Available online at www.sciencedirect.com

ScienceDirect

Journal of Complexity 23 (2007) 673–696

Journal of COMPLEXITY

www.elsevier.com/locate/jco

Simple Monte Carlo and the Metropolis algorithm

Peter Mathé[a], Erich Novak[b,*]

[a] *Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, D-10117 Berlin, Germany*
[b] *Friedrich Schiller University Jena, Mathem. Institute, Ernst-Abbe-Platz 2, D-07743 Jena, Germany*

Received 21 October 2006; accepted 14 May 2007
Dedicated to our dear colleague and friend Henryk Woźniakowski on the occasion of his 60th birthday
Available online 15 June 2007

Abstract

We study the integration of functions with respect to an unknown density. Information is available as oracle calls to the integrand and to the non-normalized density function. We are interested in analyzing the integration error of optimal algorithms (or the complexity of the problem) with emphasis on the variability of the weight function. For a corresponding large class of problem instances we show that the complexity grows linearly in the variability, and the simple Monte Carlo method provides an almost optimal algorithm. Under additional geometric restrictions (mainly log-concavity) for the density functions, we establish that a suitable adaptive local Metropolis algorithm is almost optimal and outperforms any non-adaptive algorithm.

MSC: 65C05; secondary: 65Y2068Q17; 82B80

Keywords: Monte Carlo methods; Metropolis algorithm; Log-concave density; Rapidly mixing Markov chains; Optimal algorithms; Adaptivity; Complexity

1. Introduction, problem description

In many applications one wants to compute an integral of the form

$$\int_\Omega f(x)\cdot c\varrho(x)\mu(dx) \tag{1}$$

with a density $c\varrho(x)$, $x\in\Omega$, where $c>0$ is unknown and μ is a probability measure. Of course we have $1/c=\int_\Omega \varrho(x)\mu(dx)$, but the numerical computation of the latter integral is often as hard as the original problem (1). Therefore it is desirable to have algorithms which are able

* Corresponding author.
E-mail addresses: mathe@wias-berlin.de (P. Mathé), novak@math.uni-jena.de (E. Novak).

0885-064X/$ - see front matter
doi:10.1016/j.jco.2007.05.002

to approximately compute (1) without knowing the normalizing constant, based solely on n function values of f and ϱ. In other terms, these functions are given by an *oracle*, i.e., we assume that we can compute function values of f and ϱ.

Solution operator. Assume that we are given any class $\mathcal{F}(\Omega)$ of input data (f, ϱ) defined on a set Ω. We can rewrite the integral in (1) as

$$S(f, \varrho) = \frac{\int f(x) \cdot \varrho(x)\mu(dx)}{\int \varrho(x)\mu(dx)} \quad (f, \varrho) \in \mathcal{F}(\Omega). \tag{2}$$

This *solution operator* is linear in f but not in ϱ. We discuss algorithms for the (approximate) computation of $S(f, \varrho)$.

Remark 1. This solution operator is closely related to systems in statistical mechanics, which obey a Boltzmann (or Maxwell or Gibbs) distribution, i.e., when there is a countable number $j = 1, 2, \ldots$ of microstates with energies, say E_j, and the overall system is distributed according to the Boltzmann distribution, with inverse temperature β, as

$$P_\beta(j) := \frac{e^{-\beta E_j}}{Z_\beta}, \quad j = 1, 2, \ldots .$$

In this case the normalizing constant Z_β is the *partition function*, corresponding to $1/c$ from (1) and $\varrho^\beta(j) = e^{-\beta E_j}$ for $j \in \mathbb{N}$.

In this setup, if A is any global thermodynamic quantity, then its expected value $\langle A \rangle_\beta$ is given by

$$\langle A \rangle_\beta := \frac{1}{Z_\beta} \sum_j A_j e^{-\beta E_j},$$

which can be written as $S(A, \varrho^\beta)$. Observe, however, that we use here slightly different assumptions since we use the counting measure on $\mathbb{N}$, not a probability measure.

Randomized methods. Monte Carlo methods (randomized methods) are important numerical tools for integration and simulation in science and engineering, we refer to the recent special issue [7]. The Metropolis method, or more accurately, the class of *Metropolis–Hastings algorithms* ranges among the most important methods in numerical analysis and scientific computation, see [6,23].

Here we consider randomized methods S_n that use n function evaluations of f and ϱ. Hence S_n is of the form as exhibited in Fig. 1.

In all steps, random number generators may be used to determine the consecutive node. If the nodes x_i from **Step** do not depend on previously computed values of $f(x_1), \ldots, f(x_{i-1})$ and $\varrho(x_1), \ldots, \varrho(x_{i-1})$, then the algorithm is called *non-adaptive*, otherwise it is called *adaptive*. Specifically we analyze the procedures S_n^{simple} and S_n^{mh}, introduced in (3) and (5) below.

Remark 2. The notion of *adaption* which is used here differs from the one recently used to introduce *adaptive MCMC*, see e.g. [1,3]. The Metropolis algorithm which is used in this paper is based on a *homogeneous* Markov chain, in our notation this is still an adaptive algorithm since the used nodes x_i depend on ϱ. Hence we use the concept of adaptivity from numerical analysis and information-based complexity, see [22].

Algorithm: $S_n(f,\varrho)$
Data: Functions f, ϱ, `random numbers` $\omega_1,\dots,\omega_n$;
Result: approximate value $S_n(f,\varrho)$ for $S(f,\varrho)$ from Eq. (2);
begin
  Init $x_1 := x_1(\omega_1)$, **Compute** $f(x_1)$ *and* $\varrho(x_1)$;
  for $i = 2,\dots,n$ **do**
    Step $x_i := x_i(f(x_1),\dots,f(x_{i-1}),\varrho(x_1),\varrho(x_{i-1}),\omega_i)$;
    Compute $f(x_i)$ *and* $\varrho(x_i)$;
  end
  Compute $S_n(f,\varrho) = \varphi_n(f(x_1),\dots,f(x_n),\varrho(x_1),\dots,\varrho(x_n)) \in \mathbb{R}$;
end

Fig. 1. Generic Monte Carlo algorithm based on n values of f and ϱ. The final **Compute** may use any mapping $\varphi_n : \mathbb{R}^{2n} \to \mathbb{R}$.

For details on the model of computation we refer to [20,21,27]. Here we only mention the following: We use the real number model and assume that f and ϱ are given by an oracle for function values. Our lower bounds hold under very general assumptions concerning the available random number generator. [1]

For the upper bounds we only study two algorithms in this paper, described in (3) and (5), below. Specifically we shall deal with the (non-adaptive) *simple Monte Carlo method* and a specific (adaptive) *Metropolis–Hastings method*. The former can only be applied if a random number generator for μ on Ω is available. Thus there are natural situations when this method cannot be used. The latter will be based on a suitable ball walk. Hence we need a random number generator for the uniform distribution on a (Euclidean) ball. Thus the Metropolis–Hastings methods can also be applied when a random number generator for μ on Ω is not available. Instead, we need a "membership oracle" for Ω: On input $x \in \mathbb{R}^d$ this oracle can decide with cost 1 whether $x \in \Omega$ or not.

Error criterion. We are interested in error bounds uniformly for classes $\mathcal{F}(\Omega)$ of input data. If S_n is any method that uses (at most) n values of f and ϱ then the (individual) error for the problem instance $(f,\varrho) \in \mathcal{F}(\Omega)$ is given by

$$e(S_n,(f,\varrho)) = \left(\mathbf{E}\,|S(f,\varrho) - S_n(f,\varrho)|^2\right)^{1/2},$$

where $\mathbf{E}$ means the expectation. The overall (or worst case) error on the class $\mathcal{F}(\Omega)$ is

$$e(S_n,\mathcal{F}(\Omega)) = \sup_{(f,\varrho)\in\mathcal{F}(\Omega)} e(S_n,(f,\varrho)).$$

The complexity of the problem is given by the error of the best algorithm, hence we let

$$e_n(\mathcal{F}(\Omega)) := \inf_{S_n} e(S_n,\mathcal{F}(\Omega)).$$

The classes $\mathcal{F}(\Omega)$ under consideration will always contain constant densities $\varrho = c > 0$ and all f with $\|f\|_\infty \leqslant 1$, hence

$$\mathcal{F}_1(\Omega) := \{(f,\varrho),\ |f(x)| \leqslant 1,\ x \in \Omega,\ \text{and}\ \varrho = c\} \subset \mathcal{F}(\Omega).$$

[1] Observe, however, that we cannot use a random number generator for the "target distribution'' $\mu_\varrho = \varrho\cdot\mu/\|\varrho\|_1$, since ϱ is part of the input.

On this class the problem (2) reduces to the classical integration problem for uniformly bounded functions, and it is well known that the error of any Monte Carlo method can decrease at a rate $n^{-1/2}$, at most. Precisely, it holds true that

$$e_n(\mathcal{F}_1(\Omega)) = \frac{1}{1+\sqrt{n}},$$

if the probability μ is non-atomic, see [17]. On the other hand we will only consider (f, ϱ) with $S(f, \varrho) \in [-1, 1]$, hence the trivial algorithm $S_0 = 0$ always has error 1.

For the classes $\mathcal{F}_C(\Omega)$ and $\mathcal{F}^\alpha(\Omega)$, which will be introduced in Section 2, we easily obtain the optimal order $e_n(\mathcal{F}(\Omega)) \asymp n^{-1/2}$. We will analyze how $e_n(\mathcal{F}(\Omega))$ depends on the parameters C and α, in case $\mathcal{F}(\Omega) := \mathcal{F}_C(\Omega)$ or $\mathcal{F}(\Omega) := \mathcal{F}^\alpha(\Omega)$, respectively.

We discuss some of our subsequent results and provide a short outline. In Section 2 we shall specify the methods and classes of input data to be analyzed. The classes $\mathcal{F}_C(\Omega)$, analyzed first in Section 3, contain all densities ϱ with $\sup \varrho / \inf \varrho \leqslant C$. In typical applications we may face $C = 10^{20}$. Then we cannot decrease the error of optimal methods from 1 to 0.7 even with sample size $n = 10^{15}$, see Theorem 1 for more details. Hence the classes $\mathcal{F}_C(\Omega)$ are so large that no algorithm, deterministic or Monte Carlo, adaptive or non-adaptive, can provide an acceptable error. We also prove that the simple (non-adaptive) Monte Carlo method is almost optimal, no sophisticated Markov chain Monte Carlo method can help.

