On the convergence of the Metropolis-Hastings Markov chains*

Dimiter Tsvetkov [1] **Lyubomir Hristov** [1] **and Ralitsa Angelova-Slavova** [2]

[1] *Department of "Mathematical Analysis and Applications"*
Faculty of Mathematics and Informatics
St. Cyril and St. Methodius University of Veliko Tarnovo e-mail:
dimiter.tsvetkov@yahoo.com; lyubomir.hristov@gmail.com

[2] *Department of "Communication Systems and Technologies"*
Vasil Levski National Military University at Veliko Tarnovo e-mail:
ralitsa.slavova@yahoo.com

Abstract: In this paper we consider Markov chains associated with the Metropolis-Hastings algorithm. We propose 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 target and the proposal densities is enough for the chain to converge.

Keywords and phrases: Markov chain, Metropolis-Hastings algorithm, total variation distance.

1. Introduction and the main result

The Metropolis-Hastings algorithm invented by Nicholas Metropolis at al. [10] and W. Keith Hastings [4] is one of the best recognized techniques in the statistical applications (see e.g. [2, 3, 6, 7, 14, 15, 16, 17, 20]). Throughout this paper we shall assume that the following condition 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. [5]). □

The Metropolis-Hastings algorithm 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

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

*This paper is a preprint on **viXra** url: http://vixra.org/.

and $\alpha(x,x')=1$ for $\pi(x)q(x'|x)=0$. All draws are taken from $\mathbb{X}$. This scheme defines a transition kernel

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

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 satisfies the *detailed balance condition* (*reversibility of the chain*)

$$\pi(x)\kappa(x \to x') = \pi(x')\kappa(x' \to x) \quad (1.2)$$

which has to be verified only for $x \neq x'$. Actually in this case we have

$$\pi(x)\kappa(x \to x') = \pi(x')\kappa(x' \to x) = \min\left(\pi(x)q(x'|x), \pi(x')q(x|x')\right).$$

From the detailed balance condition it follows that the target density is *invariant* 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 of $\mathbb{X}$-valued random variables $\left(X_{(n)}\right)$ 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

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

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

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

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

under the concepts of irreducibility, aperiodicity and reversibility (see e.g. [1, 2, 5, 9, 11, 12, 15, 17]). Here by $\mu[f]$ we denote the probability measure associated with the density $f(\cdot)$.

In this paper (Theorem 3.1) 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].

Concisely formulated our main practical result (Proposition 3.4) states that if both the target and the proposal densities are positive functions then (1.3) holds regardless from the shape of the initial density. Remember that we use no other constructive conditions besides the mentioned positiveness. In the theorems below we use essentially only the notion of the reversibility and positiveness.

2. 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 also e.g. [13, 18, 19]). Define the operator

$$\begin{aligned}\mathcal{K}[f](x) = \int \kappa(x\to x')f(x')\lambda(dx')\\ = \int \mathring{\kappa}(x\to x')f(x')\lambda(dx') + \phi(x)f(x) \quad (2.1)\end{aligned}$$

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

$$\begin{aligned}\hat{\mathcal{K}}[f](x') = \int f(x)\kappa(x\to x')\lambda(dx)\\ = \int f(x)\mathring{\kappa}(x\to x')\lambda(dx) + \phi(x')f(x') \quad (2.2)\end{aligned}$$

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 $\phi(\cdot):\mathbb{X}\to\bar{\mathbb{R}}_+$

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

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

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

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

$$\mathring{\kappa}_n(x \to x') = \int \mathring{\kappa}_{n-1}(x \to z)\mathring{\kappa}_1(z \to x')\lambda(dz), n = 2, 3, \dots. \tag{2.3}$$

One can find by induction that

$$\begin{aligned}\mathcal{K}^n[f](x) &= \int \kappa_n(x \to x')f(x')\lambda(dx') \\ &= \int \mathring{\kappa}_n(x \to x')f(x')\lambda(dx') + \int \chi_n(x \to x')f(x')\lambda(dx') + \phi^n(x)f(x)\end{aligned}$$

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 $\mathring{\kappa}_n(\cdot \to \cdot)$.

