



# A computable bound of the essential spectral radius of finite range Metropolis–Hastings kernels

Loïc Hervé, James Ledoux

## ► To cite this version:

Loïc Hervé, James Ledoux. A computable bound of the essential spectral radius of finite range Metropolis–Hastings kernels. *Statistics and Probability Letters*, 2016, 117, pp.72-79. 10.1016/j.spl.2016.05.007 . hal-01356804

**HAL Id: hal-01356804**

**<https://hal.science/hal-01356804v1>**

Submitted on 22 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# A computable bound of the essential spectral radius of finite range Metropolis-Hastings kernels

Loïc HERVÉ, and James LEDOUX \*

November 22, 2016

## Abstract

Let  $\pi$  be a positive continuous target density on  $\mathbb{R}$ . Let  $P$  be the Metropolis-Hastings operator on the Lebesgue space  $\mathbb{L}^2(\pi)$  corresponding to a proposal Markov kernel  $Q$  on  $\mathbb{R}$ . When using the quasi-compactness method to estimate the spectral gap of  $P$ , a mandatory first step is to obtain an accurate bound of the essential spectral radius  $r_{ess}(P)$  of  $P$ . In this paper a computable bound of  $r_{ess}(P)$  is obtained under the following assumption on the proposal kernel:  $Q$  has a bounded continuous density  $q(x, y)$  on  $\mathbb{R}^2$  satisfying the following finite range assumption :  $|u| > s \Rightarrow q(x, x + u) = 0$  (for some  $s > 0$ ). This result is illustrated with Random Walk Metropolis-Hastings kernels.

AMS subject classification : 60J10, 47B07

Keywords : Markov chain operator, Metropolis-Hastings algorithms, Spectral gap

## 1 Introduction

Let  $\pi$  be a positive distribution density on  $\mathbb{R}$ . Let  $Q(x, dy) = q(x, y)dy$  be a Markov kernel on  $\mathbb{R}$ . Throughout the paper we assume that  $q(x, y)$  satisfies the following finite range assumption: there exists  $s > 0$  such that

$$|u| > s \implies q(x, x + u) = 0. \quad (1)$$

Let  $T(x, dy) = t(x, y)dy$  be the nonnegative kernel on  $\mathbb{R}$  given by

$$t(x, y) := \min \left( q(x, y), \frac{\pi(y) q(y, x)}{\pi(x)} \right) \quad (2)$$

and define the associated Metropolis-Hastings kernel:

$$P(x, dy) := r(x) \delta_x(dy) + T(x, dy) \quad \text{with} \quad r(x) := 1 - \int_{\mathbb{R}} t(x, y) dy, \quad (3)$$

---

\*INSA de Rennes, IRMAR, F-35042, France; CNRS, UMR 6625, Rennes, F-35708, France; Université Européenne de Bretagne, France. {Loic.Herve,James.Ledoux}@insa-rennes.fr

where  $\delta_x(dy)$  denotes the Dirac distribution at  $x$ . The associated Markov operator is still denoted by  $P$ , that is we set for every bounded measurable function  $f : \mathbb{R} \rightarrow \mathbb{C}$  :

$$\forall x \in \mathbb{R}, \quad (Pf)(x) = r(x)f(x) + \int_{\mathbb{R}} f(y) t(x, y) dy. \quad (4)$$

In the context of Monte Carlo Markov Chain methods, the kernel  $Q$  is called the proposal Markov kernel. We denote by  $(\mathbb{L}^2(\pi), \|\cdot\|_2)$  the usual Lebesgue space associated with the probability measure  $\pi(y)dy$ . For convenience,  $\|\cdot\|_2$  also denotes the operator norm on  $\mathbb{L}^2(\pi)$ , namely: if  $U$  is a bounded linear operator on  $\mathbb{L}^2(\pi)$ , then  $\|U\|_2 := \sup_{\|f\|_2=1} \|Uf\|_2$ . Since

$$t(x, y)\pi(x) = t(y, x)\pi(y), \quad (5)$$

we know that  $P$  is reversible with respect to  $\pi$  and that  $\pi$  is  $P$ -invariant (e.g. see [RR04]). Consequently  $P$  is a self-adjoint operator on  $\mathbb{L}^2(\pi)$  and  $\|P\|_2 = 1$ . Now define the rank-one projector  $\Pi$  on  $\mathbb{L}^2(\pi)$  by

$$\Pi f := \pi(f)1_{\mathbb{R}} \quad \text{with} \quad \pi(f) := \int_{\mathbb{R}} f(x) \pi(x) dx.$$

Then the spectral radius of  $P - \Pi$  equals to  $\|P - \Pi\|_2$  since  $P - \Pi$  is self-adjoint, and  $P$  is said to have the spectral gap property on  $\mathbb{L}^2(\pi)$  if

$$\varrho_2 \equiv \varrho_2(P) := \|P - \Pi\|_2 < 1.$$

In this case the following property holds:

$$\forall n \geq 1, \forall f \in \mathbb{L}^2(\pi), \quad \|P^n f - \Pi f\|_2 \leq \varrho_2^n \|f\|_2. \quad (\text{SG}_2)$$

The spectral gap property on  $\mathbb{L}^2(\pi)$  of a Metropolis-Hastings kernel is of great interest, not only due to the explicit geometrical rate given by  $(\text{SG}_2)$ , but also since it ensures that a central limit theorem (CLT) holds true for additive functional of the associated Metropolis-Hastings Markov chain under the expected second-order moment conditions, see [RR97]. Furthermore, the rate of convergence in the CLT is  $O(1/\sqrt{n})$  under third-order moment conditions (as for the independent and identically distributed models), see details in [HP10, FHL12].

