DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Error Bounds for Computing the Expectation by Markov Chain Monte Carlo

D. Rudolf

Preprint 21

Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Error Bounds for Computing the Expectation by Markov Chain Monte Carlo

D. Rudolf

Preprint 21

ERROR BOUNDS FOR COMPUTING THE EXPECTATION BY MARKOV CHAIN MONTE CARLO

DANIEL RUDOLF

ABSTRACT. We study the error of reversible Markov chain Monte Carlo methods for approximating the expectation of a function. Explicit error bounds with respect to the l_2-, l_4- and l_∞-norm of the function are proven. By the estimation the well known asymptotical limit of the error is attained, i.e. there is no gap between the estimate and the asymptotical behavior. We discuss the dependence of the error on a burn-in of the Markov chain. Furthermore we suggest and justify a specific burn-in for optimizing the algorithm.

1. INTRODUCTION

We start with a probability distribution π on a finite set D and a function $f : D \to \mathbb{R}$. The goal is to compute the expectation denoted by

$$S(f) = \sum_{x \in D} f(x)\pi(x).$$

Let the cardinality of D be very large such that an exact computation of the sum is practically impossible. Furthermore suppose that the desired distribution is not explicitly given, i.e. we have no random number generator for π available. Such kind of problems arise in statistical physics, in statistics, and in financial mathematics (see for instance [GRS96, Liu08]). The idea of approximating $S(f)$ via Markov chain Monte Carlo (MCMC) is the following: Run a Markov chain on D to simulate the distribution π and compute the time average over the last n steps. Let $X_1, \dots, X_{n+n_0}$ be the chain, then we obtain as approximation

$$S_{n,n_0}(f) = \frac{1}{n} \sum_{i=1}^{n} f(X_{i+n_0}).$$

Date: Version: July 28, 2009.

Key words and phrases. Markov chain Monte Carlo methods, Markov chain Monte Carlo, error bounds, explicit error bounds, burn-in, mixing time, eigenvalue.

By n_0 the so called burn-in is given, loosely spoken this is the number of time steps taken to warm up. Afterwards the distribution of the generated Markov chain is (hopefully) close to the stationary one.

A Markov chain is identified with its initial distribution ν and its transition matrix P. We restrict ourself to ergodic chains, i.e. the second largest absolute value β of the eigenvalues of P is smaller than one. It is well known that the distribution of these chains reaches stationarity exponentially (see [Bré99, RR97, LPW09]).

The error of S_{n,n_0} for $f \in \mathbb{R}^D$ is measured by

$$e_\nu(S_{n,n_0}, f) = \left(\mathbf{E}_{\nu,P} \left|S_{n,n_0}(f) - S(f)\right|^2\right)^{1/2},$$

where $\mathbf{E}_{\nu,P}$ denotes the expectation of the Markov chain. The asymptotic behavior of the integration error can be written in terms of the eigenvalues and eigenfunctions of P. It holds true that

$$\lim_{n\to\infty} n \cdot e_\nu(S_{n,n_0}, f)^2 \leq \frac{1+\beta_1}{1-\beta_1} \|f\|_2^2,$$

where β_1 is the second largest eigenvalue (see [Sok97, Mat99]). The constant $\frac{1+\beta_1}{1-\beta_1}$ is optimal but this statement does not give an error bound for finite n and also does not include anything concerning the choice of n_0. How does an explicit error bound of the MCMC method look like where the asymptotic behavior is attained?

Let us give an outline of the structure and the main results. Section 2 contains the used notation and presents some relevant statements concerning Markov chains. Section 3 contains the new results. The explicit error bound is developed with respect to the l_2-, l_4- and l_∞-norm of the function f. For $\|f\|_\infty \leq 1$ and $C = 2\sqrt{\left\|\frac{\nu}{\pi} - 1\right\|_\infty}$ we obtain the following. The error obeys

$$e_\nu(S_{n,n_0}, f)^2 \leq \frac{2}{n(1-\beta_1)} + \frac{2C\beta^{n_0}}{n^2(1-\beta)^2}.$$

For details and estimates concerning l_2 and l_4 we refer to Theorem 11 in Section 3.3. In Section 4 it turns out that $n_0 = \max\left\{\left\lceil \frac{\log(C)}{\log(\beta^{-1})} \right\rceil, 0\right\}$ is a reasonable choice for the burn-in. Then the error bound simplifies to

$$e_\nu(S_{n,n_0}, f)^2 \leq \frac{2}{n(1-\beta_1)} + \frac{2}{n^2(1-\beta)^2}.$$

In many examples a good estimate for β can be achieved, see for instance [MR02, BD06, BL07]. Therefore it is straightforward to apply the explicit error bound.

2. Preliminaries

The Markov chain $X_1, X_2, \dots$ is a stochastic process with state space D. We identify it with initial distribution ν and transition matrix $P = (p(x,y))_{x,y\in D}$ and denote it by (ν, P). For $x, y \in D$ the entry $p(x,y)$ presents the probability of jumping from state x to state y in one step of the chain.

By $Pf(x) = \sum_{y\in D} p(x,y)f(y)$ we obtain the expectation of the value of $f \in \mathbb{R}^D$ after one step of the chain starting from $x \in D$. The expectation after k steps of the Markov chain from x is given by $P^k f(x) = \sum_{y\in D} p^k(x,y)f(y)$, where $P^k = (p^k(x,y))_{x,y\in D}$ denotes the k-th power of P. Similarly we consider the application of P to a distribution ν, i.e. $\nu P(x) = \sum_{y\in D} \nu(y)p(y,x)$. This is the distribution which arises after one step where the initial state was chosen by ν. The distribution which arises after k steps is given by $\nu P^k(x) = \sum_{y\in D} \nu(y)p^k(y,x)$.

