Weierstraß–Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Efficient mixing of product walks on product groups

Peter Mathé

submitted: 13th November 1996

Weierstrass Institute
for Applied Analysis
and Stochastics
Mohrenstraße 39
D – 10117 Berlin
Germany

Preprint No. 284
Berlin 1996

1991 Mathematics Subject Classification. 60J15.

Key words and phrases. Product random walk, mixing time.

Edited by
Weierstraß–Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
D — 10117 Berlin
Germany

Fax: + 49 30 2044975
e–mail (X.400): c=de;a=d400-gw;p=WIAS-BERLIN;s=preprint
e–mail (Internet): preprint@wias-berlin.de

ABSTRACT. We are going to study the mixing behavior of product–type random walks on product groups. This study is inspired by the investigation of the relaxation of random walks on d–dimensional grids with possibly direction dependent mesh size. Typically such walks are designed to randomly visit a coordinate direction and then to carry out a random step within the chosen component according to some random walk in this direction. We will derive a dependence of the mixing times of such random walks in terms of the component mixing times. If we are free to optimize the random visiting scheme, then we can speed up mixing in case the component mixing times vary much. In more homogeneous situations the overall mixing time is bounded by a multiple of the sum of the single ones times the logarithm of the number of components.

1. INTRODUCTION

Suppose we are given d finite groups $G_1, \dots, G_d$ with random walks driven by the respective transition matrices $P_1, \dots, P_d$. We are going to study the mixing behavior of product–type random walks on the product $G := \prod_{j=1}^d G_j$.

A product–type random walk is obtained from these components in the following way. We choose a convex combination $\rho := (\rho_1, \dots, \rho_d)$, i.e., $\rho_j \geq 0$, $\sum_{j=1}^d \rho_j = 1$, and compose

$$P_\rho := \sum_{j=1}^{d} \rho_j \tilde{P}_j, \tag{1}$$

where ˜ indicates the embedding of the component transition matrices into ones for G. In conjunction with an initial distribution ν on G we obtain a random walk on G with respective distribution νP_ρ^n at the nth step. This corresponds to a mixture of the components and means, that with a certain probability we choose a component of our product group and then we take a transition according to the random walk acting on this component. So we may think of ρ as a randomized visiting scheme being the counterpart of the visiting scheme in the context of Gibbs–type samplers, see [6], where this is called a *proposal* or *exploration* distribution. This study is inspired by the investigation of the relaxation of random walks on d–dimensional grids with possibly direction dependent mesh size. Typically such walks are designed to randomly pick a coordinate direction and then to carry out a random step within the chosen component according to some random walk responsible for this direction.

The prominent example for this is the walk on the hyper cube $\mathbb{Z}_2^d$, which is studied in [3, Ch. 3C, Ex. 2]. However, the component groups are very small, such that the overall behavior of the random walk is better than predicted by the considerations concerning the components. This is due to "degrees of symmetry". Much closer to our intentions is the case $\mathbb{Z}_n^d$ with n being large, which avoids additional degrees of symmetry. Asymptotic considerations as carried out in [1] predict relaxation on $\mathbb{Z}_n^d$ in time proportional to n^2. We shall make this more precise below.

Let us mention that a random walk as described above has to compete with a *Gibbs*–type sampler, where (as usual) the components are chosen deterministically by a visiting scheme. As it will turn out below, and this is quite obvious, the latter scheme is more efficient in a homogeneous situation, where each component looks more or less the same. In non–homogeneous cases it is desirable to visit "expensive" components more often than others, such that randomized visiting schemes are preferable. The overall performance will be optimal when the randomized visiting schemes ρ are properly chosen in correspondence with the mixing times within the components. This will be the topic of Section 5. Roughly speaking, if we denote by $K(P_j), j = 1, \dots, d$ the mixing times of the components, then we obtain a probability σ on $\{1, \dots, d\}$ by

$$\sigma_j := \frac{K(P_j)}{\sum_{j=1}^{d} K(P_j)}, \quad j = 1, \dots, d.$$

As it will turn out, the entropy $H(\sigma)$ of this probability will determine whether the product–type random walk we are going to study can compete with the Gibbs–type sampler or not. If the single mixing times do not vary between the components too much then the entropy $H(\sigma)$ will be of the order $\log(d)$, an additional effort which has to be made by our randomized visiting scheme. In case the entropy does not depend on the actual number of components, a properly chosen randomized visiting scheme allows to compete with Gibbs–type random walks.

Let us mention that the restriction to products of groups rather than to products of arbitrary finite state spaces is not necessary. Much of the results may be generalized. To the authors opinion the present setup is easier to access.

