}_1^Z] \leqslant e^{\lambda\hat{V}(Z^{(1)}) + 1}$$

when $|Z^{(1)}| > q$. This implies the geometrical recurrence.

We shall next prove the geometrical convergence of the Markov chain with any initial point.

According to Corollary A14,

$$\|P^n(X,\cdot) - \mu(\cdot)\|_{\mathrm{Var}} \leqslant M(X)\,\rho^n$$

for some $M(\cdot) \in L^1(\mu)$, $0 < \rho < 1$,

$$\|P^{(n+1)}(X, \cdot) - \mu(\cdot)\|_{\text{Var}} \leqslant \int P(X, dY)\, \|P^n(Y, \cdot) - \mu(\cdot)\|_{\text{Var}}$$

$$\leqslant \rho^n \int P(X, dY)\, M(Y). \tag{3.1}$$

The transition density $P(X, Y)$ defined by (2.2) satisfies,

$$P(X, Y) \leqslant c^d \exp\left(-\sum_{k=1}^{d} (V(Y^{(k)}) - V(\tilde{Y}^{(k)}))\right)$$
$$Y^{(k)} = (y_1, ..., y_k, x_{k+1}, ..., x_d),$$
$$\tilde{Y}^{(k)} = (y_1, ..., y_{k-1}, \phi_k(y_1, ..., y_{k-1}, x_{k+1}, ..., x_d), x_{k+1}, ..., x_d).$$

Observing,

$$V(Y^{(k-1)}) \geqslant V(\tilde{Y}^{(k)}), \qquad k = 2, ..., d,$$

we then have

$$\begin{aligned} P(X, Y) &\leqslant c^d \exp(-(V(Y) - V(\tilde{Y}^{(1)})) \\ &= c^d \exp(\hat{V}_1(x_2, ..., x_d)) \exp(-V(Y)). \end{aligned} \tag{3.2}$$

Plugging (3.2) in (3.1), we obtain

$$\|P^{(n+1)}(X, \cdot) - \mu(\cdot)\|_{\text{Var}} \leqslant c^d \exp(\hat{V}_1(x_2, ..., x_d)) \int M(Y)\, \mu(dY)\, \rho^n.$$

This completes the proof.

Remark. Assume $\mu(\cdot)$ has the property

$$c_1 \leqslant \frac{\mu(dY)}{\mu_0(dY)} \leqslant c_2$$

for some c_1, $c_2 > 0$ and $\mu_0(\cdot)$ satisfies the condition (A). Then Lemma 2.5, hence Theorem 3.5, still holds.

Proof of Lemma 3.3. It is easy to see, by the strict convexity of $V(\cdot)$, that $\hat{V}(\Psi(Z)) < \hat{V}(Z)$ for any $Z \neq 0$. Therefore it is enough to consider the case where $|Z|$ is large.

Fix $\delta > 0$ and small. Assume

$$\left|\frac{\partial V}{\partial x_{i+1}}(X^{(i)})\right| < \delta\, |Z| \tag{3.3}$$

holds for $X^{(i)} = (\phi_1(Z), \psi_2(Z), ..., \psi_i(Z), z_{i+1}, ..., z_d)$ and $i = 1, ..., d-1$. Then, by using

$$\frac{\partial V}{\partial x_{i+1}}(X^{(i)}) - \frac{\partial V}{\partial x_{i+1}}(X^{i+1})$$
$$= \int_0^1 \frac{\partial^2 V}{\partial x_{i+1}^2}(X^{(i)} + \lambda(X^{(i+1)} - X^{(i)}))(z_{i+1} - \psi_{i+1}(Z))\, d\lambda,$$

$$\frac{\partial V}{\partial x_{i+1}}(X^{(i+1)}) = 0,$$

and the condition (A), we deduce

$$|z_{i+1} - \psi_{i+1}(Z)| \leqslant \frac{1}{\alpha_1} \delta\, |Z|.$$

for $i = 1, ..., d-1$. This further implies

$$|X^{(i)} - X^{(1)}| \leqslant (i-1) \frac{1}{\alpha_1} \delta\, |Z|$$
$$\leqslant d \frac{1}{\alpha_1} \delta\, |Z|.$$