The expectation $\mathbf{E}_{\nu,P}$ of the Markov chain $X_1, \dots, X_{n+n_0}$ is taken with respect to the probability measure

$$W_{\nu,P}(x_1,\dots,x_{n+n_0}) = \nu(x_1)p(x_1,x_2)\cdot\dots\cdot p(x_{n+n_0-1},x_{n+n_0})$$

on D^{n+n_0}. Using this for $i \le j$ we obtain a characterization by the transition matrix

$$\mathbf{E}_{\nu,P}(f(X_i)f(X_j)) = \sum_{x\in D} P^i(fP^{j-i}f)(x)\nu(x). \tag{1}$$

2.1. **Reversibility and spectral structure.** We call the Markov chain with transition matrix P, or simply P, reversible with respect to a probability measure π if the detailed balance condition

$$\pi(x)p(x,y) = \pi(y)p(y,x)$$

holds true for $x, y \in D$. If P is reversible, then π is a so called stationary distribution of the Markov chain, i.e. $\pi P(x) = \pi(x)$. Note that, if P is reversible then P^k is also reversible. Let us define the weighted scalar-product

$$\langle f, g\rangle_\pi = \sum_{x\in D} f(x)g(x)\pi(x),$$

for functions $f, g \in \mathbb{R}^D$. Then let $\|f\|_2 = \langle f, f \rangle_\pi^{1/2}$. By considering the scalar-product it is easy to show, that reversibility is equivalent to P being self-adjoint. Furthermore suppose that the underlying Markov chain is irreducible and aperiodic, this is also called ergodic. For details of these conditions we refer to the literature, for instance [Häg02, Bré99, LPW09]. It is a well known fact that this implies the uniqueness of the stationary distribution. Applying the spectral theorem of self-adjoint stochastic matrices and ergodicity we obtain that P has real eigenvalues

$$1 = \beta_0 > \beta_1 \geq \beta_2 \geq \cdots \geq \beta_{|D|-1} > -1$$

with a basis of orthogonal eigenfunctions u_i for $i \in \{0, \ldots, |D| - 1\}$, i.e.

$$Pu_i = \beta_i u_i, \qquad \langle u_i, u_j \rangle_\pi = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Additionally one can see that $u_0(x) = \mathbf{1}$ and $S(u_i) = 0$ for $i > 0$.

2.2. **Convergence of the chain.** The speed of convergence of the Markov chain to stationarity is measured by the so called χ^2-contrast. Let ν, μ be distributions on D then

$$\chi^2(\nu, \mu) = \sum_{x \in D} \frac{(\nu(x) - \mu(x))^2}{\mu(x)}.$$

The χ^2-contrast is not symmetric and therefore no distance. For arbitrary distributions it can be very large, i.e. $\chi^2(\nu, \mu) \leq \left\| \frac{\nu}{\mu} - 1 \right\|_\infty$, where $\left\| \frac{\nu}{\mu} - 1 \right\|_\infty = \max_{x \in D} \left| \frac{\nu(x)}{\mu(x)} - 1 \right|$.

From [Bré99, Theorem 3.3 p. 209] we have

$$\chi^2(\nu P^k, \pi) \leq \beta^{2k}\, \chi^2(\nu, \pi), \tag{2}$$

where $\beta = \max\left\{\beta_1, \left|\beta_{|D|-1}\right|\right\}$ denotes the second largest absolute value of the eigenvalues. Let us turn to another presentation of the convergence property. We have

$$\begin{aligned}
\nu P^k(x) - \pi(x) &= \sum_{y\in D} \frac{\nu(y)}{\pi(y)} p^k(y,x)\pi(y) - \pi(x) \\
&\underset{\text{rev.}}{=} \sum_{y\in D} \frac{\nu(y)}{\pi(y)} p^k(x,y)\pi(x) - \pi(x) \\
&= \sum_{y\in D} \frac{\nu(y)}{\pi(y)} p^k(x,y)\pi(x) - \sum_{y\in D} \frac{\nu(y)}{\pi(y)} \pi(y)\pi(x) \\
&= \sum_{y\in D} \frac{\nu(y)}{\pi(y)} (p^k(x,y) - \pi(y))\pi(x).
\end{aligned}$$

The second equality follows by the reversibility of the Markov chain. For simplicity let

$$d_k(x) := \sum_{y\in D} \frac{\nu(y)}{\pi(y)} (p^k(x,y) - \pi(y)),$$

such that altogether

$$\|d_k\|_2 = \sqrt{\chi^2(\nu P^k, \pi)} \underset{(2)}{\leq} \beta^k \sqrt{\left\|\frac{\nu}{\pi} - 1\right\|_\infty}. \tag{3}$$

Since $\beta < 1$ we have an exponential decay of the norm with $k \to \infty$. We define the weighted sequence spaces for $1 \leq p \leq \infty$ by

$$l_p = l_p(D,\pi) := \left\{ f \in \mathbb{R}^D : \|f\|_p^p = \sum_{x\in D} |f(x)|^p \, \pi(x) < \infty \right\}.$$

It is clear that $l_p = \mathbb{R}^D$, since the state space has finite cardinality.

Remark 1. As we have seen the χ^2-contrast corresponds to the l_2-norm of the function d_k. Other tools for measuring the speed of convergence induce similar relations. For instance

$$\|d_k\|_1 = 2\left\|\nu P^k - \pi\right\|_{\text{tv}} \quad \text{and} \quad \|d_k\|_\infty = \left\|\frac{\nu P^k}{\pi} - 1\right\|_\infty.$$

The total variation corresponds to the l_1-norm of d_k and the l_∞-norm to the supremum-distance.

Remark 2. The constant β plays a crucial role in estimating the speed of convergence of the Markov chain to stationarity. In general it is not easy to handle β_1 or β, but there are different auxiliary tools, e.g.

canonical path technique, conductance (see [JS89] and [DS91]), log-Sobolev inequalities and path coupling. For a small survey see [Ran06].

2.3. **Norm of the transition matrix.** Let us consider P and S as operators acting on l_p. Then the functional S maps arbitrary functions to constant functions. Let

$$l_p^0 := l_p^0(D,\pi) = \{g \in l_p : S(g) = 0\} \quad \text{for } 2 \le p \le \infty.$$

The norm of P as operator on l_2^0 and l_4^0 is essential in the analysis. We state and show some results which are implied by the Theorem of Riesz-Thorin. For a proof and an introduction we refer to [BS88].