2. Auxiliary Results

As described in the Introduction, we suppose that we are given d finite groups $G_1, \dots, G_d$ with random walks driven by the respective transition matrices $P_1, \dots, P_d$. We additionally assume that each P_j is associated a probability μ_j in the manner

$$P_j(\xi_j, \eta_j) := \mu_j(\{\xi_j^{-1} \circ \eta_j\}), \quad \xi_j, \eta_j \in G_j, \quad j = 1, \dots, d.$$

It is well known and easy to check, that such random walks are irreducible with unique invariant distribution being the uniform distribution $U_j, \quad j = 1, \dots, d$, i.e., $U_j(\{\xi_j\}) = \frac{1}{|G_j|}$, if the supports of the μ_j generate the groups G_j, see [1].

A random walk on the product $G := \prod_{j=1}^{d} G_j$ is constructed as follows. We first embed the random walks $P_j, \quad j = 1, \dots, d$ into the product by letting

$$\tilde{\mu}_j(\{x\}) := \begin{cases} \mu_j(\{\xi_j\}) & \text{, if } x = (e_1, \dots, \xi_j, \dots, e_d) \\ 0 & \text{, otherwise} \end{cases}, \tag{2}$$

and the corresponding transition matrices

$$\tilde{P}_j(x, y) := \tilde{\mu}_j(\{x^{-1} \circ y\}), \quad x, y \in G. \tag{3}$$

Above, the symbols $e_j, \quad j = 1, \dots, d$ denote the corresponding neutral elements in G_j. Hence, the random walks $\tilde{P}_j$ accept transitions in the components G_j only. We mention the following

Lemma 1. *The convolution $\tilde{\mu}_i * \tilde{\mu}_j, \quad i \neq j$ is commutative, precisely we have for any $x = (\xi_1, \dots, \xi_d)$ the equality*

$$\tilde{\mu}_i * \tilde{\mu}_j(\{x\}) = \tilde{\mu}_j * \tilde{\mu}_i(\{x\}) = \mu_i(\{\xi_i\})\mu_j(\{\xi_j\}).$$

Proof. We have

$$\tilde{\mu}_i * \tilde{\mu}_j(\{x\}) = \sum_{z \in G} \tilde{\mu}_i(\{z\})\tilde{\mu}_j(\{z^{-1} \circ x\}).$$

To make the first factor non–vanishing, $z = (e_1, \dots, \zeta_i, \dots, e_d)$ is required. Analogously, the second factor is nonzero only if $\zeta_i = \xi_i$. From both requirements we conclude that the sum reduces to only one summand $z = (\zeta_1, \dots, \zeta_d)$ with

$$\zeta_l = \begin{cases} \xi_l & \text{, for } l = i \text{ or } l = j \\ e_l & \text{, else} \end{cases}, \quad l = 1, \dots, d.$$

□

We recall from the Introduction that a product–type random walk is obtained from these components by choosing a convex combination $\rho := (\rho_1, \dots, \rho_d)$, i.e., $\rho_j \geq 0, \quad \sum_{j=1}^d \rho_j = 1$, and compose

$$P_\rho := \sum_{j=1}^{d} \rho_j \tilde{P}_j. \tag{4}$$

We mention that the random walk P_ρ is again driven by a single probability on G, namely $\mu_\rho := \sum_{j=1}^d \rho_j \tilde{\mu}_j$. Again, if all μ_j have a G_j–generating support and if all ρ_j are positive, then P_ρ is ergodic with unique invariant distribution U on G.

In addition to the examples presented in Section 1 we introduce the following

Example. Let $U_j, \quad j = 1, \dots, d$, denote the uniform distributions on G_j and consider the random walk Q_j, describing an i.i.d. walk on G_j, hence

$$Q_j(\xi_j, \eta_j) := U_j(\{\xi_j^{-1} \circ \eta_j\}) = \frac{1}{|G_j|}, \quad \xi_j, \eta_j \in G_j, \quad j = 1, \dots, d.$$

Let $\tilde{U}_j, \quad j = 1, \dots, d$ denote the embeddings of U_j into G. The following observation is easily checked.

1. The convolution $\tilde{U}_j * \tilde{U}_j$ equals $\tilde{U}_j, \quad j = 1, \dots, d$.
2. In view of Lemma 1 we have $\tilde{U}_1^{r_1} * \cdots * \tilde{U}_d^{r_d} = U$ whenever all $r_1, \dots, r_d$ are positive.

As above, given ρ we construct U_ρ and Q_ρ as mixtures, cf. (4).

