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 Metropolis-Hastings Markov chains with absolutely continuous with respect to Lebesgue measure target and proposal distributions. We show that under some very general conditions the sequence of the powers of the conjugate transition operator has a strong limit in a properly defined Hilbert space described for example in Stroock [17].

Then 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. [8] and W. Keith Hastings [4] is one of the best recognized techniques in the statistical applications (see e.g. [2, 6, 12, 13, 14, 15]). Throughout all this paper we shall assume that the following condition are valid.

$\mathcal{H}$: Assume we are given a Borel subset $\mathbb{X}$ of some Euclidean space and some *target* probability distribution on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ which is absolutely continuous with respect to the Lebesque measure 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 Lebesque measure *proposal* distribution on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 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 $\mathcal{B}(\mathbb{X}) \times \mathcal{B}(\mathbb{X})$ 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

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

draw $x_{(n)}$ one should generate a candidate $x_* \sim q(x|x_{(n-1)})$ and accept the candidate with a probability

$$\alpha = \min\left(1, \frac{\pi(x_*)}{\pi(x_{(n-1)})}\frac{q(x_{(n-1)}|x_*)}{q(x_*|x_{(n-1)})}\right)$$

taking $x_{(n)} = x_*$ or reject the candidate with a probability $1-\alpha$ and take $x_{(n)} = x_{(n-1)}$. All draws are taken from $\mathbb{X}$ and to avoid an ambiguity set $\alpha = 1$ for $q(x_*|x_{(n-1)}) = 0$. This scheme defines a transition kernel

$$\kappa(x \to x') = \min\left(1, \frac{\pi(x')}{\pi(x)}\frac{q(x|x')}{q(x'|x)}\right) q(x'|x) + \delta(x - x') \int \left(1 - \min\left(1, \frac{\pi(z)}{\pi(x)}\frac{q(x|z)}{q(z|x)}\right)\right) q(z|x)dz. \quad (1.1)$$

where $\delta(\cdot)$ is the Dirac function with a property $\int \delta(x-x')\varphi(x')dx' = \varphi(x)$. The integral sign stands for the Lebesque 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. Furthermore this kernel fulfills the normalizing condition $\int \kappa(x \to x')dx' = 1$ 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 an invariant density for the kernel, i.e. it holds $\pi(x') = \int \pi(x)\kappa(x \to x')dx$. The transition kernel (1.1) defines a Metropolis-Hastings Markov chain of 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

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

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

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

under the concepts of Harris recurrence, irreducibility, aperiodicity and reversibility (see e.g. [1, 2, 5, 7, 9, 10, 13, 15]). Here by $\mu[f]$ we denote the probability measure associated with the density $f(\cdot)$.

In this paper (Theorem 3.1 and Theorem 3.2) 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 [17].

Concisely formulated our main practical result (Proposition 3.3) 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.

2. The $L^2(\pi)$ structure

Following Stroock [17] we shall consider the Hilbert space $L^2(\pi)$ with an inner product

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

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

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

Define the operator

$$\mathcal{K}[f](x) = \int \kappa(x \to x')f(x')dx' \tag{2.1}$$

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

$$\hat{\mathcal{K}}[f](x') = \int f(x)\kappa(x \to x')dx. \tag{2.2}$$

Actually $\kappa(x \to x')$ is 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 the sequence of kernels

$$\kappa_n(x \to x') = \int \kappa_{n-1}(x \to z)\kappa_1(z \to x')dz, n = 2, 3, \ldots, \tag{2.3}$$

which are just the transition kernels of the transition-like operators $\hat{\mathcal{K}}^n$ and the usual kernels of the operators $\mathcal{K}^n$. It is easy to verify that $\kappa_n(x \to x')$ also satisfies the detailed balance condition and the Chapman-Kolmogorov equation

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

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

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

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

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

3) Suppose that there exists an integer $n \geq 1$ such that $\kappa_n(x \to x') > 0$ for $(x, x') \in \mathbb{X} \times \mathbb{X}$ where $\kappa_n(x \to x')$ is a composite 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. in $\mathbb{X}$.

Proof. 1) We have

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

Here we use the well-known Cauchy-Schwarz inequality. Multiplying the latter with $\pi(x)dx$ and integrating over $\mathbb{X}$ we get

$$\|\mathcal{K}[f]\|_{2,\pi}^2 = \int |\mathcal{K}[f](x)|^2 \pi(x) dx \leq \iint \pi(x)\kappa(x \to x') |f(x')|^2 dx' dx =$$
$$\iint \pi(x')\kappa(x' \to x)|f(x')|^2 dx' dx = \int \pi(x')|f(x')|^2 dx' = \|f\|_{2,\pi}^2. \tag{2.6}$$

In (2.6) we firstly use the detailed balance condition (1.2) then secondly we use the Tonelli's theorem which allows us to reduce the double integral to the iterated integral in the case of positive integrand. In this way we prove simultaneously the inequality (2.4) and the fact that $\mathcal{K}[f] \in L^2(\pi)$.