The quasi-compactness approach can be used to compute the rate  $\varrho_2(P)$ . This method is based on the notion of essential spectral radius. Indeed, first recall that the essential spectral radius of  $P$  on  $\mathbb{L}^2(\pi)$ , denoted by  $r_{ess}(P)$ , is defined by (e.g. see [Wu04] for details):

$$r_{ess}(P) := \liminf_n (\|P^n - K\|_2)^{1/n} \quad (6)$$

where the above infimum is taken over the ideal of compact operators  $K$  on  $\mathbb{L}^2(\pi)$ . Note that the spectral radius of  $P$  is one. Then  $P$  is said to be quasi-compact on  $\mathbb{L}^2(\pi)$  if  $r_{ess}(P) < 1$ . Second, if  $r_{ess}(P) \leq \alpha$  for some  $\alpha \in (0, 1)$ , then  $P$  is quasi-compact on  $\mathbb{L}^2(\pi)$ , and the following properties hold: for every real number  $\kappa$  such that  $\alpha < \kappa < 1$ , the set  $\mathcal{U}_\kappa$  of the spectral values  $\lambda$  of  $P$  satisfying  $\kappa \leq |\lambda| \leq 1$  is composed of finitely many eigenvalues of  $P$ , each of them having a finite multiplicity (e.g. see [Hen93] for details). Third, if  $P$  is quasi-compact on  $\mathbb{L}^2(\pi)$  and satisfies usual aperiodicity and irreducibility conditions (e.g. see [MT93]), then  $\lambda = 1$  is the only spectral value of  $P$  with modulus one and  $\lambda = 1$  is a simple eigenvalue of  $P$ , so

that  $P$  has the spectral gap property on  $\mathbb{L}^2(\pi)$ . Finally the following property holds: either  $\varrho_2(P) = \max\{|\lambda|, \lambda \in \mathcal{U}_\kappa, \lambda \neq 1\}$  if  $\mathcal{U}_\kappa \neq \emptyset$ , or  $\varrho_2(P) \leq \kappa$  if  $\mathcal{U}_\kappa = \emptyset$ .

This paper only focusses on the preliminary central step of the previous spectral method, that is to find an accurate bound of  $r_{ess}(P)$ . More specifically, we prove that, if the target density  $\pi$  is positive and continuous on  $\mathbb{R}$ , and if the proposal kernel  $q(\cdot, \cdot)$  is bounded continuous on  $\mathbb{R}^2$  and satisfies (1) for some  $s > 0$ , then

$$r_{ess}(P) \leq \alpha_a \quad \text{with} \quad \alpha_a := \max(r_a, r'_a + \beta_a) \quad (7)$$

where, for every  $a > 0$ , the constants  $r_a, r'_a$  and  $\beta_a$  are defined by:

$$r_a := \sup_{|x| \leq a} r(x), \quad r'_a := \sup_{|x| > a} r(x), \quad \beta_a := \int_{-s}^s \sup_{|x| > a} \sqrt{t(x, x+u)t(x+u, x)} du. \quad (8)$$

This result is illustrated in Section 2 with Random Walk Metropolis-Hastings (RWMH) kernels for which the proposal Markov kernel is of the form  $Q(x, dy) := \Delta(|x - y|) dy$ , where  $\Delta : \mathbb{R} \rightarrow [0, +\infty)$  is an even continuous and compactly supported function.

In [AP07] the quasi-compactness of  $P$  on  $\mathbb{L}^2(\pi)$  is proved to hold provided that 1) the essential supremum of the rejection probability  $r(\cdot)$  with respect to  $\pi$  is bounded away from unity; 2) the operator  $T$  associated with the kernel  $t(x, y)dy$  is compact on  $\mathbb{L}^2(\pi)$ . Assumption 1) on the rejection probability  $r(\cdot)$  is a necessary condition for  $P$  to have the spectral gap property (SG<sub>2</sub>) (see [RT96]). But this condition, which is quite generic from the definition of  $r(\cdot)$  (see Remark 3), is far to be sufficient for  $P$  to satisfy (SG<sub>2</sub>). The compactness Assumption 2) of [AP07] is quite restrictive, for instance it is not adapted for random walk Metropolis-Hastings kernels. Here this compactness assumption is replaced by the condition  $r'_a + \beta_a < 1$ . As shown in the examples of Section 2, this condition is adapted to RWMH.

In the discrete state space case, a bound for  $r_{ess}(P)$  similar to (7) has been obtained in [HL16]. Next a bound of the spectral gap  $\varrho_2(P)$  has been derived in [HL16] from a truncation method for which the control of the essential spectral radius of  $P$  is a central step. It is expected that, in the continuous state space case, the bound (7) will provide a similar way to compute the spectral gap  $\varrho_2(P)$  of  $P$ . This issue, which is much more difficult than in the discrete case, is not addressed in this work.

## 2 An upper bound for the essential spectral radius of $P$

Let us state the main result of the paper.

**Theorem 1** *Assume that*

- (i)  $\pi$  is positive and continuous on  $\mathbb{R}$ ;
- (ii)  $q(\cdot, \cdot)$  is bounded and continuous on  $\mathbb{R}^2$ , and satisfies the finite range assumption (1).

For  $a > 0$ , set  $\alpha_a := \max(r_a, r'_a + \beta_a)$ , where the constants  $r_a, r'_a$  and  $\beta_a$  are defined in (8). Then

$$\forall a > 0, \quad r_{ess}(P) \leq \alpha_a. \quad (9)$$

Theorem 1 is proved in Section 3 from Formula (6) by using a suitable decomposition of the iterates  $P^n$  involving some Hilbert-Schmidt operators.