Proposition 2.1. *Suppose $\mathcal{H}$ holds and let $f \in L^2(\pi)$. Then for the operator defined in (2.1) it holds*
1) $\mathcal{K}[f] \in L^2(\pi)$ and also

$$\|\mathcal{K}[f]\|_{2,\pi} \le \|f\|_{2,\pi}. \tag{2.4}$$

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

$$\|\mathcal{K}\| \le 1. \tag{2.5}$$

3) Suppose that there exists an integer $n \ge 1$ 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 (2.3). 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 $\mathbb{X}$.

Proof. 1) Let $f \in L^2(\pi)$. In the beginning we assume that $f(\cdot)$ is bounded, i.e. $|f(\cdot)| \le C$ a.e. (λ) in $\mathbb{X}$ for some constant C. Put

$$A(x) = \int \kappa(x \to x')|f(x')|\lambda(dx')$$

and consider the identity

$$|f(x')|^2 = A^2(x) + 2A(x)(|f(x')| - A(x)) + (|f(x')| - A(x))^2.$$

Multiply the latter with $\kappa(x \to x')$ and integrate. Then receive

$$\int \kappa(x \to x')|f(x')|^2 \lambda(dx') = A^2(x) + \int \kappa(x \to x')(|f(x')| - A(x))^2 \lambda(dx')$$

because by construction

$$\int \kappa(x \to x') 2A(x)(|f(x')| - A(x))\lambda(dx') = 0.$$

Therefore

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

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

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

Here we firstly use the detailed balance condition (1.2) and secondly we use the Fubini's theorem which allows us to interchange the order of integration in the case of positive integrand. In this way we prove simultaneously the inequality (2.4) and the fact that $\mathcal{K}[f] \in L^2(\pi)$ for the case of bounded $f(\cdot)$.

Now let $f(\cdot)$ be any function of $L^2(\pi)$. It is enough to show the validity of 1) for the modulus $|f(\cdot)|$. Put $|f(\cdot)|_m = \min(|f(\cdot)|, m)$ where $m \ge 1$ is a positive integer. The function $|f|_m \in L^2(\pi)$ is bounded therefore by the first part of the proof we have

$$\begin{aligned}
\|\mathcal{K}[|f|_m]\|_{2,\pi}^2 &= \int \pi(x) \left(\int \kappa(x \to x')|f(x')|_m \lambda(dx') \right)^2 \lambda(dx) \\
&= \int \pi(x) \left(\int \mathring{\kappa}(x \to x')|f(x')|_m \lambda(dx') + \phi(x)|f(x)|_m \right)^2 \lambda(dx) \\
&\le \||f|_m\|_{2,\pi}^2 \le \|f\|_{2,\pi}^2
\end{aligned}$$

consequently

$$\int \pi(x) g_m(x) \lambda(dx) \le \|f\|_{2,\pi}^2 \tag{2.6}$$

where

$$g_m(x) = \left(\int \mathring{\kappa}(x \to x')|f(x')|_m \lambda(dx') + \phi(x)|f(x)|_m \right)^2$$

is an increasing sequence of positive measurable functions on $\mathbb{X}$. Now we are at position to apply the Fatou's lemma in (2.6) and after passing to a limit for $m \to \infty$ conclude that

$$\int \pi(x) \left(\int \mathring{\kappa}(x \to x')|f(x')|\lambda(dx') + \phi(x)|f(x)| \right)^2 \lambda(dx) \leq \|f\|^2_{2,\pi}$$

which proves simultaneously that $\mathcal{K}[|f|] \in L^2(\pi)$ and the inequality (2.4). 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 (2.5) follows immediately from (2.4). 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

$$\begin{aligned}\int \left(\int \pi(x)\kappa_n(x \to x')h^2(x')\lambda(dx') \right) \lambda(dx) &= \int \pi(x)h^2(x)\lambda(dx) \\ &+ \int \pi(x) \left(\int \kappa_n(x \to x')(h(x') - h(x))^2\lambda(dx') \right) \lambda(dx). \quad (2.7)\end{aligned}$$

It is easy to see that the left-hand side in (2.7) 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 $\mathbb{X}$. □

Proposition 2.2. *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.