Thus we face the question whether adaptive algorithms, such as the Metropolis algorithm, help significantly on "suitable and interesting" subclasses of $\mathcal{F}_C(\Omega)$. We give a positive answer for the classes $\mathcal{F}^\alpha(\Omega)$, analyzed in Section 4. Here we assume that $\Omega \subset \mathbb{R}^d$ is a convex body, and that μ is the normalized Lebesgue measure μ_Ω on Ω. The class $\mathcal{F}^\alpha(\Omega)$ contains log-concave densities, where α is the Lipschitz constant of $\log \varrho$. We shall establish in Section 4.1 that all non-adaptive methods (such as the simple Monte Carlo method) suffer from the curse of dimension, i.e., we get similar lower bounds as for the classes $\mathcal{F}_C(\Omega)$. However, in Section 4.2 we shall design and analyze specific (adaptive) Metropolis algorithms that are based on some underlying ball walks, tuned to the class parameters. Using such algorithms we can break the curse of dimension by adaption. The main error estimate for this algorithm is given in Theorem 5, and we conclude this study with further discussion in the final Section 5.

2. Specific methods and classes of input

We consider the approximate computation of $S(f, \varrho)$ for large classes of input data. Since with deterministic algorithms one cannot improve the trivial zero algorithm (with error 1), we study randomized or Monte Carlo algorithms.

The methods. The Monte Carlo methods under consideration fit the schematic view from Fig. 1.

Simple Monte Carlo. Here the random numbers $\omega_1, \ldots, \omega_n$ are identically and independently distributed according to μ, and the routine **Step** chooses $X_i := \omega_i$. The final routine **Compute** is the quotient of the sample means of the computed function values

$$S_n^{\text{simple}}(f, \varrho) := \frac{\sum_{j=1}^n f(X_j)\varrho(X_j)}{\sum_{j=1}^n \varrho(X_j)}. \tag{3}$$

Metropolis–Hastings method. This describes a class of (adaptive) Monte Carlo methods which are based on the ingenious idea to construct in **Step** a Markov chain having

$$\mu_\varrho := \frac{\varrho \cdot \mu}{\int \varrho(x)\mu(dx)} \tag{4}$$

as invariant distribution without knowing the normalization. Thus, if $(X_1, X_2, \dots, X_n)$ is a trajectory of such a Markov chain, then we let **Compute** be given as

$$S_n^{\mathrm{mh}}(f, \varrho) := \frac{1}{n}\sum_{j=1}^{n} f(X_j). \tag{5}$$

Hence we use n steps of the Markov chain, the number of needed (different) function values of ϱ and f might be smaller. We will further specify the Metropolis–Hastings algorithm for the problem at hand in Section 4.2, see Figs. 2 and 3 for a schematic presentation and Theorem 5 for the choice of δ. Both Monte Carlo methods construct Markov chains, i.e., the point x_i depends on x_{i-1} and $\varrho(x_{i-1})$, only. This trivially holds true for simple Monte Carlo, since x_i does not at all depend on earlier computed function values.

Remark 3. Comparisons of different Monte Carlo methods for problems similar to (2) are frequently met in the literature. We mention [5] with a comparison of *Metropolis algorithms* and *importance sampling*, where an error expansion at any instance (f, ϱ) is given in terms of certain auto-correlations. The simple Monte Carlo method, as introduced below, is also studied there as $\tilde{\mu}_I$ for $\varrho = 1$.

The (point-wise almost sure) convergence of both methods S_n^{simple} and S_n^{mh}, as $n \to \infty$, is ensured by corresponding ergodic theorems, see [14]. But, as outlined above, we are interested in the uniform error on relatively large *problem classes*.

The classes. Here we formally describe the classes of input under consideration.

The class $\mathcal{F}_C(\Omega)$. Let μ be an arbitrary probability measure on a set Ω and consider the set

$$\mathcal{F}_C(\Omega) = \left\{(f, \varrho) \,|\, \|f\|_\infty \leqslant 1,\ \varrho > 0,\ \frac{\varrho(x)}{\varrho(y)} \leqslant C,\ x, y \in \Omega\right\}.$$

Note that necessarily $C \geqslant 1$. If $C = 1$ then ϱ is constant and we almost face the ordinary integration problem, since ϱ can be recovered with only one function value.

In many applications the constant C is huge and we will establish that the complexity of the problem (the cost of an optimal algorithm) is linear in C. Therefore, for large C, the class is too large. We have to look for smaller classes that contain many interesting pairs (f, ϱ) and have smaller complexity.

The class $\mathcal{F}^\alpha(\Omega)$ *with log-concave densities*. In many applications, we have a weight ϱ with additional properties and we assume the following:

- The set $\Omega \subset \mathbb{R}^d$ is a *convex body*, that is a compact and convex set with non-empty interior. The probability $\mu = \mu_\Omega$ is the normalized Lebesgue measure on the set Ω.
- The functions f and ϱ are defined on Ω.
- The weight $\varrho > 0$ is log-concave, i.e.,

  $$\varrho(\lambda x + (1-\lambda)y) \geqslant \varrho(x)^\lambda \cdot \varrho(y)^{1-\lambda},$$

  where $x, y \in \Omega$ and $0 < \lambda < 1$.
- The logarithm of ϱ is Lipschitz, i.e., $|\log \varrho(x) - \log \varrho(y)| \leqslant \alpha\|x - y\|_2$.

Thus we consider the class of log-concave weights on $\Omega \subset \mathbb{R}^d$ given by

$$\mathcal{R}^\alpha(\Omega) = \{\varrho | \varrho > 0,\ \log \varrho \text{ is concave},\ |\log \varrho(x) - \log \varrho(y)| \leqslant \alpha \|x - y\|_2\}. \tag{6}$$

We study the following class $\mathcal{F}^\alpha(\Omega)$ of problem elements,

$$\mathcal{F}^\alpha(\Omega) = \left\{(f, \varrho) | \varrho \in \mathcal{R}^\alpha(\Omega),\ \|f\|_{2,\varrho} \leqslant 1\right\}, \tag{7}$$

where $\|\cdot\|_{2,\varrho}$ is the L_2-norm with respect to the probability measure μ_ϱ, see (4). In some places we restrict our study to the (Euclidean) unit ball, i.e., $\Omega := B^d \subset \mathbb{R}^d$.

Remark 4. Let $\mathcal{R}_C(\Omega)$ be the class of weight functions that belong to $\mathcal{F}_C(\Omega)$. Then $\mathcal{R}^\alpha(\Omega) \subset \mathcal{R}_C(\Omega)$ if $C = e^{\alpha D}$, where D is the diameter of Ω. Thus large α correspond to "exponentially large" values of C. However, the densities from the class $\mathcal{R}^\alpha(\Omega)$ have some extra (local) properties: they are log-concave and Lipschitz continuous. These properties can be used for the construction of fast adaptive methods, via rapidly mixing Markov chains.

3. Analysis for $\mathcal{F}_C(\Omega)$

We assume that Ω is an arbitrary set and μ is a probability measure on Ω, and that the functions f and ϱ are defined on Ω.

In the applications, the constant C might be very large, something like $C = 10^{20}$ is a realistic assumption. Therefore we want to know how the complexity (the cost of optimal algorithms) depends on C. Observe that the problem is correctly normalized or scaled such that $S(\mathcal{F}_C(\Omega)) = [-1, 1]$, for any $C \geqslant 1$. We will prove that the complexity of the problem is linear in C, and hence there is no way to solve the problem if C is really huge. We start with establishing a lower bound and then show that simple Monte Carlo achieves this error up to a constant.

3.1. Lower bounds

Here we prove lower bounds for all (adaptive or non-adaptive) methods that use n evaluations of f and ϱ. We use the technique of Bahvalov, i.e., we study the average error of deterministic algorithms with respect to certain discrete measures on $\mathcal{F}_C(\Omega)$.

Theorem 1. *Assume that we can partition Ω into $2n$ disjoint sets with equal measure (equal to $1/2n$). Then for any Monte Carlo method S_n that uses n values of f and ϱ we have the lower bound*

$$e(S_n, \mathcal{F}_C(\Omega)) \geqslant \frac{1}{6}\sqrt{2} \begin{cases} \sqrt{\dfrac{C}{2n}}, & 2n \geqslant C - 1, \\ \dfrac{3C}{C + 2n - 1}, & 2n < C - 1. \end{cases} \tag{8}$$

The lower bound will be obtained in two steps.

(1) We first reduce the error analysis for Monte Carlo sampling to the average case error analysis with respect to a certain prior probability on the class $\mathcal{F}_C(\Omega)$. This approach is due to Bahvalov, see [4].
(2) For the chosen prior the average case analysis can be carried out explicitly and will thus yield a lower bound.

To construct the prior let $m := 2n$ and $\Omega_1, \dots, \Omega_m$ the partition into sets of equal probability, and χ_{Ω_j} the corresponding characteristic functions. Furthermore, let

$$l := \begin{cases} \left\lceil \dfrac{m}{C-1} \right\rceil, & m \geqslant C-1, \\ 1 & \text{else.} \end{cases}$$

Denote J_l^m the set of all subsets of $\{1, \dots, m\}$ of cardinality equal to l, and $\mu_{m,l}$ the equi-distribution on J_l^m, while $\mathbf{E}_{m,l}$ denotes the expectation with respect to the prior $\mu_{m,l}$. Let $(\varepsilon_1, \dots, \varepsilon_m)$ be independent and identically distributed with $P(\varepsilon_j = -1) = P(\varepsilon_j = 1) = \frac{1}{2}$, $j = 1, \dots, m$. The overall prior is the product probability on $J_l^m \times \{\pm 1\}^m$. For any realization $\omega = (I, \varepsilon_1, \dots, \varepsilon_m)$ we assign

$$f_\omega := \sum_{j \in I} \varepsilon_j \chi_{\Omega_j} \quad \text{and} \quad \varrho_\omega := C \sum_{j \in I} \chi_{\Omega_j} + \sum_{j \notin I} \chi_{\Omega_j}.$$

The following observation is useful.