Now, given ρ the random walk Q_ρ is important as it reduces all considerations concerning mixing times to the properties of the visiting scheme ρ, since within the component one step will suffice to reach stationarity.

The mixing behavior of random walks shall be quantified in terms of the variation distance of measures. Given a (signed) measure λ on some (finite) set X we denote by

$$\|\lambda\|_X := \max_{A \subset X} |\lambda(A)| = \frac{1}{2} \sum_{x \in X} |\lambda(\{x\})|.$$

Whenever it will be clear from the context, we will suppress the subscript indicating the set the measure is living on. Let us however mention that for a measure λ_j on

G_j the corresponding embedded $\tilde{\lambda}_j$ on G obeys $\|\tilde{\lambda}_j\|_G = \|\lambda_j\|_{G_j}$. Moreover, if two measures allow a convolution then it has variation norm less then the product of the single norms.

The following facts will be important, cf. [3, Ch. 3.].

Lemma 2. *Let P be a random walk on G associated to a probability μ. Let ν be an initial distribution of a random walk driven by P. We have for $k \geq 1$*

$$\|\nu P^k - U\|_G \leq \|\mu^{*k} - U\|_G.$$

We also explicitly state an estimate, similar to the one in Lemma (7.9) in [2]:.

Let U denote the uniform distribution on G. For an arbitrary P on G and constant $0 < \alpha < 1$ set $A_\alpha := \{x \in G, \quad P(\{x\}) \leq \alpha U(\{x\})\}$.

Lemma 3. *For any $0 < \alpha < 1$ and set A_α as above we have*

$$\|P - U\| \geq (1 - \alpha) U(A_\alpha).$$

Proof. By our assumption on A_α we obtain $P(A_\alpha^c) \geq 1 - \alpha U(A_\alpha)$ and consequently

$$\begin{aligned} \|P - U\| &= \frac{1}{2} \sum_{x \in G} |P(\{x\}) - U(\{x\})| \\ &\geq \frac{1}{2} \sum_{x \in A_\alpha} (1 - \alpha) U(\{x\}) + \frac{1}{2} \sum_{x \in A_\alpha^c} |P(\{x\}) - U(\{x\})| \\ &\geq \frac{1}{2}(1 - \alpha) U(A_\alpha) + \frac{1}{2} P(A_\alpha^c) - \frac{1}{2} U(A_\alpha^c) \\ &\geq (1 - \alpha) U(A_\alpha). \end{aligned}$$

$\square$

We turn to the study of mixing times. Our approach is close to [1]. Given a random walk P on a set X with invariant distribution π we let for $k \geq 1$ the number

$$d_k(P) := \max_{x \in X} \|\delta_x P^k - \pi\|.$$

As a function of $k \in \mathbb{N}$ it is easily seen to be decreasing. Further, as will be clear below it makes sense to measure the time to reach stationarity in terms of this quantity. So we agree to let

$$K(P) := \min \left\{ k \in \mathbb{N}, \quad d_k(P) \leq \frac{1}{2\mathrm{e}} \right\}. \tag{5}$$

Moreover the quantity $d_k(P)$ is close to being submultiplicative. Precisely we have

Lemma 4. *For any $k \in \mathbb{N}$ the following inequality holds true*

$$d_{l \cdot k}(P) \leq (2 d_k(P))^l.$$

Especially, with $k := K(P)$ we obtain $d_{l \cdot K(P)}(P) \leq \mathrm{e}^{-l}$.

Proof. We introduce another auxiliary quantity, cf. [1]

$$\rho_k(P) := \max_{x,y \in X} \|\delta_x P^k - \delta_y P^k\|.$$

It is known from Lemma (3.5) in [1] that this is submultiplicative and that $\rho_k(P) \leq 2d_k(P)$. In fact we even have

$$d_k(P) \leq \rho_k(P) \leq 2d_k(P), \tag{6}$$

where it only remains to prove the left–hand side inequality. But this follows from the invariance of π under P^k. Indeed it yields

$$\begin{aligned} 2\|\delta_x P^k - \pi\| &= \sum_{z \in X} |P^k_{x,z} - \pi_z| \\ &= \sum_{z \in X} |P^k_{x,z} - \sum_{y \in X} \pi_y P^k_{y,z}| \\ &\leq \sum_{z \in X} |\sum_{y \in X} \pi_y (P^k_{x,z} - P^k_{y,z})| \\ &\leq \sum_{y \in X} \pi_y \sum_{z \in X} |P^k_{x,z} - P^k_{y,z}| \\ &\leq 2\rho_k(P). \end{aligned}$$

Since this is valid for any initial value $x \in X$ we have proved (6). Now it is straight–forward to prove the assertions of the lemma. □

Remark 1. The important though easy to prove $d_k(P) \leq \rho_k(P)$ was observed by E. Behrends and is reproduced here with kind permission.

