Canad. J. Math. Vol. **53** (5), 2001 pp. 1057–1120

# Potential Theory in Lipschitz Domains

N. Th. Varopoulos

*Abstract.* We prove comparison theorems for the probability of life in a Lipschitz domain between Brownian motion and random walks.

*Résumé.* On donne des théorèmes de comparaison pour la probabilité de vie dans un domain Lipschitzien entre le Brownien et de marches aléatoires.

## 0 Introduction

### 0.1 The Statement of the Main Theorem

Let $D \subset \mathbb{R}^d$ be some domain and let $\big(b(t) \in \mathbb{R}^d \ ; \ t > 0\big)$ be standard brownian motion (*cf.* [1]) normalized for $b(1)$ to have co-varience 1. We shall denote:

$$\tau = \inf[s \, ; b(s) \notin D], \tag{0.1}$$

$$p_t(x,y)dy = p_t^D(x,y)dy = \mathbb{P}_x[b(t) \in dy, \tau > t], \tag{0.2}$$

$$P(t,x) = P_D(t,x) = \mathbb{P}_x[\tau > t]. \tag{0.3}$$

We shall also consider:

$$\big(z(n) \in \mathbb{R}^d \ ; n \geq 0\big), \tag{0.4}$$

the random walk that is defined by

$$\mathbb{P}[z(n+1) \in dy // z(n) = x] = d\mu(y-x), \tag{0.5}$$

where $\mu \in \mathbb{P}(\mathbb{R}^d)$ is a centered probability measure with the same co-varience as $b(1)$, *i.e.*,

$$\int x \, d\mu(x) = 0; \quad \int x_i x_j \, d\mu(x) = \delta_{ij}; \quad i, j = 1, \ldots, d. \tag{0.6}$$

We shall impose on $\mu$ the following moment condition

$$\int |x|^B \, d\mu(x) \leq M_B < +\infty, \tag{$M_B$}$$

Received by the editors July 11, 2000; revised January 29, 2001.
AMS subject classification: Primary: 39A70, 35-02; secondary: 65M06.

for some appropriate $B \geq 0$. For this random walk, we shall define the analogous quantities (*cf.* [28]):

$$\tau = \inf[s = 0, 1, \dots ; z(s) \notin D], \tag{0.7}$$

$$P_\mu(n, x) = \mathbb{P}_x[\tau > n], \tag{0.8}$$

$$p_n^\mu(x, y)dy = \mathbb{P}_x[z(n) \in dy ; \tau > n]. \tag{0.9}$$

Let us now assume that there exists $A > 0$ and $\varphi$ some Lipschitz function on $\mathbb{R}^{d-1}$, such that

$$\begin{aligned} &|\varphi(x_1') - \varphi(x_2')| \leq A|x_1' - x_2'| ; \quad x_1', x_2' \in \mathbb{R}^{d-1}, \\ &D = [(x_1, x') \in \mathbb{R} \times \mathbb{R}^{d-1} = \mathbb{R}^d ; x_1 > \varphi(x')]. \end{aligned} \tag{0.10}$$

We then say that $D \subset \mathbb{R}^d$ is a globally Lipschitz domain. For any domain $D \subset \mathbb{R}^d$ the following notation will also be used throughout:

$$\delta(x) = \operatorname{dist}(x, \partial D) ; \quad \delta(t, x) = \operatorname{Min}[\delta(x), \sqrt{t}] ; \quad t > 0, x \in D.$$

***The Main Theorem*** *Let D be some globally Lipschitz domain, then for all* $0 < \varepsilon \leq \frac{4}{5}$ *there exists* $B > 0$, *a positive constant, that only depends on d and A,* $\varepsilon$, *such that for all* $\mu \in \mathbb{P}(\mathbb{R}^d)$ *that satisfies (0.6) and* ($M_B$), *we have*

$$|P(t, x) - P^\mu(t, x)| \leq C \frac{P(t, x)}{\delta(t, x)^\varepsilon} ; \quad \delta(x) \geq C, t \geq C, \tag{0.11}$$

*where* $C > 0$ *only depends on d, A, B,* $M_B$ *and* $\varepsilon$. *If the measure is compactly supported we can take* $0 < \varepsilon < 1$.

The lower estimate in (0.11) works for $0 < \varepsilon < 1$ in general (*cf.* (4.34)). The $\varepsilon \leq \frac{4}{5}$ in (0.11) is of course technical. But the possibility for $\varepsilon$ to go all the way up to 1 is related to the compactness of the support (*cf.* [26]). At any rate the optimal aspect of the above theorem will be discussed in Section 0.3.

The above theorem has a counterpart in the context of homogenization theory (*cf.* [10], [11], [9]) and in the context of Markov chains in random environment (*cf.* [12], [9]). These theorems will be stated and proved in Sections 5 and 6 below and they are important because they have applications to the theory of Lie groups (*cf.* [4], [5], [7], [9], [13], [14]). It is this work on Lie groups that lead me to the above Theorem. The above Main Theorem generalizes to *space inhomogeneous* random walks as in [5] but the proofs are just a formal and easy extention of the proofs given here, and they will be left to the interested reader.

The standard notational convention that consists in using the letter $C$ or $c$, possibly with suffixes, to indicate positive constants that only depend on the *important parameters* will be used throughout. These $C$, $c$ could differ from place to place, even in the same formula, and to avoid becoming cumbersome, I shall not always state explicitly which are these *important parameters* on which $C$, $c$ depend. Hopefully this will be clear from the context.

Before moving on to describe some of the more technical aspects of this paper I will state the following.

***The Kernel Estimates*** *The kernel (0.2) satisfies*

(0.12)
$$p_t(x, y) \le C_\varepsilon t^{-d/2} P(t, x)P(t, y) \exp\left(-\frac{|x-y|^2}{(4+\varepsilon)t}\right); \quad t > 0,\ x, y \in D,\ \varepsilon > 0,$$

*where $D$ is a globally Lipschitz and $C_\varepsilon$ depends on $d$, $A$ and $\varepsilon$.*

This is, in fact, an automatic consequence of the easy estimate:

(0.13)
$$p_t(x, y) \le Ct^{-d/2} P(t/3, x)P(t/3, y)$$

(a proof of this can be found in [7, p. 651]), and of Davies' work in [6], together with one of the main technical results in this paper (this is proved in Section 2 below):

(0.14)
$$-\frac{C}{t} P(t, x) \le \frac{\partial}{\partial t} P(t, x) \le 0, \quad t > 0, \quad x \in D.$$

A similar estimate can be proved for the kernel (0.9)

(0.15)
$$p^\mu_t(x, y) \le Ct^{-d/2} P_\mu(t, x)P_\mu(t, y) \exp\left(-\frac{|x-y|^2}{Ct}\right);$$
$$t > C,\ x, y \in D,\ \delta(x), \delta(y) > C,\ \operatorname{supp} \mu \subset [|x| \le \delta].$$

The dependence of $C$ in (0.15) is $C = C(d, A)$, where $\mu$ is as in the Main Theorem. But, of course, the estimate (0.15) *cannot* hold for an arbitrary $\mu \in \mathbb{P}(\mathbb{R}^d)$. It is clear that for (0.15) to hold, $\mu$ itself *has* to have already a Gaussian decay at infinity. But we also *have* to impose some smoothness on $\mu$ (*cf.* [5, Sections 0.7, 0.8], and Section 8.2 below). Say $d\mu = f\,dx$, with $f \in L^\infty$ or even $L^2$, or that the random walk (0.4) is a pure lattice walk and $\mu \in \mathbb{P}(\mathbb{Z}^d)$. In the second case the definition of the kernel $p^\mu(\cdot, \cdot)$ in (0.9) has to be modified in the obvious way in terms of the counting measure on $\mathbb{Z}^d$ (*cf.* [5]). The proof of (0.15) is a trifle more involved and a technique from [20] is used.

We can also consider the more general Markov process generated by some elliptic operator (*cf.* [16]).

(0.16)
$$\mathcal{A} = -\partial_i a_{ij} \partial_j,$$

where $a_{ij} \in L^\infty(\mathbb{R}^d)$. The corresponding semigroup with Dirichlet boundary conditions on $D$ can then be constructed, $T_t = e^{-t\mathcal{A}}$ (*cf.* [2] for the symmetric case), and we shall denote by $p^{\mathcal{A}}_t(\cdot, \cdot)$ the kernel of that semigroup and

(0.17)
$$P_{\mathcal{A}}(t, x) = \int p^{\mathcal{A}}_t(x, y)\, dy.$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

The Main Theorem does not make sense for $P_{\mathcal{A}}$ and (0.11) cannot possibly hold. What does hold and presents some independent interest is the Gaussian estimate

(0.18)

$$p_t^{\mathcal{A}}(x,y) \le Ct^{-d/2}P_{\mathcal{A}}(t,x)P_{\mathcal{A}^*}(t,y)\exp\left(-\frac{d^2(x,y)}{(4+\varepsilon)t}\right); \quad \varepsilon > 0,\ t > 0,\ x,y \in D,$$

where $C = C(d, A, \varepsilon, \lambda)$, where $\lambda$ is the ellipticity constant of $\mathcal{A}$:

$$(0.19) \qquad |a_{ij}| \le \lambda\,; \quad \lambda^{-1}|\xi|^2 \le a_{ij}(x)\xi_i\xi_j \le \lambda|\xi|^2\,; \quad \xi \in \mathbb{R}^d,$$

and where the distance $d(\cdot,\cdot)$ in the Gaussian of (0.18) is the distance induced by $\mathcal{A}$ (in the standard way *cf.* [6]; $\lambda^{-1}|x-y| \le d(x,y) \le \lambda|x-y|$). In (0.18) $\mathcal{A}^* = -\partial_i a_{ji}\partial_j$ denotes the formal adjoint of $\mathcal{A}$, and if $\mathcal{A} \ne \mathcal{A}^*$ the $\varepsilon > 0$ in (0.18) has to be assumed large enough $\varepsilon > \varepsilon_0$. The derivation of (0.18) (for the symmetric case) is once more contained in [6], where instead of (0.14) we use the integrated version which holds for $P_{\mathcal{A}}$:

$$(0.20) \qquad 1 \le \frac{P(t_2,x)}{P(t_1,x)} \le C\left(\frac{t_1}{t_2}\right)^C; \quad x \in D,\ t_1 \ge t_2 > 0,$$

where $C = C(d, A, \lambda)$. (0.20) is nontrivial. What is much easier to establish is the corresponding backwards "spacial" Harnack estimate

$$(0.21) \quad \frac{P(t,x)}{P(s,y)} \le C\,; \quad t,s>0,\ x,y \in D,\ |t-s| \le C\delta^2(x),\ |x-y| \le \delta(x)/2,$$

which is valid for an arbitrary domain $D$ but only for the constant coefficient operator $L = \Delta - \partial_t$. All in all, the above Gaussian estimates for the kernels are not so easy to establish, but the corresponding lower kernel estimates are easy and we have:

$$p_t^{\mathcal{A}}(x,y) \ge C^{-1}t^{-d/2}P_{\mathcal{A}^*}(t,x)P_{\mathcal{A}}(t,x)\exp\left(-C\frac{|x-y|^2}{t}\right),$$

where $\mathcal{A}$ is as in (0.16) and $C = C(d, A, \lambda)$. This will be proved in Section 8 below. Analogous lower estimates hold for the kernel of the random walk (0.9) (*cf.* Section 8 below) provided that $|x-y| \le ct$, and provided also that additional smoothness conditions are imposed on the measure $\mu$. In Section 8 I will give the main ingredient for the proof of this discrete lower estimate, but I will leave it to the interested reader to complete the proof if he so wishes.

## 0.2 Technical Results for the Green Potential

Let $D$ be a globally Lipschitz domain as in (0.10), and let $u(\cdot) > 0$ be some harmonic function in $D$ such that $u|_{\partial D} = 0$. The existence of such a (Martin point) function,

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

which also has to be (up to scalar factor) unique, is not trivial to establish, but it is well known (at least to experts *cf.* [8], [9]). Let us also denote by

$$G(y,x) = G(x,y) = C_d|x-y|^{2-d} - H(x,y)\ ; \quad d > 2, \tag{0.22}$$

(with the obvious modification of $d = 2$) the corresponding Green function where $H(\cdot,\cdot)$ is harmonic in each variable, and $G(x_0,\cdot)|_{\partial D} \equiv 0$. In [4] I gave a proof of the fact that for all $c > 0$ there exists $C > 0$ such that:

$$G(x,x_0)u(x) \le C|x-x_0|^{2-d}u(x_0)\ ; \quad x, x_0 \in D,\ |x-x_0| \ge c\delta(x_0). \tag{0.23}$$

If $d > 2$ this estimate clearly also holds for $x$ close to $x_0$.