Lemma 1. *For any subset $N \subset \{1, \dots, m\}$ of cardinality at most n it holds*

$$\mathbf{E}_{m,l} \#(I \setminus N) \geqslant \frac{l}{2}.$$

Proof. Clearly, for any fixed $k \in \{1, \dots, m\}$ we have $\mu_{m,l}(k \in I) = l/m$, thus

$$\mathbf{E}_{m,l} \#(I \setminus N) = \sum_{r \in N^c} \mathbf{E}_{m,l} \chi_I(r) = \#(N^c) \frac{l}{m} \geqslant \frac{l}{2},$$

where we denoted by N^c the complement of N. □

Proof Theorem 1. Given the above prior let us denote

$$e_n^{\mathrm{avg}}(\mathcal{F}_C(\Omega)) := \inf_q \left(\mathbf{E}_{m,l} \mathbf{E}_\varepsilon |S(f, \varrho) - q(f, \varrho)|^2 \right)^{1/2}, \tag{9}$$

where the inf is taken with respect to any (possibly adaptive) deterministic algorithm which uses at most n values from f and ϱ.

For any Monte Carlo method S_n we have, using Bahvalov's argument [4], the relation

$$e(S_n, \mathcal{F}_C(\Omega)) \geqslant e_n^{\mathrm{avg}}(\mathcal{F}_C(\Omega)). \tag{10}$$

We provide a lower bound for $e_n^{\mathrm{avg}}(\mathcal{F}_C(\Omega))^2$. To this end note that for each realization $(f_\omega, \varrho_\omega)$ the integral $\int \varrho_\omega \, d\mu$ is constant. In the first case $m \geqslant C-1$, and we can bound the integral by the choice of l as

$$c_{m,l} := \int \varrho_\omega(x) \, \mu(dx) = \frac{1}{m} \left(lC + (m-l)1 \right) \leqslant 3. \tag{11}$$

In the other case $m < C-1$, we obtain $c_{m,1} = (C - 1 + m)/m$. Now, to analyze the average case error, let q_n be any (deterministic) method, and let us assume that it uses the set N of nodes.

We have the decomposition

$$S(f_\omega, \varrho_\omega) - q_n(f_\omega, \varrho_\omega) = \left(\frac{C}{mc_{m,l}} \sum_{j \in I \setminus N} \varepsilon_j \right) - \left(\frac{C}{mc_{m,l}} \sum_{j \in I \cap N} \varepsilon_j - q_n(f_\omega, \varrho_\omega) \right).$$

Given I, the random variables in the brackets are conditionally independent, thus uncorrelated. Hence we conclude that

$$\begin{aligned} \mathbf{E}_{m,l}\mathbf{E}_\varepsilon \left| S(f_\omega, \varrho_\omega) - q_n(f_\omega, \varrho_\omega) \right|^2 &\geqslant \mathbf{E}_{m,l}\mathbf{E}_\varepsilon \left| \frac{C}{mc_{m,l}} \sum_{j \in I \setminus N} \varepsilon_j \right|^2 \\ &= \frac{C^2}{m^2 c_{m,l}^2} \mathbf{E}_{m,l} \#(J \setminus N) \geqslant \frac{C^2 l}{2m^2 c_{m,l}^2}, \end{aligned}$$

by Lemma 1. In the case $m \geqslant C - 1$ we obtain $l \geqslant m/C$ and have $c_{m,l} \leqslant 3$, such that

$$\mathbf{E}_{m,l} \left| S(f, \varrho) - q_n(f, \varrho) \right|^2 \geqslant \frac{C}{36n},$$

which in turn yields the first case bound in (8). In the other case $m < C - 1$ the value of $l = 1$ yields the second bound in (8). □

3.2. The error of the simple Monte Carlo method

The direct approach to evaluate (1) would be to use the method S_n^{simple} from (3). We will prove an upper bound for the error of this method, and we start with the following:

Lemma 2. *If the function ϱ obeys the requirements in $\mathcal{F}_C(\Omega)$, then*

(1) $0 < \inf_{x \in \Omega} \varrho(x) \leqslant \sup_{x \in \Omega} \varrho(x) < \infty$.
(2) *For every probability measure μ on Ω we have $\|\varrho\|_{2,\mu} \leqslant \sqrt{C} \|\varrho\|_{1,\mu}$.*

Proof. To prove the first assertion, fix any $y_0 \in \Omega$. Then the assumption on ϱ yields $\varrho(x) \leqslant C\varrho(y_0)$, and reversing the roles of x and y also the lower bound. Now both, the assumption on ϱ as well as the second assertion, are invariant with respect to multiplication of ϱ by a constant. In the light of the first assertion we may and do assume that $1 \leqslant \varrho(x) \leqslant C$, $x \in \Omega$, and we derive, using $1 \leqslant \int_\Omega \varrho(x)\ \mu(dx)$, that

$$\int_\Omega \varrho^2(x)\ \mu(dx) \leqslant C \int_\Omega \varrho(x)\ \mu(dx) \leqslant C \left(\int_\Omega \varrho(x)\ \mu(dx) \right)^2,$$

completing the proof of the second assertion and of the lemma. □

We turn to the bound for the simple Monte Carlo method.

Theorem 2. *For all $n \in \mathbb{N}$ we have*

$$e(S_n^{\text{simple}}, \mathcal{F}_C(\Omega)) \leqslant 2 \min \left\{ 1, \sqrt{\frac{2C}{n}} \right\}. \tag{12}$$

Proof. The upper bound 2 is trivial, it even holds deterministically. Fix any pair (f,ϱ) of input. For any sample $(X_1,\dots,X_n)$ and function g we denote the sample mean by $S_n^{\mathrm{mean}}(g) := 1/n\sum_{j=1}^n g(X_j)$. It is well known that $e(S_n^{\mathrm{mean}},g)\leqslant\|g\|_2/\sqrt{n}$. With this notation we can bound

$$\begin{aligned}\left|S(f,\varrho)-S_n^{\mathrm{simple}}(f,\varrho)\right| &\leqslant \left|S(f,\varrho)-\frac{S_n^{\mathrm{mean}}(f\varrho)}{\int\varrho(x)\mu(dx)}\right|+\left|\frac{S_n^{\mathrm{mean}}(f\varrho)}{\int\varrho(x)\mu(dx)}-\frac{S_n^{\mathrm{mean}}(f\varrho)}{S_n^{\mathrm{mean}}(\varrho)}\right|\\ &\leqslant \frac{1}{\|\varrho\|_1}\left(\left|\int f(x)\varrho(x)\mu(dx)-S_n^{\mathrm{mean}}(f\varrho)\right|\right.\\ &\quad\left.+\left|\frac{S_n^{\mathrm{mean}}(f\varrho)}{S_n^{\mathrm{mean}}(\varrho)}\right|\left|\int\varrho(x)\mu(dx)-S_n^{\mathrm{mean}}(\varrho)\right|\right)\\ &\leqslant \frac{1}{\|\varrho\|_1}\left(\left|\int f(x)\varrho(x)\mu(dx)-S_n^{\mathrm{mean}}(f\varrho)\right|\right.\\ &\quad\left.+\|f\|_\infty\left|\int\varrho(x)\mu(dx)-S_n^{\mathrm{mean}}(\varrho)\right|\right),\end{aligned}$$

where we used $\left|S_n^{\mathrm{mean}}(f\varrho)/S_n^{\mathrm{mean}}(\varrho)\right|\leqslant\|f\|_\infty$, which holds true since the enumerator and denominator use the same sample. This yields the following error bound:

$$\begin{aligned}e(S_n^{\mathrm{simple}},(f,\varrho)) &\leqslant \frac{\sqrt{2}}{\|\varrho\|_1}\left(e(S_n^{\mathrm{mean}},f\varrho)+\|f\|_\infty e(S_n^{\mathrm{mean}},\varrho)\right)\\ &\leqslant \frac{\sqrt{2}}{\|\varrho\|_1\sqrt{n}}\left(\|f\varrho\|_2+\|f\|_\infty\|\varrho\|_2\right)\leqslant\frac{2\sqrt{2}\|f\|_\infty}{\sqrt{n}}\frac{\|\varrho\|_2}{\|\varrho\|_1}\leqslant\frac{2\sqrt{2C}}{\sqrt{n}},\end{aligned}$$

where we use Lemma 2. Taking the supremum over $(f,\varrho)\in\mathcal{F}_C(\Omega)$ allows to complete the proof. □

4. Analysis for $\mathcal{F}^\alpha(\Omega)$

In this section we impose restrictions on the input data, in particular on the density, in order to improve the complexity. This class is still large enough to contain many important situations. Monte Carlo methods for problems when the target (invariant) distribution is log-concave proved to be important in many studies, we refer to [10]. One of the main intrinsic features of such classes of distributions are *isoperimetric inequalities*, see [2,13], which will also be used here in the form as used in [29]. Recall that here we always require that $\Omega\subset\mathbb{R}^d$ is a convex body, as introduced in Section 2.

We start with a lower bound for all non-adaptive algorithms to exhibit that simple Monte Carlo cannot take into account the additional structure of the underlying class of input data and adaptive methods should be used. This bound, together with Theorem 5, will show that adaptive methods can outperform any non-adaptive method, if we consider S on $\mathcal{F}^\alpha(B^d)$. Indeed, we also show that specific Metropolis algorithms, based on local underlying Markov chains are suited for this problem class.

4.1. A lower bound for non-adaptive methods

Here we prove a lower bound for all non-adaptive methods (hence in particular for the simple Monte Carlo method) for the problem on the classes $\mathcal{F}^\alpha(\Omega)$. Again, this lower bound will use Bahvalov's technique.

We start with a result on sphere packings. The Minkowski–Hlawka theorem, see [25], says that the density of the densest sphere packing in $\mathbb{R}^d$ is at least $\zeta(d) \cdot 2^{1-d} \geqslant 2^{1-d}$. It is also known, see [11], that the density (by definition of the whole $\mathbb{R}^d$) can be replaced by the density within a convex body Ω, as long as the radius r of the spheres tends to zero. Hence we obtain the following result.

Lemma 3. *There is $n_\Omega \in \mathbb{N}$ such that for all $m \geqslant n_\Omega$ there are points $y_1, \ldots, y_m \in \Omega$ such that with*

$$r := r(\Omega, m) := 2^{-1} m^{-1/d} \left(\frac{\mathrm{vol}(\Omega)}{\mathrm{vol}(B^d)} \right)^{1/d}$$

the closed balls $B_i := B(y_i, r) \subset \Omega$ are disjoint.