Then

$$\left| \frac{\partial V}{\partial x_{i+1}}(X^{(1)}) \right| \leqslant \left(1 + cd \frac{1}{\alpha_1}\right) \delta\, |Z|$$

for $i = 1, ..., d-1$, since $\partial V / \partial x_{i+1}$ is Lipschitz. We conclude,

$$|\nabla V(X^{(1)})| \leqslant d \left(1 + cd \frac{1}{\alpha}\right) \delta\, |Z|,$$

which contradicts Lemma 2.2 if δ is small enough.

We now fix such a δ. By the above result, there is an i such that

$$\left| \frac{\partial V}{\partial x_{i+1}}(X^{(i)}) \right| \geqslant \delta\, |Z|. \tag{3.4}$$

Using the following relation,

$$\frac{\partial V}{\partial x_{i+1}}(X^{(i)}) = \frac{\partial V}{\partial x_{i+1}}(X^{(i)}) - \frac{\partial V}{\partial x_{i+1}}(X^{(i+1)})$$
$$= \int_0^1 \frac{\partial^2 V}{\partial x_{i+1}^2}(X^{(i+1)} + \lambda(X^{(i)} - X^{(i+1)}))\, d\lambda (z_{i+1} - \psi_{i+1}(Z)),$$

we obtain

$$|z_{i+1} - \psi_{i+1}(Z)| \geqslant \frac{1}{\alpha_2}\, \delta\, |Z|.$$

Then

$$V(X^{(i)}) - V(X^{(i+1)})$$
$$= \int_0^1 \int_0^1 \frac{\partial^2 V}{\partial x_{i+1}}(X^{(i+1)} + \lambda\mu(X^{(i)} - X^{(i+1)}))\, \lambda\, d\mu\, d\lambda (z_{i+1} - \psi_{i+1}(Z))^2$$
$$\geqslant \frac{1}{2}\, \alpha_1 (z_{i+1} - \psi_{i+1}(Z))^2$$

implies

$$V(X^{(i)}) - V(X^{(i+1)}) \geqslant \frac{1}{2}\, \alpha_1 \left(\frac{\delta}{\alpha_2}\right)^2 |Z|^2.$$

We conclude

$$\hat{V}(Z) - \hat{V}(\Psi(Z)) = V(X^{(1)}) - V(X^{(d)})$$
$$\geqslant \frac{1}{2}\, \alpha_1 \left(\frac{\delta}{\alpha_2}\right)^2 |Z|^2.$$

Finally, from this and Lemma 3.2,

$$1 - \frac{\hat{V}(\Psi(Z))}{\hat{V}(Z)} \geqslant \frac{\alpha_1}{\alpha_2}\left(\frac{\delta}{\alpha_2}\right)^2 = \delta_0$$

i.e.,

$$\frac{\hat{V}(\Psi(Z))}{\hat{V}(Z)} \leqslant 1 - \delta_0.$$

This completes the proof.

APPENDIX

The purpose of this section is to give a brief review of some relevant results taken from the theory of Harris recurrent Markov chains. Definitions and results, except for A.15 and A.16, can be found in [13].

Let P be a transition probability on a measurable space (E, ε) and ϕ be a σ-finite measure on (E, ε).

DEFINITION A1. P is ϕ-irreducible if for all $x \in E$, $\phi(A) > 0$, there exists an n such that $P^n(x, A) > 0$.

P is irreducible if P is ϕ-irreducible for some ϕ. In this case, ϕ is called an irreducible measure for P.

An irreducible measure ψ for P is called maximal if every irreducible measure for P is absolutely continuous with respect to ψ.