Thus we may think of $K(P)$ as a threshold level starting from which the convergence to stationarity is exponential.

We close this section with some facts about *multinomial distributions*, which will occur very naturally below. Given a d–tuple $\bar{r} = (r_1, \ldots, r_d)$ of natural numbers with $r_1 + \ldots r_d = k$ we denote by $\binom{k}{\bar{r}} := \frac{k!}{r_1! \cdots r_d!}$ and $r_{min} := \min_{j=1,\ldots,d} r_j$. Let $P_{k,\rho}$ denote the multinomial distribution on $\{0, \ldots, k\}^d$ with point masses

$$P_{k,\rho}((r_1, \ldots, r_d)) = \binom{k}{\bar{r}} \prod_{j=1}^{d} \rho_j^{r_j}, \quad \text{if } r_1 + \ldots r_d = k.$$

A detailed exposition with further references can be found in [4, Ch. 11.2]. We mention that the component distributions of $P_{k,\rho}$ are respective binomial ones B_{k,ρ_j} with respective ρ_j. The following lemma is probably well known. Since we are not aware of any reference we include the proof.

Lemma 5. *For any d, convex combination ρ and $k \in \mathbb{N}$ we have*

$$1 - e^{-\sum_{j=1}^{d}(1-\rho_j)^k} \leq P_{k,\rho}(\text{"}r_{min} = 0\text{"}) \leq \sum_{j=1}^{d} (1 - \rho_j)^k.$$

Proof. An upper bound is provided with

$$\begin{aligned} P_{k,\rho}(''r_{min} = 0'') &\le \sum_{j=1}^{d} P_{k,\rho}(''r_j = 0'') \\ &= \sum_{j=1}^{d} B_{k,\rho_j}(\{0\}) = \sum_{j=1}^{d} (1-\rho_j)^k. \end{aligned}$$

A tight lower bound is obtained by using an inequality due to Mallows, [5], which says that the multinomial distribution obeys a strong (negative) correlation principle. We have

$$P_{k,\rho}(\bigcap_{j=1}^{d} \{r_j > 0\}) \le \prod_{j=1}^{d} P_{k,\rho}(\{r_j > 0\}).$$

Since each component is distributed binomially, this amounts to

$$P_{k,\rho}(''r_{min} > 0'') \le \prod_{j=1}^{d} (1 - (1-\rho_j)^k).$$

Applying the geometric – arithmetic mean inequality we arrive at

$$P_{k,\rho}(''r_{min} > 0'') \le (1 - \frac{1}{d}\sum_{j=1}^{d}(1-\rho_j)^k)^d \le \mathrm{e}^{-\sum_{j=1}^{d}(1-\rho_j)^k} \tag{7}$$

from which the proof can be completed. □

3. The Mixing of Q_ρ

Let us investigate the mixing behavior of the random walk Q_ρ introduced before. For this particular type of walk one can expect that the mixing behavior does not really depend on the underlying groups but rather on the number d of such. This is supported by Proposition 1 below.

Recall the definition of the multinomial distribution $P_{k,\rho}$ as introduced above. We have

Lemma 6. *For fixed d, convex combination ρ and natural k we have*

$$\prod_{j=1}^{d}(1 - \frac{1}{|G_j|})P_{k,\rho}(''r_{min} = 0'') \le \|U_\rho^{*k} - U\| \le 2P_{k,\rho}(''r_{min} = 0''). \tag{8}$$

Proof. According to Lemma 1 we have

$$U_\rho^{*k} = \left(\sum_{j=1}^{r} \rho_j \tilde{U}_j\right)^{*k}$$

$$= \sum_{r_1+\ldots+r_d=k} \binom{k}{\bar{r}} \prod_{j=1}^{d} \left(\rho_j \tilde{U}_j\right)^{r_j} \quad 1$$

$$(9) \qquad = \sum_{\substack{r_1+\ldots+r_d=k \\ r_{min}>0}} \binom{k}{\bar{r}} \prod_{j=1}^{d} \rho_j^{r_j} U + \sum_{\substack{r_1+\ldots+r_d=k \\ r_{min}=0}} \binom{k}{\bar{r}} \prod_{j=1}^{d} \left(\rho_j \tilde{U}_j\right)^{r_j}.$$

Consequently we obtain

$$\|U_\rho^{*k} - U\| \leq \| \left(1 - \sum_{\substack{r_1+\ldots+r_d=k \\ r_{min}>0}} \binom{k}{\bar{r}} \prod_{j=1}^{d} \rho_j^{r_j}\right) U\|$$

$$+ \| \sum_{\substack{r_1+\ldots+r_d=k \\ r_{min}=0}} \binom{k}{\bar{r}} \prod_{j=1}^{d} \left(\rho_j \tilde{U}_j\right)^{r_j} \|$$

$$\leq 2 \sum_{\substack{r_1+\ldots+r_d=k \\ r_{min}=0}} \binom{k}{\bar{r}} \prod_{j=1}^{d} \rho_j^{r_j}$$

$$(10) \qquad = 2P_{k,\rho}\left("r_{min} = 0"\right).$$

This proves the right–hand side inequality.