Proposition 1 (Theorem of Riesz-Thorin). *Let $1 \le p, q_1, q_2 \le \infty$. Further let $\theta \in (0,1)$ and*

$$\frac{1}{p} := \frac{1-\theta}{q_1} + \frac{\theta}{q_2}$$

and

$$\begin{aligned} T &: l_{q_1} \to l_{q_1} \quad \text{with} \quad \|T\|_{l_{q_1}\to l_{q_1}} \le M_1, \\ T &: l_{q_2} \to l_{q_2} \quad \text{with} \quad \|T\|_{l_{q_2}\to l_{q_2}} \le M_2. \end{aligned}$$

Then

$$\|T\|_{l_p\to l_p} \le 2M_1^{1-\theta}M_2^{\theta}.$$

Note that the factor two in the last inequality comes from the fact that we consider real valued functions f. In the following we show a relation between P, $P-S$ and β.

Lemma 2. *Let P be a reversible transition matrix with respect to π and $n \in \mathbb{N}$. Then*

$$\|P^n - S\|_{l_2\to l_2} = \|P^n\|_{l_2^0\to l_2^0} = \beta^n. \tag{4}$$

Furthermore if $2 \le p \le \infty$ then

$$\|P^n\|_{l_p^0\to l_p^0} \le \|P^n - S\|_{l_p\to l_p} \le 2. \tag{5}$$

Proof. The self-adjointness of P implies $\|P\|_{l_2^0\to l_2^0} = \max\left\{\beta_1, \left|\beta_{|E|-1}\right|\right\} = \beta$, such that $\|P^n\|_{l_2^0\to l_2^0} = \beta^n$. By

$$\begin{aligned} \|P^n - S\|_{l_2\to l_2} &= \sup_{\|f\|_2\le 1} \|(P^n - S)f\|_2 = \sup_{\|f\|_2\le 1} \|P^n(f - S(f))\|_2 \\ &\le \sup_{\|f\|_2\le 1}\ \sup_{\|g\|_2\le 1,\, S(g)=0} \|P^n g\|_2 = \|P^n\|_{l_2^0\to l_2^0} \end{aligned}$$

and

$$\|P^n\|_{l_p^0\to l_p^0} = \sup_{\|g\|_p\le 1,\, S(g)=0} \|P^n g\|_p = \sup_{\|g\|_p\le 1,\, S(g)=0} \|P^n g - S(g)\|_p$$
$$\le \sup_{\|f\|_p\le 1} \|(P^n - S)f\|_p = \|P^n - S\|_{l_p\to l_p}$$

claim (4) and the first part of (5) is shown. Finally, by applying the triangle inequality of the norm

$$\|P^n - S\|_{l_p\to l_p} = \sup_{\|f\|_p\le 1} \|P^n f - Sf\|_p \le \|P^n\|_{l_p\to l_p} + \|S\|_{l_p\to l_p} = 2.$$

□

The next statement adds the result about the matrix norm which is used in the proof of the error bound.

Lemma 3. *Let P be a reversible transition matrix with respect to π and $n \in \mathbb{N}$. Then*

$$\|P^n\|_{l_4^0\to l_4^0} \le 2\sqrt{2}\,\beta^{n/2}. \tag{6}$$

Proof. By Lemma 2 we have

$$\|P^n - S\|_{l_2\to l_2} = \beta^n \quad \text{and} \quad \|P^n - S\|_{l_\infty\to l_\infty} \le 2.$$

Then the result is an application of Proposition 1, where $T = P^n - S$ and $q_1 = 2$, $q_2 = \infty$, $p = 4$ thus $\theta = \frac{1}{2}$. □

3. Error bounds

In this section we mainly follow two steps to develop the error bound. At first a special case of method S_{n,n_0} is considered. The initial distribution is the stationary one, thus it is not necessary to do a burn-in, i.e. $n_0 = 0$. Secondly we relate the result of the first step to the general case where the chain is initialized by a distribution ν. The techniques which we will use are similar as in [Rud09].

3.1. Starting from stationarity.

This is also called starting in equilibrium, i.e. the distribution of the Markov chain does not change, it is already balanced. In the following we will always denote $S_{n,0}$ as S_n. Let us start with stating and discussing a result from [BD06, Prop. 2.1 p.3].

Proposition 4. *Let $f \in \mathbb{R}^D$. Let $X_1, \ldots, X_n$ be a reversible Markov chain with respect to π, given by (P, π). Then*

$$e_\pi(S_n,f)^2 = \frac{1}{n^2}\sum_{k=1}^{|D|-1} |a_k|^2\, W(n,\beta_k), \tag{7}$$

where

$$a_k = \langle f, u_k\rangle_\pi \quad \textit{and} \quad W(n,\beta_k) := \frac{n(1-\beta_k^2) - 2\beta_k(1-\beta_k^n)}{(1-\beta_k)^2}.$$

Proof. Let us consider $g := f - S(f) \in \mathbb{R}^D$. Because of the orthogonal basis the presentation $g(x) = \sum_{k=1}^{|D|-1} a_k u_k(x)$ is given. The error obeys

$$\begin{aligned} e(S_n,f)^2 &= \mathbf{E}_{\pi,P}\left|\frac{1}{n}\sum_{j=1}^n g(X_j)\right|^2 = \frac{1}{n^2}\mathbf{E}_{\pi,P}\left|\sum_{j=1}^n g(X_j)\right|^2 \\ &= \frac{1}{n^2}\sum_{j=1}^n \mathbf{E}_{\pi,P}\, g(X_j)^2 + \frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{i=j+1}^n \mathbf{E}_{\pi,P}\, g(X_j)g(X_i). \end{aligned}$$

