Elect. Comm. in Probab. **5** (2000) 29–37

ELECTRONIC COMMUNICATIONS in PROBABILITY

TRANSITION PROBABILITY ESTIMATES FOR REVERSIBLE MARKOV CHAINS

ANDRÁS TELCS
International Business School Budapest,
Tárogató út 2-4, 1021 Budapest, HUNGARY.
email: `atelcs@tech.ibs-b.hu`

submitted January 25, 1999; *revised* November 25, 1999;*final version accepted* January 3, 2000.

AMS 1991 Subject classification: 60J10, (60J35, 60J45)
Random walks, Markov chains, fractals, dimensions

Abstract:

This paper provides transition probability estimates of transient reversible Markov chains. The key condition of the result is the spatial symmetry and polynomial decay of the Green's function of the chain.

1 Introduction

Significant developments have taken place in the study of transition probability estimates of reversible Markov chains since the first major step by Varopoulos [V]. These results can be separated into two groups. One deals with random walks on groups, the other with random walks on fractals. The analysis of random walks on groups utilizes the group structure in several ways. Earlier results were based on the generalization of the Fourier techniques on Abelian groups, and later works extended the harmonic analysis and its consequences to several types of groups. A good reference in this area is [VSC], and further generalization to transitive and almost transitive Markov chains is given in [S]. The common consequence of this approach is that the walk travels with speed of $n^{1/2}$ similarly to the Simple Symmetric Random Walk on the integer lattice Z^d. The fundamental result of Gromov and Bass [VSC, c.f. Theorem VI.2.1. and 2.2] has the consequence in this context that the volume growth of the geometric balls is restricted to integer powers. Both limitations vanish in the context of random walks on fractals. The speed of the walk is $n^{\frac{1}{d_R}}$, where $d_R \geq 2$ is the walk exponent and other characteristic exponents related to the underlying structure and the random walk are not necessarily integers (Other works (c.f. [B]) denote the walk exponent by d_w, here we use d_R following [T1] and [XYZ]) . Several early results in this direction are rooted in the self similar structure of the fractals, which makes possible some renormalization (c.f. bibliography of [B]). The difficulties start if either the fractal cannot be renormalized with a simple operation and/or if the fractal is infinitely ramified. Barlow and Hambly [BH] have obtained a new result in this direction, but they still need strong local symmetry. Varopoulos' works and the book by Doyle and Snell [DS] raise the question of finding a setting in which the limitations of the group and fractal approach can be weakened or even neglected. This is one of the aims of this paper. Let us recall Varopoulos' celebrated result [V].

Consider a Markov chain on state space V, which is reversible with respect to measure $\mu(x) > 0$ and has transition probabilities $p_{x,y}$, $x, y \in V$. Thus

$$\mu(x)p_{x,y} = \mu(x)p_{y,x} =: \mu_{x,y}$$

for all $x, y \in V$.

Theorem (Varopoulos) *For the reversible Markov chain, if the Dirichlet inequality holds, i.e.*

$$\|f\|_r \leq c_1 \|f\|_D \tag{1}$$

for $r = \frac{2\nu}{\nu-2}$, *where* $\nu \geq 2$ *and for all* $f \in C_0(V)$ *then*

$$\sup_{x,y\in V} \mu(y)^{-1} P_n(x,y) = O(n^{-\frac{\nu}{2}}). \tag{2}$$

and conversely (2) implies (1) if $\nu > 2$.

Here, $\|f\|_D = \left(\frac{1}{2}\sum \mu_{x,y}(f(x)-f(y))^2\right)^{\frac{1}{2}}$ denotes the Dirichlet norm of the function $f \in C_0(V)$(=the space of functions of compact support on V) and $\|f\|_r = \left(\sum_{x\in V}\mu(x)|f(x)|^r\right)^{\frac{1}{r}}$. The n step transition probabilities are denoted by $P_n(x,y)$.

In this result the transition probability estimate is controlled by one parameter ν. From the earlier studies of random walks on fractals it is known that the walk is ruled by two parameters, by the fractal dimension and an exponent describing the conductivity of the structure (in the physics literature [RT] or for rigorous results ([T1]-[T4])). The second aim of this paper is to present an estimate of the transition probability of type (2) involving both parameters.