The left–hand side inequality is based on Lemma 3. Applying this estimate with $P := U_\rho^{*k}$ and

$$\alpha := P_{k,\rho}("r_{min} > 0"),$$

we infer from equation (9) above that on the set

$$A_\alpha := \prod_{j=1}^{d} (G_j \setminus \{e_j\})$$

the assumptions of Lemma 3 are fulfilled. This easily accomplishes the proof of the lemma. □

In "typical" situations with groups having many elements, say that $\prod_{j=1}^{d}(1 - \frac{1}{|G_j|})$ close to 1, the mixing behavior is completely determined by the time required to touch every component with high (occupancy) probability. This probability will be precisely estimated from both sides and leads to a description of the mixing behavior of Q_ρ.

The sharp bounds from Lemma 5 immediately yield

[1]The symbol $\prod$ in conjunction with measures denotes the convolution throughout. In view of Lemma 1 this may sometimes be identified with a product as usual.

Proposition 1. *Let ρ be a fixed convex combination. If the groups G_j are rich enough such that $\prod_{j=1}^{d}(1-\frac{1}{|G_j|}) \geq \frac{4}{5}$, then*

$$\|\delta_x Q_\rho^k - U\| \geq \frac{1}{\mathrm{e}},$$

unless $k \geq d(0.5+\log(d))$.

On the other hand, for any initial distribution ν we obtain

$$\|\nu Q_{\rho_0}^k - U\| \leq \frac{1}{2\mathrm{e}},$$

if $k \geq d(2.5+\log(d))$ by the choice of $\rho_0 = (\frac{1}{d},\ldots,\frac{1}{d})$.

Remark 2. This result has an interesting interpretation. Since the uniform distribution of a product is the product of the corresponding uniform distributions, exactly d steps are required for its generation if we sequentially choose the components. If we agree to pick components at random uniformly on $\{1,\ldots,d\}$ then the effort multiplies by approximately $1+\log(d)$.

Proof of Proposition 1. The proof is an immediate consequence of Lemmas 5 and 6. We only mention that the lower bound in (7) is minimized by letting $\rho = \rho_0 = (\frac{1}{d},\ldots,\frac{1}{d})$. In this case the sum reduces to $d\mathrm{e}^{-k/d}$ and yields with $k = d(0.5+\log(d))$ the estimate

$$1-\mathrm{e}^{-\sum_{j=1}^{d}(1-\rho_j)^k} \geq 1-\mathrm{e}^{-\mathrm{e}^{-1/2}}.$$

from which the first assertion follows by noting that under our assumptions on G we obtain

$$\|\delta_x Q_\rho^k - U\| \geq \frac{4}{5}(1-\mathrm{e}^{-\mathrm{e}^{-1/2}}) \geq \frac{1}{\mathrm{e}}.$$

On the other hand it is easy to see that with $k \geq d(2.5+\log(d))$ the desired upper bound is obtained, completing the proof of the proposition. □

4. Mixing with fixed ρ

The basic step towards determination of the mixing time on product groups is the following

Proposition 2. *Let $k \geq 1$ and ρ be fixed. For probabilities μ_ρ we have*

$$\|\mu_\rho^{*k} - U_\rho^{*k}\| \leq \sum_{j=1}^{d} \mathrm{e}^{-\frac{k\rho_j}{8}} + \sum_{j=1}^{d} \|\mu_j^{\lfloor\frac{k\rho_j}{2}\rfloor+1} - U_j\|.$$

Proof. Arguing as in the proof of Proposition 1 we obtain

$$\mu_\rho^{*k} - U_\rho^{*k} = \sum_{r_1+\ldots+r_d=k} \binom{k}{\bar{r}} \left(\prod_{j=1}^{d}(\rho_j\tilde{\mu}_j)^{r_j} - \prod_{j=1}^{d}\left(\rho_j\tilde{U}_j\right)^{r_j}\right).$$