2) To prove that the operator $\mathcal{K}$ is self-adjoint assume that $f \in L^2(\pi)$ and $g \in L^2(\pi)$ and write down by means of the Fubini's theorem

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

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 $\kappa_n(x \to x')dx'$ and integrate over $\mathbb{X}$. Then we get

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

because

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

Multiply the latter with $\pi(x)dx$ and integrate over $\mathbb{X}$. Then

$$\iint \pi(x)\kappa_n(x \to x')h^2(x')dx'dx = \int \pi(x)h^2(x)dx + \iint \pi(x)\kappa_n(x \to x')(h(x') - h(x))^2 dx'dx. \quad (2.7)$$

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

$$\iint \pi(x)\kappa_n(x \to x')(h(x') - h(x))^2 dx'dx = 0$$

from which along with the inequalities $\pi(x) > 0$ for $x \in \mathbb{X}$ and $\kappa_n(x \to x') > 0$ for $(x, x') \in \mathbb{X} \times \mathbb{X}$ it is not difficult to find that there exists a constant γ such that $h(\cdot) = \gamma$ a.e. 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. $\langle \mathcal{K}^{2\nu n+2\nu}[h], h\rangle_\pi \geq \langle \mathcal{K}^{2\nu n}[h], h\rangle_\pi$ 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 the spectrum of the self-adjoint operator $\mathcal{K}$ lies entirely in the segment $[-1, 1]$. We are able to apply the well-known spectral theorem (see e.g. [3, 11]) from which we have

$$\mathcal{K}^{2\nu n} = \int_{-1}^{1} \theta^{2\nu n} dP_\theta, n = 1, 2, \ldots, \quad (2.8)$$

for some projection-valued measure P_θ and respectively

$$\langle \mathcal{K}^{2\nu n}[h], h\rangle_\pi = \int_{-1}^{1} \theta^{2\nu n} d\langle P_\theta[h], h\rangle_\pi, n = 1, 2, \dots. \tag{2.9}$$

where $\langle P_\theta[h], h\rangle_\pi$ is increasing function of $\theta \in [-1, 1]$ for any $h \in L^2(\pi)$. The integral (2.8) has a sophisticated nature but the integral in (2.9) is taken in the very usual Stieltjes sense. Now the validity of the conclusion 2) follows immediately from (2.9) and from the fact that the sequence of positive functions $(\theta^{2\nu n})$, $\theta \in [-1, 1]$, is decreasing.

3) The validity of the conclusion 3) follows from the representation

$$\langle (\mathcal{K}^{2\nu n} - \mathcal{K}^{2\nu n+2\nu p})[h], h\rangle_\pi = \int_{-1}^{1} (\theta^{2\nu n} - \theta^{2\nu n+2\nu p}) d\langle P_\theta[h], h\rangle_\pi$$

and from the fact that $\theta^{2\nu n} - \theta^{2\nu n+2\nu p} \geq 0$ for $\theta \in [-1, 1]$. □

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.

$\mathcal{H}_q$: Assume that there exists an integer $\nu \geq 1$ for which $\kappa_\nu(x \to x') > 0$ for $(x, x') \in \mathbb{X} \times \mathbb{X}$. □

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

$$\lim_{n\to\infty} \|\mathcal{K}^n[f] - \langle f, \mathbf{1}\rangle_\pi \mathbf{1}\|_{2,\pi} = 0. \tag{2.10}$$

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 it holds the inequality

$$\|Tu\|^2 \leq \|T\| \langle Tu, u\rangle. \tag{2.11}$$

The proof of (2.11) will be given at the end of the paper. Choose arbitrary $f \in L^2(\pi)$. Applying (2.11) 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

$$\|\mathcal{K}^{2\nu n}[f] - \mathcal{K}^{2\nu n+2\nu p}[f]\|_{2,\pi}^2 \leq 2\left(\langle \mathcal{K}^{2\nu n}[f], f\rangle_\pi - \langle \mathcal{K}^{2\nu n+2\nu p}[f], f\rangle_\pi\right). \tag{2.12}$$