It is also worthwhile to remark here that in the same work [V, Proposition 1] Varopoulos has shown that (1) is equivalent to

$$\|Gf\|_r \leq c_2 \|f\|_s \tag{3}$$

where s and r are Hölder conjugates and G stands for Green's function of the Markov chain (see definition below). This shows that the Green's function plays a central role in the characterization of the chain. As it will be shown in this paper, the power of the decay to zero of the Green's function, i.e. the Greenian index and the fractal dimension together provide a good description of the Markov chain (provided they exist).

1.1 Acknowledgment

The author is indebted to B. Virág and B. Hambly for the careful reading of the manuscript and helping to improve it. The author wish to express his thanks to G. Elek for the useful, inspiring discussions. The author wishes to thank the referee and editors for their helpul suggestions.

2 Introductory definitions and results

Let X_n be a discrete time reversible Markov chain of state space V and transition matrix $P = (p_{x,y})$. Assume that the measure is bounded i.e. for all $x \in V : 0 < c \leq \mu(x) \leq C < \infty$ for a fixed c and C.

This introduction of the Markov chain fully coincides with the definition of random walk on a graph $G = (V, E)$, where $E = \{(x, y) \text{ where } \mu_{x,y} := \mu(x)p_{x,y} > 0\}$ is the edge set, $\mu_{x,y}$ the conductivity assigned to the edges (c.f. [KSK, 9.10.]). In earlier works the RW terminology was used ([T1]-[T4]), here the Markov chain language is preferred.

On V the shortest path metric will be used and denoted by $d(x, y)$. The geometric balls are defined by

$$B_{x,N} = \{y \in V : d(x, y) \leq N\}$$

and the μ-volume by

$$b_{x,N} = \sum_{z \in B_{x,N}} \mu(z)$$

also the surface of the ball by $S_{x,N} = B_{x,N} \backslash B_{x,N-1}$ for $N > 0$. If the center of the ball is not important it will be omitted and the shorter notation $b_N = b_{x,N}$ will be used. The same convention will be used for other quantities without any comment if there is no danger of confusion. The transition probability matrix restricted to $B_{x,N}$ (for $x \in V, N > 0$) will be denoted by

$$Q_N(y, z) = Q_{x,N}(y, z) = \begin{cases} P(y, z) & \text{for } y, z \in B_{x,N} \\ 0 & \text{otherwise} \end{cases}$$

and its iterates by $Q^n(y, z) = Q_N^n(y, z) = Q_{x,N}^n(y, z)$ for $n \geq 1$. It is clear that Q_N^n are substochastic matrices.

For a given label set Γ, two sequences a_γ, b_γ we use the notation $a_\gamma \asymp b_\gamma$, if there is a $c > 1$ such that for all $\gamma \in \Gamma$, $c^{-1} \leq a_\gamma / b_\gamma \leq c$.

Definition 1 A Markov chain has Greenian index $\beta > 0$ if the Green's function

$$G(x, y) = \sum_n P_n(x, y)$$

satisfies uniform convergence to zero, i.e.,

$$G(x, y) \asymp d(x, y)^{-\beta} \tag{4}$$

for $x \neq y \in V$.

Remark 1 This is a strong restriction and for instance scale irregular fractals ([BH]) do not necessarily have Greenian index, but on the other hand polynomially growing graphs which are either vertex transitive or which have Martin boundary containing only one point have Greenian index.

Theorem 1 *If a reversible Markov chain has regular growth rate $d > 0$, i.e.,*

$$b_N \asymp N^d \tag{5}$$

and has Greenian index $\beta > 0$, i.e., $G(x,y) \asymp d(x,y)^{-\beta}$, then, there is a $c_3 > 0$ such that for all $N > 0$ and $n \geq N^{d_R}$,

$$Q_N^n(x,y) \leq c_3 n^{-\frac{d}{d_R}}, \tag{6}$$

for all $x, y \in V$, where $d_R = d - \beta$.