Proof. 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

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

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

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

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). □

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 $\mathbf{1}$ 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 $\mathbb{X} \times \mathbb{X}$. □

Theorem 2.1. *Suppose $\mathcal{H}$ and $\mathcal{H}_p$ hold. Then for every $f \in L^2(\pi)$ we have*

$$\lim_{n\to\infty} \|\mathcal{K}^n[f] - \langle f, \mathbf{1}\rangle_\pi \mathbf{1}\|_{2,\pi} = 0. \tag{2.8}$$

Proof. 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

$$\|Tu\|^2 \leq \|T\|\langle Tu, u\rangle. \tag{2.9}$$

The proof of (2.9) will be given at the end of the paper. Choose arbitrary $f \in L^2(\pi)$. Applying (2.9) 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

$$\|\mathcal{K}^{2\nu n}[f] - \mathcal{K}^{2\nu n+2\nu p}[f]\|_{2,\pi}^2 \leq 2\left(\langle \mathcal{K}^{2\nu n}[f], f\rangle_\pi - \langle \mathcal{K}^{2\nu n+2\nu p}[f], f\rangle_\pi\right). \quad (2.10)$$

From Proposition 2.2 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 (2.10) 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 2.1(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 (2.8) for the subsequence of the powers $(2\nu n)_{n=1}^\infty$, i.e. that

$$\lim_{n\to\infty} \mathcal{K}^{2\nu n}[f] = \langle f, \mathbf{1}\rangle_\pi \mathbf{1}. \quad (2.11)$$

From (2.11) we obtain

$$\begin{aligned}\lim_{n\to\infty} \mathcal{K}^{2\nu n+m}[f] &= \lim_{n\to\infty} \mathcal{K}^{2\nu n}\left[\mathcal{K}^m[f]\right] \\ &= \langle \mathcal{K}^m[f], \mathbf{1}\rangle_\pi \mathbf{1} = \langle f, \mathcal{K}^m[\mathbf{1}]\rangle_\pi \mathbf{1} = \langle f, \mathbf{1}\rangle_\pi \mathbf{1} \quad (2.12)\end{aligned}$$

which proves (2.8) for the all the power subsequences $(2\nu n + m)_{n=1}^\infty$ where m, $1 \leq m < 2\nu$, is a nonzero remainder after a division by 2ν. Now it is not difficult to see that the validity of (2.8) follows from (2.11) and (2.12). □

3. 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 (2.2) 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 $L^1_{\mathbb{X}}$ 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

$$\|f\|_1 \leq \|f/\pi\|_{2,\pi} \tag{3.1}$$

because (by Cauchy-Schwarz inequality)

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

From (3.1) it follows that if $f/\pi \in L^2(\pi)$ then $f \in L^1_{\mathbb{X}}$.

Proposition 3.1. *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*

$$\lim_{n\to\infty} \|\hat{\mathcal{K}}^n[f] - \gamma\pi\|_1 = 0. \tag{3.2}$$

Proof. First of all notice that (3.1) 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

$$\begin{aligned}\|\hat{\mathcal{K}}^n[f] - \gamma\pi\|_1 &= \int |\hat{\mathcal{K}}^n[f](x') - \gamma\pi(x')|\lambda(dx') \\ &= \int \pi(x')\left|\mathcal{K}^n\left[\frac{f}{\pi}\right](x') - \gamma\mathbf{1}(x')\right|\lambda(dx') \\ &= \int \sqrt{\pi(x')}\left(\sqrt{\pi(x')}\left|\mathcal{K}^n\left[\frac{f}{\pi}\right](x') - \gamma\mathbf{1}(x')\right|\right)\lambda(dx') \\ &\leq \sqrt{\int \pi(x')\left|\mathcal{K}^n\left[\frac{f}{\pi}\right](x') - \gamma\mathbf{1}(x')\right|^2\lambda(dx')} = \left\|\mathcal{K}^n\left[\frac{f}{\pi}\right] - \gamma\mathbf{1}\right\|_{2,\pi}\end{aligned} \tag{3.3}$$

On the other hand by (2.8) it follows that

$$\lim_{n\to\infty}\left\|\mathcal{K}^n\left[\frac{f}{\pi}\right] - \gamma\mathbf{1}\right\|_{2,\pi} = 0 \tag{3.4}$$

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 (3.2) follows immediately from (3.3) and (3.4). □

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

Proposition 3.2. *Suppose $\mathcal{H}$ holds. Let $q(\cdot|\cdot) > 0$ a.e. $(\lambda \times \lambda)$ in $\mathbb{X} \times \mathbb{X}$. Let also $f/\pi \in L^2(\pi)$. Then (3.2) holds.*

Proof. 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 apply Proposition 3.1 (for $\nu = 1$ in $\mathcal{H}_p$) which completes the proof. □

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.

