Serdica

Mathematical Journal

Сердика

Математическо списание

ON THE CONVERGENCE OF THE
METROPOLIS-HASTINGS MARKOV CHAINS

Dimiter Tsvetkov, Lyubomir Hristov, Ralitsa Angelova-Slavova

Communicated by M. Savov

Abstract. In this paper we study Markov chains associated with the Metropolis-Hastings algorithm. We consider conditions under which the sequence of the successive densities of such a chain converges to the target density according to the total variation distance for any choice of the initial density. In particular we prove that the positiveness of the proposal density is enough for the chain to converge. The content of this work basically presents a stand alone proof that the reversibility along with the kernel positivity imply the convergence.

1. Introduction and main result. The Metropolis-Hastings algorithm invented by Nicholas Metropolis at al. [10] and W. Keith Hastings [5] is one of the best recognized Markov chain Monte Carlo (MCMC) techniques in the statistical applications (see e.g. [3, 4, 6, 7, 14, 15, 16, 17, 18]). Throughout this paper we shall assume that the following conditions are valid.

$\mathcal{H}$: Let $(\mathbb{X}, \mathcal{A}, \lambda)$ be some measure space with a σ-finite measure λ. Assume we are given a *target* probability distribution on $(\mathbb{X}, \mathcal{A})$ which is absolutely continuous with respect to λ with density $\pi(\cdot) : \mathbb{X} \to \mathbb{R}_+$ for which $\pi(x) > 0$ for all $x \in \mathbb{X}$. Assume also we are given an absolutely continuous with respect to λ *proposal* distribution on $(\mathbb{X}, \mathcal{A})$ which density $q(\cdot|x) : \mathbb{X} \to \mathbb{R}_+$ is set conditionally to $x \in \mathbb{X}$. It is assumed that $q(\cdot|\cdot) : \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$ is jointly $\mathcal{A} \times \mathcal{A}$ measurable (see e.g. [1]).

Note that $\mathcal{H}$ brings entirely technical nature, providing the proper container for the further treatments. The assumption that $\pi(x) > 0$ for all $x \in \mathbb{X}$ offers also some technical facilities and does not represent a limitation of the generality in the present work purposes.

All the probability densities in this paper are considered with respect to the common reference measure λ.

The Metropolis-Hastings algorithm, as a MCMC algorithm, serves for sampling from the target distribution $\pi(\cdot)$ and consists of the following steps. Generate first initial draw $x_{(0)}$. Let we know the current draw $x_{(n-1)}$. To obtain the next draw $x_{(n)}$ one should generate a candidate $x_* \sim q(x|x_{(n-1)})$ and accept the candidate with a probability $\alpha(x_{(n-1)}, x_*)$ taking $x_{(n)} = x_*$ or reject the candidate with a probability $1 - \alpha(x_{(n-1)}, x_*)$ and take $x_{(n)} = x_{(n-1)}$ where

$$\alpha(x, x') = \min\left(1, \frac{\pi(x')}{\pi(x)}\frac{q(x|x')}{q(x'|x)}\right) \text{ for } \pi(x)q(x'|x) > 0.$$

Following [17] we set $\alpha(x, x') = 1$ for $q(x'|x) = 0$ to avoid ambiguity. All draws are taken from $\mathbb{X}$. This scheme defines a transition kernel

$$(1.1) \quad \kappa(x \to x') = \alpha(x, x')q(x'|x) + \delta(x - x') \int (1 - \alpha(x, z)) \, q(z|x)\lambda(dz).$$

where $\delta(\cdot)$ is the delta function. The integral sign stands for the λ integration over $\mathbb{X}$ (including where it is necessary the delta function rule). The notation $\kappa(x \to x')$ stands for a function of two variables $(x, x') \in \mathbb{X} \times \mathbb{X}$ associated (by analogy to the discrete state space) with the conditional probability to move from state x to state x'. According to the assumptions for $\pi(\cdot)$ and $q(\cdot|\cdot)$ the kernel (1.1) is nonnegative function. This kernel fulfills the normalizing condition $\int \kappa(x \to x')\lambda(dx') = 1$ but first of all it is well known that the kernel satisfies the *detailed balance condition* (reversibility of the chain)

$$(1.2) \qquad\qquad \pi(x)\kappa(x \to x') = \pi(x')\kappa(x' \to x).$$

From the detailed balance condition it follows that the target density is *invariant* (stationary) for the kernel, i.e. it holds $\pi(x') = \int \pi(x)\kappa(x \to x')\lambda(dx)$. The transition kernel (1.1) defines a Metropolis-Hastings Markov chain (shortly MH-chain) of $\mathbb{X}$-valued random variables $(X_{(n)})$ according to the following rule. Define the initial random variable $X_{(0)}$ with some proper density $f_{(0)}(\cdot) : \mathbb{X} \to \mathbb{R}_+$. For any next random variable $X_{(n)}$ the corresponding density $f_{(n)}(\cdot)$ is defined by the recurrent formula

$$f_{(n)}(x') = \int f_{(n-1)}(x)\kappa(x \to x')\lambda(dx), n = 1, 2, \ldots.$$

One of the main problems arise here is to establish conditions under which the sequence $(f_{(n)}(\cdot))$ converges to the invariant density $\pi(\cdot)$. In the general case of stationary Markov chain usually proves that this sequence converges with respect to the total variation distance d_{TV}, i.e. that

$$(1.3) \qquad \lim_{n\to\infty} d_{TV}(\mu[f_{(n)}], \mu[\pi]) = \lim_{n\to\infty} \frac{1}{2} \int |f_{(n)}(x) - \pi(x)|\lambda(dx) = 0$$

under various specific assumptions (see e.g. [2, 3, 1, 9, 11, 12, 15, 17]). Here by $\mu[f]$ we denote the probability measure associated with the density $f(\cdot)$.