From (0.23) we shall deduce

$$w(x_0) = \int_D G(x,x_0)\frac{u(x)}{\delta^{2+\varepsilon}(x)}\,dx \le C\frac{u(x_0)}{\delta^{\varepsilon}(x_0)}\ ; \quad x_0 \in D,\ 0 < \varepsilon < 1, \tag{0.24}$$

where $C = C(d, A, \varepsilon)$. The case when $D$ is convex was proved in [4]. $w$ is the Green potential of $u/\delta^{2+\varepsilon}$. As we shall see, this generalizes to general Lipschitz domains and to the following parabolic Green potential:

$$\begin{gathered} w_\varepsilon(t,x_0) = \int_0^t \int_D p_{t-s}(x,x_0)\frac{P(s,x)}{\delta^{2+\varepsilon}(x)}\,dx\,ds \le C\frac{P(t,x_0)}{\delta(t,x_0)^{\varepsilon}}\ ; \\ 0 < \varepsilon < 1,\ x_0 \in D,\ t > 0, \end{gathered} \tag{0.25}$$

where $C = C(d, A, \varepsilon) > 0$ (what is crucial in the definition of $w$ is: $(\Delta - \partial_t)w = -P\delta^{-2-\varepsilon}$, *cf.* [1]).

It is important to observe that we can break the proof of (0.25) into two parts: For all $c > 0$ there exists $C > 0$ such that

$$w_\varepsilon(1,x_0) \le C\frac{P(1,x_0)}{\delta^{\varepsilon}(x_0)}, \quad \delta(x_0) < c, \tag{0.26}$$

$$w_\varepsilon(1,x_0) \le C,\ \delta(x_0) > c, \tag{0.27}$$

(0.26) and (0.27) put together can be scaled and are equivalent to (0.25) (*cf.* Remark at the end of Section 1.1).

The estimate (0.25) as well as the estimate of the gradients $\nabla^k w$ is the key technical tool of this paper and presents, perhaps, some independent interest. Once this has been done (in Sections 2, 3 below) the Main Theorem follows by standard methods (Section 4). The estimate (0.25) is some what surprising and much harder to establish than (0.24). In fact (0.25) can be improved and we can replace $\delta(t,x_0)$ by $\delta(x_0)$ in the right hand side. I will not give the proof of this here.

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

### 0.3 The Increasing Degree of Technical Difficulty: Further Results

If we restrict ourselves to convex domains the proof of (0.25) is elementary, and can be given without Dahlberg's theorem on the harmonic measure on Lipschitz domains [16], [17]. I have written the paper in such a way that the reader, if he so wishes (in fact, I strongly recommend this), can go through the proofs for convex domains (and to a certain extent also in the case $d = 2$) without the deeper part of the theory.

To prove (0.24) for instance, for convex domains, one only needs to go as far as (1.24) in Section 1 where I introduce the maximal function. A similar approach applies to Sections 2 and 3. The computations, especially in Section 3.2, are not necessarily simpler for the convex case (compare with Section 3.3), but they are certainly more elementary.

The estimate (0.25) proves the Main Theorem and (0.11) with the routine methods of Section 4. These methods, however, only give the Main Theorem with $\varepsilon = 4/5$. The proof for compactly supported measures and any $0 < \varepsilon < 1$ is more subtle and will appear, as part of a more general scheme, in a later publication. It turns out, at least when $\mu$ is compactly supported and smooth and when $D$ is convex, that the Main Theorem and (0.11) holds even with $\varepsilon = 1$. To prove (0.11) with $\varepsilon = 1$ for a convex domain one needs to consider a different Green potential

$$w^{(k)}(t, x_0) = \int_0^t \int_D p_{t-s}(x, x_0) |\nabla_x^k P(s, x)| \delta^{k-3}(x)\, ds\, dx\,; \quad k = 2, 3, \dots . \tag{0.28}$$

What is nontrivial, but essential if we wish to be able to set $\varepsilon = 1$ in the Main Theorem, is the estimate:

$$w^{(3)}(t, x_0) \le C \frac{P(t, x_0)}{\delta(t, x_0)}. \tag{0.29}$$

This is valid if $D$ is convex. In the integrant of (0.28) the cofactor of $p_{t-s}(\cdot, \cdot)$ is $O\big(P(s, x)\delta^{-3}(x)\big)$ (*cf.* Section 2.1 below). It follows that the $w^{(3)}$ should be though as the limiting case $\varepsilon = 1$ of (0.25), and the convergence of (0.28) is critical. Unfortunately (0.29) does not suffice to give $\varepsilon = 1$ in the Main Theorem and a number of other nontrivial complications arise. I have decided therefore to postpone the proof of the convex case with $\varepsilon = 1$ to another publication.

This attitude seems reasonable at this point, especially since in the absence of convincing counterexamples, I do not know whether for $\varepsilon = 1$ to hold in (0.11) one needs convexity, or whether it works for all Lipschitz domains.

The possibility to set $\varepsilon = 1$ in (0.11) is interesting because this renders (0.11) unimprovable (*e.g.* $D = [0, +\infty[ \subset \mathbb{R}$ and $\mu = 1/2(\delta_{-1} + \delta_1)$. The reflection principle can then be used, both for the brownian motion and for the random walk and exact formulas can be worked out for both $P$ and $P_\mu$. To estimate $P - P_\mu$ one can then use the Edgeworth expansion *cf.* [28]). $\varepsilon = 1$ in (0.11) is, also interesting because then the result is related with the discrete Harnack estimates of [26] and [29] and also, in an obvious way, with Berry-Esseen types of theorems *cf.* [28].

A number of other applications will be outlined in Section 8.3. The most significant of these applications is a version of the parabolic Harnack inequality of [3] at the boundary for discrete Markov chains (*cf.* [26], [29]) and also the following comparison of harmonic measures:

To avoid complications let us define $\omega$ $(x, E; \text{walk})$ $(x \in D \cap \mathbb{Z}^d, E \subset \partial(\mathbb{Z}^d \setminus D))$ the harmonic measure with respect to Bernoulli standard random walk on $\mathbb{Z}^d$. We have then the following comparison with the Newtonian harmonic measure:

$$C^{-1} \leq \frac{\omega(x, E + B\,;\ \text{Brownian})}{\omega(x, E\,;\ \text{walk})} \leq C,$$

where $B \subset \mathbb{R}^d$ is the ball of radius $Cd$. The proof follows easily from the results of this paper but it will not be given here.

Finally a good part of this paper extends to non divergence form operators $a_{ij}\partial_i\partial_j$ ($a_{ij} \in L^\infty$, *cf.* [38]). More details on these applications will be given elsewhere.

### 0.4 A Guide to the Reader

The heart of the matter lies in Sections 2 and 3, while Section 1 is a "warming up run" so to speak. The proof of the Main Theorem lies in Section 4, while the proof of the kernel estimates lies in Section 8.

Sections 5, 6 and 7 lie apart and are considerably more specialized. They deal with homogenization problems, random environment and random sampling. Had it not been for the applications in Lie group theory, I would not have incorporated these sections in this paper. Given, on the other hand, that it is precisely these applications that made me look at random walks in Lipschitz domains in the first place (the domains that occur in these applications are conical domains), I felt that I had to include these sections here.

The presentation in the last four sections is sketchy. The excuse for this is that the content of Sections 5, 6 and 7 is rather esoteric, and that I intend to develop further the content of Section 8 in a future publication.

## 1 Harmonic Green Potential

### 1.1 The Estimate at Infinity

In this section we shall consider $D$ a globally Lipschitz domain (0.10) and the corresponding Green function (0.22), and we shall give a proof of (0.24) for some harmonic $u(\cdot) > 0$, $u|_{\partial D} = 0$. We shall furthermore normalize as follows:

$$x_0 = (x_1, 0), \quad \varphi(0) = 0, \quad \delta(x_0) = 1. \tag{1.1}$$

The proof below is a simplification of the proof of [4] and works also in the more general setting where we replace the Green function and the harmonic function $u$ respectively, by the fundamental solution (also denoted by $G(x_0, \cdot)$) and by a positive solution $u(\cdot)$ of the general divergence form elliptic operator (*cf.* [16]):

$$\mathcal{A} = -\partial_i a_{ij}(x)\partial_j, \tag{1.2}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

where both $G(x_0, \cdot)$ and $u(\cdot)$ vanish at the boundary.

Let $R \gg 1$ and let us fix

$$x = (a, 0), \quad \zeta = (b, 0), \tag{1.3}$$

where $a, b \sim R$. The notation "$\sim$" means throughout:

$$x \sim y \Leftrightarrow C^{-1} \leq x/y \leq C, \tag{1.4}$$

where $C$ is a constant that depends only on acceptable parameters. In our case here these parameters are $d$, the dimension, $A$ the Lipschitz constant of $D$, and $\lambda$ the ellipticity constant of (1.2) (*cf.* (0.19)).

It is then a consequence of the formula (0.22) in the classical case (0.2), of the Harnack estimate, and of the Aronson's estimates for the heat diffusion kernel of $e^{-t\mathcal{A}}$ (*cf.* [36], which we can integrate in $t$) in the general case, that

$$G(x, \zeta) \sim R^{2-d}\ ; \quad |a - b| \sim R. \tag{1.5}$$

The reader should draw a picture at this point, and should also "keep an eye open" for the modifications that have to be made in the formulas when $d = 2$.

The interior Harnack estimate implies that:

$$u(x) \sim u(\zeta), \tag{1.6}$$

and the classical Carleson estimate (*cf.* [15], [16]) implies that:

$$u(\hat{x}) \leq Cu(x),\ G(\hat{x}\,; x_0) \leq CG(x, x_0)\ ; \quad \hat{x} \in D,\ |\hat{x}| \sim R. \tag{1.7}$$

We next invoke the Harnack boundary principle and deduce that

$$\frac{G(x, \zeta)}{G(x, x_0)} \sim \frac{u(\zeta)}{u(x_0)}. \tag{1.8}$$

Here we use the fact that $G(x, \cdot)$, for fixed $x$, is a solution of $\mathcal{A}$ while in (1.7) we use the fact that $G(\cdot, x_0)$ is a solution of $\mathcal{A}^*$, the formal adjoint of $\mathcal{A}$. Combining the estimates (1.5), (1.6), (1.7) and (1.8), we finally obtain the following estimate:

(1.9)
$$G(\hat{x}, x_0)u(\hat{x}) \leq CG(x, x_0)u(\zeta) \sim G(x, \zeta)u(x_0) \sim R^{2-d}u(x_0) \sim |\hat{x} - x_0|^{2-d}u(x_0).$$

We can chose $R \gg 1$ arbitrary, and (1.9) proves (0.23) even for operators $\mathcal{A}$ that are not symmetric, *i.e.*, we have:

$$G(x, x_0)u(x) \leq C|x - x_0|^{2-d}u(x_0)\ ; \quad x, x_0 \in D\ ; \quad |x - x_0| > a\delta(x_0), \tag{1.10}$$

where $C = C(d, A, a, \lambda)$.

***Remarks***

(i) The normalization (1.1) was not really used in the above proof. It is however important to observe that both (0.23) and (0.24) are scale invariant and that it suffices to prove (0.24) for $\delta(x_0) = 1$. The point is of course that if we dilate $x \mapsto \alpha x$ ($\alpha > 0$) in $\mathbb{R}^d$, the domain $D$ is replaced by $D_\alpha = \alpha D$ and $G_D(x, x_0)$ is replaced by $G_\alpha(\cdot, \cdot) = G_{D_\alpha}(\cdot, \cdot)$ and we have

$$G_\alpha(\alpha x, \alpha x_0) = G(x, x_0)\alpha^{2-d}. \tag{1.11}$$

The Lipschitz constant of $D_\alpha$ is the same as for $D$ and furthermore, in the case that we are considering, for the general operator $\mathcal{A}$, the ellipticity constant of the dilated operator does not change. In so far as the constants in (1.10) only depend in $d$, $A$, $\lambda$ we can therefore scale and assume that $\delta(x_0) = 1$.
Similarly, we can use the dilation:

$$x \mapsto \alpha x \in \mathbb{R}^d,\ t \to \alpha^2 t > 0\ ; \quad \alpha > 0, \tag{1.12}$$

and we can assume in (0.25) that $t = 1$ (but we then have no control in $\delta(x_0)$). The point to observe is then that

$$P_\alpha(\alpha^2 t, \alpha x) = P(t, x),\ p^{(\alpha)}_{\alpha^2 t}(\alpha x, \alpha y) = \alpha^{-d} p_t(x, y)\ ; \quad t > 0,\ x, y \in D, \tag{1.13}$$

with obvious notation.

(ii) The estimate (1.10) is a special case of a more general estimate

$$G(x_0, x)G(x, x_1) \le C \operatorname{Max}[|x - x_0|^{2-d},\ |x - x_1|^{2-d}]G(x_0, x_1)\ ;$$
$$x \in D,\ |x - x_0| \ge a\delta(x_0),\ |x - x_1| \ge a\delta(x_1),$$

where $x_1, x_2 \in D$. The proof of this follows the same lines but is slightly more involved. No immediate use of this will be made in this paper and therefore it will be left as an exercise to the reader. One advantage that this estimate has over (1.10) is that it makes sense and extends to bounded Lipschitz domains (*cf.* [16]). If $d > 2$ the above estimate holds for all $x \in D$.

## 1.2 The Green Potential

Let $D$ be as in (0.10), and $Q \in \partial D$ we shall use throughout the notation

$$T_R(Q) = T_{R_1,R_2}(Q) = [x = (x_1, x') \in D\ ; |x_1 - Q_1| \le R_1, |x' - Q'| \le R_2], \tag{1.14}$$

where $Q = (Q_1, Q') \in \mathbb{R} \times \mathbb{R}^{d-1}$ and $R = (R_1, R_2 > 0)$ are chosen so that $T_R(Q)$ is connected (even with room to spare so that the connectivity property of (iii) Section 3 Ch. 1 of [16] holds. At this point, in fact, the inexperienced reader would do well to study the first few pages of [16]). The following Lemma is well known [15], [21].