Our construction will use such points $y_1, \ldots, y_m \in \Omega$ and the corresponding balls $B_1, \ldots, B_m$ as follows.

For $i \in \{1, \ldots, m\}$ we assign

$$\varrho_i(y) := c_i \exp\left(-\alpha \|y - y_i\|_2\right), \quad y \in \Omega \quad \text{and}$$
$$f_i(y) := \tilde{c}_i \chi_{B_i}(y), \quad y \in \Omega,$$

with constants c_i and $\tilde{c}_i$ chosen such that

$$1 = \int_\Omega \varrho_i(y)\, dy = c_i \int_\Omega \exp(-\alpha \|y - y_i\|) dy \quad \text{and}$$
$$1 = \|f_i\|_{2,\varrho_i} = \tilde{c}_i^2 c_i \int_{B_i} \exp(-\alpha \|y - y_i\|)\, dy.$$

The corresponding values of the mapping S are computed as

$$\begin{aligned} S(f_i, \varrho_i) &= \int_\Omega f_i \varrho_i \, dy = \tilde{c}_i c_i \int_{B_i} \exp(-\alpha \|y - y_i\|)\, dy \\ &= \left(c_i \int_{B_i} \exp(-\alpha \|y - y_i\|)\, dy \right)^{1/2} = \left(c_i \int_{B(0,r)} \exp(-\alpha \|y\|)\, dy \right)^{1/2} \\ &= \left(\frac{\int_{B(0,r)} \exp(-\alpha \|y\|)\, dy}{\int_\Omega \exp(-\alpha \|y - y_i\|)\, dy} \right)^{1/2}. \end{aligned} \tag{13}$$

Again we turn to the average case setting, this time with probability measure μ^{2n} being the equidistribution on the set

$$\mathcal{F}^{2n} := \left\{ (\varepsilon_i f_i, \varrho_i),\ i = 1, \ldots, 2n,\ \varepsilon_i = \pm 1 \right\} \subset \mathcal{F}^\alpha(\Omega).$$

Similar to (10) we have for any non-adaptive Monte Carlo method $S_n(f, \varrho)$ the relation

$$e(S_n, \mathcal{F}^\alpha(\Omega)) \geqslant \min \left\{ e^{\mathrm{avg}}(q_n, \mu^{2n}),\ q_n \text{ is deterministic and non-adaptive} \right\},$$

where $e^{\mathrm{avg}}(q_n, \mu^{2n})$ denotes the average case error of the deterministic non-adaptive method q_n with respect to the probability μ^{2n}. Thus let q_n be any non-adaptive (deterministic) algorithm for S on the class $\mathcal{F}^\alpha(\Omega)$ that uses at most n values.

The average case error can then be bounded from below as

$$\mathbf{E}_{\mu^{2n}} |S(f,\varrho) - q_n(f,\varrho)|^2 = \frac{1}{2n} \sum_{i=1}^{2n} \mathbf{E}_\varepsilon \left|S(\varepsilon_i f_i, \varrho_i) - q_n(\varepsilon_i f_i, \varrho_i)\right|^2$$
$$\geqslant \frac{1}{2} \min_{i=1,\dots,2n} \mathbf{E}_\varepsilon \left|S(\varepsilon_i f_i, \varrho_i)\right|^2 \geqslant \frac{1}{2} \min_{i=1,\dots,2n} S(f_i, \varrho_i)^2.$$

Above, $\mathbf{E}_\varepsilon$ denotes the expectation with respect to the independent random variables $\varepsilon_i = \pm 1$. Together with (13) we obtain

$$e(S_n, \mathcal{F}^\alpha(\Omega)) \geqslant \frac{1}{2}\sqrt{2} \min_{i=1,\dots,2n} \left(\frac{\int_{B(0,r)} \exp(-\alpha\|y\|)\, dy}{\int_\Omega \exp(-\alpha\|y - y_i\|)\, dy} \right)^{1/2}.$$

We bound the enumerator from below and the denominator from above. For $\alpha r \leqslant \log 2$ we can bound

$$\int_{B(0,r)} \exp(-\alpha\|y\|)\, dy \geqslant \frac{1}{2} \operatorname{vol}(B(0,r)) = \frac{1}{2} r^d \operatorname{vol}(B^d).$$

For the denominator we have

$$\int_\Omega \exp(-\alpha\|y - y_i\|)\, dy \leqslant \int_{\mathbb{R}^d} \exp(-\alpha\|y - y_i\|)\, dy$$
$$= \alpha^{-d} \int_{\mathbb{R}^d} \exp(-\|y\|)\, dy = \alpha^{-d} \Gamma(d) \operatorname{vol} \partial B^d,$$

such that we finally obtain, using the well known formula $\operatorname{vol}(\partial B^d) = d \operatorname{vol}(B^d)$, that

$$e(S_n, \mathcal{F}^\alpha(\Omega)) \geqslant \frac{1}{2}\sqrt{2} \left(\frac{\alpha^d r^d}{2d!} \right)^{1/2} = \frac{1}{2} \left(\frac{\alpha^d r^d}{d!} \right)^{1/2}.$$

Using the value for $r = r(\Omega, 2n)$ from Lemma 3 we end up with

Theorem 3. *Assume that S_n is any non-adaptive Monte Carlo method for the class $\mathcal{F}^\alpha(\Omega)$. Then, with n_Ω from Lemma 3, we have for all*

$$2n \geqslant \max \left\{ n_\Omega, (\alpha/\log 4)^d \cdot \frac{\operatorname{vol}\Omega}{\operatorname{vol} B^d} \right\}$$

that

$$e(S_n, \mathcal{F}^\alpha(\Omega)) \geqslant 2^{-d/2-3/2} \cdot \left(\frac{\operatorname{vol}\Omega}{\operatorname{vol} B^d} \right)^{1/2} \cdot \frac{\alpha^{d/2}}{\sqrt{d!}}\, n^{-1/2}. \tag{14}$$

Remark 5. For fixed d this is a lower bound of the form $e(S_n) \geqslant c_\Omega \alpha^{d/2} n^{-1/2}$. It is interesting only if α is "large", otherwise the already mentioned lower bound $(1 + \sqrt{n})^{-1}$ is better.

We stress that in the above reasoning we essentially used the non-adaptivity of the method S_n. Indeed, if S_n were adaptive, then by just one appropriate function value $\varrho(x)$, we could identify the index i, since the functions ϱ_i are global. Then, knowing i, we could ask for the value of ε_i and would obtain the exact solution to $S(f,\varrho)$ for this small class $\mathcal{F}^{2n}$ for all $n \geqslant 2$.

4.2. Metropolis method with local underlying walk

The Metropolis algorithm we consider here has a specific routine **Step** in Fig. 1, whereas the final step **Compute** is exactly as given in (5). It is based on a specific ball walk and this version is sometimes called *ball walk with Metropolis filter*, see [29]. Two concepts from the theory of Markov chains turn out to be important, reversibility and uniform ergodicity. We recall these notions briefly, see [24] for further details. A Markov chain (K, π) is *reversible with respect to* π, if for all measurable subsets $A, B \subset \Omega$ the balance

$$\int_A K(x, B)\pi(dx) = \int_B K(x, A)\pi(dx) \tag{15}$$

holds true. Notice that in this case necessarily π is an invariant distribution.

A Markov chain is *uniformly ergodic* if there are $n_0 \in \mathbb{N}$, a constant $c > 0$ and a probability measure ν on Ω such that

$$K^{n_0}(x, A) \geqslant c\nu(A) \quad \text{for all } A \subset \Omega \text{ and } x \in \Omega. \tag{16}$$

Markov chains which are uniformly ergodic have a unique invariant probability distribution.

Our analysis will be based on conductance arguments and we recall the basic notions, see [12,16]. If (K, π) is a Markov chain with transition kernel K and invariant distribution π then we assign the

(1) *local conductance* at $x \in \Omega$ by $l_K(x) := K(x, \Omega \setminus \{x\})$,
(2) and the *conductance* as

$$\varphi(K, \pi) := \inf_{0<\pi(A)<1} \frac{\int_A K(x, A^c)\pi(dx)}{\min\{\pi(A), \pi(A^c)\}}, \tag{17}$$

where $A^c = \Omega \setminus A$.

Below we call $l > 0$ a *lower bound for the local conductance*, if $l_K(x) \geqslant l$ for all $x \in \Omega$.

The ball walk and some of its properties. Here we gather some properties of the ball walk, see [16,29], which will serve as ingredients for the analysis of Metropolis chains using this as the underlying proposal. In particular we prove that on convex bodies in $\mathbb{R}^d$ the ball walk is uniformly ergodic and we bound its conductance from below, in terms of bounds $l > 0$ for the local conductance.

We abbreviate $B(0, \delta) = \delta B^d$. Let Q_δ be the transition kernel of a local random walk having transitions within δ-balls of its current position, i.e., we let

$$Q_\delta(x, \{x\}) := 1 - \frac{\mathrm{vol}(B(x, \delta) \cap \Omega)}{\mathrm{vol}(\delta B^d)}, \tag{18}$$

and

$$Q_\delta(x, A) := \begin{cases} \dfrac{\mathrm{vol}(B(x, \delta) \cap A)}{\mathrm{vol}(\delta B^d)}, & A \subset \Omega \quad \text{and} \quad x \notin A, \\ Q_\delta(x, A \setminus \{x\}) + Q_\delta(x, \{x\}), & A \subset \Omega \quad \text{and} \quad x \in A. \end{cases} \tag{19}$$

Schematically, the transition kernel may be viewed as in Fig. 2.

Procedure `Ball-walk-step`(x, δ)

Input : current position x; $\delta > 0$;
Output: next position;
Propose: *Choose* $y \in B(x, \delta)$ *uniformly*;
Accept: if $y \in \Omega$ **then**
  return y;
else
  return x;
end

Fig. 2. Schematic view of ball walk step.

Clearly we may restrict to $\delta \leqslant D$, the diameter of Ω. The following observation is important and explains why we restrict ourselves to convex bodies..

Lemma 4. *If $\Omega \subset \mathbb{R}^d$ is a convex body, then the ball walk Q_δ has a (non-trivial) lower bound $l > 0$ for the local conductance.*

Proof. It is well-known that convex bodies satisfy the cone condition (see [9, Section 3.2, Lemma 3]). Therefore we obtain that for each $\delta > 0$ there is $l > 0$ such that for each $x \in \Omega$ we have $l_{Q_\delta}(x) \geqslant l$. □

Remark 6. Observe however, that l might be very small. For $\Omega = [0, 1]^d$, for example, we get $l = 2^{-d}$, even if δ is very small. In contrast, we will see that a large l is possible for $\Omega = B^d$ and $\delta \leqslant 1/\sqrt{d+1}$, see Lemma 7.