In this paper (Theorem 4.1) we propose conditions under which (1.3) holds but we follow a somewhat different approach by means of the properly defined Hilbert space described for example in Stroock [21].

2. Some preliminaries and notes. This study was motivated by the well-known and clearly proven result in the discrete case where the simple positivity of the transition matrix (or some its power) guarantees the convergence of the corresponding Markov chain to the stationary distribution, i.e. the positiveness occurs the only constructive condition needed for the convergence. So a natural question arises whether some proper positiveness condition is also sufficient for the convergence in the general state discrete time Markov chains? The answer turns out to be positive in the case of the Markov chains, associated with the Metropolis-Hastings algorithm.

The general state discrete time Markov chains convergence is well investigated (see e.g. [2, 3, 1, 9, 11, 12, 15, 17]) and very common advanced results were achieved by using of some specific notions as reversibility, irreducibility and aperiodicity. In the Metropolis-Hastings Markov chains we have two important particular advantages, sourced by the nature of the chain. Such a chain is always reversible and the target distribution stands for the (known) stationary distribution. These facts simplify the environment of the proof scenario.

The most famous convergence result for the MH-chains, formulated in easily verifiable conditions, is announced for example in [15] (Theorem 7.4 along with Corollary 7.5). Therein is shown that the positivity of the proposal distribution

(2.1) $$q(x|x') > 0 \ for \ all \ (x, x') \in \mathbb{X} \times \mathbb{X},$$

provides the irreducibility of the corresponding MH-chain and also if the algorithm admits the event $x_{(n)} = x_{(n-1)}$ with nonzero probability then the MH-chain is aperiodic. The second claim of Theorem 7.4 in [15] says that both irreducibility and aperiodicity imply the total variation convergence. Also in [15] is pointed out the existence of classes of examples in which the event $x_{(n)} = x_{(n-1)}$ never occurs.

Concisely formulated our main practical result (Corollary 4.1) states that if the proposal density $q(\cdot|\cdot)$ is positive then (1.3) holds regardless from the shape of the initial density. Therefore (2.1) provides also the aperiodicity of the chain because it is a necessary condition for an irreducible chain to converge.

3. The $L^2(\pi)$ structure. Following Stroock [21] we shall consider the Hilbert space $L^2(\pi)$ with an inner product

$$\langle f, g \rangle_\pi = \int f(x)g(x)\pi(x)\lambda(dx).$$

The space $L^2(\pi)$ consists of the measurable functions $f(\cdot) : \mathbb{X} \to \bar{\mathbb{R}}$ for which

$$\|f\|_{2,\pi} = \sqrt{\int |f(x)|^2 \pi(x)\lambda(dx)} < \infty$$

(see e.g. [13, 19, 20]). Define the operator

(3.1) $$\mathcal{K}[f](x) = \int \kappa(x \to x')f(x')\lambda(dx')$$
$$= \int \mathring{\kappa}(x \to x')f(x')\lambda(dx') + \phi(x)f(x)$$

which is formally conjugate to the basic transition operator of the chain

(3.2) $$\hat{\mathcal{K}}[f](x') = \int f(x)\kappa(x \to x')\lambda(dx)$$

$$= \int f(x) \mathring{\kappa}(x \to x') \lambda(dx) + \phi(x') f(x')$$

where the sub-kernel $\mathring{\kappa}(\cdot \to \cdot) : \mathbb{X} \times \mathbb{X} \to \bar{\mathbb{R}}_+$

$$\mathring{\kappa}(x \to x') = \min \left(q(x'|x), \frac{\pi(x')}{\pi(x)} q(x|x') \right)$$

is nonnegative $\mathcal{A} \times \mathcal{A}$ measurable function and the function $\phi(\cdot) : \mathbb{X} \to \bar{\mathbb{R}}_+$

$$\phi(x) = \int \left(1 - \alpha(x, z)\right) q(z|x) \lambda(dz)$$

is measurable with $0 \le \phi(x) \le 1$ for $x \in \mathbb{X}$. Actually $\kappa(\cdot \to \cdot)$ stands for a transition kernel of the transition operator $\hat{\mathcal{K}}$ and simply is a kernel of the conjugate operator $\mathcal{K}$.

Put $\kappa_1(x \to x') = \kappa(x \to x')$ and compose formally the sequence of kernels

$$\kappa_n(x \to x') = \int \kappa_{n-1}(x \to z) \kappa_1(z \to x') \lambda(dz), n = 2, 3, \ldots,$$

which are just the transition kernels of the transition-like operators $\hat{\mathcal{K}}^n$ in a sense that

$$\hat{\mathcal{K}}^n[f](x') = \int f(x) \kappa_n(x \to x') \lambda(dx)$$

and the usual kernels of the operators $\mathcal{K}^n$, i.e.

$$\mathcal{K}^n[f](x) = \int \kappa_n(x \to x') f(x') \lambda(dx').$$

Put also $\mathring{\kappa}_1(x \to x') = \mathring{\kappa}(x \to x')$ and compose the sub-kernels

(3.3) $$\mathring{\kappa}_n(x \to x') = \int \mathring{\kappa}_{n-1}(x \to z) \mathring{\kappa}_1(z \to x') \lambda(dz), n = 2, 3, \ldots.$$

One can find by induction that

$$\mathcal{K}^n[f](x) = \int \kappa_n(x \to x') f(x') \lambda(dx')$$

$$= \int \mathring{\kappa}_n(x \to x') f(x') \lambda(dx') + \int \chi_n(x \to x') f(x') \lambda(dx') + \phi^n(x) f(x)$$

where $\chi_n(\cdot \to \cdot) : \mathbb{X} \times \mathbb{X} \to \bar{\mathbb{R}}_+$ is some nonnegative $\mathcal{A} \times \mathcal{A}$ measurable function. One can verify that $\kappa_n(\cdot \to \cdot)$ also satisfies the detailed balance condition and the Chapman-Kolmogorov equation

$$\kappa_{m+n}(x \to x') = \int \kappa_m(x \to z)\kappa_n(z \to x')\lambda(dz), m = 1, 2, \dots, n = 1, 2, \dots,$$

and the same is true for the sub-kernel $\mathring{\kappa}_n(\cdot \to \cdot)$.