Proposition 3.3. *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

$$\|\hat{\mathcal{K}}[f]\|_1 \leq \|f\|_1 \tag{3.5}$$

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

$$\|\hat{\mathcal{K}}^n[f] - \hat{\mathcal{K}}^n[g]\|_1 \leq \|f - g\|_1. \tag{3.6}$$

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

$$\lim_{m \to \infty} \|f - f_{[m]}\|_1 = 0. \tag{3.7}$$

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

$$\|\gamma_{[m]}\pi - \gamma\pi\|_1 \leq \|f - f_{[m]}\|_1. \tag{3.8}$$

Proof. 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 (3.5). The validity of (3.6) follows immediately from (3.5) 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

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

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 (3.7) 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 (3.8). □

We are ready to give more general conditions under which (3.2) holds.

Theorem 3.1. *Suppose $\mathcal{H}$ and $\mathcal{H}_p$ hold. Let $f \in L^1_{\mathbb{X}}$ and put $\gamma = \int f(x)\lambda(dx)$. Then*

$$\lim_{n\to\infty} \|\hat{\mathcal{K}}^n[f] - \gamma\pi\|_1 = 0. \tag{3.9}$$

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

$$\lim_{n\to\infty} d_{TV}(\mu[\hat{\mathcal{K}}^n[f]], \mu[\pi]) = 0. \tag{3.10}$$

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

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

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

$$\|\hat{\mathcal{K}}^n[f]-\gamma\pi\|_1 \leq \|\hat{\mathcal{K}}^n[f]-\hat{\mathcal{K}}^n[f_{[m]}]\|_1+\|\hat{\mathcal{K}}^n[f_{[m]}]-\gamma_{[m]}\pi\|_1+\|\gamma_{[m]}\pi-\gamma\pi\|_1 \quad (3.11)$$

Choose some $\varepsilon > 0$. By (3.7) fix an integer $m \geq 1$ such that $\|f - f_{[m]}\|_1 < \varepsilon/3$. Then according to (3.6) we obtain

$$\|\hat{\mathcal{K}}^n[f] - \hat{\mathcal{K}}^n[f_{[m]}]\|_1 < \frac{\varepsilon}{3} \quad (3.12)$$

for any $n = 1, 2, \dots$ and according to (3.8) we obtain

$$\|\gamma_{[m]}\pi - \gamma\pi\|_1 < \frac{\varepsilon}{3}. \quad (3.13)$$

For such a fixed m we have from Proposition 3.3(2) that $f_{[m]}/\pi \in L^2(\pi)$ therefore by Proposition 3.1 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

$$\|\hat{\mathcal{K}}^n[f_{[m]}] - \gamma_{[m]}\pi\|_1 < \frac{\varepsilon}{3} \quad (3.14)$$

for any $n > n_0$. Replacing the inequalities (3.12), (3.13) and (3.14) in (3.11) 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 (3.9) for the case of bounded $f \in L^1_{\mathbb{X}}$.

Choose now arbitrary $f \in L^1_{\mathbb{X}}$ and put $f_m(x) = f(x)$ where $|f(x)| \leq m$ and $f_m(x) = 0$ elsewhere, $m = 1, 2, \dots$. In the same way as for (3.11) 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 (3.9) is a consequence of the well-known fact that

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

The validity of (3.10) follows immediately from (3.9). □

Theorem 3.1 allows us to enforce the Proposition 3.2.

Proposition 3.4. *Suppose $\mathcal{H}$ holds. Let $q(\cdot|\cdot) > 0$ a.e $(\lambda \times \lambda)$ in $\mathbb{X} \times \mathbb{X}$. Let also $f \in L^1_{\mathbb{X}}$. Then (3.2) holds.*

Proof. Analogously to the proof of Proposition 3.2 we conclude that here we can apply Theorem 3.1 for $\nu = 1$ in $\mathcal{H}_p$. □

4. Remarks

The convergence results according to the total variation distance described by Robert and Casella in [15] (see also e.g. [1, 2, 5, 9, 11, 12, 17]) use essentially the concepts of aperiodicity. The approach used here does not need the notion of aperiodicity.