Notice that $l_{Q_\delta}(x) = \mathrm{vol}(B(x, \delta) \cap \Omega)/\mathrm{vol}(\delta B^d)$, hence in the following we use the inequality:

$$\mathrm{vol}(B(x, \delta) \cap \Omega) \geqslant l \,\mathrm{vol}(\delta B^d), \tag{20}$$

where $l > 0$ is a lower bound for the local conductance of the ball walk.

The following result is *folklore*, but for a lack of reference we sketch a proof.

Proposition 1. *The ball walk Q_δ is reversible with respect to the uniform distribution μ_Ω and uniformly ergodic.*

The crucial tool for proving this is provided by the notion of small and petite sets, where we refer to [19, Sections 5.2 and 5.5] for details and properties. To this end we introduce a *sampled* chain, say $(Q_\delta)_a$, where a is some probability $a = (a_0, a_1, \ldots)$ on $\{0, 1, 2, \ldots\}$ and $(Q_\delta)_a$ is defined by $(Q_\delta)_a(x, C) := \sum_{j=0}^{\infty} a_j Q_\delta^j(x, C)$. We recall that a (measurable) subset $C \subset \Omega$ is *petite* (for Q_δ), if there are a probability a and a probability measure v on Ω such that

$$(Q_\delta)_a(y, A) \geqslant \varepsilon v(A), \quad A \subset \Omega, \quad y \in C. \tag{21}$$

A set $C \subset \Omega$ is *small*, if the same property holds true for some Dirac probability $a := \delta_n$, such that obviously small sets are petite. We first show that certain balls are small.

Lemma 5. *The sets $B(x, \delta/2) \cap \Omega$, $x \in \Omega$ are small for Q_δ.*

Proof. First, we note that $y \in B(x, \delta/2)$ implies $B(x, \delta/2) \subset B(y, \delta)$. Let $l > 0$ be a lower bound for the local conductance of $Q_{\delta/2}$. Using (20) for $Q_{\delta/2}$, we obtain for any set $A \subset \Omega$ that

$$Q_\delta(y, A) \geqslant Q_\delta(y, A \setminus \{y\}) = \frac{\mathrm{vol}(B(y, \delta) \cap A)}{\mathrm{vol}(B(y, \delta))} \geqslant 2^{-d} \frac{\mathrm{vol}(B(x, \delta/2) \cap A)}{\mathrm{vol}(\delta/2 B^d)}$$
$$\geqslant l \cdot 2^{-d} \frac{\mathrm{vol}(A \cap B(x, \delta/2) \cap \Omega)}{\mathrm{vol}(B(x, \delta/2) \cap \Omega)}.$$

Hence estimate (21) holds true with $n_0 := 1,\ \varepsilon := l \cdot 2^{-d}$ and

$$\nu(A) := \frac{\mathrm{vol}(A \cap B(x, \delta/2) \cap \Omega)}{\mathrm{vol}(B(x, \delta/2) \cap \Omega)}, \quad A \subset \Omega.$$

This completes the proof. □

Proof Proposition 1. We first prove reversibility with respect to μ_Ω. Notice that it is enough to verify (15) for disjoint sets $A, B \subset \Omega$. Furthermore we observe that for any pair $A, B \subset \Omega$ of measurable subsets the characteristic function of the set

$$\{(x, y) \in \Omega \times \Omega, \quad x \in A,\ y \in B,\ \|x - y\| \leqslant \delta\}$$

can equivalently be rewritten as

$$\chi_B(y)\chi_{B(y,\delta)\cap A}(x) \quad \text{or} \quad \chi_A(x)\chi_{B(x,\delta)\cap B}(y).$$

Hence, letting temporarily $c := \mathrm{vol}(\Omega)\,\mathrm{vol}(\delta B^d)$ we obtain

$$\int_A Q_\delta(x, B)\ \mu_\Omega(dx) = \frac{1}{c} \int_A \mathrm{vol}(B(x, \delta) \cap B)\, dx$$
$$= \frac{1}{c} \int_\Omega \int_\Omega \chi_A(x)\chi_{B(x,\delta)\cap B}(y)\, dy\, dx$$
$$= \frac{1}{c} \int_\Omega \int_\Omega \chi_B(y)\chi_{B(y,\delta)\cap A}(x)\, dx\, dy = \int_B Q_\delta(y, A)\mu_\Omega(dy),$$

proving reversibility.

By Lemma 5 each set $B(x, \delta/2) \cap \Omega$ is small, thus also petite. Petiteness is inherited by taking finite unions. Since Ω, being compact, can be covered by finitely many sets $B(x, \delta/2) \cap \Omega$, this implies that Ω is petite. By [19, Theorem 16.2.2] this yields uniform ergodicity of the ball walk (see [19, Theorem 16.0.2(v)]). □

We mention the following conductance bound of the ball walk, which is a slight improvement of [29, Theorem 5.2]. This will be a special case of Theorem 4, below, and we omit the proof.

Proposition 2. *Let (Q_δ, μ_Ω) be the ball walk from above, and let $\varphi(Q_\delta, \mu_\Omega)$ be its conductance. Let D be the diameter of Ω and let l be a lower bound for the local conductance. Then*

$$\varphi(Q_\delta, \mu_\Omega) \geqslant \sqrt{\frac{\pi}{2}} \frac{l^2 \delta}{8D\sqrt{d+1}}. \tag{22}$$

The local conductance may be arbitrarily small if the domain Ω has sharp corners. For specific sets Ω we can explicitly provide lower bounds for the local conductance, and this will be used in the later convergence analysis. In the following we mainly discuss the case $\Omega = B^d$.

We start with a technical result, related to the Gamma function on $\mathbb{R}^+$. We use the well-known formula

$$\mathrm{vol}(B^d) = \pi^{d/2}/\Gamma(d/2+1). \tag{23}$$

Lemma 6. *For any $z > 0$ we have*

$$\frac{\Gamma(z+1/2)}{\Gamma(z)} \leqslant \sqrt{z}. \tag{24}$$

Consequently,

$$\frac{\mathrm{vol}(B^{d-1})}{\mathrm{vol}(B^d)} \leqslant \sqrt{\frac{d+1}{2\pi}}. \tag{25}$$

Proof. By [8, Chapter VII, Eq. (11)] we know that the function $z \mapsto \log\Gamma(z)$ is convex for $z > 0$. Thus we conclude

$$\begin{aligned}\log\Gamma(z+1/2) &\leqslant \tfrac{1}{2}\,(\log\Gamma(z+1) + \log\Gamma(z)) \\ &= \tfrac{1}{2}\,(\log z + 2\log\Gamma(z)) = \log\sqrt{z} + \log\Gamma(z),\end{aligned}$$

from which the proof of assertion (24) can be completed. Using the representation for the volume from (23) and applying the above bound with $z := (d+1)/2$ we obtain

$$\frac{\mathrm{vol}(B^{d-1})}{\mathrm{vol}(B^d)} \leqslant \frac{\Gamma(d/2+1)}{\sqrt{\pi}\Gamma((d+1)/2)} \leqslant \sqrt{\frac{d+1}{2\pi}},$$

and the proof is complete. □

Using Lemma 6, we can prove the following lower bound for the local conductance of the ball walk on B^d.

Lemma 7. *Let (Q_δ, μ_Ω) be the local ball walk on $B^d \subset \mathbb{R}^d$. If $\delta \leqslant 1/\sqrt{d+1}$, then its local conductance obeys $l \geqslant 0.3$.*

Proof. The proof is based on some geometric reasoning. It is clear that the local conductance $l(x)$ is minimal for points x at the boundary of B^d, and in this case its value equals the portion, say $\widetilde{V}$, of the volume of $B(x,\delta)$ inside B^d. If H is the hyperplane at x to B^d, then this cuts off $B(x,\delta)$ exactly one half of its volume. Thus we let $Z(h)$ be the cylinder with base being the $(d-1)$-ball around x in the hyperplane H of radius δ. Its height h is the distance of H to the hyperplane determined by the intersection of $B^d \cap B(x,\delta)$. This height h is exactly determined from the quotient $h/\delta = \delta/2$, by similarity, hence $h := \delta^2/2$. By construction we have $\widetilde{V} \geqslant \frac{1}{2} - \mathrm{vol}(Z(h))/\mathrm{vol}(B(x,\delta))$ and we can lower bound the local conductance $l(x)$ by

$$l(x) \geqslant \frac{1}{2} - \frac{\mathrm{vol}(Z(h))}{\mathrm{vol}(B(x,\delta))}.$$

We can evaluate $\mathrm{vol}(Z(h))$ as $\mathrm{vol}(Z(h)) = h\delta^{d-1}\,\mathrm{vol}(B^{d-1})$, and we obtain

$$l(x) \geqslant \frac{1}{2} - \frac{\delta^{d+1}\,\mathrm{vol}(B^{d-1})}{2\delta^d\, vol(B^d)} = \frac{1}{2}\left(1 - \frac{\delta\,\mathrm{vol}(B^{d-1})}{\mathrm{vol}(B^d)}\right).$$

The bound (25) from Lemma 6 implies

$$l(x) \geqslant \frac{1}{2}\left(1 - \frac{\delta\sqrt{d+1}}{\sqrt{2\pi}}\right).$$

For $\delta \leqslant 1/(\sqrt{d+1})$ we get $l(x) \geqslant \frac{1}{2}(1 - 1/\sqrt{2\pi}) \geqslant 0.3$, completing the proof. □

We close this subsection with the following technical lemma, which can be extracted from the unpublished seminar note [28]. For the convenience of the reader we present its proof. In addition we will slightly improve the statement.

Lemma 8. *Let $l > 0$ be a lower bound for the local conductance of the ball walk (Q_δ, μ_Ω). For any $0 < t < l$ and any set $A \subset \Omega$ with related sets*

$$A_1 := \left\{x \in A, \quad Q_\delta(x, A^c) < \frac{l-t}{2}\right\} \subset A, \tag{26}$$

$$A_2 := \left\{y \in A^c, \quad Q_\delta(y, A) < \frac{l-t}{2}\right\} \subset A^c, \tag{27}$$

we have $d(A_1, A_2) > t\delta\sqrt{2\pi/(d+1)}$.