For $i \le j$,

$$\begin{aligned} \mathbf{E}_{\pi,P}\, g(X_i)g(X_j) &= \sum_{k=1}^{|D|-1}\sum_{l=1}^{|D|-1} a_k a_l\ \mathbf{E}_{\pi,P}\, u_k(X_i)u_l(X_j) \\ &\underset{(1)}{=} \sum_{k=1}^{|D|-1}\sum_{l=1}^{|D|-1} a_k a_l\ \left\langle u_k, P^{j-i}u_l\right\rangle_\pi \\ &= \sum_{k=1}^{|D|-1}\sum_{l=1}^{|D|-1} a_k a_l\ \beta_l^{j-i}\left\langle u_k, u_l\right\rangle_\pi = \sum_{k=1}^{|D|-1} a_k^2\ \beta_k^{j-i}, \end{aligned}$$

where the equality of the second line is due to the fact that the initial step is chosen from the stationary distribution. The last two equalities follow from the orthonormality of the basis of the eigenvectors. Altogether we have

$$\begin{aligned} e(S_n,f)^2 &= \frac{1}{n^2}\sum_{k=1}^{|D|-1} a_k^2\left[n + 2\sum_{j=1}^{n-1}\sum_{i=j+1}^n \beta_k^{i-j}\right] \\ &= \frac{1}{n^2}\sum_{k=1}^{|D|-1} a_k^2\left[n + 2\frac{(n-1)\beta_k - n\beta_k^2 + \beta_k^{n+1}}{(1-\beta_k)^2}\right] \\ &= \frac{1}{n^2}\sum_{k=1}^{|D|-1} |a_k|^2\, W(n,\beta_k). \end{aligned}$$

$\square$

Let us consider $W(n,\beta_k)$ to simplify and interpret Proposition 4.

Lemma 5. *For all $n \in \mathbb{N}$ and $k \in \{1,\dots,|D|-1\}$ we have*

$$W(n,\beta_k) \leq W(n,\beta_1) \leq \frac{2n}{1-\beta_1}. \tag{8}$$

Proof. Let $x \in [-1,1)$, then we are going to show that $W(n,x)$ is monotone increasing, i.e. $W(n,\beta_k) \leq W(n,\beta_1)$. For $i \in \{0,\dots,n-1\}$ it is true that

$$x^{n-i} \leq 1 \quad \Longleftrightarrow \quad (1-x^i)\,x^{n-i} \leq 1-x^i \quad \Longleftrightarrow \quad x^{n-i}+x^i \leq 1+x^n.$$

Therefore

$$x^i + x^{i+1} + x^{n-i-1} + x^{n-i} \leq 2(1+x^n),$$

and

$$(1+x)\sum_{i=0}^{n-1} x^i = \frac{1}{2}\sum_{i=0}^{n-1} x^i + x^{i+1} + x^{n-i-1} + x^{n-i} \leq n(1+x^n).$$

Now

$$\frac{dW}{dx}(n,x) = -2\frac{(1+x)\sum_{i=0}^{n-1} x^i - n(1+x^n)}{(1-x)^2} \geq 0$$

and the first inequality is shown. By

$$W(n,x) \leq \begin{cases} \frac{n(1+x)-2xn}{1-x} & x \in [-1,0] \\ \frac{n(1+x)}{1-x} & x \in (0,1) \end{cases} \leq \frac{2n}{1-x}$$

the claim is proven. □

An explicit formula of the error if the initial state is chosen by the stationary distribution is established. Let us discuss the worst case error of S_n.

Proposition 6. *Let $X_1,\dots,X_n$ be a reversible Markov chain with respect to π, given by (P,π). Then*

$$\sup_{\|f\|_2 \leq 1} e(S_n,f)^2 = \frac{1+\beta_1}{n(1-\beta_1)} - \frac{2\beta_1(1-\beta_1^n)}{n^2(1-\beta_1)^2} \leq \frac{2}{n(1-\beta_1)}. \tag{9}$$

Proof. The individual error of f is

$$\begin{aligned} e(S_n,f)^2 &\underset{(7)}{=} \frac{1}{n^2}\sum_{k=1}^{|D|-1} |a_k|^2\, W(n,\beta_k) \leq \frac{\|f\|_2^2}{n^2} \max_{k=1,\dots,|D|-1} W(n,\beta_k) \\ &\underset{(8)}{=} \frac{\|f\|_2^2}{n^2} W(n,\beta_1) = \frac{1+\beta_1}{n(1-\beta_1)}\,\|f\|_2^2 - \frac{2\beta_1(1-\beta_1^n)}{n^2(1-\beta_1)^2}\,\|f\|_2^2\,, \end{aligned}$$

where a_k is chosen as in Proposition 4 and therefore $\sum_{k=1}^{|D|-1} |a_k|^2 \leq \|f\|_2^2$. From the preceding analysis of the individual error we have an upper error bound. Now we consider $f = u_1$, where obviously $\|u_1\|_2 = 1$ and get by applying (7) that

$$e(S_n, u_1)^2 = \frac{1+\beta_1}{n(1-\beta_1)} - \frac{2\beta_1(1-\beta_1^n)}{n^2(1-\beta_1)^2}.$$

Thus the error bound is attained for u_1 and by (8) everything is shown. □

Finally an explicit presentation for the worst case error on the class of bounded functions with respect to $\|\cdot\|_2$ is shown. Notice, that (9) is an equality, which means that the integration error is completely known if we start with the stationary distribution. In some artificial cases this method even beats direct simulation, e.g. if all eigenvalues are smaller than zero or if one specific $\beta_i < 0$ and the goal is to approximate $S(u_i)$. In [FHY92, Remark 3, p.617] the authors state a simple transition matrix where $\beta_i = -\frac{1}{|D|-1}$ for all i. Now one could think to construct a transition matrix where β_1 is close to -1 and therefore damp the integration error. But it is well known that this is not possible for large $|D|$, since $\beta_1 \geq -\frac{1}{|D|-1}$.

In the next subsection we link the results to a more general framework, where the unrealistic assumption that the initial distribution is the stationary one is abandoned.

3.2. **Starting from somewhere else.** In the next statement a relation between the error of starting by π and the error of starting not by the invariant distribution is established.

Proposition 7. *Let $f \in \mathbb{R}^D$ and $g := f - S(f)$. Let $X_1, \dots, X_{n+n_0}$ be a reversible Markov chain with respect to π, given by (P, ν). Then*

(10)
$$e_\nu(S_{n,n_0}, f)^2 = e_\pi(S_n, f)^2 + \frac{1}{n^2}\sum_{j=1}^{n} L_{j+n_0}(g^2) + \frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} L_{j+n_0}(gP^{k-j}g),$$

where

$$L_i(h) = \sum_{x\in D} d_i(x)h(x)\pi(x) = \sum_{x\in D}\sum_{y\in D}\frac{\nu(y)}{\pi(y)}(p^i(x,y)-\pi(y))h(x)\pi(x).$$

Remark 3. The proof of this identity is similar as in [Rud09], except for the fact that we study a discrete state space and therefore integrals become sums.