**Remark 1** Assume that the assumptions (i)-(ii) of Theorem 1 hold. Then, if there exists some  $a > 0$  such that  $\alpha_a < 1$ ,  $P$  is quasi-compact on  $\mathbb{L}^2(\pi)$ . Suppose moreover that the proposal Markov kernel  $Q(x, dy)$  satisfies usual irreducibility and aperiodicity conditions. Then  $P$  has the spectral gap property on  $\mathbb{L}^2(\pi)$ . Actually, if  $q$  is symmetric (i.e.  $q(x, y) = q(y, x)$ ), it can be easily proved that, under the condition  $r'_a + \beta_a < 1$ ,  $P$  satisfies the so-called drift condition with respect to  $V(x) := 1/\sqrt{\pi(x)}$ , so that  $P$  is  $V$ -geometrically ergodic, that is  $P$  has the spectral gap property on the space  $(\mathcal{B}_V, \|\cdot\|_V)$  composed of the functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  such that  $\|f\|_V := \sup_{x \in \mathbb{R}} |f(x)|/V(x) < \infty$ . If furthermore  $\int_{\mathbb{R}} \sqrt{\pi(x)} dx < \infty$ , then the spectral gap property of  $P$  on  $\mathbb{L}^2(\pi)$  can be deduced from the  $V$ -geometrical ergodicity since  $P$  is reversible (see [RR97, Bax05]). However this fact does not provide a priori any precise bound on the essential spectral radius of  $P$  on  $\mathbb{L}^2(\pi)$ . Indeed, mention that the results [Wu04, Th. 5.5] provide a comparison between  $r_{\text{ess}}(P)$  and  $r_{\text{ess}}(P|_{\mathcal{B}_V})$ , but unfortunately, to the best of our knowledge, no accurate bound of  $r_{\text{ess}}(P|_{\mathcal{B}_V})$  is known for Metropolis-Hasting kernels. In particular note that the general bound of  $r_{\text{ess}}(P|_{\mathcal{B}_V})$  given in [HL14, Th. 5.2] is of theoretical interest but is not precise, and that the more accurate bound of  $r_{\text{ess}}(P|_{\mathcal{B}_V})$  given in [HL14, Th. 5.4] cannot be used here since in general no iterate of  $P$  is compact from  $\mathcal{B}_0$  to  $\mathcal{B}_V$ , where  $\mathcal{B}_0$  denotes the space of bounded measurable functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  equipped with the supremum norm. Therefore, the  $V$ -geometrical ergodicity of  $P$  is not discussed here since the purpose is to bound the essential spectral radius of  $P$  on  $\mathbb{L}^2(\pi)$ .

**Remark 2** If  $\pi$  and  $q$  satisfy the assumptions (i)-(ii) of Theorem 1, and if moreover  $q$  satisfies the following mild additional condition

$$\forall x \in \mathbb{R}, \exists y \in [x - s, x + s], \quad q(x, y) q(y, x) \neq 0, \quad (10)$$

then, for every  $a > 0$ , we have  $r_a < 1$ , so that the quasi-compactness of  $P$  on  $\mathbb{L}^2(\pi)$  holds provided that there exists some  $a > 0$  such that  $r'_a + \beta_a < 1$ . Note that Condition (10) is clearly fulfilled if  $q$  is symmetric. To prove the previous assertion on  $r_a$ , observe that  $r(\cdot)$  is continuous on  $\mathbb{R}$  (use Lebesgue's theorem). Consequently, if  $r_a = 1$  for some  $a > 0$ , then  $r(x_0) = 1$  for some  $x_0 \in [-a, a]$ , but this is impossible from the definition of  $r(x_0)$  and Condition (10).

**Remark 3** Actually, under the assumptions (i)-(ii) of Theorem 1, the fact that  $r_a < 1$  for every  $a > 0$ , and even the stronger property  $\sup_{x \in \mathbb{R}} r(x) < 1$ , seem to be quite generic. For instance, if  $q$  is of the form  $q(x, y) = \Delta(|x - y|)$  for some function  $\Delta$  and if there exists  $\theta > 0$  such that  $\pi$  is increasing on  $(-\infty, -\theta]$  and decreasing on  $[\theta, +\infty)$ , then  $\sup_{x \in \mathbb{R}} r(x) < 1$ . Thus, for every  $a > 0$ , we have  $r_a < 1$  and  $r'_a < 1$ . Indeed, first observe that  $r_a < 1$  for every  $a > 0$  from Remark 2. Consequently, if  $\sup_{x \in \mathbb{R}} r(x) = 1$ , then there exists  $(x_n)_n \in \mathbb{R}^{\mathbb{N}}$  such that  $\lim_n |x_n| = +\infty$  and  $\lim_n r(x_n) = 1$ . Let us prove that this property is impossible under our assumptions. To simplify, suppose that  $\lim_n x_n = +\infty$ . Then, from the definition of  $r(\cdot)$ , from our assumptions on  $q$ , and finally from Fatou's Lemma, it follows that, for almost every  $u \in [-s, s]$  such that  $\Delta(u) \neq 0$ , we have  $\liminf_n \min(1, \pi(x_n + u)/\pi(x_n)) = 0$ . But this is impossible since, if  $u \in [-s, 0]$  and  $x_n \geq \theta + s$ , then  $\pi(x_n + u) \geq \pi(x_n)$ .