PROPOSITION A2. *Assume P is irreducible. Then*

(i) *there exists a maximal irreducible measure for P,*

(ii) *ψ is maximal irreducible iff $\psi P \ll \psi$.*

Now fix a maximal irreducible measure ψ for P and denote

$$\varepsilon^+ = \{f \geqslant 0 \text{ and measurable, } \psi(f) > 0\}$$

$$M^+ = \{\lambda \text{ nonnegative measure on } (E, \varepsilon), \lambda(E) > 0\}.$$

DEFINITION A3. Assume $S \in \varepsilon^+$, $\nu \in M^+$ are such that

$$P^{m_0}(x, A) \geqslant \beta S(x)[\mathrm{n}]\nu(A), \qquad \forall x \in E, \qquad A \in \varepsilon$$

for some m_0 and $\beta > 0$. Then S is called a small function and ν is called a small measure.

THEOREM A4. *If P is irreducible, then there exist a small function and a small measure.*

DEFINITION A5. P is recurrent if $h_N^\infty > 0$ everywhere, $h_B^\infty = 1$ ψ-a.e. for $B \in \varepsilon^+$.

P is Harris recurrent if $h_B^\infty \equiv 1$. Here $h_B^\infty(x) = P_x\{X^{(n)} \in B \text{ i.o.}\}$, $X^{(0)}$, $X^{(1)}$, $X^{(2)}$, ..., is the Markov chain generated by P.

THEOREM A6. *Assume P is irreducible and recurrent and*

$$P^{m_0}(x, \cdot) \geqslant S(x)\, \nu(\cdot).$$

Then there exists a unique invariant measure π_S such that $\pi_S(S) = 1$.

DEFINITION A7. Assume that P is irreducible and recurrent with a unique invariant measure π_S given in Theorem A6. We call P positive recurrent if $\pi_S(E) < \infty$. Otherwise P is null recurrent.

THEOREM A8. *P is Harris recurrent iff $P_x\{S_C < \infty\} = 1$ for all $x \in C$ for some small set C.*

If $\sup_x E_x S_C < \infty$ for some small set C (i.e., I_C is small), then P is positive recurrent. Here $S_C = \inf\{n \geqslant 1;\, X^{(n)} \in C\}$.

THEOREM A9. *Assume that P is positive Harris recurrent and aperiodic with the unique invariant distribution π. Then*

$$\|\lambda P^n - \pi\|_{\text{var}} \to 0 \qquad \text{as} \quad n \to \infty$$

for any initial distribution λ. Here $\|\cdot\|_{\text{var}}$ is the variational norm for measures.

The following define the aperiodicity of P.

DEFINITION A10. The disjoint sets $E_0, E_1, ..., E_{m-1}$ in ε, are called m-cycle (for P) if for $x \in E_i$, $i = 0, 1, ..., m-1$, $j = i+1 \pmod{m}$, $P(x, E_j) = 1$.

DEFINITION A11. Assume that S is a small function and ν is a small measure, such that $\nu(S) > 0$ and $P^{m_0}(x, \cdot) \geqslant S(x)\, \nu(\cdot)$. It is easy to see that the set $\{m \geqslant 1, P^m(x, \cdot) \geqslant \beta_m S(x)\, \nu(\cdot)$ for some $\beta_m > 0\}$ is closed under addition. The greatest common divisor of this set is independent of the choice of S and ν and is called the period of P. P is aperiodic if its period is 1.

THEOREM A12. *Let d be the period. Then*

(i) *If there is a m-cycle, then d divides m,*

(ii) *There is a d-cycle.*

DEFINITION A13. P is irreducible and Harris recurrent. Then P is geometrically recurrent if for some small set C and $\gamma > 1$

$$\sup_{x \in C} E_x[r^{S_C}] < \infty.$$

THEOREM A14. *Assume that P is aperiodic, geometrically recurrent. Then there are $M(\cdot) \in L^1(\pi)$, $\rho < 1$ such that*

$$\|P^n(x, \cdot) - \pi\|_{\text{Var}} \leqslant M(x)\, \rho^n.$$

Now assume on R^d,

$$\mu(dx) = f(x)\, dx,$$

and f is positive and continuous. Let P be the transition density of the Gibbs sample for μ defined in Section 2. Then it is easy to see that P is irreducible with Lebesque measure as a maximal irreducible measure. Remark that μ is an invariant measure for P. The following can be proved using an argument in [18, pp. 235–241].