Let us pay more attention to some valuable facts that stand back in the proofs above. From the proof of Theorem 2.1 one can see that for any $f \in L^2(\pi)$ the sequence $(\mathcal{K}^{2n}[f])_{n=1}^{\infty}$ is a Cauchy sequence in the $L^2(\pi)$ norm. For the limit $h \in L^2(\pi)$ of this sequence it holds $\mathcal{K}^2[h] = h$ and from the proof of Proposition 2.1 as for (2.7) one can see that

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

for any $n = 1, 2, \ldots$. This fact allows us to show (as in the proof of Theorem 3.1) that for any $f \in L^1_{\mathbb{X}}$ the sequence $(\hat{\mathcal{K}}^{2n}[f])_{n=1}^{\infty}$ is a Cauchy sequence in the $L^1_{\mathbb{X}}$ norm. Consequently for any probability density function $f(\cdot)$ on $\mathbb{X}$ the sequence $(\hat{\mathcal{K}}^{2n}[f])_{n=1}^{\infty}$ is a Cauchy sequence with respect to the total variation distance.

Let us prove the validity of (2.9). 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

$$|\langle Tu, Tu\rangle|^2 = \|Tu\|^4 \leq \langle Tu, u\rangle\langle T^2u, Tu\rangle. \tag{4.1}$$

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

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

Replacing the latter in (4.1) 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. (2.9). Certainly various proofs of (2.9) can be found in many places but we present a proof here for the sake of completeness taking into account the importance of this inequality in our construction.

For example our approach comprises with nominal adaptation the case of the random sweep Gibbs sampler and the case of the random sweep Metropolis within Gibbs sampler.

References

[1] Billingsley, P. (1995). *Probability and Measure.* John Wiley & Sons, New York. MR1324786

[2] Chen, M.-H., Shao, Q.-M. and Ibrahim, J. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag Inc., New York. MR1742311

[3] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (2004). *Bayesian Data Analysis*. Chapman & Hall, London. MR2027492

[4] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97-109.

[5] Krishna, B. and Soumendra, N. (2006). *Measure Theory and Probability Theory*. Springer Science+Business Media, New York. MR2247694

[6] Liang, F., Liu, C. and Carroll, R. (2010). *Advanced Markov Chain Monte Carlo Methods*. John Wiley and Sons Ltd, Chichester. MR2828488

[7] Link, W. and Barker, R. (2010). *Bayesian Inference: with ecological applications*. Elsevier Ltd.

[8] Liu, J.S., Wong, W.H. and Kong, A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *Journal of the Royal Statistical Society B* 57:157-169. MR1325382

[9] Madras, N. and Sezer, D. (2010). Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances. *Bernoulli* 16(3):882-908. MR2730652

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

[11] Meyn, S. and Tweedie, R. (2009). *Markov chains and stochastic stability*. Cambridge University Press, Cambridge. MR2509253

[12] Nummelin, E. (1984). *General Irreducible Markov Chains and Nonnegative Operators*. Cambridge University Press, Cambridge. MR0776608

[13] Reed, M. and Simon, B. (1980). *Methods of modern mathematical physics. I. Functional analysis*. Academic Press, New York-London. MR0493419

[14] Robert, C. (2007). *The Bayesian Choice - From Decision-Theoretic Foundations to Computational Implementation*. Springer Science + Business Media, New York. MR2723361

[15] Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer Science + Business Media, New York. MR2080278

[16] Robert, C. and Casella, G. (2010). *Introducing Monte Carlo Methods with R*. Springer Science + Business Media, New York. MR2572239

[17] Roberts, G. and Rosenthal, J. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:2071. MR2095565

[18] Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill, New York. MR0385023

[19] Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill, New York. MR0924157

[20] Rubinstein, R.Y. and Kroese, D.P. (2008). *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York. MR2365210

[21] Stroock, D. (2005). *An Introduction to Markov Processes*. Springer-Verlag Inc., Berlin. MR2126069