Theorem 1 is illustrated with symmetric proposal Markov kernels of the form

$$Q(x, dy) := \Delta(x - y) dy$$

where  $\Delta : \mathbb{R} \rightarrow [0, +\infty)$  is : 1) an even continuous function ; 2) assumed to be compactly supported on  $[-s, s]$  and positive on  $(-s, s)$  for some  $s > 0$ . Then  $q(x, y) := \Delta(x - y)$  satisfies (1) and  $t(\cdot, \cdot)$  is given by

$$\forall u \in [-s, s], \quad t(x, x + u) := \Delta(u) \min \left( 1, \frac{\pi(x + u)}{\pi(x)} \right).$$

**Corollary 1** Assume that  $q(x, y) := \Delta(x - y)$  with  $\Delta(\cdot)$  satisfying the above assumptions and that  $\pi$  is an even positive continuous distribution density such that the following limit exists:

$$\forall u \in [0, s], \quad \tau(u) := \lim_{x \rightarrow +\infty} \frac{\pi(x + u)}{\pi(x)} \in [0, 1]. \quad (11)$$

Assume that the set  $\{u \in [0, s] : \tau(u) \neq 1\}$  has a positive Lebesgue-measure. Then  $P$  is quasi-compact on  $\mathbb{L}^2(\pi)$  with

$$r_{ess}(P) \leq \alpha_\infty := \max(r_\infty, \gamma_\infty) < 1 \quad \text{where} \quad \gamma_\infty := 1 - \int_0^s \Delta(u) (1 - \tau(u)^{1/2})^2 du.$$

*Proof.* We know from Theorem 1 that, for any  $a > 0$ ,  $r_{ess}(P) \leq \max(r_a, r'_a + \beta_a)$  with  $r_a := \sup_{|x| \leq a} r(x)$ ,  $r'_a := \sup_{|x| > a} r(x)$ . It is easily checked that

$$\beta_a = \int_{-s}^s \Delta(u) \sup_{|x| > a} \min \left( \sqrt{\frac{\pi(x + u)}{\pi(x)}}, \sqrt{\frac{\pi(x)}{\pi(x + u)}} \right) du.$$

Note that

$$\forall x \in \mathbb{R}, \quad r(x) = 1 - \int_{-s}^s \Delta(u) \min \left( 1, \frac{\pi(x + u)}{\pi(x)} \right) du.$$

For  $u \in [-s, 0]$ ,  $\tau(u)$  is defined as in (11). Then

$$\forall u \in [-s, s], \quad \tau(u) = \lim_{y \rightarrow +\infty} \frac{\pi(y)}{\pi(y - u)} = \frac{1}{\tau(-u)}$$

with the convention  $1/0 = +\infty$ . Thus, for every  $u \in [-s, 0]$ , we have  $\tau(u) \in [1, +\infty]$ . Moreover we obtain for every  $u \in [-s, s]$ :

$$\lim_{x \rightarrow -\infty} \frac{\pi(x + u)}{\pi(x)} = \tau(-u).$$

since  $\pi$  is an even function. We have for every  $a > 0$

$$r'_a = 1 - \min \left( \inf_{x < -a} \int_{-s}^s \Delta(u) \min \left( 1, \frac{\pi(x + u)}{\pi(x)} \right) du, \inf_{x > a} \int_{-s}^s \Delta(u) \min \left( 1, \frac{\pi(x + u)}{\pi(x)} \right) du \right).$$

Moreover it follows from dominated convergence theorem and from the above remarks that

$$\lim_{x \rightarrow \pm\infty} \int_{-s}^s \Delta(u) \min \left( 1, \frac{\pi(x+u)}{\pi(x)} \right) du = \int_{-s}^s \Delta(u) \min (1, \tau(\pm u)) du$$

from which we deduce that

$$\begin{aligned} r'_\infty &:= \lim_{a \rightarrow +\infty} r'_a \\ &= 1 - \min \left( \int_{-s}^s \Delta(u) \min (1, \tau(-u)) du, \int_{-s}^s \Delta(u) \min (1, \tau(u)) du \right) \\ &= 1 - \int_{-s}^s \Delta(u) \min (1, \tau(u)) du \quad (\text{since } \Delta \text{ is an even function}) \\ &= 1 - \int_{-s}^0 \Delta(u) du - \int_0^s \Delta(u) \tau(u) du \quad (\text{since } \tau(u) \leq 1 \text{ for } u \in [0, s], \tau(u) \geq 1 \text{ for } u \in [-s, 0]) \\ &= 1 - \int_0^s \Delta(u) [1 + \tau(u)] du. \end{aligned}$$

Note that, for every  $a > 0$ , we have  $r_a < 1$  from Remark 2. Moreover  $r'_\infty \leq 1/2$  from the last equality. Thus  $r_\infty := \sup_{x \in \mathbb{R}} r(x) < 1$ . Next we obtain for every  $a > 0$

$$\beta_a = \int_{-s}^s \Delta(u) \max \left[ \sup_{x < -a} \min \left( \sqrt{\frac{\pi(x+u)}{\pi(x)}}, \sqrt{\frac{\pi(x)}{\pi(x+u)}} \right), \sup_{x > a} \min \left( \sqrt{\frac{\pi(x+u)}{\pi(x)}}, \sqrt{\frac{\pi(x)}{\pi(x+u)}} \right) \right] du$$

and again we deduce from dominated convergence theorem and from the above remarks that

$$\begin{aligned} \beta_\infty &:= \lim_{a \rightarrow +\infty} \beta_a = \int_{-s}^s \Delta(u) \max \left[ \min \left( \tau(-u)^{1/2}, \frac{1}{\tau(-u)^{1/2}} \right), \min \left( \tau(u)^{1/2}, \frac{1}{\tau(u)^{1/2}} \right) \right] du \\ &= \int_{-s}^s \Delta(u) \min \left( \tau(u)^{1/2}, \frac{1}{\tau(u)^{1/2}} \right) du \quad (\text{since } \tau(-u) = \frac{1}{\tau(u)}) \\ &= \int_{-s}^0 \Delta(u) \tau(u)^{-1/2} du + \int_0^s \Delta(u) \tau(u)^{1/2} du \\ &= \int_{-s}^0 \Delta(u) \tau(-u)^{1/2} du + \int_0^s \Delta(u) \tau(u)^{1/2} du \\ &= 2 \int_0^s \Delta(u) \tau(u)^{1/2} du. \end{aligned}$$