THEOREM A15. *Let K be a compact set in R^d with nonempty interior. Then*

$$h_K^\infty(x) \equiv 1 \quad \forall x, \quad \forall K$$

or

$$h_K^\infty(x) \equiv 0 \quad \forall x, \quad \forall K.$$

PROPOSITION A16. *P is Harris recurrent. Therefore*

$$\|\lambda P^n - \mu\|_{\text{Var}} \to 0 \qquad as \quad n \to \infty$$

for any initial distribution λ.

Proof. See [18, p. 242]. Let $h_K^m(x) = P_x\{x^{(n)} \in K \text{ for some } n \geqslant m\}$. Then $h_K^m(x) \downarrow h_K^\infty(x)$. Since $h_K^m(x) \geqslant P^m(x, K)$ and

$$\mu(K) = \int P^m(x, K)\, \mu(dx)$$

$$\leqslant \int h_K^m(x)\, \mu(dx),$$

we have

$$0 < \mu(K) \leqslant \int h_K^\infty(x)\, \mu(dx).$$

Therefore we can not have $h_K^\infty \equiv 0$, i.e., $h_K^\infty \equiv 1$ and P is Harris recurrent.

REFERENCES

1. Y. Amit, On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions, *J. Multivariate Anal.* **38** (1991), 82–99.
2. Y. Amit, Convergence properties of the Gibbs sampler for perturbations of Gaussians, *Ann. Statist.* **24** (1996), 122–140.
3. Y. Amit and U. Grenander, Comparing sweep strategies for stochastic relaxation, *J. Multivariate Anal.* **37** (1991), 197–222.
4. H. Z. An and F. C. Huang, The geometrical ergodicity of nonlinear autoregressive models, *Statist. Sinica* **6** (1996), 943–956.
5. K. B. Athreya, H. Doss, and J. Sethuraman, On the convergence of the Markov chain simulation method, *Ann. Statist.* **24** (1996), 69–100.
6. P. Barone and A. Frigessi, Improving Stochastic relaxation for Gaussian random fields, *Probab. Engrg. Inform. Sci.* **4** (1990), 233–247.
7. A. Frigessi, C.-R. Hwang, S.-J. Sheu, and P. Di Stefano, Convergence rates of the Gibbs sampler, the Metropolis algorithm, and other single-site updating dynamics, *J. Roy. Statist. Soc. Ser. B* **55** (1993), 205–219.
8. S. B. Gelfand and S. K. Mitter, Metropolis type annealing algorithms for global optimization in R^d, *SIAM J. Control Optim.* **31** (1993), 111–131.
9. S. Geman and D. Geman, Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intel.* **6** (1984), 721–741.
10. C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu, Accelerating Gaussian diffusions, *Ann. Appl. Probab.* **3** (1993), 897–913.
11. A. P. Knoerr, "Global Model of Natural Boundaries: Theory and Applications," Ph.D Thesis, Div. Appl. Math., Brown University, 1988.
12. K. L. Mengersen and R. L. Tweedie, Rates of convergence of the Hastings and Metropolis algorithms, *Ann. Statist.* **24** (1996), 101–121.
13. E. Nummelin, "General Irreducible Markov Chains and Non-negative Operators," Cambridge Univ. Press, Cambridge, UK, 1984.
14. G. O. Roberts, A. Gelman, and W. R. Gilks, Weak convergence and optimal scaling of random walk Metropolis algorithm, *Ann. Appl. Probab.* **7** (1997), 110–120.
15. G. O. Roberts and R. L. Tweedie, Geometric convergence and central limit theorems for multidimensial Hasting and Metropolis algorithms, *Biomerika* **83** (1996), 95–110.
16. D. Tjøstheim, Non-linear time series and Markov chains, *Adv. in Appl. Probab.* **22** (1990), 587–611.
17. H. Tong, "Non-linear Time Series, A Dynamical System Approach," Oxford Statistical Science Series 6, Oxford Univ. Press, London, 1990.
18. S. R. S. Varadhan, "Lectures on Diffusion Problems and Partial Differential Equations," Tata Institute, 1980.