***Lemma*** *Let $D$ and $Q \in \partial D$ be as above and $c = (c_1, c_2)$, and let $u(\cdot) > 0$ be harmonic in $T_{2c}(Q)$ and such that $u|_{\partial D} \equiv 0$. Then there exists $\alpha = \alpha(d, A) > 0$ and $C = C(d, A, c)$ such that*

$$u(x) \leq C\delta^{\alpha}(x)u\big(A_c(Q)\big) \ ; \quad x \in T_c(Q). \tag{1.15}$$

In (1.15) we adopt the notation in [16] and set

$$A_r(Q) = Q + r(1, 0, \dots, 0) \in D. \tag{1.16}$$

The estimate (1.15) holds more generally for solutions of $\mathcal{A}u = 0$ where $\mathcal{A}$ is as in (0.16), in which case the ellipticity constant $\lambda$ (0.19) has to be incorporated in the dependence of the constants. If however $\mathcal{A} = \Delta = -\sum \frac{\partial^2}{\partial x_i^2}$, *i.e.*, for classical harmonic functions, then

(i) If $d = 2$ we can choose

$$\alpha > 1/2. \tag{1.17}$$

(ii) If $D$ is convex we can choose

$$\alpha = 1. \tag{1.18}$$

The above estimate can of course be scaled to any domain $T_R(Q) = T_{Rc}(Q)$, where $R > 0$ is arbitrary and $Rc = (Rc_1, Rc_2)$, where we shall fix $c = (c_1, c_2)$ and where we shall assume, for notational convenience, that $c_1 = c_2 = 1$. The estimate (1.15) becomes then:

$$u(x) \leq C\left(\frac{\delta(x)}{R}\right)^{\alpha} u\big(A_R(Q)\big) \ ; \quad R > 0,\ x \in T_R(Q), \tag{1.19}$$

where $u > 0$ is harmonic in $T_{2R}(Q)$.

Before we move on, let us recall how (1.19) and (1.10) can be used to supply a proof of (0.24) in the special cases (i) $d = 2$ or (ii) $D$ is convex.

We assume as we may that $\delta(x_0) = 1$ $u(x_0) = 1$ in (0.24) and we have

$$\int G(x, x_0)\frac{u(x)}{\delta^{2+\varepsilon}(x)}\, dx \cong \sum_{j\geq 1} \int_{|x-x_0|\sim 2^j} = \sum I_j. \tag{1.20}$$

In the above decomposition we ignore the contribution coming from $|x - x_0| \leq C$ because this can be handled trivially (*cf.* [4]). The estimate (1.19) with $R \sim 2^j$ can be easily adapted to the "Π shape" region:

$$[|x - x_0| \sim 2^j] = T_{2^{j+1}}(0) \setminus T_{2^j}(0), \tag{1.21}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

and by an elementary computation we obtain:

$$I_j \leq CG(x_0, x_j)u(x_j)2^{j(d-2-\varepsilon)} \tag{1.22}$$

where $x_j \in T_{2^{j+1}} \setminus T_{2^j}$ is appropriately chosen, and where we assume that

$$0 < \varepsilon < 2\alpha - 1. \tag{1.23}$$

Because of (1.23) the series in (1.20) can therefore be summed in the following two cases:

(i) $d = 2$ and $\varepsilon > 0$ is small enough.
(ii) $D$ is convex and $0 < \varepsilon < 1$ is arbitrary.

This completes the proof.

Let now $F(x) = F(x_1, x')$ $(x \in T_R(Q))$ be an arbitrary function we shall define then the following maximal function

$$F^* = F^*_R(x') = \sup_y \left[ \frac{|F(x)|}{\delta(x)} \; ; x = (y, x') \in T_R(Q) \right]. \tag{1.24}$$

$F^*$ can be thought either as a function of $x' \in \mathbb{R}^{d-1}$ $(|x' - Q'| \leq R)$ or, with the obvious identifications, as a function on $T_R(Q) \cap \partial D$.

We shall now establish the following average estimate that replaces our pointwise estimate in the previous Lemma.

***Key Lemma*** *Let $Q \in \partial D$ and $R > 0$ be fixed, and let $u(\cdot)$ be some positive harmonic function in $T_{2R}(Q)$ such that $u|_{\partial D} \equiv 0$. We have then $u^*_R \in L^2(\mathbb{R}^{d-1})$ and*

$$\|u^*_R\|_2 \leq C_{d,A} R^{\frac{d-3}{2}} u(x_R), \tag{1.25}$$

*where $x_R \in T_R(Q)$ is any point such that $\delta(x_R) \sim R$.*

For the proof we can scale and normalize as in (1.1) so that $R = 1$, $Q = 0$. By the boundary Harnack principle [8], [16], we may also assume that

$$u(\cdot) = G(X, \cdot), \tag{1.26}$$

is the Green function of $T_{10}(0)$ at the point $X = A_5(0)$. The standard estimate given in Lemma 1.3.3 of [16] implies then that

$$u^*(\xi) \leq C \sup_{r \leq 1} r^{1-d} \omega_X[\partial D \cap |x - \xi| \leq r] \; ; \quad \xi \in \partial D, \tag{1.27}$$

where I use the same notation as in [16] and $\omega_X(\cdot)$ denotes harmonic measure at $X$.

If we denote by $\sigma(\cdot)$ the Lebesgue $[(d-1)$-dimensional Hausdorff measure$]$ on $\partial D$ we have by a result of Dahlberg [17], [16]

$$k = \frac{d\omega}{d\sigma} \in L^2_{\text{loc}}(\partial D \; ; d\sigma) \underset{\text{(identification)}}{=} L^2_{\text{loc}}(\mathbb{R}^{d-1}). \tag{1.28}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

Our lemma follows therefore from (1.27), the Hardy-Littlewood Maximal Theorem and the reverse Hölder inequality $B_2$ (notation of [16, Sections 1.4.13, 2.1–5]) that is verified by $k$.

Let us go back to the proof of (1.24) in the general case. We consider again the decomposition (1.20) and the corresponding "Π-shaped" regions $T_{2^{j+1}} \setminus T_{2^j}$. The estimate of $I_j$ in the "top part" of the "Π" is easy and is done as before. It is the "side regions" of the "Π" that create problems.

In these "side regions" we clearly have:

$$\frac{G(x_0,x)u(x)}{\delta^{2+\varepsilon}(x)} \le C\frac{G^*(x_0,x')u^*(x')}{\delta^{\varepsilon}(x)}\ ; \quad x=(x_1,x')\in D. \tag{1.29}$$

So in these side regions if we first integrate in the $x_1$-variable we obtain a contribution $\sim 2^{(1-\varepsilon)j}$. The integration in the $x'$-variable is then estimated by Hölder with the help of the Key Lemma. We thus obtain the estimate

$$I_j \le G(x_0,x_j)u(x_j)2^{j(d-2-\varepsilon)}, \tag{1.30}$$

where $x_j$ are as in (1.22). The summation can be carried out as before and we are done. Observe, however, that here the key Lemma has to be used also to control the missing term in (1.20) that comes from $[|x-x_0|\sim C]$.

Observe that if we use the Remark 1.1 (ii) and the above argument, we can easily prove the varient

$$\int_D \frac{G(x,x_0)G(x,x_1)}{\delta(x)^{2+\varepsilon}}\,dx \le CG(x_0,x_1)[\delta^{-\varepsilon}(x_1)+\delta^{-\varepsilon}(x_0)]\ ; \quad x_0,x_1\in D, \tag{1.31}$$

with $0<\varepsilon<1$. No use of this will be made in this paper.

## 2 The Parabolic Harnack Estimates

### 2.1 The Harnack Estimate for $p_t(x,y)$

Let $D\subset\mathbb{R}^d$ be an arbitrary domain and let $u(t,x)=p_t(x,x_0)$ be as in (0.2) for some fixed $x_0\in D$. The function $u(\cdot)$ can be extended to be identically zero for $[t\le 0,(t,x)\ne(0,x_0)]$ (*cf.* [1]) and to give a non-negative solution of the Heat Equation $\dot\Delta = L = \Delta-\partial_t$ in the domain:

$$\dot D = \dot D_1\cup\dot D_2 = [t>0, x\in D]\cup[t\in\mathbb{R}, x\in D, x\ne x_0]\subset\mathbb{R}^{d+1}=\dot{\mathbb{R}}^d, \tag{2.1}$$

where the dots correspond to Doob's notation in [1]. From the definition of $\dot D$ it follows that if $(t,x)\in\dot D$ $(t>0)$ then

$$[t-\ell^2, t+\ell^2]\times B_\ell(x)\subset\dot D\ ; \quad \ell>0,$$

provided that:

$$\ell\le 1/10\,\mathrm{Min}[\delta(x),\sqrt{t}+|x-x_0|]=\ell_0(t\ ;x,x_0)=\ell_0,$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

where $B_\ell(x)$ denotes throughout the Euclidean ball centered at $x$ and radius $\ell$. The basic Harnack principle [1] implies therefore that for all $c > 0$ and integers $a, b \geq 0$, there exists $C > 0$ such that:

$$\left| \frac{\partial^a}{\partial t^a} \nabla_x^b p_t(x, x_0) \right| \leq C \ell^{-2a-b} p_{t+c\ell^2}(x', x_0) \; ; \qquad 0 < \ell < \ell_0, \; t > 0, \; x \in D, \; |x' - x| \leq \ell. \tag{2.2}$$

## 2.2 The Backwards Harnack Estimates

The function $P(t, x)$ in (0.3) also satisfies the Harnack estimate

$$|\nabla_y^k P(t, y)| \leq CP(t+1, x) \leq CP(t, x) \; ; \quad k = 0, 1, \dots \; ; \quad t > 1/10 \; ; \qquad x, y \in D, \; \delta(x) = 1, \; |x - y| \leq 1/2,$$

because $P(\cdot, x)$ is decreasing in $t$. We have on the other hand

$$|\nabla_y^k P(t, y)| \leq C, P(t, x) \geq C \; ; \quad 0 < t \leq 1/10, \; x, y \in D, \; \delta(x) = 1, \; |x - y| \leq 1/2,$$

by the regularity properties up to the smooth boundary at $t = 0$ for the solutions of $Lu = (\Delta - \partial_t)u = 0$. Combining the above two estimates we have

$$|\nabla^k P(t, y)| \leq CP(t, x) \; ; \quad t > 0, \; x, y \in D, \; \delta(x) = 1, \; |x - y| \leq 1/2,$$

It follows by scaling, that for an arbitrary domain $D$, the function $P(t, x)$ satisfies the "backwards" (*cf.* [3]) Harnack estimate

$$|\nabla_y^k P(t, y)| \leq C_{d,k} \frac{P(t, x)}{\delta^k(x)} \; ; \quad t > 0, \; x, y \in D, \; |x - y| \leq \delta(x)/2. \tag{2.3}$$

An identical argument proves the estimate (0.21) but no essential use of this will be made in this paper.

## 2.3 Lipschitz Domains and the Time Derivative

Here and for the rest of this paper we shall restrict ourselves to domains that are globally Lipschitz as in (0.10). The results of this section present an independent interest and they will be formulated in the general setting of a time-dependent parabolic operator (*cf.* [3]):

$$L = \partial/\partial x_i a_{ij}(t, x) \partial/\partial x_j - \partial/\partial t \; ; \qquad a_{ij} \in L^\infty, \; \|a_{ij}\|_\infty \leq \lambda, \; \lambda^{-1}|\xi|^2 \leq a_{ij}\xi_i\xi_j \leq \lambda|\xi|^2, \; \xi \in \mathbb{R}^d. \tag{2.4}$$

We shall denote by

$$G(s, x \, ; t, y) dy = \mathbb{P}[z(t) \in dy // z(s) = x] \; ; \quad x, y \in D, \; t \geq s \geq 0,$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

where $G(\cdot\,;\,\cdot)$ is the fundamental solution of $L$ and the probabilities $\mathbb{P}(\cdot)$ refer to the time inhomogeneous diffusion $(z(t) \in \mathbb{R}^d, t > 0)$ generated by (2.4) in $D$ (*cf.* [22]). Smoothness on the coefficients has to be imposed if $(z(t);\, t > 0)$ is indeed to be a diffusion (*i.e.*, continuous paths). But this should not worry us because what will be obtained here will simply be "*a priori* estimates". These will be uniform with respect to the smoothness of the coefficients of (2.4) which will be assumed to be $C^\infty$, and will only depend on $\|a_{ij}\|_\infty$ and on the ellipticity constant $\lambda > 0$ in (2.4).