Remark 2 The conditions are stronger than those assumed in Varopoulos' Theorem. On the other hand as it is pointed out at the end of the introduction (3) shows that Varopoulos' result is also based on a condition on the Green's functions. As a return a relatively simple proof can be given and the role of d and β is clear. A large class of Markov chains has this property, such random walks defined on vertex transitive graphs (e.g. polynomially growing Cayley graphs) with $\mu \equiv 1$. The next theorem provides an on-diagonal lower bound and "non-Gaussian" upper bound for small times (relative to $d(x,y)^{d_R}$).

Theorem 2 *Under the conditions of Theorem 1 there are $c_i > 0, i = 4,5,6$ constants such that*

$$P_{2n}(x,x) \geq c_4 n^{-\frac{d}{d_R}}$$

for all $x \in V$, $n \geq 0$ where $d_R = d - \beta$, and

$$P_n(x,y) \leq c_5 \exp\left(-c_6 \left(\frac{d(x,y)^{d_R}}{n}\right)^{\frac{1}{d_R - 1}}\right) \tag{7}$$

for all $x, y \in V, n \geq 0$.

Remark 3 The restriction $\beta > 0$ ensures transience but the results are also true for strongly recurrent reversible Markov chains.(c.f. [T3]). The case of $\beta = 0$ i.e. the weak recurrence would not need any new concepts but involves technical difficulties and hence is not studied here.

3 Further definitions and remarks

The stopping time $T_{x,N} = \min\{k : X_k \in S_{x,N}\}$ is an important object in the analysis of random walks. The expected escape time is denoted by

$$e_N = e_{x,N} = E(T_{x,N}).$$

where the starting point $x \in V$ is suppressed if this does not lead to confusion.

The killed random walk (at time $T_{x,N}$) has the substochastic transition matrix $Q_{x,N} = P|_{B_{x,N}}$ which is the restriction of P to $B_{x,N}$. The study of the Markov chain needs the eigenvalue properties of the Laplacian $I - Q_N$ (where $I = I_N$ is also restricted to the same ball as P). The smallest eigenvalue of the Laplacian can be defined as

$$\mu_N = \inf_{(\Phi,\Phi)\leq 1} \frac{(\Phi, (I - Q_N)\Phi)}{(\Phi, \Phi)}. \tag{8}$$

where (f,h) is the inner product of the space $l_2(B_N)$

Proposition 1 *Under the condition of Theorem 1.*

$$e_N \asymp N^{d-\beta} \tag{9}$$

Proof The proof is easy, based on summation of the local times of states in $B_{x,N}$ using the property of G, (The statement also follows from the combination of [T1] and [T4, Theorem 2.].) $\square$

The relation (9) shows that the Einstein relation

$$d_R = d - \beta,$$

holds, whereas in earlier works [T1-T4] $\beta = d_\Omega - 2$ was used.

Lemma 1 *Under the condition of Theorem 1. there is a constant $c_7 > 0$ such that*

$$\mu_{x,N} \geq \frac{c_7}{E_x(T_{x,N})}. \tag{10}$$

Proof Consider the normed left eigenvector α ($\sum_{y\in B_{x,N}} \alpha_y = 1$) of $I - Q_N$ belonging to μ_N.

Starting the process from α it follows that

$$\mu_{x,N}^{-1} = \sum_{y\in B_{x,N}} \alpha_y E_y(T_{x,N}) \leq \max_{y\in B_{x,N}} E_y(T_{x,N}) \tag{11}$$

$$\leq \max_{y\in B_{x,N}} E_y(T_{y,2N}) \leq c_8 E_x\left(T_{x,N}\right) \tag{12}$$

where the last inequality uses (9) the uniform upper and lower estimates of the mean escape time i.e. the whole set of the conditions of Theorem 1. $\square$

The Markov chains with transition probability matrixes P^2 and $Q_{w,N}^2$ are also reversible. It is also worth mentioning that the Laplacian $I - Q_{w,N}^2$ has smallest eigenvalue $\rho_{w,N}$

$$\rho = \rho_{w,N} = 1 - (1-\mu_{w,N})^2 = \mu(2-\mu) > \mu \tag{13}$$

if $1 - \mu = 1 - \mu_{w,N}$ is the largest eigenvalue of the substochastic matrix $Q_{w,N}$.