For its proof we need the following:

Lemma 9. *Let $\delta > 0$. If $x, y \in \mathbb{R}^d$ are two points with distance $t\delta\sqrt{2\pi/(d+1)}$ at most, then*

$$\mathrm{vol}(B(x,\delta) \cap B(y,\delta)) \geqslant (1-t)\,\mathrm{vol}(\delta B^d). \tag{28}$$

Proof. Let $u := \|x - y\|_2$. If $u < \delta$ then the volume of the intersection of $B(x,\delta)$ and $B(y,\delta)$ is exactly the same as the volume of the ball δB^d minus the volume of the middle slice with distance u as thickness. The volume of this slice is bounded from above by the volume of the cylinder with base δB^{d-1} and thickness u. Thus we obtain

$$\mathrm{vol}(B(x,\delta) \cap B(y,\delta)) \geqslant \mathrm{vol}(\delta B^d) - u\,\mathrm{vol}(\delta B^{d-1}) = \mathrm{vol}(\delta B^d)\left(1 - u\frac{\mathrm{vol}(\delta B^{d-1})}{\mathrm{vol}(\delta B^d)}\right).$$

Applying Lemma 6 we obtain

$$\frac{\mathrm{vol}(\delta B^{d-1})}{\mathrm{vol}(\delta B^d)} = \frac{\mathrm{vol}(B^{d-1})}{\delta\,\mathrm{vol}(B^d)} \leqslant \frac{1}{\delta}\sqrt{\frac{d+1}{2\pi}},$$

thus by the choice of $u \leqslant \sqrt{2\pi}t\delta/\sqrt{d+1}$ we conclude that

$$u\frac{\mathrm{vol}(\delta B^{d-1})}{\mathrm{vol}(\delta B^d)} \leqslant \frac{\sqrt{2\pi}t\delta\sqrt{d+1}}{\delta\sqrt{2\pi}\sqrt{d+1}} \leqslant t,$$

and the proof is complete. □

We turn to the

Proof of Lemma 8. Let $x \in A_1$ and $y \in A_2$ be in Ω, and suppose that their distance is at most $t\delta\sqrt{2\pi/(d+1)}$. Simple set theoretic reasoning shows that

$$\begin{aligned}\mathrm{vol}(B(x,\delta) \cap B(y,\delta) \cap \Omega) &\geqslant \mathrm{vol}(B(x,\delta) \cap \Omega) - \mathrm{vol}(B(x,\delta) \setminus B(y,\delta)) \\ &\geqslant \mathrm{vol}(B(x,\delta) \cap \Omega) - \mathrm{vol}(B(x,\delta) \setminus (B(x,\delta) \cap B(y,\delta))) \\ &= \mathrm{vol}(B(x,\delta) \cap \Omega) - \mathrm{vol}(\delta B^d) + \mathrm{vol}(B(x,\delta) \cap B(y,\delta)).\end{aligned}$$

Since l is a lower bound for the conductance $l(x)$ we have that

$$\mathrm{vol}(B(x,\delta)\cap\Omega)\geqslant l\,\mathrm{vol}(B(x,\delta))=l\,\mathrm{vol}(\delta B^d).$$

Taking this into account and using (28) we end up with

$$\begin{aligned}\mathrm{vol}(B(x,\delta)\cap B(y,\delta)\cap\Omega)&\geqslant l\,\mathrm{vol}(\delta B^d)-\mathrm{vol}(\delta B^d)+(1-t)\,\mathrm{vol}(\delta B^d)\\&=(l-t)\,\mathrm{vol}(\delta B^d).\end{aligned}$$

In probabilistic terms this rewrites as $Q_\delta(x, B(x,\delta)\cap B(y,\delta)\cap\Omega)\geqslant l-t$, and similarly $Q_\delta(y, B(x,\delta)\cap B(y,\delta)\cap\Omega)\geqslant l-t$. Now, if $A\subset\Omega$ is any measurable subset with complement A^c then for $x\in A$ and $y\in A^c$ we obtain

$$B(x,\delta)\cap B(y,\delta)\cap\Omega\subset\left(B(x,\delta)\cap A^c\cap\Omega\right)\bigcup\left(B(y,\delta)\cap A\cap\Omega\right),$$

which in turn yields $Q_\delta(x,A^c)+Q_\delta(y,A)\geqslant l-t$, but this contradicts the definition of the sets A_1 and A_2. Hence any two points from A_1 and A_2, respectively, must have distance larger than $t\delta\sqrt{2\pi/(d+1)}$, and the proof is complete. □

Properties of the related Metropolis method. We analyze Metropolis Markov chains which are based on the ball walk, introduced above, for some appropriately chosen δ. As it will turn out, the related Metropolis chains are *perturbations* of the underlying ball walk, and its properties, as established in Propositions 1 and 2 extend in a natural way.

For $\varrho\in\mathcal{R}^\alpha(\Omega)$ we define the *acceptance probabilities* as

$$\theta(x,y):=\min\left\{1,\frac{\varrho(y)}{\varrho(x)}\right\}. \tag{29}$$

The corresponding Metropolis kernel is given by

$$K_{\varrho,\delta}(x,dy):=\theta(x,y)Q_\delta(x,dy)+(1-\int\theta(x,y)Q_\delta(x,dy))\delta_x(dy). \tag{30}$$

Note that for $x\notin A$ we obtain

$$K_{\varrho,\delta}(x,A)=\int_A\theta(x,y)Q_\delta(x,dy)=\frac{1}{\mathrm{vol}(\delta B^d)}\int_{A\cap B(x,\delta)}\theta(x,y)\,dy.$$

Below we sketch a single Metropolis **Step** from the present position $x\in\Omega$ with kernel $K_{\varrho,\delta}(x,\cdot)$ (Fig. 3). The procedure **Ball-walk-step** was described in Fig. 2.

We start with the following observation.

Lemma 10. *Let α be the Lipschitz constant in $\mathcal{R}^\alpha(\Omega)$ and $\beta:=\exp(-\alpha\delta)$. Uniformly for $\varrho\in\mathcal{R}^\alpha(\Omega)$ the following bound for the related Metropolis chain holds true*:

$$K_{\varrho,\delta}(x,dy)\geqslant\beta Q_\delta(x,dy). \tag{31}$$

Proof. Let $A\subset\Omega$. If $\mathrm{dist}(x,A)>\delta$ then there is nothing to prove. Otherwise, for $y\in A\cap B(x,\delta)$ we find from (6) and (29) that

$$\theta(x,y)\geqslant\exp(-\alpha\|x-y\|_2)\geqslant e^{-\alpha\delta}=\beta.$$

Procedure `Metropolis-step`(x, ϱ, δ)

Input : current position x, $\delta > 0$, function ϱ;
Output: next position;
Propose: $y := \textbf{Ball-walk-step}(x, \delta)$;
Accept:
if $\varrho(y) \geq \varrho(x)$ **then**
  return y
else if $\varrho(y) \geq \textbf{rand}() \cdot \varrho(x)$ **then**
  return y
else
  return x
end

Fig. 3. Schematic view of the Metropolis step. Note that the Acceptance step results in an acceptance probability of $\theta(x, y) = \min\{1, \varrho(y)/\varrho(x)\}$.

By definition of the transition kernel $K_{\varrho,\delta}$ from (30) we can use β to bound

$$K_{\varrho,\delta}(x, A) \geqslant \min\{\theta(x, y),\ y \in A \cap B(x, \delta)\}\, Q_\delta(x, A) \geqslant \beta Q_\delta(x, A).$$

The proof is complete. □

The assertion of Proposition 1 extends to the family of Metropolis chains as follows.

Proposition 3 (*cf. Mathé [18, Proposition 1]*). *Let Q_δ be the ball walk from* (19) *on Ω. For each $\varrho \in \mathcal{R}^\alpha(\Omega)$ and $\delta \leqslant D$ the corresponding Metropolis chains from* (30) *are uniformly ergodic and reversible with respect to the related μ_ϱ.*

Proof. Reversibility with respect to μ_ϱ is clear by the choice of the function θ. To prove uniform ergodicity, let β be from Lemma 10 and c from (16). As established in Lemma 10 we have $K_{\varrho,\delta}(x, dy) \geqslant \beta Q_\delta(x, dy)$. It is easy to see, and was established in [18, Proof of Theorem 2], that this extends to all iterates as

$$K^n_{\varrho,\delta}(x, dy) \geqslant \beta^n Q^n_\delta(x, dy).$$

Recall that under the assumptions made, the ball walk is uniformly ergodic, and from Proposition 1 we obtain n_0 such that for all $x \in \Omega$ we have

$$K^{n_0}_{\varrho,\delta}(x, A) \geqslant \beta^{n_0} c v(A), \quad A \subset \Omega, \tag{32}$$

proving uniform ergodicity. □

Remark 7. Notice that (32) is obtained with right-hand side *uniformly* for all $\varrho \in \mathcal{R}^\alpha(\Omega)$, a fact which will prove useful later.

Finally we prove lower bounds for the conductance of the Metropolis chains.

Theorem 4. *Let $(K_{\varrho,\delta}, \mu_\varrho)$ be the Metropolis chain based on the local ball walk (Q_δ, μ_Ω) and let $\varphi(K_{\varrho,\delta}, \mu_\varrho)$ be its conductance, where $\varrho \in \mathcal{R}^\alpha(\Omega)$. Let l be a lower bound for the local conductance of Q_δ. For $\varrho \in \mathcal{R}^\alpha(\Omega)$ we have*

$$\varphi(K_{\varrho,\delta}, \mu_\varrho) \geqslant \frac{l e^{-\alpha\delta}}{8} \min\left\{\sqrt{\frac{\pi}{2}} \frac{l\delta}{D\sqrt{d+1}}, 1\right\}, \tag{33}$$

where D is the diameter of Ω.

Remark 8. As mentioned above, Proposition 2 is a special case of Theorem 4 for $\alpha = 0$.

The proof of Theorem 4 will be based on Lemma 8 for the underlying ball walk, specifying $t := l/2$. This extends to the Metropolis walk as follows.