The proofs of the first two claims of the following proposition can be found substantially for example in [21], but here we propose our ones for the sake of completeness.

Proposition 3.1. *Suppose $\mathcal{H}$ holds and let $f \in L^2(\pi)$. Then the following assertions are valid for the operator defined in (3.1) .*
1) $\mathcal{K}[f] \in L^2(\pi)$ *and also*

(3.4) $$\|\mathcal{K}[f]\|_{2,\pi} \leq \|f\|_{2,\pi}.$$

2) *The operator $\mathcal{K} : L^2(\pi) \to L^2(\pi)$ is self-adjoint and for its norm we have*

(3.5) $$\|\mathcal{K}\| \leq 1.$$

3) *Suppose that there exists an integer $n \geq 1$ such that $\mathring{\kappa}_n(\cdot \to \cdot) > 0$ a.e. $(\lambda \times \lambda)$ in $\mathbb{X} \times \mathbb{X}$ where $\mathring{\kappa}_n(\cdot \to \cdot)$ is a composite sub-kernel defined in (3.3). Let also $h \in L^2(\pi)$ be a function for which $\mathcal{K}^n[h] = h$. Then there exists a constant γ such that $h(\cdot) = \gamma$ a.e. (λ) in $\mathbb{X}$.*

P r o o f.   1) Let $f \in L^2(\pi)$. Applying the Holder's inequality to the functions $|f|\sqrt{\kappa}$ and $\sqrt{\kappa}$ we find

$$|\mathcal{K}[f](x)|^2 \leq \int \kappa(x \to x')|f(x')|^2\lambda(dx').$$

Multiplying the latter with $\pi(x)$ and integrating over $\mathbb{X}$ we get

$$\|\mathcal{K}[f]\|_{2,\pi}^2 \leq \int \left(\int \pi(x)\kappa(x \to x')|f(x')|^2\lambda(dx') \right) \lambda(dx)$$
$$= \int \left(\int \pi(x')\kappa(x' \to x)|f(x')|^2\lambda(dx') \right) \lambda(dx)$$
$$= \int \pi(x')|f(x')|^2\lambda(dx') = \|f\|_{2,\pi}^2.$$

Here we use the detailed balance condition (1.2) and the Tonelli's theorem which allows us to interchange the order of integration. Thus we prove simultaneously the inequality (3.4) and the fact that $\mathcal{K}[f] \in L^2(\pi)$.

2) Let $f \in L^2(\pi)$ and $g \in L^2(\pi)$ and write by means of the Fubini's theorem and by the detailed balance condition

$$
\begin{aligned}
\langle \mathcal{K}[f], g \rangle_\pi &= \int \left(\int \kappa(x \to x') f(x') \lambda(dx') \right) g(x) \pi(x) \lambda(dx) \\
&= \int \left(\int \pi(x) \kappa(x \to x') f(x') \lambda(dx') \right) g(x) \lambda(dx) \\
&= \int f(x') \left(\int \kappa(x' \to x) g(x) \lambda(dx) \right) \pi(x') \lambda(dx') = \langle f, \mathcal{K}[g] \rangle_\pi.
\end{aligned}
$$

which proves that the operator $\mathcal{K}$ is self-adjoint. The inequality (3.5) follows immediately from (3.4).

3) Write the identity

$$
h^2(x') = h^2(x) + 2h(x)(h(x') - h(x)) + (h(x') - h(x))^2
$$

multiply with $\kappa_n(x \to x')$ and integrate. Then we get

$$
\int \kappa_n(x \to x') h^2(x') \lambda(dx') = h^2(x) + \int \kappa_n(x \to x')(h(x') - h(x))^2 \lambda(dx')
$$

because

$$
\int \kappa_n(x \to x') 2h(x)(h(x') - h(x)) \lambda(dx') = 2h(x) \left(\mathcal{K}^n[h](x) - h(x) \right) = 0.
$$

Multiply with $\pi(x)$ and integrate. Then

$$
(3.6) \quad \int \left(\int \pi(x) \kappa_n(x \to x') h^2(x') \lambda(dx') \right) \lambda(dx) = \int \pi(x) h^2(x) \lambda(dx)
$$

$$
+ \int \pi(x) \left(\int \kappa_n(x \to x')(h(x') - h(x))^2 \lambda(dx') \right) \lambda(dx).
$$

It is easy to see that the left-hand side in (3.6) is equal to the first addend in the right-hand side. Therefore

$$
\int \pi(x) \left(\int \kappa_n(x \to x')(h(x') - h(x))^2 \lambda(dx') \right) \lambda(dx) = 0
$$

which implies immediately that also

$$\int \pi(x) \left(\int \mathring{\kappa}_n(x \to x')(h(x') - h(x))^2 \lambda(dx') \right) \lambda(dx) = 0.$$

Now the inequalities $\pi(\cdot) > 0$ and $\kappa_n(\cdot \to \cdot) > 0$ give that there exists a constant γ such that $h(\cdot) = \gamma$ a.e. (λ) in $\mathbb{X}$. $\quad\square$