I shall denote by

$$P(t,x) = \int G(0,x\,;\,t,y)\,dy = \mathbb{P}[z(s) \in D, 0 < s < t//z(0) = x],$$

and by

$$J_h = P(1,x) - P(1+h,x) = \int_D G(0,x\,;\,1,y)M(y,h)\,dy, \quad x \in D,\ h > 0,$$

where

(2.5)
$$M(y,h) = \mathbb{P}[z(s) \notin D \text{ some } 1 \le s \le 1+h//z(1) = y] \le C \exp\left(-\frac{cd^2(y,\partial D)}{h}\right).$$

The Gaussian estimate (2.5) is the direct analogue of the corresponding result for brownian motion which is quite standard (*cf.* [24]). One can derive (2.5) from the Gaussian estimates of the kernel $G(\cdot,\cdot)$ (*cf.* [36]) by a well known procedure (*e.g. cf.* [7, Section 1.3] where we estimate $\mathbb{P}[\sup_{0<s<t} |2(s)| \ge r]$).

The integration in $J_h$ will be split in the two ranges

$$\int_{d(y,\partial D)\ge 1} + \int_{d(y,\partial D)\le 1} = I_1 + I_2,$$

and by (2.5), the estimate

$$I_1 \le CP(1,x)\exp\left(-\frac{c}{h}\right); \quad h > 0, \tag{2.6}$$

is automatic. To estimate $I_2$ we shall use the "Carleson" estimate for parabolic equations (*cf.* [15], [18]). I shall also use the notation of (0.11) and write $y = (y_1, y')$. I shall denote

$$y_\mu = \big(\varphi(y') + \mu, y'\big)\ ; \quad \mu > 0,\ y' \in \mathbb{R}^{d-1}. \tag{2.7}$$

From [18] it follows that in the range $d(y,\partial D) \le 1$ there exist $c, C, C_1 > 0$ depending only on the dimension $d$, on the Lipschitz constant $A$ (0.11), and on the ellipticity constant $\lambda$ (2.4), such that for all $c_1, c_2, c_3 > 0$ there exists $C > 0$

$$\sup_{0<\mu<c_1} G(0,x\,;\,1,y_\mu) = m(y') \le CG(0,x\,;\,1+c_2,y_{c_3}). \tag{2.8}$$

The reader should draw a picture at this point. By integrating in $I_2$, first in the $y_1$ variable then in $y'$, we deduce that

$$
\begin{aligned}
I_2 &\le C \int_{\delta(y)\le 1} m(y') \exp\left(-\frac{\delta^2(y)}{Ch}\right) dy \\
&\le Ch^{1/2} \int_{\delta(y)\le 1} m(y')\, dy \le Ch^{1/2} \int G(0,x\,;1+c,y_c)\, dy \qquad (2.9)\\
&\le Ch^{1/2} P(1+c,x) \le Ch^{1/2} P(1,x).
\end{aligned}
$$

This combined with (2.6) gives the estimate

$$
P(1,x) - P(1+h,x) \le Ch^{1/2} P(1,x)\,; \quad x \in D,\ h > 0. \tag{2.10}
$$

The backwards Harnack estimate (0.20) follows by the scaling (1.12) and by an obvious elementary argument.

The key of the above proof is Salsa's "Carleson" estimate in (2.8):

$$
m(y') \le CG(0,x\,;1+c,y_c).
$$

This estimate admits the following improvements:

For the general situation we can assert that there exists some positive $\alpha > 0$ such that for all $a > 0$ there exists $C, c > 0$ such that

$$
G(0,x\,;1+a,y) \le C\delta(y)^\alpha G(0,x\,;1+c,y_c). \tag{2.11}
$$

This allows us to improve (2.10) to

$$
P(1,x) - P(1+h,x) \le Ch^{1/2+\alpha/2} P(1,x)\,; \quad x \in D, h > 0. \tag{2.12}
$$

For the special case when $L$ is the constant coefficient operator $L = \Delta - \partial_t$, and $d = 2$ we can even assume in (2.11) that $\alpha > 1/2$. What is more to the point, however, is that in the special case when $L = \Delta - \partial_t$, and when in addition we assume that $D$ is convex (*and* globally Lipschitz) then we can set $\alpha = 1$ in (2.11). This improvement together with the previous method suffices therefore to prove, in the convex case, the stronger estimate (0.14)

$$
0 \le -\frac{\partial}{\partial t} P(t,x) \le \frac{c}{t} P(t,x). \tag{2.13}
$$

In the time homogeneous case, *i.e.*, when the operator $L$ (2.4) is independent of $t$, the above improvements (2.11) are a consequence of the corresponding results for positive solutions of $L$ that vanish at the boundary ($u > 0$, $Lu = 0$ *cf.* [3] and the two lemmas in Section 1.2). Indeed we can then use the parabolic Harnack boundary principle to compare such solutions with positive parabolic functions that vanish at the boundary.

The estimate (2.11) holds also for the more general (time inhomogeneous) case. Standard methods can be used to obtain this (*cf.* [21], [18]). The result is well known in that generality, but I was unable to find an explicit reference. It is also of interest to observe that the above argument extends to the non-cylindrical domains (8.33), (8.34) below.

## 2.4 The Classical Potential Theory

In this section we shall go back to the constant coefficient operator:

$$L = \Delta - \partial_t,$$

and to $P(t,x)$ as (0.3), and we shall re-examine the proof of the estimate (0.14)

$$0 \le -\frac{\partial}{\partial t} P(t,x) \le \frac{c}{t} P(t,x), \tag{2.14}$$

in the light of Section 1.2. The domain $D$ is assumed to be a general global Lipschitz domain as in (0.10).

We shall follow and adapt the notation of Section 0.1 so that now:

$$G(0,x\,;t,y) = p_t(x,y) = u(t,y)\,; \quad t > 0,\ x, y \in D, \tag{2.15}$$

which is a parabolic function in $(t,y)$. With the same notation ($y = (y_1, y')$) as in (2.7) and in (1.24) we then consider:

$$u^*(y') = \sup_{0<\mu<C} \frac{u(1, y_\mu)}{\mu}. \tag{2.16}$$

It follows that the argument that lead to (2.9) can be adapted and it gives:

$$I_2 \le C \int_{\delta(y)\le 1} u^*(y')\delta(y) \exp\left(-\frac{\delta^2(y)}{Ch}\right) dy \le Ch \int_{d(y,\partial D)\le 1} u^*(y')\,dy. \tag{2.17}$$

We shall use now the Key Lemma of Section 1.2 and make a comparison, as in the previous section, of $u(t,y)$ with some harmonic function defined in some neighborhood of the point $Q = \big(\varphi(Q'), Q'\big) \in \partial D$ ($Q' \in \mathbb{R}^{d-1}$) that vanishes at the boundary. We deduce that

$$\int_{\delta(y), |y'-Q| \le 1} u^*(y')\,dy \le Cu(2, \overline{Q}), \quad \overline{Q} = \big(\varphi(Q') + 1, Q'\big) \in D. \tag{2.18}$$

By a simple use of the parabolic Carleson principle (*cf.* [18]) we deduce that

$$u(2, \overline{Q}) \le C \int_{E_Q} u(3,y)\,dy, \tag{2.19}$$

where:

$$E_Q = \{y = (y_1, y') \in D\,; |y' - Q| \le 1, 0 < y_1 < 10\}. \tag{2.20}$$

If we combine (2.15), (2.18) and (2.19) and sum over all $Q' \in \mathbb{Z}^{d-1} \subset \mathbb{R}^{d-1}$ we obtain

$$\int_{\delta(y)\le 1} u^*(y')\,dy \le CP(3,x) \le CP(1,x). \tag{2.21}$$

The estimates (2.17) and (2.21) put together finally give $I_2 \le ChP(1,x)$ and this combined with (2.6) completes the proof of (2.14) in the general case ($d \ge 2$ and $D$ an arbitrary Lipschitz domain as in Section 0.2, and $L = \Delta - \partial_t$). It is, of course, well known that such an estimate cannot possibly hold with $\Delta$ replaced by a more general operator $\mathcal{A}$ as in (0.16). In view of the counterexample [23] the estimate (2.14) does not extend to the domains (8.34).

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

### 2.5 The General Gradient Estimate

It is of some interest that the spacial and the time estimates for the gradient can be combined to give:

$$\left|\frac{\partial}{\partial t}\nabla_x^b P(t,x)\right| \le \frac{c}{t}\delta^{-b}(x)P(t,x)\ ; \quad t>0,\ x\in D, \tag{2.22}$$

and, of course, since $\frac{\partial}{\partial t} = \Delta$, we can combine this with Section 2.2 to write down the general estimate:

$$\left|\frac{\partial^{a+1}}{\partial t^{a+1}}\nabla_x^b P(t,x)\right| \le \frac{C}{t+\delta^2(x)}\delta^{-b-2a}(x)P(t,x)\ ; \quad t>0,\ x\in D. \tag{2.23}$$

Because of (2.2) for the proof we may assume that $\delta(x) \le \sqrt{t}$. As before, to prove (2.22) it suffices to show that

$$\left|\nabla_x^b\big(P(1+h,x)-P(1,x)\big)\right| \le Ch\delta^{-b}(x)P(1,x)\ ; \quad x\in D. \tag{2.24}$$

Writing, therefore, $P(1+h,x)-P(1,x)$ as in Section 2.3 we must prove that:

(2.25)
$$\int |\nabla_x^b p_1(x,y)| M(y,h)\,dy \le c\delta^{-b}(x)\int p_{1+\delta^2(x)}(x,y)M(y,h)\,dy$$
$$= \int_{d(y,\partial D)\ge 1} + \int_{d(y,\partial D)\le 1} = I_1 + I_2 \le Ch\delta^{-b}(x)P(1,x).$$

The proof of this runs as before.

We have outlined here the proof for the classical case $L = \Delta - \partial_t$. For the more general operator (2.4) something of the above argument can be rescued, but one has to bear in mind that in general we have to replace (2.2) with a Hölder estimate. We shall leave matters at that.

An alternative argument based on the positivity $-\frac{\partial}{\partial t}P(t,x) = Q(t,x) = -\Delta P(t,x) \ge 0$ can also be used. We can scale and set $\delta(x) = 1$. Then we can consider separately $t<1$ as in Section 2.2, and $t>1$ where we use Harnack. *E.g.*:

$$\left|\nabla_x\frac{\partial}{\partial t}P(t,x)\right| \le C\ ; \quad P(t,x)\ge C\ ; \quad \delta(x)\ge 1/2,\ 0<t<1.$$

$$\left|\nabla_x\frac{\partial}{\partial t}P(t,x)\right| = |\nabla_x Q(t,x)| \le CQ(t+1,x)$$
$$\le \frac{c}{t}P(t+1,x) \le \frac{c}{t}P(t,x)\ ; \quad t>1,\ \delta(x)\ge 1/2.$$

## 3 The Parabolic Green Function

In this section I shall restrict myself to the case where $L = \Delta - \partial_t$ is the classical heat diffusion operator and $D$ will be a general Lipschitz domain as in (0.10). The estimate at infinity in Section 3.1 below hold for a general operator as in (2.4) but it did not seem worth the notational complications that this would involve to spell the proof out in that generality.

### 3.1 The Estimate at Infinity

In this subsection I shall prove the parabolic analogue of (0.23). The function (*cf.* Section 2)

$$G(s, x\,; t, y) = p_{t-s}(y, x)\,; \quad (t, y) \neq (s, x), \tag{3.1}$$

is parabolic in $(t, y)$ for every fixed $(s, x)$. The function $P(t, y)$ is also parabolic.

We shall normalize throughout by $\varphi(0) = 0$ in (0.10) and fix a reference point $(s_0, x_0) \in \dot{\mathbb{R}}^d$ with

$$x_0 = (r, 0),\ \delta(x_0) \sim r \ll 1\,; \quad c < s_0 < C, \tag{3.2}$$

(we could even normalize and set $s_0 = 1$). We shall use the same notation as in (1.3) and consider

$$x = (a, 0),\ \zeta = (b, 0)\,; \quad cR \leq a, b \leq CR, \tag{3.3}$$

where now $R$ will be a new parameter that satisfies:

$$cr \leq R \leq C. \tag{3.4}$$

A "summation" (as in (1.20)) will be performed on $R$ and the behaviour at infinity will now be captured by the fact that $r$ will be allowed to tend to zero.

It will also be convenient to introduce the following notation:

$$X_R = [(s, y) \in \mathbb{R} \times D\,; |s_0 - s| \leq cR^2, y \in T_R(Q)], \tag{3.5}$$

where now $Q = 0$, $T_R(Q)$ is as in (1.14), and $R$ is as in (3.4). We shall also define the following parabolic analii:

$$Y_R = X_{C_0R} \setminus X_R\,; \quad cr \leq R \leq C, \tag{3.6}$$

where $C_0$ will be chosen large enough. With the above notation we will chose the constants in (3.2), (3.4), (3.5) and (3.6) so that

$$(s, y) \in X_{C_0R} \Rightarrow c < s < C. \tag{3.7}$$

If $x$, $\zeta$ are as in (3.3), by the parabolic Carleson principle (*cf.* [18]) and (0.20), we have:

$$P(s_1, y) \leq CP(s_2, x) \sim CP(s_3, \zeta)\,; \quad y \in T_{C_1R}(0)\,; \quad c \leq s_1, s_2, s_3 \leq C. \tag{3.8}$$

Similarly, the Carleson principle for the parabolic function $u(t, x) = p_t(x, x_0)$ in the region $\dot{D}_1 \cup \dot{D}_2$ of (2.1) applies, and we see that for all $R$ as in (3.4) we have

$$p_{t-c_0R^2}(y, x_0) \leq Cp_t(x, x_0), \tag{3.9}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

where $x$, $x_0$ are as (3.2) and (3.3), and either, or both, the following two conditions hold:

$$y \in T_{c_0^{-1}R}(0), \quad t - c_0R^2 > c_0R^2, \tag{3.10}$$

$$y \in T_{c_0^{-1}R}(0), \quad |y - x_0| \geq c_0R, \quad t \in \mathbb{R}, \tag{3.11}$$

where $R$ is as in (3.3), and where the $c_0$ that appears in (3.9), (3.10) and (3.11) are now all *identical* ($c_0$ should be thought as "small" and the $C$ in (3.9) depends on $c_0$).