Proposition 2 *For the transition probabilities, there is a $c_9 > 0$, such that for all $k, l > 0$, $x, y \in V$,*

$$Q_N^{k+l}(x,y) \leq c_9 \left[Q_N^{2k}(x,x)\right]^{1/2} \left[Q_N^{2l}(y,y)\right]^{1/2}. \tag{14}$$

Proof The statement follows from the Schwarz inequality, from the reversibility of the chain and from $\mu(z) \leq C$ for all $z \in V$.

$$\begin{aligned} Q_N^{k+l}(x,y) &= \sum_{z\in B_N} Q_N^k(x,z) Q_N^l(z,y) \\ &\leq \left(\sum_{z\in B_N} Q_N^k(x,z)^2\right)^{\frac{1}{2}} \left(\sum_{z\in B_N} Q_N^l(z,y)^2\right)^{\frac{1}{2}} \\ &\leq c_9 Q_N^{2k}(x,x)^{1/2} Q_N^{2l}(y,y)^{1/2} \end{aligned}$$

□

Lemma 2 *For all $x \in V$, $n > 0$, $N \geq 0$*

$$Q_N^{2n+1}(x,x) \leq Q_N^{2n}(x,x) \tag{15}$$

Proof This is immediate consequence of the eigenvector decomposition. □

4 Proof of the results

Proof of Theorem 1 Let us fix $w \in V$ and $N > 0$ and $x \in B_{w,N}$. Let $u_k(y) = Q^k(x,y) = Q^k_{w,N}(x,y)$ for all $y \in B_{w,N}$. Define

$$g(k) = g_N(k) = Q_N^{2k}(x,x)$$

and

$$\overline{u}_N = \overline{u}_{w,N}(y) = \begin{cases} b_{w,N}^{-1} \sum_{z \in B_{w,N}} u_k(z) : & y \in B_{w,N} \\ 0 : & \text{elsewhere} \end{cases}$$

From the eigenvalue decomposition it follows that $g(k)$ is decreasing in k.

$$\begin{aligned} g(2k) - g(2k+1) &= \sum_{y \in B_{w,N}} \left(Q^{2k}(x,y) - Q^{2k+2}(x,y)\right) Q^{2k}(y,x) \\ &\geq \sum_{y \in B_{w,N}} \left(u_{2k}(y) - u_{2k+2}(y)\right)(1+c_9)\frac{c_x}{c_y} u_{2k}(y) \\ &= c_{10}(u_{2k}(I-Q^2), u_{2k}) \end{aligned}$$

In order to use the definition of the eigenvalue (8) of the Laplacian one has to observe that $\|u_{2k}\|_1 \leq 1$ and consequently

$$(u_{2k}, u_{2k}) \leq 1.$$

The estimation can then proceed as follows using (13)

$$\begin{aligned} g(2k) - g(2k+1) &\geq c_{10}(u_{2k}(I-Q^2), u_{2k}) \\ &\geq c_{10}\mu_N(u_{2k}, u_{2k}) \geq c_{10}\mu_N(u_{2k} - \overline{u}_N, u_{2k} - \overline{u}_N) \\ &\geq c_{10}\mu_N\left((u_{2k}, u_{2k}) - b_{w,N}^{-1}\right) \\ &\geq c_{10}\mu_N\left(c_{11}g(2k) - b_{w,N}^{-1}\right), \end{aligned}$$

where the last inequality uses the reversal of the transition probabilities. Introducing $s_J = \inf\{k : c_{11}g_N(2k) \leq b_{w,J}^{-1}\}$ and

$$h(k) = c_{11}g_N(k) - b_{w,N}^{-1},$$

it can be written as

$$h(2k) - h(2k+1) \geq c_{11}\mu_N h(2k).$$

For $h(2k) > 0$, the following holds:

$$\frac{h(2k) - h(2k+1)}{h(2k)} \geq c_{11}\mu_N \tag{16}$$

Denote $c_b = \max_{w \in V, N>0}\{\frac{b_{w,N}}{N^d}, \frac{N^d}{b_{w,N}}\}$. Now an exponential sequence is introduced. For a fixed $q \geq c_b^2 > 1$, let $N_i = q^i$ which provides easy handling of $b_i = b_{N_i}$, $s_i = s_{N_i}$, $e_i = e_{N_i}$ and $g_n = g_{N_n}$. From the monotonicity of g and h we have that $s_{i-1} \leq s_i$ for all $i \geq 1$ and (16) is true for all $0 < k < s_n$,with $N = N_n$. Summing (16) from s_{n-2} to s_{n-1}, one has

$$\begin{aligned}
\mu_n(s_{n-1} - s_{n-2}) &\leq c_{12} \sum_{i=s_{n-2}}^{s_{n-1}} \frac{h(2i) - h(2i+1)}{h(i)} \\
&\leq c_{12} \sum_{i=s_{n-2}}^{s_{n-1}} \frac{g(2i) - g(2i+2)}{c_{13}g(2i) - b_n^{-1}}
\end{aligned}$$

Using the trivial lower estimate of the denominator and collapsing the sum it follows that

$$\begin{aligned}
\mu_n(s_{n-1} - s_{n-2}) &\leq c_{12} \frac{g(2s_{n-2})}{c_{13}g(2s_{n-1} - 2) - b_n^{-1}} \\
&\leq c_{13} \frac{b_{n-2}^{-1}}{c_{13}b_{n-1}^{-1} - b_n^{-1}} \leq c_{14}
\end{aligned}$$

Hence, by Lemma 1, $s_{n-1} \leq c_{15}\mu_n^{-1} \leq c_{16}e_n$, which means that for all $k : c_{16}e_n \leq k \leq c_{16}e_{n+1}$,$g(2k) \leq c_{17}b_n^{-1}$. I.e., for all $c_{21}e_n \leq k \leq c_{21}e_{n+1}$,

$$g_n(2k) \leq c_{18}k^{-\frac{d}{d_R}}$$

Using the definition of $g(k) = g_{N_n}(k)$,

$$g_n(2k) \geq c_{19}Q^{4k}_{w,N_n}(x,x).$$

Rescaling this with $e_{N_n} \asymp N_n^{d_R}$ and $b_n^{-1} \asymp N_n^{-d}$ and adjusting the constants it follows that for $k > e_N$

$$Q^{2k}_{w,N}(x,x) \leq c_{20}k^{-\frac{d}{d_R}}.$$

Using the observations (14),(15) the proof is complete. □

Proposition 3 *Under the conditions of Theorem 1,*

$$P_x(T_{x,N} > n) > c_{21} \tag{17}$$

if $n \leq \frac{1}{2}E_x(T_{x,N})$ *where* $c_{21} > 0$ *is independent of* N *and* x*.*

Proof From the easy estimate,

$$T_N \leq n + I(T_N \geq n)(T_N - n),$$

we have

$$E(T_N) \le n + E\left(I(T_N > n)E_{X_n}(T_N)\right);$$

we have used the strong Markov property for $T_N > n$ for $X_n \in B_{x,N}$. Finally, by the uniform upper estimate of the mean exit time, one has

$$P(T > n) \ge \frac{1}{c_{22}} - \frac{n}{c_{22}E(T)} =: c_{22}. \tag{18}$$

This concludes our proof. □

Proof of Theorem 2 The lower diagonal estimate for the Markov chain follows simply from (17) as in [T3]. Let N be chosen to ensure $c_{23}N^{d_R} \ge \frac{1}{2}E_x(T_{xN}) \ge n \ge c_{24}N^{d_R}$, then

$$\begin{aligned} P_{2n}(x,x) &\ge Q_N^{2n}(x,x) \\ &\ge c_{25} \sum_{y \in B_{x,N}} Q_N^n(x,y)^2 \\ &\ge b_N^{-1}(\sum_{y \in B_{x,N}} Q_N^n(x,y))^2 = b_N^{-1} P(T_N > n)^2 \\ &\ge c_{22}^2 b_N^{-1} \ge c_{26} n^{-\frac{d}{d_R}} \end{aligned}$$