Thus

$$r'_\infty + \beta_\infty = 1 - \int_0^s \Delta(u) [1 + \tau(u) - 2\tau(u)^{1/2}] du = 1 - \int_0^s \Delta(u) (1 - \sqrt{\tau(u)})^2 du < 1$$

since by hypothesis the set  $\{u \in [0, s] : \tau(u) \neq 1\}$  has a positive Lebesgue-measure.

Since  $r_{ess}(P) \leq \max(r_a, r'_a + \beta_a)$  holds for every  $a > 0$ , we obtain that  $r_{ess}(P) \leq \max(r_\infty, r'_\infty + \beta_\infty) < 1$ . Thus  $P$  is quasi-compact on  $\mathbb{L}^2(\pi)$ .  $\square$

**Example 2.1 (Laplace distribution)** Let  $\pi(x) = e^{-|x|}/2$  be the Laplace distribution density, and set  $q(x, y) := \Delta(x - y)$  with  $\Delta(u) := (1 - |u|) 1_{[-1,1]}(u)$ . Then

$$\forall u \in [0, 1], \quad \tau(u) := \lim_{x \rightarrow +\infty} \frac{\pi(x + u)}{\pi(x)} = e^{-u}.$$

Then

$$\gamma_\infty = 1 - \int_0^1 (1 - u)(1 - e^{-u/2})^2 du = 8e^{-1/2} - e^{-1} - 7/2.$$

From Corollary 1,  $P$  is quasi-compact on  $\mathbb{L}^2(\pi)$  with  $r_{ess}(P) \leq \max(1 - 1/e, 8e^{-1/2} - e^{-1} - 7/2) = 8e^{-1/2} - e^{-1} - 7/2 \approx 0.9843$  since  $r_\infty := \sup_{x \in \mathbb{R}} r(x) \leq 1 - 1/e$ .

**Example 2.2 (Gauss distribution)** Let  $\pi(x) = e^{-x^2/2}/\sqrt{2\pi}$  be the Gauss distribution density, and set  $q(x, y) := \Delta(|x - y|)$  with  $\Delta(u) := (1 - |u|) 1_{[-1,1]}(u)$ . Then

$$\forall u \in (0, 1], \quad \tau(u) := \lim_{x \rightarrow +\infty} \frac{\pi(x + u)}{\pi(x)} = 0,$$

so that

$$\gamma_\infty = 1 - \int_0^1 (1 - u) du = \frac{1}{2}.$$

From Corollary 1,  $P$  is quasi-compact on  $\mathbb{L}^2(\pi)$  with  $r_{ess}(P) \leq \max(0.156, 0.5) = 0.5$  since  $r_\infty \leq 1 - e^{-1/2} - e^{1/8} \int_0^1 (1 - u)e^{-(u+1)^2/2} du \leq 0.156$ .

In view of the quasi-compactness approach presented in Introduction for computing the rate  $\varrho_2(P)$  in (SG<sub>2</sub>), the bound  $r_{ess}(P) \leq 0.5$  obtained for Gauss distribution (for instance) implies that, for every  $\kappa \in (0.5, 1)$ , the set of the spectral values  $\lambda$  of  $P$  on  $\mathbb{L}^2(\pi)$  satisfying  $\kappa \leq |\lambda| \leq 1$  is composed of finitely many eigenvalues of finite multiplicity. Moreover, from aperiodicity and irreducibility,  $\lambda = 1$  is the only eigenvalue of  $P$  with modulus one and it is a simple eigenvalue of  $P$ . Consequently the spectral gap property (SG<sub>2</sub>) holds with  $\varrho_2(P)$  given by

- $\varrho_2(P) = \max\{|\lambda|, \lambda \in \mathcal{U}_\kappa, \lambda \neq 1\}$  if  $\mathcal{U}_\kappa \neq \emptyset$ ,
- $\varrho_2(P) \leq \kappa$  if  $\mathcal{U}_\kappa = \emptyset$  (in particular, if for every  $\kappa \in (0.5, 1)$  we have  $\mathcal{U}_\kappa = \emptyset$ , then we could conclude that  $\varrho_2(P) \leq 0.5$ ).

The numerical computation of the eigenvalues  $\lambda \in \mathcal{U}_\kappa$ ,  $\lambda \neq 1$ , is a difficult issue. Even to know whether the set  $\mathcal{U}_\kappa \setminus \{1\}$  is empty or not seems to be difficult. In the discrete state space case (i.e  $P = (P(i, j))_{i, j \in \mathbb{N}}$ ), this problem has been solved by using a weak perturbation method involving some finite truncated matrices derived from  $P$  (see [HL16]). In the continuous state space case, a perturbation method could be also considered, but it raises a priori difficult theoretical and numerical issues.