We come now to the principal ingredient of our argument. We shall apply the boundary parabolic Harnack principle (*cf.* [3], [25]) to the two functions $G(s, x, \cdot, \cdot)$ and $P(\cdot, \cdot)$ (for fixed $s, x$). We obtain:

$$\frac{p_{s_0-s}(x, x_0)}{p_{t-s}(x, \zeta)} = \frac{G(s, x\,;\, s_0, x_0)}{G(s, x\,;\, t, \zeta)} \leq C\frac{P(s_0, x_0)}{P(u, \zeta)}, \tag{3.12}$$

where $s_0$, $x_0$, $x$, $\zeta$ (and $r$, $R$) are as in (3.2) (3.3) and (3.4) and in addition:

$$s < s_0 < t = s_0 + cR^2\,; \quad c < u < s_0 - cR^2. \tag{3.13}$$

For the proof of (3.12) we also use (3.8), and to avoid the singularity $\big((\cdot, \cdot) = (s, x)\big)$ of $G(s, x\,;\, \cdot, \cdot)$ we must impose in addition on the variable $(\cdot, \cdot) = (s_0, x_0)$, one (or both) of the following two conditions:

$$|x - x_0| \geq cR, \tag{3.14}$$

$$s_0 - s \geq cR^2. \tag{3.15}$$

The correct way to think of the above is that the variable point $(\cdot, \cdot)$ lies in $(\cdot, \cdot) \in [c, C] \times T_{C_1R}(0)$ where $P(\cdot, \cdot)$ is certainly parabolic and that the additional conditions (3.14) and (3.15) take care of the "fixed" singularity $(s, x)$ of $G(s, x\,;\, \cdot, \cdot)$ as in Section 2.1.

In the above range of $x$, $\zeta$, $s$, $t$ we have therefore

$$p_{t-s}(x, \zeta) \leq c(t - s)^{-d/2} \leq C(t - s_0)^{-d/2} \sim R^{-d}, \tag{3.16}$$

$$p_{s_0-s}(x, x_0)P(u, \zeta) \leq CP(s_0, x_0)R^{-d}. \tag{3.17}$$

By (3.8), the $(u, \zeta)$ in the left hand side of (3.17) can be replaced by any point

$$(s_1, y_1) \in [c, C] \times T_{C_1R}(0). \tag{3.18}$$

We shall also apply (3.9) (with $t = s_0 - s$) and replace in (3.17) $p_{s_0-s}(x, x_0)$ by $p_{s_0-s-c_0R^2}(y, x_0)$ so as to obtain:

$$p_{s_0-s-c_0R^2}(y, x_0)P(s_1, y_1) \leq CP(s_0, x_0)R^{-d}, \tag{3.19}$$

with $(s_1, y_1)$ as in (3.18) and where $y$ is such that one of the conditions (3.10), (3.11) is verified. We shall choose the geometric constants $c$, $C$ and $C_0$ in (3.2)–(3.6) appropriately first, and then $c_0$ in (3.9), (3.10), (3.11) small enough. If this is done we have:

$$(s + c_0R^2, y) \in X_{C_0R} \Rightarrow y \in T_{c_0^{-1}R}. \tag{3.20}$$

But again if $c_0$ is small enough, and if $(s, y)$ are such that $(s', y) = (s + c_0R^2, y) \notin X_R$, then one or the other (or both) of the following will hold

$$s_0 - s - c_0R^2 = s_0 - s' \geq c_0R^2 \ ; \quad |y - x_0| \geq c_0R. \tag{3.21}$$

The final upshot is therefore:

$$\begin{aligned} &p_{s_0-s_2}(y_2, x_0)P(s_1, y_1) \leq CP(s_0, x_0)R^{-d} \ ; \\ &(s_1, y_1) \in [c, C] \times T_{C_1R}(0), \ (s_2, y_2) \in Y_R. \end{aligned} \tag{3.22}$$

By (3.7) and an appropriate choice of the constants, we see in particular that (3.22) holds for

$$(s_1, y_1), (s_2, y_2) \in Y_R. \tag{3.23}$$

What is important here is that $s_0 - s_2$ could $\sim 0$ (or even take negative values with the extension of $p_t(\cdot, \cdot)$ as in Section 2.1).

The estimates (3.22) will now be used to obtain an estimate of:

$$J_R = \int_{Y_R} p_{s_0-s}(y, x_0) \frac{P(s, y)}{\delta^{2+\varepsilon}(y)} \, ds \, dy. \tag{3.24}$$

To illustrate how one goes about this we shall, just as in Section 1.2, first consider the case when $D$ is convex (or $d = 2$) and we shall use the parabolic boundary Harnack principle to compare the parabolic function

$$u(t, y) = p_t(y, x_0) \ ; \quad \text{or } u(t, y) = P(t, y),$$

with some positive harmonic function $U(y)$ in $T_{C_0R}(0)$ that vanishes at the boundary.

For the use of the parabolic Harnack boundary principle for $P(s, y)$ we simply use the fact that in $Y_R$ we have $s > c$ (*cf.* (3.7)) and $y \in T_{C_0R}(0)$. For $p_{s_0-s}(y, x_0)$ on the other hand, as before, we have to use the larger domain $\dot{D} = \dot{D}_1 \cup \dot{D}_2$ (*cf.* (2.1), (3.21)). Once this is done, a simple use of the Lemma in Section 1.2 yields

$$u(s_0 - s, y) \leq Cu(s_0 - s + cR^2, y_R) \left(\frac{\delta(y)}{R}\right)^{\alpha}, \tag{3.25}$$

$$P(s, y) \leq CP(s + cR^2, y_R) \left(\frac{\delta(y)}{R}\right)^{\alpha}, \tag{3.26}$$

where $y_R = (cR, 0)$, and where $\alpha = 1$ in the convex case, and $\alpha > 1/2$ if $d = 2$. A use of (0.20) and (3.22) gives therefore the estimate:

$$J_R \leq C \left( \int \frac{ds dy}{\delta^{2+\varepsilon-2\alpha}(y)} \right) P(s_0, x_0) R^{-d-2\alpha}. \tag{3.27}$$

As in Section 1.2, if $0 < \varepsilon < 2\alpha - 1$ the $dy$ integration in (3.27) can be estimated by $O(R^{d-2-\varepsilon+2\alpha})$. The integration in $s$ gives an extra factor $R^2$. If we put everything together, we obtain

$$J_R \leq C P(s_0, x_0) R^{-\varepsilon}. \tag{3.28}$$

The important thing about the estimate (3.28) is that it can be summed through a geometric sequence $R = R_j$:

$$R_j \sim \delta(x_0) c^j \; ; \quad j = 1, \ldots, N \; ; \quad R_N \sim 1. \tag{3.29}$$

The constant $c > 0$ and $N$ in (3.29) will be chosen so that:

$$\bigcup_{j=1}^{N} Y_{R_j} \supset X_C \setminus X_{c\delta(x_0)} = Y_\infty. \tag{3.30}$$

Once this is done, we obtain the following estimate for the parabolic Green function:

$$\int_{Y_\infty} p_{s_0-s}(x, x_0) \frac{P(s,x)}{\delta^{2+\varepsilon}(x)} \, dx \, ds \leq C \delta^{-\varepsilon}(x_0) P(s_0, x_0). \tag{3.31}$$

In fact from this we can obtain an estimate for the following *critical* region of integration in (0.25)

$$\int_\eta^{s_0} \int_{|x-x_0| \leq C} p_{s_0-s}(x, x_0) \frac{P(s,x)}{\delta^{2+\varepsilon}(x)} \, dx \, ds \leq C_2 \frac{P(s_0, x_0)}{\delta^\varepsilon(x_0)}, \quad 0 < \eta < 1. \tag{3.32}$$

Indeed the difference between the regions of integration in (3.31) and in (3.32) is the region:

$$[s_0 - c\delta^2(x_0), s_0] \times [|x - x_0| \leq 1/10\delta(x_0)], \tag{3.33}$$

and the following estimates

$$\delta(x) \sim \delta(x_0) \; ; \quad \int p_{s_0-s}(x_0, x) \, dx \leq 1 \; ; \quad P(s, x) \leq C P(s_0, x_0), \tag{3.34}$$

that are either trivial or a consequence of (0.20) and (0.21) in that region, give at once the required result (3.32). What has at least been proved is that there exist $C, C_2, C_3$ and $0 < \eta < 1$ s.t. (3.32) holds for $\delta(x_0) \leq C_3$. If we think a bit about the choice of the constants in the course of the above proof we see that $C, C_3$ and $\eta$ can be chosen at will and that $C_2$ will depend on that choice. In view of Section 3.2 we do not need this more precise statement.

The modifications of the above proof that are needed to handle the general domain $D$ when $d > 2$ are not trivial, but they follow strictly identical lines as in the elliptic case in Section 1.2, or as in Section 2.3. We shall use the maximal operator $F^*$ of (1.24), for fixed $s$, on the two functions $F(y) = p_{s_0-s}(y, x_0)$ and $P(s, y)$, where the sup in the definition of $F^*$, for fixed $s$, is restricted to be in the range where $(s, y) \in Y_R$. With the notation $y = (y_1, y')$ of (0.10), we can control then:

$$J_R \leq C \int p^*_{s_0-s}(y', x_0) P^*(s, y') \delta^{-\varepsilon}(y)\, ds\, dy, \tag{3.35}$$

where the integration in $(s, y)$ is taken in the same range as in (3.24). The integration in $y_1$ brings out a factor $R^{1-\varepsilon}$ as long as $0 < \varepsilon < 1$. We are therefore left with the estimate

$$J_R \leq CR^{1-\varepsilon} \int p^*_{s_0-s}(y', x_0) P^*(s, y')\, ds\, dy'. \tag{3.36}$$

The Harnack boundary comparison with harmonic functions $U_s(y)$, $V_s(y)$ appropriately normalized at $y_R = (cR, 0)$ gives then, with the help of the key lemma of Section 1.2, the estimate:

$$J_R \leq CR^{1-\varepsilon} \iint U^*_s(y') V^*_s(y')\, ds\, dy' \leq CR^{1-\varepsilon} \int R^{d-3} p_{s_0-s}(y_R, x_0) P(s, y_R)\, ds, \tag{3.37}$$

(what we obtain in the integrant by the parabolic boundary Harnack principle is $\cdots p_{s_0-s\pm cR^2}(\cdot,\cdot) P(s \pm cR^2, \cdot) \cdots$ but, of course, a change of variable and (0.20) reduces this to the above). This together with the estimate (3.22), and the fact that the integration range in $s$ is of the order $R^2$, finally gives the same estimate (3.28) and the proof finishes as before. We obtain thus the estimate (3.32) in this general situation for any $0 < \varepsilon < 1$.

### 3.2 The Proof of (0.25): The Convex Case

The estimates (0.26) and (0.27) are a consequence of (3.32) and of the following estimates: For all $c > 0$, $\eta < 1$ there exists $C > 0$ such that

$$I = \int_0^1 \int_{|x-x_0|>c} \cdots \leq CP(1, x_0)\,; \quad x_0 \in D, \tag{3.38}$$

$$J = \int_0^\eta \int_{|x-x_0|<c} \cdots \leq CP(1, x_0)\,; \quad x_0 \in D, \tag{3.39}$$

$$K = \int_0^1 \int_{|x-x_0|\leq\delta(x_0)/2} \cdots \leq CP(1, x_0)\,; \quad x_0 \in D,\ \delta(x_0) \geq c, \tag{3.40}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

where the integrand $\cdots$ in the above integrals is:

$$p_{1-s}(x, x_0)\frac{P(s, x)}{\delta^{2+\varepsilon}(x)}\,dx\,ds.$$

The estimate of $K$ follows from the more general estimate:

$$K \le \int_0^1 \int_{\delta(x)>c} \cdots \le C \int_0^1 \int_D p_{1-s}(x, x_0)P(s, x)\,ds\,dx = CP(1, x_0), \tag{3.41}$$

where we use the evolution equation:

$$e^{-t\Delta}P(s, x) = P(s + t, x). \tag{3.42}$$

Alternatively, using the fact that the integrant of (3.41) is $\le Cp_{1-s}(x, x_0)$, it follows that $K \le C$. And since for $\delta(x_0) \ge 1$ we have $P(1, x_0) \ge c$, the estimate (3.40) follows.