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 (2.12) 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) we get that $h \equiv \gamma\mathbf{1}$ with some constant γ because

$$\kappa_{2\nu}(x \to x') = \int \kappa_\nu(x \to z)\kappa_\nu(z \to x')dz > 0$$

for $(x, x') \in \mathbb{X}\times\mathbb{X}$. 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 (2.10) for the subsequence of the powers $(2\nu n)_{n=1}^{\infty}$, i.e. that

$$\lim_{n\to\infty} \mathcal{K}^{2\nu n}[f] = \langle f, \mathbf{1}\rangle_\pi \mathbf{1}. \tag{2.13}$$

From (2.13) we obtain

$$\lim_{n\to\infty} \mathcal{K}^{2\nu n+m}[f] = \lim_{n\to\infty} \mathcal{K}^{2\nu n}\left[\mathcal{K}^m[f]\right] = \langle \mathcal{K}^m[f], \mathbf{1}\rangle_\pi \mathbf{1} = \langle f, \mathcal{K}^m[\mathbf{1}]\rangle_\pi \mathbf{1} = \langle f, \mathbf{1}\rangle_\pi \mathbf{1} \tag{2.14}$$

which proves (2.10) for the all the subsequences $(2\nu n + m)_{n=1}^{\infty}$ where m, $1 \le m < 2\nu$, is the nonzero remainder after a division by 2ν. Now it is not difficult to see that the validity of (2.10) follows from (2.13) and (2.14). □

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. Remember that for the absolutely continuous probability measures μ_1 and μ_2 with densities $f_1(\cdot)$ and $f_2(\cdot)$ for the total variation distance $d_{TV}(\mu_1, \mu_2)$ it holds (see e.g. [7, 15])

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

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

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

provided with the usual norm. We have

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

because

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

In the latter we use the Cauchy-Schwarz inequality. From (3.1) it follows that if $f/\pi \in L^2(\pi)$ then $f \in L^1_{\mathbb{X}}$.

Theorem 3.1. *Suppose $\mathcal{H}$ and $\mathcal{H}_q$ hold. Let $f(\cdot)$ be a function such that $f/\pi \in L^2(\pi)$ and put $\gamma = \int f(x)dx$. Then*

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

Therefore if $f(\cdot)$ is a probability density function on $\mathbb{X}$ such that $f/\pi \in L^2(\pi)$ then

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

Proof. First of all notice that (3.1) guarantees the existence of the constant γ. It can be shown straightforwardly by induction that

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

therefore by means of the Cauchy-Schwarz inequality we obtain

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

On the other hand by (2.10) 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

$$\left[\frac{f}{\pi}, \mathbf{1}\right]_\pi = \int \frac{f(x)}{\pi(x)}\pi(x)dx = \int f(x)dx = \gamma.$$

Now the validity of (3.2) follows immediately from (3.3) and (3.4). □

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

Proposition 3.1. Suppose $\mathcal{H}$ holds. Let $q(x'|x) > 0$ for $(x', x) \in \mathbb{X} \times \mathbb{X}$. Let also $f(\cdot)$ be a probability density function on $\mathbb{X}$ such that $f/\pi \in L^2(\pi)$. Then

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

Proof. By conditions we have $\pi(x) > 0$ for $x \in \mathbb{X}$ and $q(x'|x) > 0$ for $(x', x) \in \mathbb{X} \times \mathbb{X}$. These inequalities guarantee that $\kappa(x \to x') > 0$ for $(x', x) \in \mathbb{X} \times \mathbb{X}$ therefore we can apply Theorem 3.1 for $\nu = 1$ (in $\mathcal{H}_q$) 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 $f : \mathbb{X} \to \mathbb{R}$ put $f_{[m]}(x) = f(x)$ where $x \in \mathbb{X}_m$ and $f_{[m]}(x) = 0$ elsewhere.

Proposition 3.2. Suppose $\mathcal{H}$ holds. Then the following assertions are true.
1) Let $f \in Ł^1_{\mathbb{X}}$. Then

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

consequently for any $f \in Ł^1_{\mathbb{X}}$ and $g \in Ł^1_{\mathbb{X}}$ 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 $f \in Ł^1_{\mathbb{X}}$ 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 $f \in Ł^1_{\mathbb{X}}$ and put

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

Then

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

Proof. 1) We have

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

whence

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

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, i.e. 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

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

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

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

which proves (3.7) because $\mathbb{X}_m \nearrow \mathbb{X}$.
3) We have

$$|\gamma_m \pi(x') - \gamma\pi(x')| = \left| \pi(x') \int (f(x) - f_{[m]}(x)) dx \right| \le \pi(x') \|f - f_{[m]}\|_1$$

therefore

$$\|\gamma_m\pi - \gamma\pi\|_1 = \int |d_m\pi(x') - d\pi(x')| dx' \le \|f - f_{[m]}\|_1 \int \pi(x') dx' = \|f - f_{[m]}\|_1$$

which proves (3.8). □

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

Theorem 3.2. *Suppose $\mathcal{H}$ and $\mathcal{H}_q$ hold. Let $f \in L^1_{\mathbb{X}}$ and put $\gamma = \int f(x)dx$. 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 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