Proposition 3.2. *Suppose $\mathcal{H}$ holds and let $\mathcal{K} : L^2(\pi) \to L^2(\pi)$ be the operator defined above. Then for any integer $\nu \geq 1$ the following assertions are valid.*

1) *Every power $\mathcal{K}^{2\nu n}$, $n = 1, 2, \ldots$, is positive operator, i.e. $\langle \mathcal{K}^{2\nu n}[h], h \rangle_\pi \geq 0$ for any $h \in L^2(\pi)$.*

2) *The sequence $(\mathcal{K}^{2\nu n})$ is decreasing, i.e. $\langle \mathcal{K}^{2\nu n + 2\nu}[h], h \rangle_\pi \leq \langle \mathcal{K}^{2\nu n}[h], h \rangle_\pi$ for any $h \in L^2(\pi)$, $n = 1, 2, \ldots$.*

3) *All the operators $\left(\mathcal{K}^{2\nu n} - \mathcal{K}^{2\nu n + 2\nu p} \right)$ for $n = 1, 2, \ldots$ and $p = 1, 2, \ldots$ are also positive.*

P r o o f. 1) Let $h \in L^2(\pi)$. The operator $\mathcal{K}$ is self-adjoint therefore

$$\langle \mathcal{K}^{2\nu n}[h], h \rangle_\pi = \langle \mathcal{K}^{\nu n}[h], \mathcal{K}^{\nu n}[h] \rangle_\pi \geq 0.$$

2) We have $\|\mathcal{K}\| \leq 1$ therefore

$$\begin{aligned}
\langle \mathcal{K}^{2\nu n + 2\nu}[h], h \rangle_\pi &= \langle \mathcal{K}^{\nu n + \nu}[h], \mathcal{K}^{\nu n + \nu}[h] \rangle_\pi \\
&= \|\mathcal{K}^{\nu n + \nu}[h]\|_{2,\pi}^2 = \|\mathcal{K}^\nu \left[\mathcal{K}^{\nu n}[h] \right]\|_{2,\pi}^2 \leq \|\mathcal{K}^{\nu n}[h]\|_{2,\pi}^2 \\
&= \langle \mathcal{K}^{\nu n}[h], \mathcal{K}^{\nu n}[h] \rangle_\pi = \langle \mathcal{K}^{2\nu n}[h], h \rangle_\pi.
\end{aligned}$$

3) The positiveness of the operator $\left(\mathcal{K}^{2\nu n} - \mathcal{K}^{2\nu n + 2\nu p} \right)$ means that

$$\left\langle \left(\mathcal{K}^{2\nu n} - \mathcal{K}^{2\nu n + 2\nu p} \right)[h], h \right\rangle_\pi \geq 0$$

for any $h \in L^2(\pi)$ that is equivalent to

$$\langle \mathcal{K}^{2\nu n}[h], h \rangle_\pi \geq \langle \mathcal{K}^{2\nu n + 2\nu p}[h], h \rangle_\pi$$

which follows immediately from 2). $\quad\square$

Here we are at position to prove that the operator sequence $(\mathcal{K}^n)$ has a strong limit. More precisely we are going to prove that for every $f \in L^2(\pi)$ there exists the limit

$$\lim_{n \to \infty} \mathcal{K}^n[f] = \langle f, \mathbf{1} \rangle_\pi \mathbf{1}$$

where **1** denotes the constant function which equals to one.

Further we shall need the following condition of positiveness.

$\mathcal{H}_p$: Assume that there exists an integer $\nu \geq 1$ for which $\mathring{\kappa}_\nu(\cdot \to \cdot) > 0$ a.e. $(\lambda \times \lambda)$ in $\mathbb{X} \times \mathbb{X}$. $\square$

The condition $\mathcal{H}_p$ is met for example when the proposal density is positive. Remember that the target density is positive by condition $\mathcal{H}$.

Theorem 3.1. *Suppose $\mathcal{H}$ and $\mathcal{H}_p$ hold. Then for every $f \in L^2(\pi)$ we have*

$$(3.7) \qquad \lim_{n \to \infty} \|\mathcal{K}^n[f] - \langle f, \mathbf{1} \rangle_\pi \mathbf{1}\|_{2,\pi} = 0.$$

P r o o f.   It is not difficult to find out that for any real Hilbert space H with an inner product $\langle \cdot, \cdot \rangle$ and a norm $\|u\| = \sqrt{\langle u, u \rangle}, u \in H$, with a given linear bounded self-adjoint positive operator $T : H \to H$ it holds the inequality

$$(3.8) \qquad \|Tu\|^2 \leq \|T\| \langle Tu, u \rangle.$$

The proof of (3.8) will be given at the end of the paper. Choose arbitrary $f \in L^2(\pi)$. Applying (3.8) to the positive operators $(\mathcal{K}^{2\nu n} - \mathcal{K}^{2\nu n + 2\nu p})$ for $n = 1, 2, \dots$ and $p = 1, 2, \dots$ we get

$$\|(\mathcal{K}^{2\nu n} - \mathcal{K}^{2\nu n + 2\nu p})[f]\|_{2,\pi}^2 \leq \|\mathcal{K}^{2\nu n} - \mathcal{K}^{2\nu n + 2\nu p}\| \langle \mathcal{K}^{2\nu n}[f] - \mathcal{K}^{2\nu n + 2\nu p}[f], f \rangle_\pi$$

from which follows that

$$(3.9) \qquad \|\mathcal{K}^{2\nu n}[f] - \mathcal{K}^{2\nu n + 2\nu p}[f]\|_{2,\pi}^2 \leq 2 \left(\langle \mathcal{K}^{2\nu n}[f], f \rangle_\pi - \langle \mathcal{K}^{2\nu n + 2\nu p}[f], f \rangle_\pi \right).$$