To estimate $I$ and $J$ we shall make essential use of the kernel estimate (0.18). This implies that the integrant of $I$, $J$ satisfies

(3.43)
$$\cdots \le C(1 - s)^{-d/2}P(1 - s, x)P(1 - s, x_0)P(s, x)\delta^{-2-\varepsilon}(x)\exp\left(-c\frac{|x - x_0|^2}{1 - s}\right).$$

From this if $D$ is convex the estimate (3.39) follows at once. Indeed, by comparing $P(t, x)$ with probability of life in a half-space, we see that in the convex case we have:

$$P(t, x) \le C\frac{\delta(x)}{\sqrt{t}}\,; \quad x \in D,\ t > 0. \tag{3.44}$$

Combining (3.43), (3.44) and (0.20) we see that the integrant of $J$ in (3.39) can be estimated by

$$\frac{P(1 - s, x_0)P(1 - s, x)P(s, x)}{\delta^{2+\varepsilon}(x)} \le CP(1, x_0)\frac{1}{\delta^{\varepsilon}(x)\sqrt{s(1 - s)}}.$$

This gives a finite integral and proves (3.39). A similar argument can be used for a general Lipschitz domain (0.10) provided that the dimension $d = 2$. (3.44) is replaced then by the estimate

$$P(t, x) \le C\left(\frac{\delta(x)}{\sqrt{t}}\right)^{\alpha}; \quad x \in D,\ t > 0, \tag{3.45}$$

where $C > 0$ and $\alpha = \alpha(A) > 1/2$. The estimate (3.39) no longer holds for any $0 < \varepsilon < 1$, but for $0 < \varepsilon < 2\alpha - 1$ small enough.

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

The estimate of $I$ is more involved. We shall write

$$I = \int_0^{1/2} \int_{|x-x_0|\geq c} \cdots + \int_{1/2}^1 \int_{|x-x_0|\geq c} \cdots = I_1 + I_2, \tag{3.46}$$

and with the help of (0.20) (3.43) we see that:

$$I_1 \leq CP(1,x_0) \int_0^{1/2} \int_D \frac{P(1,x)P(s,x)}{\delta^{2+\varepsilon}} \exp(-c|x-x_0|^2)\, ds\, dx. \tag{3.47}$$

Now if $D$ is convex we can use (3.44) (and if $d = 2$ we can use (3.45)) to estimate $P(s,x)$ in the integrant. The first part of the following Lemma does the rest.

***Lemma*** *For all $c > 0$ and $0 \leq \beta < 1$ we have:*

$$\int_D \delta^{-\beta}(x) \exp\left(-\frac{|x-x_0|^2}{cs}\right) dx \leq C(d,A,\beta,c)s^{d/2-\beta/2}\,; \quad 0 < s < 1. \tag{3.48}$$

*For all $c, c_1 > 0$ and all $0 \leq \beta < 1$, $\mu \geq 0$ we have:*

$$\int_{D\cap[|x-x_0|\geq c_1]} \delta^{-\beta}(x) \exp\left(-\frac{|x-x_0|^2}{cs}\right) dx \leq Cs^{\mu}\,; \quad 0 < s < 1, \tag{3.49}$$

*where $C = C(d,A,\beta,\mu,c,c_1)$.*

If we scale and denote by $D_s = \sqrt{s}D$ the corresponding scaled domain, we see that (3.48) and (3.49) amounts to estimates of

$$s^{d/2-\beta/2} \int_{D_s} \delta^{-\beta}(x) \exp(-c|x-x_0|^2)\, dx, \tag{3.50}$$

$$s^{d/2-\beta/2} \int_{D_s\cap[|x-x_0|\geq c_1/\sqrt{s}]} \delta^{-\beta}(x) \exp(-c|x-x_0|^2)\, dx. \tag{3.51}$$

To estimate the above two integrals we use our previous notation $x = (x_1, x')$, $x_0 = (x_1^0, x_0')$ and factorize

$$\exp(-c|x-x_0|^2) = \exp(-c|x_1 - x_1^0|^2) \exp(-c|x' - x_0'|^2). \tag{3.52}$$

The integration in (3.50) is performed first in $x_1$ (this is possible since $\beta < 1$), and then in $x' \in \mathbb{R}^{d-1}$. This bounds the integral (3.50) and proves (3.48). To estimate the integral (3.51) and prove (3.49), we must make this additional observation that when $|x - x_0| \geq c_1/\sqrt{s}$, then either or both:

$$|x_1 - x_1^0| \geq c_1/\sqrt{s}\,; \quad |x' - x_0'| \geq c_1/\sqrt{s},$$

have to hold. This completes the proof of the Lemma.

The Lemma gives us therefore the estimate (3.38) for $I_1$ when $D$ is convex (or when $d = 2$ provided that we assume that $0 < \varepsilon < 2\alpha - 1$ with $\alpha \in \left]1/2, 1\right]$ as in (3.45)).

The integral $I_2$ is more subtle and we have to use (0.20) which implies that

$$P(t,x) \leq Ct^{-c}P(1,x)\ ; \quad 0 < t < 1,\ x \in D, \tag{3.53}$$

for some $c = c(d, A)$. The estimates (3.43) and (3.53) can then be used to estimate the integrant of $I_2$ by:

$$s^{-c}(1-s)^{-c}P(1,x_0)\frac{P^2(1,x)}{\delta^{2+\varepsilon}(x)}\exp\left(-c\frac{|x-x_0|^2}{1-s}\right).$$

Under the estimates (3.44) or (3.45) we can replace $P^2(1,x)\delta^{-2-\varepsilon}(x)$ by $\delta^{-\lambda}(x)$ ($\lambda < 1$), and the integration in $x$ in $I_2$ can be estimated by (3.49). If $\mu \geq c$ the integration in $s \in [1/2, 1]$ can be performed and the estimate (3.38) follows.

### 3.3 Proof of (0.25): General Case

As we already pointed out, an easy consequence of (0.20) is that:

$$P(t,x) \leq Ct^{-c}P(1,x)\ ; \quad 0 < t < 1,\ x \in D, \tag{3.54}$$

with $C, c > 0$. This combined with (3.43) allows us to estimate the integrant $\cdots$ in (3.38) and (3.39) by

$$\begin{aligned}&\delta^{-2-\varepsilon}(x)P(1,x_0)P(s,x)P(1-s,x)(1-s)^{-c}\exp\left(-c\frac{|x-x_0|^2}{1-s}\right)\\&\qquad\leq \delta^{-\varepsilon}(x)P(1,x_0)\frac{P(s,x)P(1-s,x)}{\delta^2(x)}\exp(-c|x-x_0|^2),\end{aligned}\tag{3.55}$$

(this, of course, only holds in the integration range of $I$ and $J$).

If we integrate in $s$ and take (0.20) into account, we see that the contribution coming from the middle factor of the right hand side of (3.55) can be estimated by (we split: $\int_0^1 = \int_0^{1/2} + \int_{1/2}^1$)

$$\delta^{-2}(x)P(1,x)\int_0^1 P(s,x)\,ds \leq CP(1,x)Q(x)\delta^{-2}(x)\ ; \quad x \in D,\ \delta(x) \leq 1, \tag{3.56}$$

where

$$Q(x) = \int_0^1 P(s,x)\,ds. \tag{3.57}$$

If we restrict the $dx$-integration, as we may (*cf.* (3.41)), to $\delta(x) \leq 1$ we see that what has to be proved is:

$$\int_{\delta(x)\leq 1} P(1,x)Q(x)\delta^{-2-\varepsilon}(x)\exp(-c|x-x_0|^2)\,dx \leq C(d,A,\varepsilon)\ ; \quad 0 < \varepsilon < 1. \tag{3.58}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

This in one stroke proves (3.38) and (3.39).

By splitting as before:

$$\exp(-c|x-x_0|^2) = \exp(-c|x_1 - x_1^0|^2)\exp(-c|x' - x_0'|^2), \tag{3.59}$$

where $x_0 = (x_1^0, x_0')$, and then integrating first in $x_1$ we end up with having to control the integral

$$\int_{\delta(x)\le 1} P(1,x)Q(x)\delta^{-2-\varepsilon}(x)\exp(-c|x'-x_0'|^2)\,dx. \tag{3.60}$$

To achieve this we just have to prove

$$\int_{T(Q)} P(1,x)Q(x)\delta^{-2-\varepsilon}(x)\,dx \le C\,; \quad Q \in \partial D \tag{3.61}$$

where $T(Q) = T_5(Q)$ is as in (1.14). This will be done in the next few lines and will complete the proof of (0.25).

We clearly have

$$\Delta Q(x) = \int_0^1 \frac{\partial}{\partial s} P(s,x)\,ds = P(1,x) - 1 = -F(x), \tag{3.62}$$

where $0 \le F \le 1$ and

$$0 \le Q \le 1, \quad Q|_{\partial D} = 0. \tag{3.63}$$

We shall use the notation of Section 1.2 and the normalization (1.1). From (3.62) and (3.63) it follows that

$$Q(x) = \int_T G(x,y)F(y)\,dy + U(x) = G(x) + U(x)\,; \quad x \in T,$$

where $T = T_{10}(0)$ (as in (1.14)) where $G$ is as in (0.22) and where $U(\cdot)$ is a harmonic function in $T$, bounded by $C = C(d)$, and such that $U|_{\partial D} = 0$. The contribution of $U$ in (3.61) can be controlled in (*cf.* Section 1.2 and (3.37)) by the Key-Lemma.

What remains to be proved is the following:

$$\int_T P(1,x)G(x)\delta^{-2-\varepsilon}(x)\,dx \le C. \tag{3.64}$$

Towards that we can use the parabolic Harnack boundary principle and (0.20) to compare $P(1,x)$ with the Martin point $u$ of (0.23) normalized to be 1 at $(x_1,x') = (1,0)$. The integral (3.64) can thus be estimated by:

$$\int_T \int_T u(x)G(x,y)\delta^{-2-\varepsilon}(x)\,dx\,dy. \tag{3.65}$$

We shall perform the $x$-integration first in (3.65) and use (0.24). The integration in $y$ can then be performed because $\varepsilon < 1$. This completes the proof. ■

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

### 3.4 The Gradient Estimates

The function $w(t,x)$ is not, *a priori*, a smooth function, simply because $\delta(x)$ is not. So the first step is to modify $\delta(x)$ into $\delta_0(x) \in C^\infty(D)$ so that

$$C^{-1} < \delta/\delta_0 \le C\,; \quad |\nabla^k \delta_0| \le C\delta_0^{1-k},\ k = 0, 1, \dots. \tag{3.66}$$

The construction of $\delta_0$ is standard and it can be done by considering a Whitney decomposition of $D$ (*cf.* [4] where I spelled out the details). I shall then define $w(t,x)$ as in (0.25) with this modified $\delta_0(x)$ and I shall prove in this section that:

$$\left|\frac{\partial^a}{\partial t^a}\nabla_x^b w(t,x)\right| \le C\frac{P(t,x)}{\delta^{2a+b}(x)\delta^\varepsilon(t,x)}\,; \quad x \in D,\ t > 0, \tag{3.67}$$

where $C = C(d, A, \varepsilon, a, b)$. These estimates should be compared with the estimates of Section 2.5, and by scaling we can even assume $t = 1$ in (3.67). For technical reasons we shall assume as we may that $d \ge 3$. (Indeed in the definition of $w$ we can replace $D \subset \mathbb{R}^d$ by $D_{\text{new}} = D \times \mathbb{R} \subset \mathbb{R}^{d+1}$ and it is clear that $w_{\text{new}} = w_{\text{old}}$, $P_{\text{new}} = P_{\text{old}}$.)

The problem lies in the integration near the singularity. Indeed for any $r \le \delta(x)/2$ by using the Harnack estimate that "dips in $t < 0$" (*cf.* Section 2.1), we obtain (for typographical reason I drop the suffix and denote $\delta_0$ by $\delta$)

$$\begin{aligned}
&\int_0^1 \int_{|x-y|>r} |\nabla_x^k p_{1-s}(x,y)| \frac{P(s,y)}{\delta^{2+\varepsilon}(y)}\, ds\, dy \\
&\qquad \le \frac{C}{r^k}\int_0^1 \int p_{1-s+r^2}(x,y)\frac{P(s,y)}{\delta^{2+\varepsilon}(y)}\, ds\, dy \\
&\qquad \le Cr^{-k} w(1+r^2, x) \\
&\qquad \le Cr^{-k}\frac{P(1+r^2,x)}{\text{Min}[\delta(x), \sqrt{1+r^2}]^\varepsilon} \\
&\qquad \le Cr^{-k}\frac{P(1,x)}{\text{Min}[\delta(x), 1]^\varepsilon}.
\end{aligned} \tag{3.68}$$

The obvious scaling gives the estimate (3.67) for the space derivative, in the integration range:

$$y \in D, \quad |x-y| \ge \delta(x)/2. \tag{3.69}$$

Observe on the other hand, that quite generally we have (*cf.* the line below (0.25)):

$$\begin{aligned}
&\frac{\partial}{\partial t}\int_{-\infty}^t \int_D p_{t-s}(x,y)F(s,y)\, dy\, ds \\
&\qquad = F(t,x) + \int_{-\infty}^t \int_D \Delta_x p_{t-s}(x,y)F(s,y)\, dy\, ds\,; \quad t > 0,
\end{aligned} \tag{3.70}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

for any "reasonable" function $F(t, y)$, $(t, y) \in \mathbb{R} \times D$. If we set $F(t, y) = \frac{P(t,y)}{\delta^{2+\varepsilon}(y)}\mathbb{I}$ $(t > 0)$ and use Section 2 we see that the above formula reduces every $\frac{\partial}{\partial t}$ into $\sum \frac{\partial^2}{\partial x_i^2}$. This formula can be iterated (*cf.* (2.23)) and we see that in proving (3.67) we can replace $a$ by 0 and $b$ by $b + 2a$.