Proof. It is easy to see, that

$$\mathbf{E}_{\nu,P}\left|S(f)-S_{n,n_0}(f)\right|^2=\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\mathbf{E}_{\nu,P}(g(X_{n_0+j})g(X_{n_0+i}))$$

$$=\frac{1}{n^2}\sum_{j=1}^{n}\sum_{x\in D}P^{n_0+j}g^2(x)\,\nu(x)+\frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\sum_{x\in D}P^{n_0+j}(gP^{k-j}g)(x)\,\nu(x).$$

For every function $h\in\mathbb{R}^D$ and $i\in\mathbb{N}$ under applying the reversibility the following transformation holds true

$$\sum_{x\in D}(P^ih)(x)\,\nu(x)=\sum_{x\in D}\sum_{y\in D}h(y)\,p^i(x,y)\,\frac{\nu(x)}{\pi(x)}\,\pi(x)$$

$$\underset{\text{rev.}}{=}\sum_{x\in D}\sum_{y\in D}\frac{\nu(y)}{\pi(y)}p^i(x,y)\,h(x)\,\pi(x)$$

$$=\sum_{x\in D}h(x)\,\pi(x)+\sum_{x\in D}\sum_{y\in D}\frac{\nu(y)}{\pi(y)}\left(p^i(x,y)-\pi(y)\right)h(x)\,\pi(x)$$

$$\underset{\text{rev.}}{=}\sum_{x\in D}(P^ih)(x)\pi(x)+\sum_{x\in D}\sum_{y\in D}\frac{\nu(y)}{\pi(y)}\left(p^i(x,y)-\pi(y)\right)h(x)\,\pi(x).$$

Using this in the setting above, formula (10) is shown. □

Equation (10) is still an error characterization where equality holds. We will estimate $L_k(h)$ to derive an upper bound. This depends very much on the speed of convergence from the chain to stationarity.

Lemma 8. *Let $h\in\mathbb{R}^D$, let again $\beta=\max\left\{\beta_1,\left|\beta_{|D|-1}\right|\right\}$. Then*

$$|L_k(h)|\le\beta^k\sqrt{\left\|\frac{1}{\pi}\right\|_\infty}\sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\cdot\|h\|_1\,,\tag{11}$$

$$|L_k(h)|\le\beta^k\sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\cdot\|h\|_2\,.\tag{12}$$

Proof. Let us consider $L_k(h)=\langle d_k,h\rangle_\pi$. After applying Cauchy-Schwarz inequality we obtain

$$|L_k(h)|\le\|d_k\|_2\,\|h\|_2\,.$$

By applying (3) we showed (12). Inequality (12) and $\|h\|_2\le\sqrt{\left\|\frac{1}{\pi}\right\|_\infty}\,\|h\|_1$ implies (11). □

The ingredients for getting an explicit error bound for S_{n,n_0} are gathered together. Mainly the last Lemma ensures an exponential decay of $L_k(h)$ which is used in the next Proposition.

Proposition 9. *Let $X_1,\dots,X_{n+n_0}$ be a reversible Markov chain with respect to π, given by (P,ν). Let $f\in\mathbb{R}^D$, $g:=f-S(f)$ and*

$$V(\beta,n)=\sum_{j=1}^{n}\beta^j+2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\beta^k,$$
$$U(\beta,n)=\sum_{j=1}^{n}\beta^j+4\sqrt{2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\beta^{\frac{k+j}{2}}.$$

(i) Then for $g\in l_2^0$ we have

$$e_\nu(S_{n,n_0},f)^2\le e_\pi(S_n,f)^2+\frac{V(\beta,n)}{n^2}\beta^{n_0}\sqrt{\left\|\frac{1}{\pi}\right\|_\infty}\sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\left\|g\right\|_2^2.$$

(ii) Then for $g\in l_4^0$ we have

$$e_\nu(S_{n,n_0},f)^2\le e_\pi(S_n,f)^2+\frac{U(\beta,n)}{n^2}\beta^{n_0}\sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\left\|g\right\|_4^2.$$

(iii) Then for $g\in l_\infty^0$ we have

$$e_\nu(S_{n,n_0},f)^2\le e_\pi(S_n,f)^2+\frac{V(\beta,n)}{n^2}\beta^{n_0}\sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\left\|g\right\|_\infty^2.$$

Proof. As we have seen in (10) the error obeys

(13)
$$e_\nu(S_{n,n_0},f)^2=e_\pi(S_n,f)^2+\frac{1}{n^2}\sum_{j=1}^{n}L_{j+n_0}(g^2)+\frac{2}{n^2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}L_{j+n_0}(gP^{k-j}g).$$

Then by (11), Cauchy-Schwarz inequality and $\left\|P^{k-j}\right\|_{l_2^0\to l_2^0}=\beta^{k-j}$ we get

$$\left|L_{j+n_0}(g^2)\right|\le\sqrt{\left\|\frac{1}{\pi}\right\|_\infty}\sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\beta^{j+n_0}\left\|g\right\|_2^2,$$
$$\left|L_{j+n_0}(gP^{k-j}g)\right|\le\sqrt{\left\|\frac{1}{\pi}\right\|_\infty}\sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\beta^{k+n_0}\left\|g\right\|_2^2.$$

Putting this in the sums of equation (13) and let $\varepsilon_0 = \sqrt{\left\|\frac{1}{\pi}\right\|_\infty}\sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\beta^{n_0}$ we obtain

$$\begin{aligned}
&\sum_{j=1}^{n}\left|L_{j+n_0}(g^2)\right| + 2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\left|L_{j+n_0}(gP^{k-j}g)\right| \\
&\quad\leq \varepsilon_0\left\|g\right\|_2^2\sum_{j=1}^{n}\beta^j + \varepsilon_0\left\|g\right\|_2^2\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}2\beta^k \\
&\quad= \varepsilon_0\left\|g\right\|_2^2\left(\sum_{j=1}^{n}\beta^j + \sum_{j=1}^{n-1}\sum_{k=j+1}^{n}2\beta^k\right) = V(\beta,n)\cdot\varepsilon_0\left\|g\right\|_2^2.
\end{aligned}$$