The proof of (7) the tail estimate for the stopping time $T_{x,N}$ contained in [T4] but included here for sake of completeness. Denote $N = d(x,y)$. Clearly $P_n(x,y) \le P(T_{x,N} \le n)$. Using the triangle inequality for the graph distance metric T_N the hitting time of $S_{0,N}$ can be decomposed. Let $r > 0$ and $l = \{N/r\}$ integers and define r hitting times in the following way. Let

$$\begin{aligned} \tau_1 = s_1 &= \min\{k : X_k \in S_{0,l}\}, \\ Y_1 &= X_{\tau_1} \\ \tau_i &= \min\{k > \tau_{i-1} : X_k \in S_{y_{i-1},N/r}\}, \\ Y_i &= X_{\tau_i}, \\ s_i &= \tau_i - \tau_{i-1} \end{aligned}$$

for $r \ge i > 1$. It is clear that

$$T_N > \sum_1^r s_i.$$

from (17) and ([B]) we have

$$P_x(T_{x,l} \le s) < 1 - c_{27}(1 - \frac{1}{c_{28} e_l})^s < 1 - c_7 + c_8 \frac{s}{e_l} =: p + \frac{s}{b}$$

with $0 \le p < 1, b > 0$. From this using [BB, Lemma 1.1] it follows that

$$P(T_N < t) < \exp\left(2(\frac{brt}{p})^2 - n \log \frac{1}{p}\right)$$

if $t = \epsilon e_N$, $r = c_{28}\epsilon^{-\frac{1}{d_R - 1}}$. With the proper choice of c_{28} it results that in the above exponent the second term is the dominant. □

References

[B] Barlow, M.T. Random Walks and Diffusion on Fractals, Proc. Int. Congress Math. Kyoto, 1990

[BB] Barlow, M.T., Bass, F.R., The Construction of the Brownian Motion on the Sierpisnki Carpet, Ann. Inst. H. Poincaré, 25, (1989) 225-257

[BH] Barlow, M.T., Hambly, B., Transition density estimates for Brownian motion on scale irregular Sierpinski gaskets, Ann. I. H. Poincaré,33,1997, 531-559

[BP] Barlow, M.T., Perkins, E.A. Brownian Motion on the Sierpinski Gasket, Prob. Th. Rel. Fields, 79, (1988) 543-623

[CKS] Carlen, E.A., Kusuoka, S., Stroock, D.W., Upper bounds for symmetric Markov transition functions, Ann. Inst. H. Poincaré, (1987), 23, 245-287

[DS] Doyle, P.G., Snell, J.L., Random walks and electric networks, The Carus Mathematical Monographs, 22, (1984)

[KSK] Kemeny, G.J., Snell, J.L., Knapp, A.W., Denumerable Markov Chains, 2. ed. Graduate text in Mathematics 40, Springer, 1976

[KZ] Kusuoka, S., Zhou, X.Y., Dirichlet form on fractals: Poincaré constant and resistance, J. Theor. Prob. (1992), 93, 169-196

[RT] Rammal, R., Toulouse, G., random Walks on fractal structures and percolation clusters, J. physique Lett.(1983), 44, L13 - L22

[S] Saloff-Coste, L., Isoperimetric Inequalities and decay of iterated kernels for almost-transitive Markov chains, Combinatorics, Probability and Computing, 4., (1995), 419-442

[T1] Telcs, A., Random Walks on Graphs, Electric Networks and Fractals, J. Prob. Theo. and Rel. Fields, (1989), 82, 435-449

[T2] Telcs, A., Spectra of Graphs and Fractal Dimensions I., Prob. Theo. and Rel. Fields, (1990), 85, 489-497

[T3] Telcs, A., Spectra of graphs and Fractal dimensions II, J. Theor. Prob. (1995), 8, 77-96

[T4] Telcs, A. Fractal Dimension and Martin Boundary of Graphs, to appear in Studia Sci. Math. Hu.

[VSC] Varopoulos, N., Saloff-Coste, L., Coulhon, Th., Analysis and geometry on Groups, Cambridge University Press, 1993

[V] Varopoulos, N., Isoperimetric inequalities and Markov Chains, J. Functional Analysis 63, (1985) 215-239

[XYZ] Zhou, Z.Y., Resistance Dimension, Random Walk Dimension and Fractal Dimension, J. Theor. Prob., 1993, 6., 4., 635-652