### 3 Proof of Theorem 1

For any bounded linear operator  $U$  on  $\mathbb{L}^2(\pi)$  we define

$$\forall f \in \mathbb{L}^2(\pi), \quad U_a f := 1_{[-a,a]} \cdot Uf \quad \text{and} \quad U_{a^c} f := 1_{\mathbb{R} \setminus [-a,a]} \cdot Uf.$$

Obviously  $U_a$  and  $U_{a^c}$  are bounded linear operators on  $\mathbb{L}^2(\pi)$ , and  $U = U_a + U_{a^c}$ . Define  $Rf = rf$  with function  $r(\cdot)$  given in (3). Recall that  $T$  is the operator associated with kernel  $T(x, dy) = t(x, y)dy$ . Then the M-H kernel  $P$  defined in (4) writes as follows:

$$P = R + T = R_a + R_{a^c} + T_a + T_{a^c}$$

with  $R_{a^c}R_a = R_aR_{a^c} = 0$  and  $R_aT_{a^c} = 0$ .

**Lemma 1** *The operators  $T_a, T_{a^c}R_a$  and  $(R_{a^c} + T_{a^c})^n R_a$  for any  $n \geq 1$  are compact on  $\mathbb{L}^2(\pi)$ .*

*Proof.* Using the detailed balance equation (5), we obtain for any  $f \in \mathbb{L}^2(\pi)$

$$\begin{aligned} (T_a f)(x) &= 1_{[-a,a]}(x) \int_{\mathbb{R}} f(y) t(x, y) dy = \int_{\mathbb{R}} f(y) 1_{[-a,a]}(x) \frac{t(x, y)}{\pi(y)} \pi(y) dy \\ &= \int_{\mathbb{R}} f(y) t_a(x, y) \pi(y) dy \quad \text{with} \quad t_a(x, y) := 1_{[-a,a]}(x) \frac{t(y, x)}{\pi(x)}. \end{aligned}$$

Function  $q(\cdot, \cdot)$  is supposed to be bounded on  $\mathbb{R}^2$ , so is  $t(\cdot, \cdot)$ . From  $\inf_{|x| \leq a} \pi(x) > 0$  it follows that  $t_a(\cdot, \cdot)$  is bounded on  $\mathbb{R}^2$ . Consequently  $t_a \in \mathbb{L}^2(\pi \otimes \pi)$ , so that  $T_a$  is a Hilbert-Schmidt operator on  $\mathbb{L}^2(\pi)$ . In particular  $T_a$  is compact on  $\mathbb{L}^2(\pi)$ .

Now observe that

$$(T_{a^c} R_a f)(x) = 1_{\mathbb{R} \setminus [-a,a]}(x) \int_{\mathbb{R}} 1_{[-a,a]}(y) r(y) f(y) t(x, y) dy = \int_{\mathbb{R}} f(y) k_a(x, y) \pi(y) dy$$

where  $k_a(x, y) := 1_{[-a,a]}(y) 1_{\mathbb{R} \setminus [-a,a]}(x) r(y) t(x, y) \pi(y)^{-1}$ . Then  $T_{a^c} R_a$  is a Hilbert-Schmidt operator on  $\mathbb{L}^2(\pi)$  since  $k_a(\cdot, \cdot)$  is bounded on  $\mathbb{R}^2$  from our assumptions. Thus  $T_{a^c} R_a$  is compact on  $\mathbb{L}^2(\pi)$ .

Let us prove by induction that  $(R_{a^c} + T_{a^c})^n R_a$  is compact on  $\mathbb{L}^2(\pi)$  for any  $n \geq 1$ . For  $n = 1$ ,  $(R_{a^c} + T_{a^c}) R_a = T_{a^c} R_a$  is compact. Next,

$$(R_{a^c} + T_{a^c})^n R_a = (R_{a^c} + T_{a^c})^{n-1} (R_{a^c} + T_{a^c}) R_a = (R_{a^c} + T_{a^c})^{n-1} T_{a^c} R_a.$$

Since  $T_{a^c} R_a$  is compact on  $\mathbb{L}^2(\pi)$  and the set of compact operators on  $\mathbb{L}^2(\pi)$  is an ideal,  $(R_{a^c} + T_{a^c})^n R_a$  is compact on  $\mathbb{L}^2(\pi)$ .  $\square$

**Lemma 2** *For every  $n \geq 1$ , there exists a compact operator  $K_n$  on  $\mathbb{L}^2(\pi)$  such that*

$$P^n = K_n + R_a^n + (R_{a^c} + T_{a^c})^n.$$

*Proof.* For  $n = 1$  we have  $P = K_1 + R_a + (R_{a^c} + T_{a^c})$  with  $K_1 := T_a$  compact by Lemma 1. Now assume that the conclusion of Lemma 2 holds for some  $n \geq 1$ . Since the set of compact operators on  $\mathbb{L}^2(\pi)$  forms a two-sided operator ideal, we obtain the following equalities for some compact operator  $K'_{n+1}$  on  $\mathbb{L}^2(\pi)$ :

$$\begin{aligned} P^{n+1} = P^n P &= (K_n + R_a^n + (R_{a^c} + T_{a^c})^n)(K_1 + R_a + (R_{a^c} + T_{a^c})) \\ &= K'_{n+1} + R_a^{n+1} + R_a^n R_{a^c} + R_a^n T_{a^c} + (R_{a^c} + T_{a^c})^n R_a + (R_{a^c} + T_{a^c})^{n+1} \\ &= [K'_{n+1} + (R_{a^c} + T_{a^c})^n R_a] + R_a^{n+1} + (R_{a^c} + T_{a^c})^{n+1}. \end{aligned}$$

Then the expected conclusion holds true for  $P^{n+1}$  since  $K_{n+1} := K'_{n+1} + (R_{a^c} + T_{a^c})^n R_a$  is compact on  $\mathbb{L}^2(\pi)$  from Lemma 1.  $\square$

Theorem 1 is deduced from the next proposition which states that  $\|T_{a^c}\|_2 \leq \beta_a$ . Indeed, observe that  $\|R_a\|_2 \leq r_a$  and  $\|R_{a^c}\|_2 \leq r'_a$ . Set  $\alpha_a := \max(r_a, r'_a + \beta_a)$ . Then Lemma 2 and  $\|T_{a^c}\|_2 \leq \beta_a$  give

$$\|P^n - K_n\|_2 \leq \|R_a\|_2^n + (\|R_{a^c}\|_2 + \|T_{a^c}\|_2)^n \leq 2\alpha_a^n.$$

The expected inequality  $r_{ess}(P) \leq \alpha_a$  in Theorem 1 then follows from Formula (6).