From Proposition 3.2 we know that the numerical sequence $(\langle \mathcal{K}^{2\nu n}[f], f \rangle_\pi)_{n=1}^\infty$ is decreasing an bounded from below by zero therefore this sequence is convergent. Now from (3.9) it follows that the sequence of the powers $(\mathcal{K}^{2\nu n}[f])_{n=1}^\infty$ is a Cauchy sequence in $L^2(\pi)$ therefore it has a limit $h \in L^2(\pi)$ for which obviously it holds $\mathcal{K}^{2\nu}[h] = h$. From Proposition 3.1(3) (with $n = 2\nu$) we get that $h(\cdot) = \gamma$ a.e. (λ) in $\mathbb{X}$ with some constant γ because the stated positiveness of the sub-kernel $\mathring{\kappa}_\nu(\cdot \to \cdot)$ in $\mathcal{H}_p$ provides that

$$\mathring{\kappa}_{2\nu}(x \to x') = \int \mathring{\kappa}_\nu(x \to z) \mathring{\kappa}_\nu(z \to x') \lambda(dz) > 0$$

a.e. $(\lambda \times \lambda)$ in $\mathbb{X} \times \mathbb{X}$. We have $\mathcal{K}^{2\nu n}[f] \to \gamma \mathbf{1}$ whence $\langle \mathcal{K}^{2\nu n}[f], \mathbf{1} \rangle_\pi \to \gamma \langle \mathbf{1}, \mathbf{1} \rangle_\pi = \gamma$ which gives

$$\langle f, \mathbf{1} \rangle_\pi = \langle f, \mathcal{K}^{2\nu n}[\mathbf{1}] \rangle_\pi = \langle \mathcal{K}^{2\nu n}[f], \mathbf{1} \rangle_\pi \to \gamma$$

therefore $\gamma = \langle f, \mathbf{1} \rangle_\pi$ which proves (3.7) for the subsequence of the powers $(2\nu n)_{n=1}^\infty$, i.e. that

$$(3.10) \qquad \lim_{n\to\infty} \mathcal{K}^{2\nu n}[f] = \langle f, \mathbf{1} \rangle_\pi \mathbf{1}.$$

From (3.10) we obtain

$$(3.11) \quad \lim_{n\to\infty} \mathcal{K}^{2\nu n + m}[f] = \lim_{n\to\infty} \mathcal{K}^{2\nu n}[\mathcal{K}^m[f]]$$
$$= \langle \mathcal{K}^m[f], \mathbf{1} \rangle_\pi \mathbf{1} = \langle f, \mathcal{K}^m[\mathbf{1}] \rangle_\pi \mathbf{1} = \langle f, \mathbf{1} \rangle_\pi \mathbf{1}$$

which proves (3.7) for the all the power subsequences $(2\nu n + m)_{n=1}^\infty$ where m, $1 \le m < 2\nu$, is a nonzero remainder after a division by 2ν. Now it is not difficult to see that the validity of (3.7) follows from (3.10) and (3.11). □

4. Convergence with respect to TV distance. Our main purpose is to investigate the behavior of the operator sequence $(\hat{\mathcal{K}}^n)$ rather than the sequence $(\mathcal{K}^n)$ where the transition operator $\hat{\mathcal{K}}$ is defined in (3.2) because it actually corresponds to the Markov chain. Note that if μ_1 and μ_2 are absolutely continuous probability measures (w.r.t. λ) with densities $f_1(\cdot)$ and $f_2(\cdot)$ then for the total variation distance $d_{TV}(\mu_1, \mu_2)$ it holds (see e.g. [9, 17])

$$d_{TV}(\mu_1, \mu_2) = \frac{1}{2} \int |f_1(x) - f_2(x)| \lambda(dx).$$

Let $L_{\mathbb{X}}^1$ be the Banach space of the measurable functions $f(\cdot) : \mathbb{X} \to \bar{\mathbb{R}}$ for which

$$\|f\|_1 = \int |f(x)| \lambda(dx) < \infty$$

provided with the usual norm $\| \cdot \|_1$. We have

$$(4.1) \qquad \|f\|_1 \le \|f/\pi\|_{2,\pi}$$

because (by Cauchy-Schwarz inequality)

$$\|f\|_1 = \int |f(x)| \lambda(dx)$$
$$= \int \pi(x) \left| \frac{f(x)}{\pi(x)} \right| \lambda(dx) = \int \sqrt{\pi(x)} \left| \sqrt{\pi(x)} \frac{f(x)}{\pi(x)} \right| \lambda(dx)$$
$$\le \sqrt{\int \pi(x) \lambda(dx)} \sqrt{\int \pi(x) \left| \frac{f(x)}{\pi(x)} \right|^2 \lambda(dx)} = \|f/\pi\|_{2,\pi}.$$

From (4.1) it follows that if $f/\pi \in L^2(\pi)$ then $f \in L_{\mathbb{X}}^1$.

Proposition 4.1. *Suppose $\mathcal{H}$ and $\mathcal{H}_p$ hold. Let $f(\cdot)$ be a function such that $f/\pi \in L^2(\pi)$ and put $\gamma = \int f(x)\lambda(dx)$. Then*

$$(4.2) \qquad\qquad \lim_{n\to\infty} \|\hat{\mathcal{K}}^n[f] - \gamma\pi\|_1 = 0.$$