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

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

$$\|\hat{\mathcal{K}}^n[f] - \gamma\pi\|_1 \le \|\hat{\mathcal{K}}^n[f] - \hat{\mathcal{K}}^n[f_{[m]}]\|_1 + \|\hat{\mathcal{K}}^n[f_{[m]}] - \gamma_m\pi\|_1 + \|\gamma_m\pi - \gamma\pi\|_1 \tag{3.11}$$

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

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

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

$$\|\gamma_m\pi - \gamma\pi\|_1 \le \|f - f_{[m]}\|_1 < \frac{\varepsilon}{3}. \tag{3.13}$$

For such a fixed m we have from the proof of Proposition 3.2 that $f_{[m]}/\pi \in L^2(\pi)$ therefore by Theorem 3.1 we get that

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

therefore we can chose an positive integer n_0 such that

$$\|\hat{\mathcal{K}}^n[f_{[m]}] - \gamma_m \pi\|_1 < \frac{\varepsilon}{3} \tag{3.14}$$

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 $f \in L^1_{\mathbb{X}}$.

Chose now arbitrary $f \in L^1_{\mathbb{X}}$ and put $f_m(x) = f(x)$ where $|f(x)| \le m$ and $f_m(x) = 0$ where $|f(x)| > m$, $m = 1, 2, \ldots$. In the same way as in (3.11) one can get

$$\|\hat{\mathcal{K}}^n[f] - \gamma\pi\|_1 \le \|\hat{\mathcal{K}}^n[f_m] - \gamma_m\pi\|_1 + 2\|f_m - f\|_1. \tag{3.15}$$

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

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

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 3.2 allows us to enforce the Proposition 3.1.

Proposition 3.3. Suppose $\mathcal{H}$ holds. Let $q(x'|x) > 0$ for $(x', x) \in \mathbb{X} \times \mathbb{X}$. Let also $f \in L^1_{\mathbb{X}}$ be a probability density function. Then

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

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

4. Remarks

• The convergence results according to the total variation distance presented by Robert and Casella in [13] (see also e.g. [1, 2, 5, 7, 9, 10, 13, 15]) use essentially the concepts of aperiodicity without requirements for the initial distribution of the chain. In contrast with this the essential condition in Theorem 3.1 concerns namely the initial distribution density. On the other hand in the practical realization of a drawing process the meaning of the initial distribution (the distribution from which we draw first) vanishes in the case when we use one long chain instead of multiple short chains. Using of multiple (very) short chains already took into account (at least formally) the shape of the initial density. We can say that Theorem 3.1 advocates the using of one long chain with sufficiently long burn-in period.

If we are going to run multiple chains and the choice of the initial distribution become somewhat significant then Theorem 3.1 teaches us that one good choice is to use uniform distribution (i.e. $f_0(\cdot) = const$) with $supp(f_0)$ near to some mode of the target distribution. In this case it is natural to have $\pi(x) \geq \varepsilon > 0$ for $x \in supp(f_0)$ and therefore $f_0/\pi \in L^2(\pi)$.

• Our results say nothing about the convergence speed.

• Here we assume that the target and the proposal distributions are absolutely continuous. This restrictive assumption allows us to write concise and clear expressions in the formulas where kernels are used. But our careful review of the proofs presented above does not find irresistible obstacles for them to be generalized at least in the case when the probability measures include both absolutely continuous and discrete components. Moreover if we are allowed to use the Dirac delta function to express the probability mass functions then we are able to keep the same notation for the kernel and in this way the key points in the corresponding proofs will remain intact.

• At this point we shall 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 of this sequence $h \in L^2(\pi)$ it holds $\mathcal{K}^2[h] = h$ and from the proof of Proposition 2.1 as for (2.7) one can see that

$$\iint \pi(x)\kappa_{2n}(x \to x')(h(x') - h(x))^2 dx' dx = 0$$

for any $n = 1, 2, \ldots$. This fact allows us to show (as in the proof of Theorem 3.2) that for any $f \in L^1(\mathbb{S})$ 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 $f \in L^1(\mathbb{S})$ 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 (2.11). The operator T is assumed to be linear bounded self-adjoint positive operator 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. (2.11). Certainly various proofs of (2.11) 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.

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] Dunford, N. and Schwartz, J. (1963). *Linear operators*. Interscience Publishers, New York.

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

[8] 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.

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

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

[11] Reed, M. and Simon, B. (1980). *Functional Analysis*. Academic Press, San Diego.

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

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

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

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

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

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