We see finally that what has to be estimated is:

$$\int_0^1 \int_{|x-y| \le \delta(x)/2} \nabla_x^k p_{1-s}(x, y) \frac{P(s, y)}{\delta^{2+\varepsilon}(y)} \, dy \, ds. \tag{3.71}$$

Observe that, as long as we estimate (3.71) by replacing $\nabla_x^k$ by $|\nabla_x^k|$ we can use the fact that in the integration range (3.71) we have:

$$\delta^{-2-\varepsilon}(y) \sim \delta^{-2-\varepsilon}(x) = \delta^{-2-\varepsilon}, \tag{3.72}$$

and bring that factor outside of the integral.

To estimate (3.71) we shall use the Harnack estimate (*cf.* Section 2.1), (0.13) and (0.20), and we have

$$\begin{gathered} |\nabla_x^k p_t(x, y)| \le C\delta^{-k}(t, x) p_{t+\delta^2(t,x)}(x, y) \le C\delta^{-k} t^{-d/2} P(t, x) P(t, y) \; ; \\ t \ge \delta^2(x) = \delta^2, \; \delta(t, x) = \mathrm{Min}[\delta(x), \sqrt{t}]. \end{gathered} \tag{3.73}$$

So if we assume that $\delta = \delta(x) < 1$ (as we may, since the case $\delta \ge 1$ is covered by (3.77) below), the part of (3.71) that is away from the singularity $s = 1$, can be estimated by

$$\begin{aligned} &\int_0^{1-\delta^2} \int_{|x-y| \le \delta/2} |\nabla_x^k p_{1-s}(x, y)| P(s, y) \, ds \, dy \\ &\qquad \le \delta^{-k} \int_0^{1-\delta^2} \int_{|x-y| \le \delta/2} (1-s)^{-d/2} P(1-s, x) P(1-s, y) P(s, y) \, ds \, dy. \end{aligned} \tag{3.74}$$

We also have:

$$P(1-s, x) P(1-s, y) P(s, y) \le C P(1, x). \tag{3.75}$$

To see (3.75) we choose $s$ or $1 - s$, whichever is closest to 1, to apply (0.20). This estimates the left hand side of (3.75) by $P(1, y)$. Then use Harnack and the fact that $\delta < 1$ and (3.75) follows.

Since $d \ge 3$ the integration

$$\int_0^{1-\delta^2} \int_{|x-y| < \delta/2} (1-s)^{-d/2} \, ds \, dy \le c\delta^2, \tag{3.76}$$

gives the estimate $P(1, x)\delta^{-k+2}$ for (3.74). If we insert the cofactor (3.72) we obtain for (3.71) the required estimate $P(1, x)\delta^{-k-\varepsilon}$.

We are therefore left with the task of analyzing (3.71) in the integration range:

$$\int_{\mathrm{Max}[0,1-\delta^2]}^{1} \int_{|x-y|<\delta/2} \cdots\,; \quad x \in D. \tag{3.77}$$

This will need new ideas. To get a clue of how to proceed we shall first give the analogous estimate for the elliptic Green potential $w(x)$ in (0.24), where a similar analysis applies. What corresponds to (3.77) is then:

$$\int_{|x-y|\le\delta/2} \nabla_x^k G(x,y) u(y)\, dy, \tag{3.78}$$

where (*cf.* (0.22))

$$\nabla_x^k G(x,y) = O(|x-y|^{2-d-k}) + \nabla_x^k H(x,y). \tag{3.79}$$

For $\nabla_x^k H(x,y)$ we can use Harnack and the fact that $H(x,y) \le C\delta^{2-d}$. But for the first term in (3.79) it is essential to use integration by parts.

The same procedure can be applied to the parabolic Green potential where (: $t = 1-s$):

$$p_t(x,y) = ct^{-d/2} \exp\left(-\frac{|x-y|^2}{4t}\right) + K(t\,;x,y) = g_t(x,y) + K(t\,;x,y), \tag{3.80}$$

and where $K$ is the appropriate correcting term (*cf.* [1]).

In view of the integration by parts we shall write:

$$\nabla_x^k p_{1-s}(x,y) = \nabla_y^k p_{1-s}(x,y) + [\nabla_x^k - \nabla_y^k] p_{1-s}. \tag{3.81}$$

Integration by parts will be used to deal with the first term in the right hand side of (3.81). What has to be estimated is

$$\iint p_{1-s}(x,y) \nabla_y^p P(s,y)\, ds\, dy\,; \quad p = 0,1,\ldots,k, \tag{3.82}$$

integrated in the range of (3.77). And in addition to that we have to estimate the boundary terms:

$$\int_{\mathrm{Max}[0,1-\delta^2]}^{1} \int_{|x-y|=\delta/2} \cdots\,. \tag{3.83}$$

The second term in the right hand side of (3.80) gives rise to an additional error terms

$$\iint (|\nabla_x^p| + |\nabla_y^p|) K(1-s,x,y) P(s,y)\, ds\, dy\,; \quad p = 1,2,\ldots \tag{3.84}$$

integrated in the same range as (3.77). To see that (3.84) and (3.82) suffice, we use (3.81), (3.80) and the fact that $\nabla_x^k g_t = \pm \nabla_y^k g_t$.

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

Observe finally that the factor (3.72) now cannot come out of the integral before, but only after we perform the integration by parts (and we put absolute values inside the integrals). This means that in the consideration that follows this cofactor is not $\delta^{-2-\varepsilon}$ but $\delta^{-2-\varepsilon-m}$ ($m = 0, 1, \ldots$, *cf.* (3.66)), and a certain amount of "bookkeeping" has to be done in keeping track of the various terms that appear in the various integrations by parts. This is also the reason why in (3.82) we consider all the gradients $\nabla^p$, $0 \le p \le k$.

To estimate (3.82) we use the spacial gradient estimate (2.3)

$$|\nabla_y^p P(s, y)| \le C\delta^{-p} P(s, y), \tag{3.85}$$

and the evolution equation (*cf.* (3.42))

$$\int_D p_{1-s}(x, y) P(s, y)\, dy = P(1, x). \tag{3.86}$$

The time integration in the range of (3.83) gives an additional factor $\le \delta^2$ (consider the two cases $\delta < 1$, $\delta > 1$). If we put everything together we obtain for (3.82) the estimate

$$\frac{P(1, x)}{\delta^{p-2}}.$$

This together with the cofactor $\delta^{-2-\varepsilon-m}$ (3.72) gives the required result.

For the proof of the boundary estimate (3.83) we shall need the additional estimate:

$$|\nabla_x^p p_t(x, y)| \le C\delta^{-p-d}\,; \quad t > 0,\ p = 0, 1, \ldots,\ |x - y| = \delta(x)/2 = \delta/2. \tag{3.87}$$

To see this we first use the Harnack estimates of Section 2.1 (we dip once more in the negative time range). This reduces the estimate (3.87) to the case $p = 0$. But then $p_t(x, y) \le C t^{-d/2} \exp(-\frac{|x-y|^2}{4t})$ and (3.87) follows.

The boundary term (3.83) that has to be estimated is

$$\iint_{|x-y|=\delta/2} |\nabla_y^\alpha p_{1-s}(x, y)|\, |\nabla_y^\beta P(s, y)|\, ds\, d\sigma_{d-1}, \tag{3.88}$$

where $\alpha + \beta = p \le k - 1$ and $\sigma_{d-1}$ is surface, $d - 1$ dimensional, measure on $|x - y| = \delta/2$. The use of the Harnack estimates of Section 2.1 (once more we have to dip into the negative time region) and of the estimates of $\nabla_y^k P$ (*cf.* (0.21), (2.3)) allows us to estimate the surface integral in (3.88) by (*cf.* (3.86)):

$$\begin{aligned} \delta^{-p}\frac{1}{\delta}\int_{1/4\delta \le |x-y| \le 3/4\delta} p_{1+\delta^2-s}(x, y) P(s, y)\, dy &\le C\delta^{-p-1} P(1 + \delta^2, x) \\ &\le C\delta^{-p-1} P(1, x). \end{aligned}$$

This together with the cofactor (3.72) gives the estimate:

$$\left(\int_{\mathrm{Max}[0,1-\delta^2]}^1 ds\right) \delta^{-2-\epsilon-k} P(1, x) = \frac{\mathrm{Min}[1, \delta^2]}{\delta^2} \delta^{-k-\varepsilon} P(1, x),$$

which is again the required result.

For the final correcting term (3.84) we shall need an "explicit" formula for $K(t, x, y)$ in (3.80). Probabilistically this formula is obtained by conditioning on (notation of (0.1))

$$\big(\tau, b(\tau)\big) \in [0, t] \times \partial D, \tag{3.89}$$

which means that, if $\mu \in M\big[[0, t] \times \partial D\big]$ denotes the hitting measure that corresponds to (3.89), we have

$$K(t, x, y) = -\int_0^t \int_{\partial D} g_{t-s}(y, \xi)\, d\mu(s, \xi), \tag{3.90}$$

where in both (3.89) and in the definition of the hitting measure we use $\mathbb{P}_x(\cdots)$ on $(b(t), t > 0)$. This means that the integrant of (3.84) can be estimated directly (as in (3.87) only simpler) and we obtain

$$|\nabla_y^p K(t, x, y)| \leq C\delta^{-p-d}.$$

The $\int_{|x-y|\leq\delta/2} dy$ integration in (3.84) gives $\delta^d$, the $s$-integration gives $(1 - \mathrm{Max}[0, 1-\delta^2]) \leq \delta^2$, and the cofactor (3.72) (or rather the corresponding gradients (3.66)) does the rest. The estimate of $|\nabla_x^p K(t, x, y)|$ is identical because the reversibility of the process guarantees that:

$$K(t\,; x, y) = K(t\,; y, x).$$

The proof of (3.67) is finally complete.

## 3.5 The Lower Bound

By comparing $p_t^D(x, y)$ of (0.2) with the corresponding $p_t^B(x, y)$, where $B = B_\delta(x_0)$ $\delta = \delta(x_0)$ is the Euclidean ball, we see that for all $c > 0$ there exists $C > 0$ such that

$$p_t^D(x, y) \geq C\delta^{-d}\,; \quad x_0 \in D,\ x, y \in B_{\delta/10}(x_0),\ t \leq c\delta^2\,; \quad \delta = \delta(x_0).$$

From this it clearly follows that for all $c > 0$ there exists $C > 0$ such that

$$\begin{aligned} w(1, x_0) &\geq \int_{\mathrm{Max}[0,(1-\delta^2)]}^1 \int_{|x-x_0|<\delta/10} p_{1-s}(x, x_0) \frac{P(s, x)}{\delta^{2+\varepsilon}(x)}\, ds\, dx \\ &\geq C\frac{P(1, x_0)}{\delta^\varepsilon(x_0)}, \quad \delta(x_0) \leq c. \end{aligned}$$

Upon scaling this gives the lower bound:

$$w(t, x_0) \geq C\frac{P(t, x_0)}{\delta^\varepsilon(x_0)}\,; \quad x_0 \in D,\ t > 0,\ \delta(x_0) \leq c\sqrt{t},$$

($c$ and $C$ as above).

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

### 3.6 A Further Estimate

From the results that we have proved up to here, it is not hard to deduce that for all $\alpha > 1, 0 < \varepsilon < 1, t > 0, x_0, x_1 \in D$ we have:

$$\int_0^t \int_D \frac{p_{t-s}(x, x_0) p_s(x, x_1)}{\delta^{2+\varepsilon}(x)} \, ds \, dx \leq C(d, A, \alpha, \varepsilon) \frac{p_{\alpha t}(x_0, x_1)}{\big(\mathrm{Min}[\delta(x_0), \delta(x_1)]\big)^{\varepsilon}}.$$

The details of the proof will not be given here because no use of this estimate will be made in the paper. This, however, is the parabolic version of (1.31) and it "contains" all the other previous estimates *e.g.* (0.25), (1.31) and others.