Proof. First of all notice that (4.1) guarantees the existence of the constant γ. It can be shown by induction that

$$\hat{\mathcal{K}}^n[f](x') = \pi(x')\mathcal{K}^n\left[\frac{f}{\pi}\right](x'), n = 1, 2, \ldots,$$

whence by means of the Cauchy-Schwarz inequality we obtain

$$(4.3) \quad \|\hat{\mathcal{K}}^n[f] - \gamma\pi\|_1 = \int |\hat{\mathcal{K}}^n[f](x') - \gamma\pi(x')|\lambda(dx')$$

$$= \int \pi(x')\left|\mathcal{K}^n\left[\frac{f}{\pi}\right](x') - \gamma\mathbf{1}(x')\right|\lambda(dx')$$

$$= \int \sqrt{\pi(x')}\left(\sqrt{\pi(x')}\left|\mathcal{K}^n\left[\frac{f}{\pi}\right](x') - \gamma\mathbf{1}(x')\right|\right)\lambda(dx')$$

$$\leq \sqrt{\int \pi(x')\left|\mathcal{K}^n\left[\frac{f}{\pi}\right](x') - \gamma\mathbf{1}(x')\right|^2\lambda(dx')} = \left\|\mathcal{K}^n\left[\frac{f}{\pi}\right] - \gamma\mathbf{1}\right\|_{2,\pi}$$

On the other hand by (3.7) it follows that

$$(4.4) \qquad\qquad \lim_{n\to\infty} \left\|\mathcal{K}^n\left[\frac{f}{\pi}\right] - \gamma\mathbf{1}\right\|_{2,\pi} = 0$$

because

$$\left\langle \frac{f}{\pi}, \mathbf{1} \right\rangle_\pi = \int \frac{f(x)}{\pi(x)}\pi(x)\lambda(dx) = \int f(x)\lambda(dx) = \gamma.$$

Now the validity of (4.2) follows immediately from (4.3) and (4.4). □

Hereafter we shall prepare for the final results. Put

$$\mathbb{X}_m = \left(x \in \mathbb{X} | \pi(x) \geq \frac{1}{m}\right), m = 1, 2, \ldots.$$

Obviously $\mathbb{X}_m \subseteq \mathbb{X}_{m+1}$ and $\mathbb{X} = \cup_{m=1}^\infty \mathbb{X}_m$. For $f(\cdot) : \mathbb{X} \to \bar{\mathbb{R}}$ put $f_{[m]}(x) = f(x)$ where $x \in \mathbb{X}_m$ and $f_{[m]}(x) = 0$ elsewhere.

The following proposition helps to prove Theorem 4.1.

Proposition 4.2. *Suppose $\mathcal{H}$ holds. Then the following assertions are true.*

1) Let $f \in L^1_{\mathbb{X}}$. Then $\hat{\mathcal{K}}[f] \in L^1_{\mathbb{X}}$ and

$$(4.5) \qquad\qquad \|\hat{\mathcal{K}}[f]\|_1 \leq \|f\|_1$$

consequently for any $f \in L^1_{\mathbb{X}}$ and $g \in L^1_{\mathbb{X}}$ and any $n = 1, 2, \ldots$ we have

$$(4.6) \qquad\qquad \|\hat{\mathcal{K}}^n[f] - \hat{\mathcal{K}}^n[g]\|_1 \leq \|f - g\|_1.$$

2) Let $f \in L^1_{\mathbb{X}}$ be a bounded function. Then $f_{[m]}/\pi \in L^2(\pi)$ and

$$(4.7) \qquad\qquad \lim_{m \to \infty} \|f - f_{[m]}\|_1 = 0.$$

3) Let $f \in L^1_{\mathbb{X}}$ and put

$$\gamma = \int f(x)\lambda(dx), \gamma_{[m]} = \int f_{[m]}(x)\lambda(dx), m = 1, 2, \ldots.$$

Then

$$(4.8) \qquad\qquad \|\gamma_{[m]}\pi - \gamma\pi\|_1 \leq \|f - f_{[m]}\|_1.$$

P r o o f. 1) We have

$$|\hat{\mathcal{K}}[f](x')| \leq \int |f(x)|\kappa(x \to x')\lambda(dx)$$

whence (again by means of the Fubini's theorem)

$$\|\hat{\mathcal{K}}[f]\|_1 = \int |\hat{\mathcal{K}}[f](x')|\lambda(dx')$$
$$\leq \int \left(\int |f(x)|\kappa(x \to x')\lambda(dx) \right)\lambda(dx') = \int |f(x)|\lambda(dx) = \|f\|_1$$

which proves (4.5). The validity of (4.6) follows immediately from (4.5) and the linearity of $\hat{\mathcal{K}}$.

2) According to the assumption $f(\cdot)$ is bounded consequently for some constant C it holds $|f_{[m]}(x)| \leq C$ for $x \in \mathbb{X}_m$. Then

$$\left| \frac{f_{[m]}(x)}{\pi(x)} \right| \leq Cm, x \in \mathbb{X}_m,$$

therefore

$$\int \pi(x) \left| \frac{f_{[m]}(x)}{\pi(x)} \right|^2 \lambda(dx) \leq \int_{\mathbb{X}_m} \pi(x) |Cm|^2 \lambda(dx) \leq C^2 m^2 < \infty$$

which proves that $f_{[m]}/\pi \in L^2(\pi)$. By the definition

$$\|f - f_{[m]}\|_1 = \int |f(x) - f_{[m]}(x)|\lambda(dx) = \int_{\mathbb{X}\setminus\mathbb{X}_m} |f(x)|\lambda(dx)$$

which proves (4.7) because $\mathbb{X}_m \nearrow \mathbb{X}$.

3) We have

$$\|\gamma_{[m]}\pi - \gamma\pi\|_1 = \int |\gamma_{[m]}\pi(x') - \gamma\pi(x')|\lambda(dx')$$

$$= |\gamma_{[m]} - \gamma| \int \pi(x')\lambda(dx') = \left| \int_{\mathbb{X}\setminus\mathbb{X}_m} f(x)\lambda(dx) \right| \leq \|f - f_{[m]}\|_1$$

which proves (4.8). □

We are ready to give more general conditions under which (4.2) is valid.