Taking into account that for measures $\lambda_1,\ldots,\lambda_d$ and $\nu_1,\ldots,\nu_d$ on G we have

$$\prod_{j=1}^{d}\lambda_j - \prod_{j=1}^{d}\nu_j = \sum_{l=1}^{d}\left(\prod_{j=1}^{l-1}\lambda_j\right)(\lambda_l-\nu_l)\left(\prod_{j=l+1}^{d}\nu_j\right).$$

We arrive at

$$\begin{aligned}
\|\mu_\rho^{*k} - U_\rho^{*k}\| &\le \sum_{r_1+\ldots+r_d=k} \binom{k}{\bar{r}} \prod_{j=1}^{d} \rho_j^{r_j} \sum_{l=1}^{d} \|\tilde{\mu}_l^{r_l} - \tilde{U}_l^{r_l}\| \\
&= \sum_{l=1}^{d} \left(\sum_{r_l=0}^{k} \binom{k}{r_l} \rho_l^{r_l} (1-\rho_l)^{k-r_l} \|\tilde{\mu}_l^{r_l} - \tilde{U}_l^{r_l}\| \right) \\
&\le \sum_{l=1}^{d} \left(B_{k,\rho_l}(\left\{0,\ldots,\lfloor \frac{k\rho_l}{2} \rfloor \right\}) + \max_{r_l > \lfloor \frac{k\rho_l}{2} \rfloor} \|\tilde{\mu}_l^{r_l} - \tilde{U}_l\| \right) \\
&\le \sum_{l=1}^{d} \mathrm{e}^{-\frac{k\rho_l}{8}} + \sum_{l=1}^{d} \|\mu_j^{\lfloor \frac{k\rho_j}{2} \rfloor + 1} - U_j\|. \qquad (11)
\end{aligned}$$

To derive the first sum in (11) we used the well known estimate

$$B_{k,p}(\left\{0,\ldots,\lfloor \frac{kp}{2} \rfloor \right\}) \le \mathrm{e}^{-\frac{kp}{8}},$$

which is a consequence of Okamoto's result, see [?, Ch. 3.8]. The proof is complete. □

To proceed recall the definition of the mixing times $K(P)$ in (5). The main result in this section is

Theorem 1. *Assume we are given finite groups $G_1, \ldots, G_d$ with random walks $P_1, \ldots, P_d$ acting on them. Suppose that these random walks are associated probabilities $\mu_1, \ldots, \mu_d$. For a convex combination ρ we have*

$$K(P_\rho) \le 8 \left(\max_{j=1,\ldots,d} \frac{K(P_j)}{\rho_j} \right) (1 + \lfloor 1 + \log(d) \rfloor). \qquad (12)$$

Proof. Let $k \ge 8 \left(\max_{j=1,\ldots,d} \frac{K(P_j)}{\rho_j} \right) (1 + \lfloor 1 + \log(8d) \rfloor)$ be fixed. We have for any initial μ

$$\begin{aligned}
\|\mu_\rho^{*k} - U\| &\le \|\mu_\rho^{*k} - U_\rho^{*k}\| + \|U_\rho^{*k} - U\| \\
&\le 2\sum_{l=1}^{d} \mathrm{e}^{-k\rho_l} + \sum_{l=1}^{d} \mathrm{e}^{-\frac{k\rho_l}{8}} + \sum_{l=1}^{d} \|\mu_l^{\lfloor \frac{k\rho_l}{2} \rfloor + 1} - U_l\|. \qquad (13)
\end{aligned}$$

This last estimate is based on Propositions 1 and 2. By our assumption on k the first and second summands above can be bounded by $\frac{1}{8\mathrm{e}}$. It can further be deduced from this assumption that $\frac{K\rho_l}{2} \ge (1 + \lfloor 1 + \log(8d) \rfloor) K(P_l)$, such that an application of Lemma 4 yields

$$\|\mu_l^{\lfloor \frac{k\rho_l}{2} \rfloor + 1} - U_l\| \le \frac{1}{8d\mathrm{e}}$$

from which the proof can be completed. □

We study two applications. The first one is an "α–lazy" walk on $\mathbb{Z}_n$.