Thus claim (i) is shown. Now we use (12) and

$$\left\|gP^{k-j}g\right\|_2 \leq \left\|g\right\|_\infty\left\|P^{k-j}g\right\|_2 \leq \left\|g\right\|_\infty^2\left\|P^{k-j}\right\|_{l_2^0\to l_2^0} \leq \left\|g\right\|_\infty^2\beta^{k-j}$$

to obtain

$$\begin{aligned}
\left|L_{j+n_0}(g^2)\right| &\leq \sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\beta^{j+n_0}\left\|g\right\|_\infty^2, \\
\left|L_{j+n_0}(gP^{k-j}g)\right| &\leq \sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\beta^{k+n_0}\left\|g\right\|_\infty^2.
\end{aligned}$$

Exactly the same steps as in the proof of (i) follow, except for a different $\varepsilon_0 = \sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\beta^{n_0}$ and the supremum norm, i.e. assertion (iii) is proven. Let us turn to (ii). Again we use (12) and estimate

$$\left\|gP^{k-j}g\right\|_2 \leq \left\|g\right\|_4\left\|P^{k-j}g\right\|_4 \leq \left\|P^{k-j}\right\|_{l_4^0\to l_4^0}\left\|g\right\|_4^2 \underset{(6)}{\leq} 2\sqrt{2}\left\|g\right\|_4^2\beta^{\frac{k-j}{2}}.$$

Thus

$$\begin{aligned}
\left|L_{j+n_0}(g^2)\right| &\leq \sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\beta^{j+n_0}\left\|g\right\|_4^2, \\
\left|L_{j+n_0}(gP^{k-j}g)\right| &\leq 2\sqrt{2}\sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty}\beta^{\frac{k+j}{2}+n_0}\left\|g\right\|_4^2.
\end{aligned}$$

For $\varepsilon_0 = \sqrt{\left\|\frac{\nu}{\pi} - 1\right\|_\infty \beta^{n_0}}$ we obtain

$$\begin{aligned}
&\sum_{j=1}^{n} \left|L_{j+n_0}(g^2)\right| + 4\sqrt{2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} \left|L_{j+n_0}(gP^{k-j}g)\right| \\
&\quad\leq \varepsilon_0 \left\|g\right\|_4^2 \sum_{j=1}^{n} \beta^j + \varepsilon_0 \left\|g\right\|_4^2 \sum_{j=1}^{n-1}\sum_{k=j+1}^{n} 4\sqrt{2}\,\beta^{\frac{k+j}{2}} \\
&\quad= \varepsilon_0 \left\|g\right\|_4^2 \left(\sum_{j=1}^{n} \beta^j + 4\sqrt{2}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} \beta^{\frac{k+j}{2}}\right) = U(\beta, n)\cdot \varepsilon_0 \left\|g\right\|_4^2.
\end{aligned}$$

Finally by substituting this in equation (13) everything is shown. □

In the last Proposition we introduced $V(\beta, n)$ and $U(\beta, n)$. These functions are bounded if $\beta < 1$. By applying the infinite geometric series several times the following is proven.

Lemma 10. *For $n \in \mathbb{N}$ and $x \in [0, 1)$ we have*

$$V(x, n) \leq \frac{2}{(1-x)^2}, \qquad U(x, n) \leq \frac{4\sqrt{2}}{(1-x)(1-\sqrt{x})}. \tag{14}$$

This implies that the asymptotic optimality is reached.

3.3. **Main Theorem.** The following is the main result.

Theorem 11. *Let $X_1, \dots, X_{n+n_0}$ be a reversible Markov chain with respect to π, given by (P, ν). Let $f \in \mathbb{R}^D$ and $a_k = \langle f, u_k\rangle_\pi$.*
Then

$$\lim_{n\to\infty} n \cdot e_\nu(S_{n,n_0}, f)^2 = \lim_{n\to\infty} n \cdot e_\pi(S_n, f)^2 = \sum_{k=1}^{|D|-1} |a_k|^2 \frac{1+\beta_k}{1-\beta_k}.$$

(i) If we consider $f \in l_2$ then

$$e_\nu(S_{n,n_0}, f)^2 \leq \frac{2}{n(1-\beta_1)} \left\|f\right\|_2^2 + \frac{2\sqrt{\left\|\frac{1}{\pi}\right\|_\infty}\sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty \beta^{n_0}}}{n^2(1-\beta)^2} \left\|f\right\|_2^2.$$

(ii) If we consider $f \in l_4$ then

$$e_\nu(S_{n,n_0}, f)^2 \leq \frac{2}{n(1-\beta_1)} \left\|f\right\|_4^2 + \frac{16\sqrt{2}\sqrt{\left\|\frac{\nu}{\pi}-1\right\|_\infty \beta^{n_0}}}{n^2(1-\beta)(1-\sqrt{\beta})} \left\|f\right\|_4^2.$$

(iii) If we consider $f \in l_\infty$ then

$$e_\nu(S_{n,n_0}, f)^2 \leq \frac{2}{n(1-\beta_1)} \left\| f \right\|_\infty^2 + \frac{4\sqrt{\left\| \frac{\nu}{\pi} - 1 \right\|_\infty} \beta^{n_0}}{n^2(1-\beta)^2} \left\| f \right\|_\infty^2$$

Proof. By (10) and the fact that the remaining terms are going quadratic to zero as n goes to infinity, we see that the asymptotic result holds true. For $f \in l_2$ we have $\|f - S(f)\|_2 \leq \|f\|_2$ and furthermore if $p \neq 2$ then

$$\|f - S(f)\|_p \leq \|f\|_p + |S(f)| \leq \|f\|_p + \|f\|_1 \leq 2\|f\|_p .$$

Thus, via Proposition 9, Proposition 6 and Lemma 10 everything is shown. □

Notice, that from the estimate of Proposition 9 it follows immediately that

$$\lim_{n\to\infty} n \cdot e_\nu(S_{n,n_0}, f)^2 \leq \lim_{n\to\infty} n \cdot e_\pi(S_{n,n_0}, f)^2 \leq \frac{1+\beta_1}{1-\beta_1} \|f\|_2^2 .$$

Thus there is no gap between the estimate and the asymptotical behavior. Also notice, that the upper bounds are continuous in the sense that if the initial distribution ν is π then we obtain the bound of Proposition 6. The dependence of the bounds of (ii) and (iii) in Theorem 11 on the initial distribution is encouraging for an extension to general state spaces. (For an introduction to MCMC on general state spaces we refer to [RR04].) But the dependence of the initial distribution on the estimate in the l_2-case is disillusioning because of the additional factor of $\left\| \frac{1}{\pi} \right\|_\infty$.