Lemma 11. *Let α from* (6) *and l be the local conductance of the ball walk. We let $\beta := \exp(-\alpha\delta)$. For $A \subset \Omega$ we assign*

$$T_1 := \left\{ x \in A, \quad K_{\varrho,\delta}(x, A^c) < \frac{\beta l}{4} \right\} \subset A, \tag{34}$$

$$T_2 := \left\{ y \in A^c, \quad K_{\varrho,\delta}(y, A) < \frac{\beta l}{4} \right\} \subset A^c. \tag{35}$$

Then $d(T_1, T_2) > \delta l\sqrt{\pi/(2d+2)}$.

Proof. It is enough to prove $T_1 \subset A_1$ and $T_2 \subset A_2$. If $x \in T_1$ then Lemma 10 implies $K_{\varrho,\delta}(x, A^c) < \beta l/4$, hence

$$Q_\delta(x, A^c) \leqslant \frac{1}{\beta} K_{\varrho,\delta}(x, A^c) \leqslant \frac{l}{4}.$$

The other inclusion is proved similarly. □

We turn to the

Proof of Theorem 4. Let $A \subset \Omega$ be the set for which the conductance is attained. We assign sets T_1 and T_2 as in Lemma 11 and distinguish two cases. If $\mu_\varrho(T_1) < \mu_\varrho(A)/2$ *or* $\mu_\varrho(T_2) < \mu_\varrho(A^c)/2$, then the estimate (33) follows easily. For instance, if $\mu_\varrho(T_1) < \mu_\varrho(A)/2$ then

$$\begin{aligned} \int_A K_{\varrho,\delta}(x, A^c)\mu_\varrho(dx) &\geqslant \int_{A\setminus T_1} K_{\varrho,\delta}(x, A^c)\mu_\varrho(dx) \\ &\geqslant \frac{\beta l}{4}\mu_\varrho(A \setminus T_1) \geqslant \frac{\beta l}{8}\mu_\varrho(A) \geqslant \frac{\beta l}{8} \min\left\{\mu_\varrho(A), \mu_\varrho(A^c)\right\}, \end{aligned}$$

thus $\varphi(K_{\varrho,\delta}, \mu_\varrho) \geqslant \beta l/8$ in this case, which proves (33).

Otherwise we have $\mu_\varrho(T_1) \geqslant \mu_\varrho(A)/2$ *and* $\mu_\varrho(T_2) \geqslant \mu_\varrho(A^c)/2$. In this case we apply an isoperimetric inequality, see [29, Theorem 4.2] to the triple (T_1, T_2, T_3) with $T_3 := \Omega \setminus (T_1 \cup T_2)$ to conclude that

$$\mu_\varrho(T_3) \geqslant \frac{2d(T_1, T_2)}{D} \min\left\{\mu_\varrho(T_1), \mu_\varrho(T_2)\right\}, \tag{36}$$

hence under the size constraints in this case it holds true that

$$\mu_\varrho(T_3) \geqslant \frac{d(T_1, T_2)}{D} \min\left\{\mu_\varrho(A), \mu_\varrho(A^c)\right\}. \tag{37}$$

Using the reversibility of the Metropolis chain $(K_{\varrho,\delta}, \mu_\varrho)$ we have

$$\int_A K_{\varrho,\delta}(x, A^c)\mu_\varrho(dx) = \int_{A^c} K_{\varrho,\delta}(y, A)\mu_\varrho(dy),$$

which implies

$$\begin{aligned}\int_A K_{\varrho,\delta}(x, A^c)\mu_\varrho(dx) &= \frac{1}{2}\left(\int_A K_{\varrho,\delta}(x, A^c)\mu_\varrho(dx) + \int_{A^c} K_{\varrho,\delta}(y, A)\mu_\varrho(dy)\right)\\ &\geqslant \frac{1}{2}\left(\int_{A\cap T_3} K_{\varrho,\delta}(x, A^c)\mu_\varrho(dx) + \int_{A^c\cap T_3} K_{\varrho,\delta}(y, A)\mu_\varrho(dy)\right)\\ &\geqslant \frac{1}{2}\left(\frac{\beta l}{4}\mu_\varrho(A\cap T_3) + \frac{\beta l}{4}\mu_\varrho(A^c\cap T_3)\right)\\ &= \frac{\beta l}{8}\left(\mu_\varrho(A\cap T_3) + \mu_\varrho(A^c\cap T_3)\right) = \frac{\beta l}{8}\mu_\varrho(T_3).\end{aligned}$$

Since by Lemma 11 we can bound $d(T_1, T_2)\geqslant \delta l\sqrt{\pi/(2d+2)}$ we use (37) to complete the proof. □

If we restrict ourselves to Metropolis chains on B^d, then Lemma 7 provides a lower bound for the local conductance which is independent of the dimension d. As a simple consequence of Theorem 4 we then obtain the following:

Corollary 1. *Assume that $\varrho \in \mathcal{R}^\alpha(B^d)$ and $\delta\leqslant(d+1)^{-1/2}$. Then we obtain*

$$\varphi(K_{\varrho,\delta}, \mu_\varrho)\geqslant\sqrt{\frac{\pi}{2}}\frac{9\delta}{1600\sqrt{d+1}}e^{-\alpha\delta}.$$

To maximize φ we define $\delta^* = \min\left\{1/\sqrt{d+1}, 1/\alpha\right\}$ and obtain

$$\varphi(K_{\varrho,\delta^*}, \mu_\varrho)\geqslant 0.0025\frac{1}{\sqrt{d+1}}\min\left\{\frac{1}{\sqrt{d+1}}, \frac{1}{\alpha}\right\}.$$

Error bounds. For the class $\mathcal{F}^\alpha(\Omega)$ the above lower conductance bound (33) will yield an error estimate for the problem (2).

Let S_n^δ be the estimator based on a sample of the local Metropolis Markov chain with transition $K_{\varrho,\delta}$, starting at zero. To estimate its error we combine the estimates of the conductance of $K_{\varrho,\delta}$ with two results, partially known from the literature. To formulate the results we note the following. The Markov kernel $K_{\varrho,\delta}$ is reversible with respect to μ_ϱ and hence induces a self-adjoint operator

$$K_{\varrho,\delta} : L_2(\Omega, \mu_\varrho) \to L_2(\Omega, \mu_\varrho).$$

The spectrum $\sigma(K_{\varrho,\delta})$ is contained in $[-1, 1]$ and $1 \in \sigma(K_{\varrho,\delta})$ and we are interested in the second largest eigenvalue

$$\beta_{\varrho,\delta} := \sup\{\sigma \in \sigma(K_{\varrho,\delta})|\sigma \neq 1\}$$

of $K_{\varrho,\delta}$. This is motivated by the extension of a result from [18, Corollary 1] about the worst case error of S_n^δ, uniformly for $(f, \varrho) \in \mathcal{F}^\alpha(\Omega)$.

Lemma 12.

$$\lim_{n\to\infty}\sup_{(f,\varrho)\in\mathcal{F}^\alpha(\Omega)} e(S_n^\delta, (f, \varrho))^2\cdot n = \sup_{\varrho\in\mathcal{R}^\alpha(\Omega)}\frac{1+\beta_{\varrho,\delta}}{1-\beta_{\varrho,\delta}}.$$

The proof is given in the Appendix. For Markov chains which start according to the invariant distribution μ_ϱ the bound is similar, but more explicit and was given in [26] and [16, Theorem 1.9].

The relation of the second largest eigenvalue $\beta_{\varrho,\delta}$ to the conductance is given in

Lemma 13 (*Cheeger's Inequality, see [12,15,16]*).

$$\lambda_{\varrho,\delta} := 1 - \beta_{\varrho,\delta} \geqslant \varphi^2(K_{\varrho,\delta}, \mu_\varrho)/2.$$

We are ready to state our main result for the Metropolis algorithm S_n^δ, based on the Markov chain $K_{\varrho,\delta}$, for the class $\mathcal{F}^\alpha(B^d)$, i.e., when $\Omega \subset \mathbb{R}^d$ is the Euclidean unit ball.

Theorem 5. *Let $S_n^\delta = 1/n \sum_{j=1}^n f(X_j)$ be the estimator based on a sample $(X_1, \ldots, X_n)$ of the local Metropolis Markov chain with transition $K_{\varrho,\delta}$, where $\delta \leqslant (d+1)^{-1/2}$. Then*

$$\lim_{n\to\infty} \sup_{(f,\varrho)\in\mathcal{F}^\alpha(B^d)} e(S_n^\delta, (f,\varrho))^2 \cdot n \leqslant \frac{8 \cdot 1600^2}{81\pi}(d+1) \cdot \frac{e^{2\alpha\delta}}{\delta^2}. \tag{38}$$

Again we may choose $\delta^* = \min\left\{(d+1)^{-1/2}, \alpha^{-1}\right\}$ and obtain

$$\lim_{n\to\infty} \sup_{(f,\varrho)\in\mathcal{F}^\alpha(B^d)} e(S_n^{\delta^*}, (f,\varrho))^2 \cdot n \leqslant 594700 \cdot (d+1) \max\left\{d+1, \alpha^2\right\}. \tag{39}$$

Proof. This follows from Corollary 1, and Lemmas 12 and 13. □

5. Summary

Let us discuss our findings. The results from Section 3 clearly indicate that the superiority of Metropolis algorithms upon simpler (non-adaptive) Monte Carlo methods does not hold in general. Specifically, it does not hold for the large classes $\mathcal{F}_C(\Omega)$ of input without additional structure.

On the other hand, for the class $\mathcal{F}^\alpha(B^d)$, specific Metropolis algorithms that are based on local underlying walks are superior to all non-adaptive methods. Even more, on B^d the cost of the algorithm $S_n^{\delta^*}$, roughly given by the number n of evaluations of ϱ and f, increases like a polynomial in d and α. More precisely, according to (39), the asymptotic constant $\lim_{n\to\infty} e(S_n^{\delta^*}, \mathcal{F}^\alpha(B^d))^2 \cdot n$ is bounded by a constant times $\max\left\{d^2, d\alpha^2\right\}$, i.e., the complexity grows polynomially in d and α and, for fixed d, increases (at most) as α^2. If we only allow non-adaptive methods then this asymptotic constant, again for fixed d, increases at least as α^d, see (14).

We believe that this problem is *tractable* in the sense that the number of function values to achieve an error ε can be bounded by

$$n(\varepsilon, \mathcal{F}^\alpha(B^d)) \leqslant C\varepsilon^{-2}\, d \max(d, \alpha^2). \tag{40}$$

We did not prove (40), however, since Theorem 5 is only a statement for large n.