Example. Given $0 \leq \alpha \leq 1$, the α–lazy walk on $\mathbb{Z}_n$ is supposed to have a transition matrix given by

$$P^\alpha(\xi,\eta) = \begin{cases} \frac{1-\alpha}{2} & \text{, if } \eta = \xi \pm 1 \\ \alpha & \text{, if } \eta = \xi \end{cases},$$

hence it rests with a certain rate α. Although this walk could be studied directly using the Upper Bound Lemma, cf. [3, Ch. 3C], the following simple argument immediately provides for $n \geq 7$ and odd the bound

$$K(P^\alpha) \leq 8\frac{n^2}{1-\alpha}. \tag{14}$$

Indeed, the α–lazy walk on $\mathbb{Z}_n$ may be regarded as a product walk on $\mathbb{Z}_n \times \{0\}$ with $\rho = (\frac{1-\alpha}{2}, \alpha)$, where the first component walk is just the "busy", i.e.,

$$P_1(\xi,\eta) = \begin{cases} \frac{1}{2} & \text{, if } \eta = \xi \pm 1 \\ 0 & \text{, else} \end{cases}.$$

Now Theorem 1 together with the bound provided in [3, Ch. 3C] yields estimate (14).

Example. We further extend the busy walk to a lazy one on $\mathbb{Z}_n^d$ by letting

$$P^\alpha(x,y) = \begin{cases} \frac{1-\alpha}{2d} & \text{, if } \sum_{j=1}^d |\xi_j - \eta_j| = 1 \\ \alpha & \text{, if } y = x \end{cases}.$$

This walk also may be regarded as a product one on $\underbrace{\mathbb{Z}_n \times \ldots \times \mathbb{Z}_n}_{d\text{–fold}} \times \{0\}$ for $\rho = (\frac{1-\alpha}{2d}, \ldots, \frac{1-\alpha}{2d}, \alpha)$, such that Theorem 1 yields (for some absolute constant $C > 0$)

$$K(P^\alpha) \leq C\frac{n^2}{1-\alpha} d\log(d).$$

For the most natural choice $\alpha = \frac{1}{2d+1}$ we infer that such lazy walk does not lead to significantly longer times to approach stationarity.

5. Optimizing Efficiency with Respect to ρ

Below we allow to design our random walk P_ρ to fit the mixing properties of the components by varying ρ. A crude look at the bound provided in Theorem 1 tells that it is good to choose ρ_j proportional to the mixing times $K(P_j)$. A more closer look leads to improved estimates in case the mixing times of the components are very different.

For this purpose we introduce the following notation. Given groups G_j with random walks P_j having mixing times $K(P_j)$ we let

$$\kappa := \sum_{j=1}^d K(P_j) \quad \text{and} \quad \sigma_j := \frac{K(P_j)}{\kappa}, \quad j = 1, \ldots, d.$$

The d–tuple $\sigma = (\sigma_1, \dots, \sigma_d)$ gives rise to a probability and we let

$$H(\sigma) := -\sum_{j=1}^{d} \sigma_j \log(\sigma_j)$$

denote the entropy of σ.

The main result is

Theorem 2. *Let*

$$\rho_j := \frac{\sigma_j(3 - \log(\sigma_j))}{H(\sigma) + 3},$$

such that this provides a convex combination $\bar{\rho}$. This specific combination $\bar{\rho}$ leads to

$$\inf_{\rho} K(P_\rho) \leq K(P_{\bar{\rho}}) \leq \lfloor 8\kappa(H(\sigma) + 3)\rfloor + 1. \tag{15}$$

Proof. Let $k \geq \lfloor 8\kappa(H(\sigma) + 3)\rfloor + 1$. With the choice of $\bar{\rho}$ we arrive at

$$k\rho_j \geq \frac{k\rho_j}{8} \geq 3 - \log(\sigma_j), \quad j = 1, \dots, d.$$

Arguing as in the proof of Theorem 1 we arrive at an estimate like in (13) and can bound the first two sums by e^{-3}. To bound the third sum we observe

$$\lfloor \frac{k\rho_j}{2} \rfloor + 1 \geq \frac{k\rho_j}{2} \geq 4\lfloor 3 - \log(\sigma_j)\rfloor K(P_j),$$

such that Lemma 4 yields

$$\sum_{j=1}^{d} \|\mu_l^{\lfloor \frac{k\rho_l}{2} \rfloor + 1} - U_l\| \leq \sum_{j=1}^{d} \mathrm{e}^{-4(3-\log(\sigma_j))+1} \leq \mathrm{e}^{-11},$$

thus the overall error can be bounded by $\frac{1}{2\mathrm{e}}$. □

Of course, the above result lacks of an appropriate lower bound. As the discussion concerning $\mathbb{Z}_2^d$ in the Introduction and Lemma 6 suggest, some assumption on the richness of the components has to be made. This has to be clarified in future research.

Let us close mentioning that optimizing the α–lazy walk on $\mathbb{Z}_n^d$ with respect to the choice of α yields quickest convergence for $\alpha = 0$, hence the busy one.