Theorem 4.1. *Suppose $\mathcal{H}$ and $\mathcal{H}_p$ hold. Let $f \in L^1_{\mathbb{X}}$ and put $\gamma = \int f(x)\lambda(dx)$. Then*

(4.9)
$$\lim_{n\to\infty} \|\hat{\mathcal{K}}^n[f] - \gamma\pi\|_1 = 0.$$

Therefore if $f \in L^1_{\mathbb{X}}$ is a probability density function (w.r.t. λ) on $\mathbb{X}$ then

(4.10)
$$\lim_{n\to\infty} d_{TV}(\mu[\hat{\mathcal{K}}^n[f]], \mu[\pi]) = 0.$$

P r o o f.   In the beginning of this proof we shall assume that the function $f \in L^1_{\mathbb{X}}$ is bounded. Put again

$$\gamma_{[m]} = \int f_{[m]}(x)\lambda(dx), m = 1, 2, \ldots.$$

For any $n = 1, 2, \ldots$ and $m = 1, 2, \ldots$ we can write

(4.11) $\|\hat{\mathcal{K}}^n[f] - \gamma\pi\|_1 \leq \|\hat{\mathcal{K}}^n[f] - \hat{\mathcal{K}}^n[f_{[m]}]\|_1 + \|\hat{\mathcal{K}}^n[f_{[m]}] - \gamma_{[m]}\pi\|_1 + \|\gamma_{[m]}\pi - \gamma\pi\|_1$

Choose some $\varepsilon > 0$. By (4.7) fix an integer $m \geq 1$ such that $\|f - f_{[m]}\|_1 < \varepsilon/3$. Then according to (4.6) we obtain

$$(4.12) \qquad \|\hat{\mathcal{K}}^n[f] - \hat{\mathcal{K}}^n[f_{[m]}]\|_1 < \frac{\varepsilon}{3}$$

for any $n = 1, 2, \ldots$ and according to (4.8) we obtain

$$(4.13) \qquad \|\gamma_{[m]}\pi - \gamma\pi\|_1 < \frac{\varepsilon}{3}.$$

For such a fixed m we have from Proposition 4.2(2) that $f_{[m]}/\pi \in L^2(\pi)$ therefore by Proposition 4.1 we get that

$$\lim_{n \to \infty} \|\hat{\mathcal{K}}^n[f_{[m]}] - \gamma_{[m]}\pi\|_1 = 0$$

consequently we can choose an positive integer n_0 such that

$$(4.14) \qquad \|\hat{\mathcal{K}}^n[f_{[m]}] - \gamma_{[m]}\pi\|_1 < \frac{\varepsilon}{3}$$

for any $n > n_0$. Replacing the inequalities (4.12), (4.13) and (4.14) in (4.11) we receive that

$$\|\hat{\mathcal{K}}^n[f] - \gamma\pi\|_1 < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

for any $n > n_0$ which by definition proves the validity of (4.9) for the case of bounded $f \in L_{\mathbb{X}}^1$.

Choose now arbitrary $f \in L_{\mathbb{X}}^1$ and put $f_m(x) = f(x)$ where $|f(x)| \leq m$ and $f_m(x) = 0$ elsewhere, $m = 1, 2, \ldots$. In the same way as for (4.11) one can get

$$\|\hat{\mathcal{K}}^n[f] - \gamma\pi\|_1 \leq \|\hat{\mathcal{K}}^n[f_m] - \gamma_m\pi\|_1 + 2\|f_m - f\|_1$$

where

$$\gamma_m = \int f_m(x)\lambda(dx).$$

From the first part of the proof we already know that for arbitrary fixed positive integer m it holds

$$\lim_{n \to \infty} \|\hat{\mathcal{K}}^n[f_m] - \gamma_m\pi\|_1 = 0$$

because $f_m(\cdot)$ is bounded. Now the validity of (4.9) is a consequence of the well-known fact that

$$\lim_{m \to \infty} \|f_m - f\|_1 = 0.$$

The validity of (4.10) follows immediately from (4.9). □

Theorem 4.1 allows us to prove the following practically valuable result with easily verifiable conditions.

Corollary 4.1. *Suppose $\mathcal{H}$ holds. Let $q(x|x') > 0$ a.e. $(\lambda \times \lambda)$ in $\mathbb{X} \times \mathbb{X}$. Then (1.3) is valid for every initial density $f_0(\cdot) \in L^1_{\mathbb{X}}$.*

P r o o f. By conditions we have $\pi(x) > 0$ for all $x \in \mathbb{X}$ and $q(\cdot|\cdot) > 0$ a.e. $(\lambda \times \lambda)$ in $\mathbb{X} \times \mathbb{X}$. These inequalities guarantee that $\mathring{\kappa}(\cdot \to \cdot) > 0$ a.e. $(\lambda \times \lambda)$ in $\mathbb{X} \times \mathbb{X}$ therefore we can use $\mathcal{H}_p$ with $\nu = 1$. At this point we can apply Theorem 4.1 which implies consequently the validity of (4.9), (4.10) and (1.3). $\quad\square$

Now we can conclude that the condition (2.1) (besides the irreducibility shown for example in [15]) provides also the aperiodicity of the chain, because it is a necessary convergence condition for the irreducible chains (as it follows from Theorem 5.4.4 [11]).

5. Remarks. In this paper we offer a possible way to prove the total variance convergence of MH-chains, under the kernel positivity condition, by means referred more to the functional analysis rather than to the classical probability constructions. The offered method is backgrounded by some schemes described in [21] and does not utilize the notions of the irreducibility and aperiodicity. However it has an essential drawback, because the overall proof scheme lays heavily on the detailed balance condition property of the MH-chains. Thus our approach basically cannot apply to MCMC algorithms which do not fulfill the detailed balance condition.