**Proposition 1** *For any  $a > 0$ , we have  $\|T_{a^c}\|_2 \leq \beta_a$ .*

*Proof.* Lemma 3 below shows that, for any bounded and continuous function  $f : \mathbb{R} \rightarrow \mathbb{C}$ , we have  $\|T_{a^c} f\|_{\mathbb{L}^2(\pi)} \leq \beta_a \|f\|_{\mathbb{L}^2(\pi)}$ . Then Inequality  $\|T_{a^c}\|_2 \leq \beta_a$  of Proposition 1 follows from a standard density argument using that  $\|T\|_2 \leq \|P\|_2 = 1$  and that the space of bounded and continuous functions from  $\mathbb{R}$  to  $\mathbb{C}$  is dense in  $\mathbb{L}^2(\pi)$ .  $\square$

**Lemma 3** *For any bounded and continuous function  $f : \mathbb{R} \rightarrow \mathbb{C}$ , we have*

$$\|T_{a^c} f\|_{\mathbb{L}^2(\pi)} \leq \int_{-s}^s \left[ \int_{\{|x|>a\}} |f(x+u)|^2 t(x, x+u)^2 \pi(x) dx \right]^{\frac{1}{2}} du \leq \beta_a \|f\|_{\mathbb{L}^2(\pi)}. \quad (12)$$

*Proof.* Let  $f : \mathbb{R} \rightarrow \mathbb{C}$  be a bounded and continuous function. Set  $B := \mathbb{R} \setminus [-a, a]$ . Then it follows from (1) that

$$(T_{a^c} f)(x) = 1_B(x) \int_{\mathbb{R}} f(y) t(x, y) dy = 1_B(x) \int_{-s}^s f(x+u) t(x, x+u) du. \quad (13)$$

For  $n \geq 1$  and for  $k = 0, \dots, n$ , set  $u_k := -s + 2sk/n$  and define the following functions:  $h_k(x) := 1_B(x) f(x+u_k) t(x, x+u_k)$ . Then

$$\begin{aligned} \left[ \int_B \left| \frac{2s}{n} \sum_{k=1}^n h_k(x) \right|^2 \pi(x) dx \right]^{\frac{1}{2}} &= \left\| \frac{2s}{n} \sum_{k=1}^n h_k \right\|_{\mathbb{L}^2(\pi)} \leq \frac{2s}{n} \sum_{k=1}^n \|h_k\|_{\mathbb{L}^2(\pi)} \\ &\leq \frac{2s}{n} \sum_{k=1}^n \left[ \int_B |f(x+u_k)|^2 t(x, x+u_k)^2 \pi(x) dx \right]^{\frac{1}{2}}. \end{aligned} \quad (14)$$

Below we prove that, when  $n \rightarrow +\infty$ , the left hand side of (14) converges to  $\|T_{a^c} f\|_{\mathbb{L}^2(\pi)}$  and that the right hand side of (14) converges to the right hand side of the first inequality in (12). Define

$$\forall x \in B, \quad \chi_n(x) := \frac{2s}{n} \sum_{k=1}^n h_k(x) = \frac{2s}{n} \sum_{k=1}^n f(x + u_k) t(x, x + u_k).$$

From Riemann's integral it follows that

$$\forall x \in B, \quad \lim_{n \rightarrow +\infty} \chi_n(x) = \int_{-s}^s f(x + u) t(x, x + u) du$$

since the function  $u \mapsto f(x + u) t(x, x + u)$  is continuous on  $[-s, s]$  from the assumptions of Theorem 1. Note that  $\sup_n \sup_{x \in B} |\chi_n(x)| < \infty$  since  $f$  and  $t$  are bounded functions. From Lebesgue's theorem and from (13), it follows that

$$\begin{aligned} \lim_{n \rightarrow +\infty} \int_B \left| \frac{2s}{n} \sum_{k=1}^n h_k(x) \right|^2 \pi(x) dx &= \int_B \left| \int_{-s}^s f(x + u) t(x, x + u) du \right|^2 \pi(x) dx \\ &= \|T_{a^c} f\|_{\mathbb{L}^2(\pi)}^2. \end{aligned} \quad (15)$$

Next, observe that

$$\frac{2s}{n} \sum_{k=1}^n \left[ \int_B |f(x + u_k)|^2 t(x, x + u_k)^2 \pi(x) dx \right]^{\frac{1}{2}} = \frac{2s}{n} \sum_{k=1}^n \psi(u_k)$$

with  $\psi$  defined by

$$\psi(u) := \left[ \int_B |f(x + u)|^2 t(x, x + u)^2 \pi(x) dx \right]^{\frac{1}{2}}.$$

Using the assumptions Theorem 1, it follows from Lebesgue's theorem that  $\psi$  is continuous. Consequently Riemann integral gives

$$\lim_{n \rightarrow +\infty} \frac{2s}{n} \sum_{k=1}^n \left[ \int_B |f(x + u_k)|^2 t(x, x + u_k)^2 \pi(x) dx \right]^{\frac{1}{2}} = \int_{-s}^s \psi(u) du. \quad (16)$$

The first inequality in (12) follows from (14) by using (15) and (16).