In [Rud09, Theorem 8, p.10] a similar l_∞-bound of S_{n,n_0} for general state spaces is developed. This result holds for lazy, reversible Markov chains and may also be applied in the present setting, i.e. if the state space is discrete. In [Rud09] the asymptotic error limit is not attained. Thus we could improve the error bound and weaken the laziness condition, i.e. it is enough that $\beta_1 = \beta$. In [LPW09, Thm. 12.19, p.165] the authors obtained for another error term a comparable bound where the chain starts deterministically.

4. Burn-in

Let us assume that computer resources for the MCMC method for N time steps are available, i.e. $N = n + n_0$. We want to choose the burn-in n_0 and the number of n such that the error is as small as possible. The burn-in n_0 should be large but this implies that n is possibly quite small depending on how much resources we have. On the other hand n should be large which again implies that n_0 is possibly small. There is obviously a trade-off between choosing the parameters. In the next statement we consider the error for an explicitly given burn-in, where for simplicity $\beta_1 = \beta$.

Corollary 12. *Let $f \in \mathbb{R}^D$ be given and let*
$$n_0 = \max\left\{ \left\lceil \frac{\log(C)}{\log(\beta^{-1})} \right\rceil, 0 \right\}.$$

(i) Let $C = \sqrt{\left\| \frac{1}{\pi} \right\|_\infty} \sqrt{\left\| \frac{\nu}{\pi} - 1 \right\|_\infty}$, then
$$\sup_{\|f\|_2 \le 1} e_\nu(S_{n,n_0}, f)^2 \le \frac{2}{n(1-\beta)} + \frac{2}{n^2(1-\beta)^2}.$$

(ii) Let $C = 16\sqrt{2}\sqrt{\left\| \frac{\nu}{\pi} - 1 \right\|_\infty}$, then
$$\sup_{\|f\|_4 \le 1} e_\nu(S_{n,n_0}, f)^2 \le \frac{2}{n(1-\beta)} + \frac{1}{n^2(1-\beta)(1-\sqrt{\beta})}.$$

(iii) Let $C = 2\sqrt{\left\| \frac{\nu}{\pi} - 1 \right\|_\infty}$, then
$$\sup_{\|f\|_\infty \le 1} e_\nu(S_{n,n_0}, f)^2 \le \frac{2}{n(1-\beta)} + \frac{2}{n^2(1-\beta)^2}.$$

Note, that in the l_∞- and l_2-case the error bound is the same. Just the constant C which comes in by the density is different. This suggestion of the burn-in is justified in the following.

4.1. **Numerical experiments.** Suppose C (very large), β (close to one) and resources N are given. The worst case error for $\|f\|_2 \le 1$ or $\|f\|_\infty \le 1$ is bounded by
$$b_\infty(n, n_0) := \sqrt{\frac{2}{n(1-\beta)} + \frac{2C\beta^{n_0}}{n(1-\beta)^2}}$$
and if we consider $\|f\|_4 \le 1$ it is bounded by
$$b_4(n, n_0) := \sqrt{\frac{2}{n(1-\beta)} + \frac{C\beta^{n_0}}{n(1-\beta)(1-\sqrt{\beta})}}.$$
Since $N = n + n_0$ we can compute with a numerical procedure (here using Maple) the optimal choice of the burn-in denoted by n^4_{opt}, n^∞_{opt} to minimize the upper error bounds. (This is a simple one dimensional minimization problem with different parameters.)

N	β	n_{opt}^{4} (by Maple)	n_{opt}^{∞} (by Maple)	$n_0 = \lceil \log(C)/\log(\beta^{-1}) \rceil$ (suggested above)
10^4	0.9	656	656	656
10^5	0.9	656	656	656
10^4	0.99	6867	6867	6873
10^5	0.99	6873	6873	6873
10^4	0.999	8001	8001	69043
10^5	0.999	68977	68977	69043

TABLE 1. For $C = 10^{30}$ where n_{opt}^{i} minimizes $b_i(N - n_{\text{opt}}^{i}, n_{\text{opt}}^{i})$, $i = 4, \infty$.

Table 1 gives a collection of typical results. It turned out that the above suggested lower bound is almost the optimal choice. The computed value n_{opt}^{4} and n_{opt}^{∞} is almost the same as $n_0 = \lceil \log(C)/\log(\beta^{-1}) \rceil$. In the case $C = 10^{30}$ and $\beta = 0.999$ Theorem 11 gives for no choice of n and n_0 with $N = 10^4$ an error smaller than one.

In Figure 1 we plotted $b_4(N-n_0, n_0)$ for different n_0 and $e_\pi(S_N, u_1) = \sqrt{\frac{1+\beta_1}{N(1-\beta_1)} - \frac{2\beta_1(1-\beta_1^N)}{N^2(1-\beta_1)^2}}$. Roughly spoken one may see in Figure 1 that if the burn-in is chosen too small a vertical shifting takes place and if the burn-in is chosen to large a horizontal shifting takes place. The