It is not difficult to prove that the Kullback-Leibler divergence between the current MH-chain density $f_{(n)}(\cdot)$ and the target distribution $\pi(\cdot)$ do not increase as $n \to \infty$. The same is valid for the $d_{TV}(\mu[f_{(n)}], \mu[\pi])$, which is shown for example in [11] (Proposition 13.3.2). The latter can be considered as a good testimony for the validity of (1.3) at all under some natural requirements.

For the sake of completeness let us prove the validity of (3.8). The operator T is linear bounded self-adjoint and positive in the real Hilbert space H therefore it holds the Cauchy-like inequality

$$|\langle Tu, v \rangle|^2 \leq \langle Tu, u \rangle \langle Tv, v \rangle$$

for all $u \in H$ and all $v \in H$. Putting in the latter $v = Tu$ we obtain

(5.1) $$|\langle Tu, Tu \rangle|^2 = \|Tu\|^4 \leq \langle Tu, u \rangle \langle T^2 u, Tu \rangle.$$

Now applying the classical Cauchy inequality and the inequality for the norm we get

$$\langle T^2 u, Tu \rangle \leq \|T^2 u\| \|Tu\| \leq \|T\| \|Tu\| \|Tu\| = \|T\| \|Tu\|^2.$$

Replacing the latter in (5.1) we get the inequality

$$\|Tu\|^4 \leq \langle Tu, u \rangle \|T\| \|Tu\|^2$$

that is equivalent to $\|Tu\|^2 \leq \|T\| \langle Tu, u \rangle$, i.e. (3.8). Perhaps various proofs of (3.8) can be found in other places but we present here an explicit proof taking into account the importance of this inequality in our construction.

The operator $\hat{\mathcal{K}}$ considered in (3.1) represents an interesting nontrivial example for a self-adjoint operator. Also a careful inspection of the proofs in the discrete case leads to considering of a very interesting example of positive definite matrices of the type $A = (\min(\alpha_i, \alpha_j))$ where (α_k) are mutually different positive numbers.

Finally note that in the general case (2.1) is the only easily verifiable condition which provides the assumption $\mathcal{H}_p$.

REFERENCES

[1] K. B. ATHREYA, S. N. LAHIRI. Measure Theory and Probability Theory. Springer Texts in Statistics. New York, Springer, 2006.

[2] P. BILLINGSLEY. Probability and Measure. New York, John Wiley & Sons, 1995.

[3] M.-H. CHEN, Q.-M. SHAO, J. IBRAHIM. Monte Carlo Methods in Bayesian Computation. New York, Springer-Verlag Inc., 2000.

[4] A. GELMAN, J. B. CARLIN, H. S. STERN, D. B. RUBIN. Bayesian Data Analysis. London, Chapman & Hall, 2004.

[5] W. K. HASTINGS. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, *1* (1970), 97–109.

[6] F. LIANG, C. LIU, R. J. CARROLL. Advanced Markov Chain Monte Carlo Methods. Chichester, John Wiley and Sons Ltd, 2010.

[7] W. A. LINK, R. J. BARKER. Bayesian Inference: With Ecological Applications. Elsevier Ltd., 2010.

[8] J. S. Liu, W. H. Wong, A. Kong. Covariance structure and convergence rate of the Gibbs sampler with various scans. *J. Roy. Statist. Soc. Ser. B* **57**, *1* (1995), 157–169.

[9] N. Madras, D. Sezer. Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances. *Bernoulli* **16**, *3* (2010), 882–908.

[10] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, *6* (1953), 1087–1091.

[11] S. Meyn, R. L. Tweedie. Markov Chains and Stochastic Stability. Cambridge, Cambridge University Press, 2009.

[12] E. Nummelin. General Irreducible Markov Chains and Non-negative Operators. Cambridge, Cambridge University Press, 1984.

[13] M. Reed, B. Simon. Methods of Modern Mathematical Physics. I. Functional analysis. New York-London, Academic Press, 1972.

[14] C. P. Robert. The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation. Springer Texts in Statistics. New York, Springer, 2007.

[15] C. P. Robert, G. Casella. Monte Carlo Statistical Methods, 2nd edition. Springer Texts in Statistics. New York, Springer-Verlag, 2004.

[16] C. P. Robert, G. Casella. Introducing Monte Carlo Methods with R. Use R!. New York, Springer, 2010.

[17] G. O. Roberts, J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1**(2004), 20–71.

[18] R. Y. Rubinstein, D. P. Kroese. Simulation and the Monte Carlo Method, 2nd edition. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2008.

[19] W. Rudin. Principles of Mathematical Analysis, 3rd edition. International Series in Pure and Applied Mathematics. New York-Auckland-Düsseldorf, McGraw-Hill Book Co., 1976.

[20] W. Rudin. Real and Complex Analysis, 3rd edition. New York, McGraw-Hill Book Co., 1987.

[21] D. W. STROOCK. An Introduction to Markov Processes. Graduate Texts in Mathematics, vol. **230**. Berlin, Springer-Verlag, 2005.

Department of Mathematical Analysis and Applications
Faculty of Mathematics and Informatics
St. Cyril and St. Methodius University of Veliko Tarnovo
Veliko Tarnovo, Bulgaria
e-mail: `dimiter.p.tsvetkov@gmail.com` *(Dimiter Tsvetkov)*
e-mail: `lyubomir.hristov@gmail.com` *(Lyubomir Hristov)*

Ralitsa Angelova-Slavova
Department of Communication Systems and Technologies
Vasil Levski National Military University at Veliko Tarnovo
Veliko Tarnovo, Bulgaria
e-mail: `ralitsa.slavova@yahoo.com`