References

[1] D. Aldous. Random walks on finite groups and rapidly mixing markov chains. In *Seminaire de Probabilites XVII*, volume 986 of *Lect. Notes Math.*, pages 243 – 297, Berlin, Heidelberg, New York, London, Paris, Tokyo, 1983. Springer.

[2] D. Aldous and P. Diaconis. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, 8:69 – 97, 1987.

[3] P. Diaconis. *Group Representations in Probability and Statistics*, volume 11 of *Lect. Notes–Monogr. Series.* Instit. Math. Statist., Hayward, 1988.

[4] N. L. Johnson and S. Kotz. *Distributions in Statistics: Discrete Distributions.* John Wiley & Sons, New York, Chichester, Brisbane, 1969.

[5] C. L. Mallows. An inequality involving multinomial distributions. *Biometrika*, 55:422–424, 1968.

[6] G. Winkler. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, volume 27 of *Appl. Math.* Springer, Berlin, Heidelberg, 1995.

Peter Mathé, Weierstraß Institute for Applied Analysis and Stochastics, Mohrenstraße 39, D–10117 Berlin, Germany

E-mail address: mathe@wias-berlin.de

Recent publications of the Weierstraß–Institut für Angewandte Analysis und Stochastik

Preprints 1996

255. Grigori N. Milstein: Random walk for elliptic equations and boundary layer.

256. Lutz Recke, Daniela Peterhof: Abstract forced symmetry breaking.

257. Lutz Recke, Daniela Peterhof: Forced frequency locking in $\mathbf{S}^1$–equivariant differential equations.

258. Udo Krause: Idealkristalle als Abelsche Varietäten.

259. Nikolaus Bubner, Jürgen Sprekels: Optimal control of martensitic phase transitions in a deformation–driven experiment on shape memory alloys.

260. Christof Külske: Metastates in disordered mean field models: random field and Hopfield models.

261. Donald A. Dawson, Klaus Fleischmann: Longtime behavior of a branching process controlled by branching catalysts.

262. Tino Michael, Jürgen Borchardt: Convergence criteria for waveform iteration methods applied to partitioned DAE systems in chemical process simulation.

263. Michael H. Neumann, Jens–Peter Kreiss: Bootstrap confidence bands for the autoregression function.

264. Silvia Caprino, Mario Pulvirenti, Wolfgang Wagner: Stationary particle systems approximating stationary solutions to the Boltzmann equation.

265. Wolfgang Dahmen, Angela Kunoth, Karsten Urban: Biorthogonal spline–wavelets on the interval – Stability and moment conditions.

266. Daniela Peterhof, Björn Sandstede, Arnd Scheel: Exponential dichotomies for solitary–wave solutions of semilinear elliptic equations on infinite cylinders.

267. Andreas Rathsfeld: A wavelet algorithm for the solution of a singular integral equation over a smooth two–dimensional manifold.

268. Jörg Schmeling, Serge E. Troubetzkoy: Dimension and invertibility of hyperbolic endomorphisms with singularities.

269. Erwin Bolthausen, Dmitry Ioffe: Harmonic crystal on the wall: a microscopic approach.

270. Nikolai N. Nefedov, Klaus R. Schneider: Delayed exchange of stabilities in singularly perturbed systems.

271. Michael S. Ermakov: On large and moderate large deviations of empirical bootstrap measure.

272. Priscilla E. Greenwood, Jiaming Sun: On criticality for competing influences of boundary and external field in the Ising model.

273. Michael S. Ermakov: On distinguishability of two nonparametric sets of hypothesis.

274. Henri Schurz: Preservation of probabilistic laws through Euler methods for Ornstein–Uhlenbeck process.

275. Lev B. Ryashko, Henri Schurz: Mean square stability analysis of some linear stochastic systems.

276. Nikolaus Bubner, Jan Sokołowski, Jürgen Sprekels: Optimal boundary control problems for shape memory alloys under state constraints for stress and temperature.

277. Alison M. Etheridge, Klaus Fleischmann: Persistence of a two–dimensional super–Brownian motion in a catalytic medium.

278. Sergej Rjasanow, Wolfgang Wagner: Stochastic interacting particle systems as a numerical tool.

279. Vadim G. Bondarevsky: Energetic systems and global attractors for the 3D Navier–Stokes equations.

280. Henri Schurz: The invariance of asymptotic laws of stochastic systems under discretization.

281. Michael Nussbaum: The Pinsker bound: a review.

282. Pierluigi Colli, Maurizio Grasselli, Jürgen Sprekels: Automatic control via thermostats of a hyperbolic Stefan problem with memory.

283. Anton Bovier, Véronique Gayrard: An almost sure central limit theorem for the Hopfield model.