Notice that according to Theorem 5 the size δ^* of the underlying balls walk needs to be adjusted both to the spatial dimension d and the Lipschitz constant α.

The analysis of the Metropolis algorithm is based on properties of the underlying ball walk; in particular we establish uniform ergodicity of the ball walk for convex bodies $\Omega \subset \mathbb{R}^d$. Also, based on conductance arguments, we provide lower bounds for the spectral gap of the ball walk.

As a consequence, in the case $\alpha = 0$ the estimate (38) provides an error bound for the ball walk (Q_δ, μ), which is asymptotically of the form $e(S_n^\delta, L_2(B^d, \mu)) \leqslant C\delta^{-1}(d/n)^{1/2}$.

The results extend in a similar way to any family $\Omega_d \subset \mathbb{R}^d$ for which the underlying local ball walk Q_δ has (for $\delta \leqslant \delta_d$) a non-trivial lower bound for the local conductance that is independent of the dimension.

Finally, from the results of Section 3 we can conclude that adaption does not help much for the classes $\mathcal{F}_C(\Omega)$. Hence we have new results concerning the *power of adaption*, see [22] for a survey of earlier results, in particular that it may help to break the *curse of dimensionality* for the classes $\mathcal{F}^\alpha(B^d)$.

Acknowledgment

We thank two anonymous referees and Daniel Rudolf for their comments.

Appendix A. Proof of Lemma 12

Lemma 12 extends the bound from [18, Theorem 1], which deals with a single uniformly ergodic chain. It was obtained from on a contraction property, as stated in [18, Proposition 1]. The goal of the present analysis is to establish this asymptotic result *uniformly* for all Metropolis chains with density from $\mathcal{R}^\alpha(\Omega)$, by showing that this contractivity holds true uniformly.

Contractivity of the Markov operator. We assign to each transition kernel K on Ω with corresponding invariant distribution μ the bounded linear mapping P, given by

$$(Pf)(x) := \int f(y)K(x, dy). \tag{41}$$

Also we let E denote the mapping which assigns any integrable function its expectation as a constant function $E(f) := \int_\Omega f(x)\mu(dx)$. For each K the mapping $P - E$ is bounded in $L_\infty(\Omega, \mu)$, with norm less than or equal to one and we shall strengthen this uniformly for kernels $K_{\varrho,\delta}$ with $\varrho \in \mathcal{R}^\alpha(\Omega)$. Within this operator context *uniform ergodicity* is equivalent to a specific form of quasi-compactness, namely there are $0 < \eta < 1$ and $n_0 \in \mathbb{N}$ for which

$$\|P^n - E : L_\infty(\Omega) \to L_\infty(\Omega)\| \leqslant \eta \quad \text{for } n \geqslant n_0. \tag{42}$$

We first show that reversibility allows to transfer this to the spaces $L_1(\Omega, \mu_\varrho)$.

Lemma 14. *Suppose that the transition kernel K with corresponding mapping P is reversible. Then for all $n \in \mathbb{N}$ we have*

$$\|P^n - E : L_1(\Omega, \mu) \to L_1(\Omega, \mu)\| \leqslant \|P^n - E : L_\infty(\Omega, \mu) \to L_\infty(\Omega, \mu)\|. \tag{43}$$

Proof. If K is reversible, then so are all iterates K^n. Thus for arbitrary functions $f \in L_1(\Omega, \mu)$ and $h \in L_\infty(\Omega, \mu)$ we have, using the scalar product on $L_2(\Omega, \mu)$, that

$$\langle (P^n - E)f, h\rangle = \langle f, (P^n - E)h\rangle.$$

Consequently, for any $f \in L_1(\Omega, \mu)$ we have

$$\|(P^n - E)f\|_1 = \sup_{\|h\|_\infty \leqslant 1} \left|\langle (P^n - E)f, h\rangle\right| = \sup_{\|h\|_\infty \leqslant 1} \left|\langle f, (P^n - E)h\rangle\right| \leqslant \|f\|_1 \sup_{\|h\|_\infty \leqslant 1} \|(P^n - E)h\|_\infty,$$

from which the proof can be completed. □

Proposition 4. *For any convex body $\Omega \subset \mathbb{R}^d$ there are an integer n_0 and a constant $0 < \eta < 1$ such that uniformly for $\varrho \in \mathcal{R}^\alpha(\Omega)$ we have*

$$\|P^{n_0}_{\varrho,\delta} - E\colon L_1(\Omega, \mu_\varrho) \to L_1(\Omega, \mu_\varrho)\| \leqslant \eta. \tag{44}$$

Proof. This is an immediate consequence of the bound (32). As mentioned in Remark 7 uniform ergodicity was established uniformly for $\varrho \in \mathcal{R}^\alpha(\Omega)$. It is well known (see [19, Theorem 16.2.4]) that this implies that there is an $\eta < 1$ such that uniformly for $\varrho \in \mathcal{R}^\alpha(\Omega)$ we have

$$\|P^{n_0}_{\varrho,\delta} - E\colon L_\infty(\Omega) \to L_\infty(\Omega)\| \leqslant \eta \quad \text{for } n \geqslant n_0. \tag{45}$$

In the light of Lemma 14 this yields (44). □

Finally we sketch the

Proof of Lemma 12. Using Proposition 4 we can extend the proof of [18, Theorem 1]. In particular, the bounds from Eqs. (13)–(15) in [18] tend to zero uniformly for $\varrho \in \mathcal{R}^\alpha(\Omega)$. Moreover, starting at zero, after one step according to the underlying ball walk, the (new) initial distribution is uniformly bounded with respect to the uniform distribution on Ω, hence also with respect to μ_ϱ, such that we establish the asymptotics in Lemma 12. □

References

[1] C. Andrieu, É. Moulines, On the ergodicity properties of some adaptive MCMC algorithms, Ann. Appl. Probab. 16 (3) (2006) 1462–1505.

[2] D. Applegate, R. Kannan, Sampling and integration of near log-concave functions, in: STOC '91: Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, ACM Press, New York, NY, USA, 1991, pp. 156–163.

[3] Y.F. Atchadé, J.S. Rosenthal, On adaptive Markov chain Monte Carlo algorithms, Bernoulli 11 (5) (2005) 815–828.

[4] N.S. Bahvalov, Approximate computation of multiple integrals, Vestnik Moskov. Univ. Ser. Mat. Meh. Astr. Fiz. Him. 1959 (4) (1959) 3–18.

[5] F. Bassetti, P. Diaconis, Examples comparing importance sampling and the Metropolis algorithm, Illinois J. Math. 50 (2006) 67–91.

[6] I. Beichl, F. Sullivan, The Metropolis algorithm, Comput. Sci. Eng. 2 (1) (2000) 65–69.

[7] I. Beichl, F. Sullivan, Guest editors' introduction: Monte Carlo methods, Comput. Sci. Eng. 8 (2) (2006) 7–8.

[8] N. Bourbaki, Functions of a real variable, Elements of Mathematics (Berlin), Springer, Berlin, 2004.

[9] V.I. Burenkov, Sobolev Spaces on Domains, Teubner-Texte zur Mathematik, vol. 137, Teubner Verlag Stuttgart, 1998.

[10] A. Frieze, R. Kannan, N. Polson, Sampling from log-concave distributions, Ann. Appl. Probab. 4 (3) (1994) 812–837.

[11] E. Hlawka, Ausfüllung und Überdeckung konvexer Körper durch konvexe Körper, Monatsh. Math. Phys. 53 (1949) 81–131.

[12] M. Jerrum, A. Sinclair, Approximating the permanent, SIAM J. Comput. 18 (6) (1989) 1149–1178.

[13] R. Kannan, L. Lovász, M. Simonovits, Isoperimetric problems for convex bodies and a localization lemma, Discrete Comput. Geom. 13 (3–4) (1995) 541–559.

[14] U. Krengel, Ergodic theorems, de Gruyter Studies in Mathematics, vol. 6, Walter de Gruyter & Co., Berlin, 1985.
[15] G.F. Lawler, A.D. Sokal, Bounds on the L^2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality, Trans. Amer. Math. Soc. 309 (2) (1988) 557–580.
[16] L. Lovász, M. Simonovits, Random walks in a convex body and an improved volume algorithm, Random Structures Algorithms 4 (4) (1993) 359–412.
[17] P. Mathé, The optimal error of Monte Carlo integration, J. Complexity 11 (4) (1995) 394–415.
[18] P. Mathé, Numerical integration using Markov chains, Monte Carlo Methods Appl. 5 (4) (1999) 325–343.
[19] S.P. Meyn, R.L. Tweedie, Markov Chains and Stochastic Stability, Springer, London, 1993.
[20] E. Novak, Deterministic and stochastic error bounds in numerical analysis, Lecture Notes in Mathematics, vol. 1349, Springer, Berlin, 1988.
[21] E. Novak, The real number model in numerical analysis, J. Complexity 11 (1) (1995) 57–73.
[22] E. Novak, On the power of adaption, J. Complexity 12 (3) (1996) 199–237.
[23] D. Randall, Rapidly mixing Markov chains with applications in computer science and physics, Comput. Sci. Eng. 8 (2) (2006) 30–41.
[24] G.O. Roberts, R.L. Tweedie, Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms, Biometrika 83 (1) (1996) 95–110.
[25] C.A. Rogers, Packing and covering, Cambridge Tracts in Mathematics and Mathematical Physics, No. 54, Cambridge University Press, New York, 1964.
[26] A. Sokal, Monte Carlo methods in statistical mechanics: foundations and new algorithms, in: Functional integration (Cargèse, 1996), Plenum, New York, 1997, pp. 131–192
[27] J.F. Traub, G.W. Wasilkowski, H. Woźniakowski, Information-based complexity, Academic Press Inc., Boston, MA, 1988 with contributions by A.G. Werschulz and T. Boult.
[28] S. Vempala, Lecture 17, Random walks and polynomial time algorithms, ⟨http://www-math.mit.edu/~vempala/random/course.html⟩, 2002.
[29] S. Vempala, Geometric random walks: a survey, Combinatorial and computational geometry, Math. Sci. Res. Inst. Publ., vol. 52, Cambridge University Press, Cambridge, 2005, pp. 577–616.