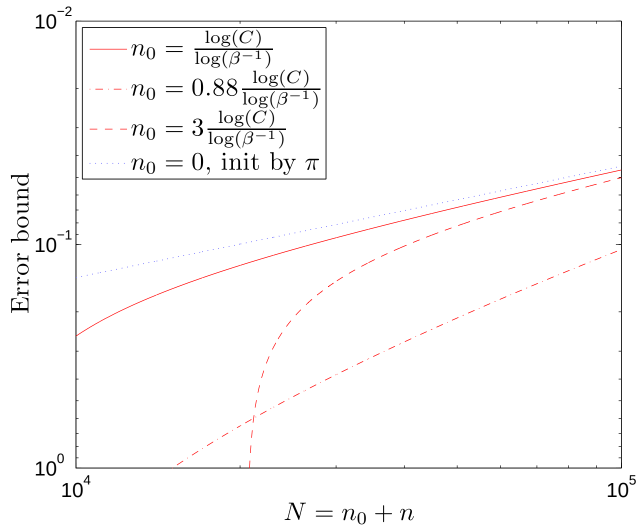

FIGURE 1. For $\beta = 0.99$ and $D = 10^{30}$.

asymptotic behavior is the same, i.e. for the long run the error of S_{n,n_0} converges to the error of S_n. If β and C are given we chose the burn-in as suggested above. If there is an estimate of $\log(C)/\log(\beta^{-1})$ one should ensure that it is not smaller than the real ratio. As seen in Figure 1 if it is slightly smaller there is already strong influence. By choosing the burn-in too large the influence is less heavy.

Finally if there is no estimation or computation of the parameters β or C a simple but very efficient strategy is given by choosing $n = n_0 = \frac{N}{2}$ (for even N). In Figure 2 we see $b_4(\frac{N}{2}, \frac{N}{2})$, $b_4(N - n_0, n_0)$ and $e_\pi(S_N, u_1)$. In the asymptotic behavior we pay the price of a factor

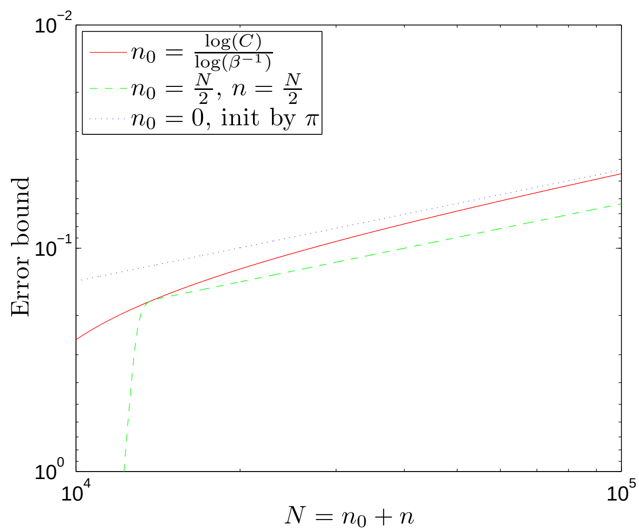

FIGURE 2. For $\beta = 0.99$ and $D = 10^{30}$.

of $\sqrt{2}$, i.e. the asymptotic error is $\sqrt{2}$ times larger than $e_\pi(S_N, u_1)$ where we started in equilibrium. This strategy works well and reaches the same convergence rate as choosing the burn-in as suggested above, which is seen in Figure 2.

Acknowledgements

The author thanks Erich Novak and Aicke Hinrichs for their valuable comments.

References

[BD06] F. Bassetti and P. Diaconis, *Examples comparing importance sampling and the Metropolis algorithm*, Illinois J. Math. **50** (2006), no. 1-4, 67–91.

[BL07] F. Bassetti and F. Leisen, *Metropolis algorithm and equienergy sampling for two mean field spin systems*, 2007.

[Bré99] P. Brémaud, *Markov chains*, Texts in Applied Mathematics, vol. 31, Springer-Verlag, New York, 1999, Gibbs fields, Monte Carlo simulation, and queues.

[BS88] C. Bennett and R. Sharpley, *Interpolation of operators*, Pure and Applied Mathematics, vol. 129, Academic Press Inc., Boston, MA, 1988.

[DS91] P. Diaconis and D. Stroock, *Geometric bounds for eigenvalues of Markov chains*, Ann. Appl. Probab. **1** (1991), no. 1, 36–61.

[FHY92] A. Frigessi, C. R. Hwang, and L. Younes, *Optimal spectral structure of reversible stochastic matrices, Monte Carlo methods and the simulation of Markov random fields*, Ann. Appl. Probab. **2** (1992), no. 3, 610–628.

[GRS96] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall, 1996.

[Häg02] O. Häggström, *Finite Markov chains and algorithmic applications*, London Mathematical Society Student Texts, vol. 52, Cambridge University Press, Cambridge, 2002.

[JS89] M. Jerrum and A. Sinclair, *Approximating the permanent*, SIAM J. Comput. **18** (1989), no. 6, 1149–1178.

[Liu08] Jun S. Liu, *Monte Carlo strategies in scientific computing*, Springer Series in Statistics, Springer, New York, 2008.

[LPW09] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*, AMS, Providence, RI, 2009.

[Mat99] P. Mathé, *Numerical integration using Markov chains*, Monte Carlo Methods Appl. **5** (1999), no. 4, 325–343.

[MR02] N. Madras and D. Randall, *Markov chain decomposition for convergence rate analysis*, Ann. Appl. Probab. **12** (2002), no. 2, 581–606.

[Ran06] D. Randall, *Rapidly Mixing Markov Chains with Applications in computer Science and Physics*, Computing in Science and Engineering **8** (2006), no. 2, 30–41.

[RR97] G. O. Roberts and J. S. Rosenthal, *Geometric ergodicity and hybrid Markov chains*, Electron. Comm. Probab. **2** (1997), no. 2, 13–25.

[RR04] G. O. Roberts and J. S. Rosenthal, *General state space Markov chains and MCMC algorithms*, Probability Surveys **1** (2004), 20–71.

[Rud09] D. Rudolf, *Explicit error bounds for lazy reversible Markov chain Monte Carlo*, J. Complexity **25** (2009), no. 1, 11–24.

[Sok97] A. Sokal, *Monte Carlo methods in statistical mechanics: foundations and new algorithms*, Functional integration (Cargèse, 1996), NATO Adv. Sci. Inst. Ser. B Phys., vol. 361, Plenum, New York, 1997, pp. 131–192.

Friedrich Schiller University Jena, Mathem. Institute, Ernst-Abbe-Platz 2, D-07743 Jena, Germany

E-mail address: `daniel.rudolf@uni-jena.de`

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.