Let us prove the second inequality in (12). The detailed balance equation (5) gives

$$\begin{aligned} & \int_{-s}^s \left[ \int_{\{|x|>a\}} |f(x + u)|^2 t(x, x + u)^2 \pi(x) dx \right]^{\frac{1}{2}} du \\ &= \int_{-s}^s \left[ \int_{\{|x|>a\}} |f(x + u)|^2 t(x, x + u) t(x, x + u) \pi(x) dx \right]^{\frac{1}{2}} du \\ &= \int_{-s}^s \left[ \int_{\{|x|>a\}} t(x, x + u) t(x + u, x) |f(x + u)|^2 \pi(x + u) dx \right]^{\frac{1}{2}} du \\ &\leq \|f\|_{\mathbb{L}^2(\pi)} \int_{-s}^s \sup_{|x|>a} \sqrt{t(x, x + u) t(x + u, x)} du = \|f\|_{\mathbb{L}^2(\pi)} \beta_a. \end{aligned}$$

□

## 4 Conclusion

The study of the iterates of a Metropolis-Hasting kernel  $P$  is of great interest to estimate the numbers of iterations required to achieve the convergence in the Metropolis-Hasting algorithm. In conclusion we discuss this issue by comparing the expected results depending on whether  $P$  acts on  $\mathcal{B}_V$  or on  $\mathbb{L}^2(\pi)$ . Recall that the  $V$ -geometrical ergodicity for  $P$  (see Remark 1) writes as: there exist  $\rho \in (0, 1)$  and  $C_\rho > 0$  such that

$$\forall n \geq 1, \forall f \in \mathcal{B}_V, \quad \|P^n f - \pi(f)\|_V \leq C_\rho \rho^n \|f\|_V. \quad (\text{SG}_V)$$

Let  $\varrho_V(P)$  be the infimum bound of the real numbers  $\rho$  such that  $(\text{SG}_V)$  holds true.

1. In most of cases, the number  $\varrho_V(P)$  is not known for Metropolis-Hasting kernels. The upper bounds of  $\varrho_V(P)$  derived from drift and minorization inequalities seem to be poor and difficult to improve, excepted in stochastically monotone case (e.g. see [MT96, Sec. 6] and [Bax05]). Consequently the inequality  $\varrho_2(P) \leq \varrho_V(P)$  (see [Bax05, Th. 6.1]) is not relevant here. Observe that applying the quasi-compactness approach on  $\mathcal{B}_V$  would allow us to estimate the value of  $\varrho_V(P)$ , but in practice this method cannot be efficient since no accurate bound of the essential spectral radius of  $P$  on  $\mathcal{B}_V$  is known.
2. The present paper shows that considering the action on  $\mathbb{L}^2(\pi)$  rather than on  $\mathcal{B}_V$  of a Metropolis-Hasting kernel  $P$  enables us to benefit from the richness of Hilbert spaces. The notion of Hilbert-Schmidt operators plays an important role for obtaining our bound (9). The reversibility of  $P$ , that is  $P$  is self-adjoint on  $\mathbb{L}^2(\pi)$ , implies that any upper bound  $\rho$  of  $\varrho_2(P)$  gives the inequality  $\|P^n f - \Pi f\|_2 \leq \rho^n \|f\|_2$  for every  $n \geq 1$  and every  $f \in \mathbb{L}^2(\pi)$ . Consequently any such  $\rho$  provides an efficient information to estimate the numbers of iterations required to achieve the convergence in the Metropolis-Hasting algorithm.

From our bound (7) it can be expected that the quasi-compactness method (cf. Introduction) will give a numerical procedure for estimating  $\varrho_2(P)$  in the continuous state space case.

## References

- [AP07] Yves F. Atchadé and François Perron. On the geometric ergodicity of Metropolis-Hastings algorithms. *Statistics*, 41(1):77–84, 2007.
- [Bax05] P. H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.*, 15(1B):700–738, 2005.
- [FHL12] D. Ferré, L. Hervé, and J. Ledoux. Limit theorems for stationary Markov processes with  $L^2$ -spectral gap. *Ann. Inst. H. Poincaré Probab. Statist.*, 48:396–423, 2012.
- [Hen93] H. Hennion. Sur un théorème spectral et son application aux noyaux lipchitziens. *Proc. Amer. Math. Soc.*, 118:627–634, 1993.

- [HL14] L. Hervé and J. Ledoux. Approximating Markov chains and  $V$ -geometric ergodicity via weak perturbation theory. *Stochastic Process. Appl.*, 124(1):613–638, 2014.
- [HL16] L. Hervé and J. Ledoux. Computable bounds of  $\ell^2$ -spectral gap for discrete Markov chains with band transition matrices. *JAP*, To appear in 2016.
- [HP10] L. Hervé and F. Pène. The Nagaev-Guivarc’h method via the Keller-Liverani theorem. *Bull. Soc. Math. France*, 138:415–489, 2010.
- [MT93] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Springer-Verlag London Ltd., London, 1993.
- [MT96] K. L. Mengersen and R. L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996.
- [RR97] G. O. Roberts and J. S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Elect. Comm. in Probab.*, 2:13–25, 1997.
- [RR04] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71 (electronic), 2004.
- [RT96] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996.
- [Wu04] L. Wu. Essential spectral radius for Markov semigroups. I. Discrete time case. *Probab. Theory Related Fields*, 128(2):255–321, 2004.