## 4 The Discrete Potential Theorem

### 4.1 The Time-Space Process

To the process $(z(n) \in \mathbb{R}^d; n = 0, 1, \dots)$ in (0.5) we shall associate the corresponding time space process:

$$\dot{z}(n) = \big(t_0 - n, z(n)\big) \in \dot{\mathbb{R}}^d = \mathbb{R} \times \mathbb{R}^d\,; \quad n \geq 0,\ t_0 \in \mathbb{R}, \tag{4.1}$$

where I use Doob's notation for time-space (*cf.* [1]). For any domain $\dot{D} \subset \dot{\mathbb{R}}^d$ I shall consider as before the first exit time:

$$\dot{\tau} = \inf[n \geq 0\,; \quad \dot{z}(n) \notin \dot{D}]. \tag{4.2}$$

To the process (0.5) (4.1) we can associate the "parabolic" operator:

$$\begin{aligned} Lf = L_\mu f(t, x) &= f * (\mu - \delta) - \delta_t f \\ &= \int_{\mathbb{R}^d} f(t, x - y) \, d(\mu - \delta)(y) - [f(t+1, x) - f(t, x)]\,; \quad (t, x) \in \dot{\mathbb{R}}^d. \end{aligned} \tag{4.3}$$

This is defined for any function $f$ on $\dot{\mathbb{R}}^d$ for which the above integral is absolutely convergent. A function $f$ as above will be called a super- [resp.: sub-] solution of $L_\mu$ on some domain $\dot{D} \subset \dot{\mathbb{R}}^d$ if

$$Lf(t, x) \leq 0 \text{ [resp.: } Lf \geq 0]\,; \quad (t, x) \in \dot{D}. \tag{4.4}$$

If on the process (0.5) or (4.1) I fix the starting points I shall denote the corresponding probabilities by:

$$\mathbb{P}_{x_0}[z(0) = x_0] = 1 = \mathbb{P}_{t_0, x_0}[\dot{z}(0) = (t_0, x_0)],$$

for some fixed $t_0 \in \mathbb{R}$, $x_0 \in \mathbb{R}^d$. Any positive function $f \geq 0$ on $\dot{\mathbb{R}}^d$ that is a supersolution in $\dot{D} \subset \dot{\mathbb{R}}^d$ satisfies

$$\mathbb{E}_{t_0, x_0}\big[f\big(\dot{z}(\dot{\tau})\big)\big] \leq f(t_0, x_0)\,; \quad (t_0, x_0) \in \dot{D}. \tag{4.5}$$

Similarly if $f$ is a bounded subsolution then:

(4.6) $$\mathbb{E}_{t_0,x_0}\big[f\big(\dot{z}(\dot{\tau})\big)\big] \geq f(t_0,x_0)\,; \quad (t_0,x_0)\in\dot{D}.$$

(4.5) [resp.: (4.6)] is of course the optional super- [resp.: sub-] martingale stopping theorem [19].

We shall now specialize and consider domains of the form:

(4.7) $$\dot{D} = D(b,a) = (t>b)\times D_a \subset \dot{\mathbb{R}}^d\,; \quad D_a = D + a(1,0,\dots,0)\,; \quad a,b\geq 0,$$

where $D$ is a globally Lipschitz domain as in (0.10).

For these domains if we denote by $\tau=\tau_a$ the first exit time of $z(n)$ from $D_a$ as in (0.7), and $\dot{\tau}$ the first exit time of $\dot{z}(n)$ from $D(b,a)$, and if $b,t_0\geq 0$ with $t_0\geq b$ and $t_0-b\in\mathbb{Z}$, we clearly have:

(4.8)
$$\dot{\tau}\leq t_0-b\,; \quad [\tau<t_0-b]=[\dot{\tau}<t_0-b]\,; \quad [\tau\geq t_0-b]=[\dot{\tau}=t_0-b],$$

where the last assertion in (4.8) is a consequence of the first two.

Observe also that if we denote by

(4.9) $$x_a = x - a(1,0,0,\dots,0), \quad x\in\mathbb{R}^d,$$

we have:

(4.10) $$\mathbb{P}_{x_a}[\tau_0>t] = \mathbb{P}_x[\tau_a>t]\,; \quad t>0,\ a>0.$$

## 4.2 Super- [resp.: Sub-] Solutions

In this section we shall give the key technical result needed for the proof of the Main Theorem. We shall consider $P(t,x)$ and $w_\varepsilon = w(t,x)$ as in (0.3) and (0.25) for some fixed $0<\varepsilon<1$, where $w$ is modified to be smooth as in (3.66), and define

(4.11) $$F_\theta^{\pm} = F^{\pm}(t,x) = P(t,x)\pm\theta w(t,x)\,; \quad t>0,\ x\in D,$$

we shall also extend the definition of $F^{\pm}$ in $(t>0,\ x\in\mathbb{R}^d)$ by making it $\equiv 0$ if $x\notin D$. In (4.11) $\theta>0$ is a positive parameter, and $F^{\pm}$ is a bounded function in $D(b,1)$ $(b>0)$.

In the proposition that follows $D$, $d$ and $A$ are as before and $0<\varepsilon<1$ is fixed but arbitrary and to make the proof as clear as possible I shall first consider the case where the measure $\mu$ in (0.6) has compact support

(4.12) $$\operatorname{supp}\mu\subset B_R(0),$$

for some $R>0$.

***Proposition*** *There exists $C_0 = C_0(d,A,\varepsilon,R)>0$ such that the function $F_\theta^+$ (resp.: $F_\theta^-$) is a super- (resp.: sub-) solution of $L_\mu$ in $D(C_0,C_0)$ for any $\theta\geq C_0$.*

To prove the proposition we apply the operator $L = L_\Delta = 1/2\Delta - \partial_t$ of Section 2.4 to $F^\pm$ and we obtain

$$LF^\pm_\theta(t,x) = \mp\theta \frac{P(t,x)}{\delta(x)^{2+\varepsilon}}\ ; \quad t > 0,\ x \in D. \tag{4.13}$$

For the proof of the proposition it suffices to estimate the error term that is obtained by Taylor's theorem:

$$\begin{aligned} |(L_\Delta - L_\mu)F^\pm_\theta(t,x)| &\le C\theta \sup_{|y-x|\le R} |\nabla^3_y w(t,y)| + C\theta \sup_{t<s<t+1} \left|\frac{\partial^2}{\partial s^2} w(s,x)\right| \\ &\quad + C \sup_{|y-x|\le R} |\nabla^3_y P(t,y)| + C \sup_{t<s<t+1} \left|\frac{\partial^2}{\partial s^2} P(s,x)\right|, \end{aligned} \tag{4.14}$$

where $C$ is independent of $\theta$. The first two terms in the right hand side of (4.14) that involve $w$ are absorbed by (4.13), uniformly in $\theta$, as long as $(t,x) \in D(C,C)$ with $C > 0$ large enough (depending on $R$, *cf.* Sections 3.4 and 2.5). As for the third and fourth term they are also absorbed by (4.13) if $(t,x) \in D(C,C)$ and $\theta > C$ with $C$ large enough. This proves the proposition in the special case (4.12). The above proposition will suffice for the proof of (0.11) when the measure satisfies (4.12). The reader could therefore skip the rest of this section and go directly to Section 4.3 where I give the proof of the theorem.

Let us also point out that the above proof gives us in fact a more precise proposition:

***Proposition (precise version)*** *There exists $C_0 = C_0(d, A, \varepsilon, R)$ such that the functions $F^\pm_\theta$ are super- sub- solutions of $L_\mu$ in $D(b,a)$, for $\theta \ge \theta_0$, provided that:*

$$a^{1-\varepsilon}, \quad a^{1-\varepsilon}(\mathrm{Min}[a, \sqrt{b}])^\varepsilon, \quad \theta_0 a^{1-\varepsilon} \ge C_0.$$

Let us now consider the case of a general measure $\mu$ as in (0.6), $(M_B)$, $B \ge B_0$, where $B_0 = B_0(d, A, \varepsilon) > 0$ is some appropriate function of its arguments.

The first observation is that for any $B_1 > 0$ and any $\mu$ that satisfies $(M_B)$ with $B$ large enough, and for any $R > 0$, we can split $\mu$ as follows:

$$\mu = \mu_R + \nu_R, \quad \mu \in \mathbb{P}(\mathbb{R}^d), \quad \operatorname{supp} \mu_R \subset B_R(0), \tag{4.15}$$

$$\mu \text{ satisfies } (0.6), (M_{B_1}), \tag{4.16}$$

$$\int (1 + |x|)^{B_1}\, d|\nu_R| = O(R^{-B_1}), \tag{4.17}$$

where $\nu_R$ is not assumed to be a positive measure.

To achieve this we first consider

$$\mu_1 = \mu\chi[|x| \le R] + \alpha\delta_\xi,$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

where $\xi \in \mathbb{R}^d$, $|\xi| \leq 1$ and $0 \leq \alpha = O(R^{-B_2})$, so that $\int x\, d\mu_1 = 0$. It is clear then that $\|\mu_1\| = 1 + O(R^{-B_3})$. We can find then $e \in \mathbb{P}(\mathbb{R}^d)$ which is symmetric with support in $[x \leq 2]$ such that

$$\lambda_R = \mu_1 + \beta e\,; \quad 0 < \beta = O(R^{-B_4}),$$

is such that:

$$\int x_i x_j\, d\lambda_R = \left(1 + O(R^{-B_5})\right) \delta_{ij}.$$

The final correction is done with the help of $e_\rho = \bigotimes 1/2(\delta_{-\rho e_i} + \delta_{\rho e_i})$, where $e_1, \dots, e_d \in \mathbb{R}^d$ orthonormal unit vectors and $\rho > 0$ so that $\int x_i x_j\, de_\rho = \rho^2 \delta_{ij}$. We set then

$$\tau_R = \lambda_R + \gamma\rho^{-2} e_\rho\,; \quad \gamma > 0,\ \rho \sim 1,$$
$$\|\tau_R\| = 1 + O(R^{-B_7}) + \gamma\rho^{-2},$$
$$\int x_i x_j\, d\tau_R = \delta_{ij}\left(1 + O(R^{-B_5}) + \gamma\right),$$

and we simply have to choose $\gamma$ and $\rho$ so that

$$\gamma = O(R^{-B_6})\,; \quad \int x_i x_j\, d\tau_R = \delta_{ij} \|\tau_R\|.$$

This is clearly possible (by first choosing $\rho = 2$ or $1/2$ to have the correct sign). A renormalization finishes the construction of $\mu_R = \left(1 + O(R^{-B_8})\right) \tau_R$. The details are elementary and they will be left as an exercise to the reader.

The next observation is the following.

***Lemma*** *There exists $C > 0$ such that*

$$\left|\frac{P(t,x_1)}{P(t,x_2)}\right| \leq (C + |x_1 - x_2|)^C, \tag{4.18}$$

*for $t > 0$, $x_i \in D$, $\delta(x_i) \geq 1$, $i = 1, 2$.*

This is nothing but the standard chain condition (*i.e.*, the "Hyperbolicity" of the natural distance in $D$ *cf.* [16, Section 3 (ii)]). This condition adapts to the parabolic function $P(t,x)$ simply because of the backwards Harnack estimates (0.21).

To state the proposition in the general case when $\mu$ does not have compact support we must also modify:

$$F^{\pm}_{\text{new}} = F^{\pm}_{\theta} \text{ if } x \in D,\ \delta(x) \geq 1\,; \quad F^{\pm}_{\text{new}} = 0 \text{ if } x \notin D,\ \text{or if } x \in D,\ \delta(x) < 1.$$

The reason for this modification is that we need to have a global function that is bounded for $t > b > 0$ (if $D$ is convex, but not necessarily otherwise, $F^{\pm}_{\theta}$ is bounded in $D(b,0)$, $b > 0$).

***Proposition*** *There exists a constant $B_0 = B_0(d, A, \varepsilon)$ such that $F^{+}_{\text{new}}$ (resp.: $F^{-}_{\text{new}}$) is a super- (resp.: sub-) solution of $L_\mu$ for any $\mu$ that satisfies (0.6) and $(M_B)$ for $B \geq B_0$. This holds in the domain $D(C_0, C_0)$ and for $\theta \geq C_0$ where*

$$C_0 = C_0(d, A, \varepsilon, B, M_B).$$

To prove the proposition in this general case we have to go back to error term (4.14). We have to account now for the two additional error terms

$$J_1 = \theta w(t,\cdot) * |\nu_R|(x)\ ; \qquad J_2 = P(t,\cdot) * |\nu_R|(x). \tag{4.19}$$

(4.17) and (4.18) put together give at once

$$J_2 \leq CP(t,x)R^{-B_8}, \tag{4.20}$$

with $B_8 > 0$ arbitrary large, as long as $B_1$ is large enough. It follows that this is absorbed by (4.13) as long as

$$R \sim \delta(x). \tag{4.21}$$

This is the choice of $R$ that will be used in (4.15), (4.16), (4.17) and it depends on $x$. This is still compatible with the estimates of the first and third term in the right hand side of (4.14) provided that $B_1$ is large enough (depending on the $C$ of (4.18)).

The estimate of $J_1$ follows identical lines. But we also have to use here the fact that because of the modification $F^{\pm}_{\text{new}}$ in the definition of $J_1$ in (4.19) we are actually using a modified function $w(t,\cdot)$ which is bounded in $\delta(x) \leq 1$. The details will be left to the reader. An alternative approach for the non-compact support case is outlined in Section 6.6.

### 4.3 The Proof of the Main Theorem

We shall use here the Proposition of Section 4.2 and the positivity

$$F^{+} \geq P \geq 0. \tag{4.22}$$

From this and from (4.5) and (4.8), with $\tau = \tau_a$ and $\dot{\tau}$ as in (4.8), we deduce that

$$\begin{aligned} F^{+}(t_0,x_0) &\geq \mathbb{E}_{t_0,x_0}\Big(F^{+}\big(\dot{z}(\dot{\tau})\big)\Big) \\ &\geq \mathbb{E}_{t_0,x_0}\Big(F^{+}\big(b, z(t_0-b)\big)\,\mathbb{I}(\tau > t_0 - b)\Big) \\ &\geq \mathbb{E}_{t_0,x_0}\Big(P\big(b, z(t_0-b)\big)\,\mathbb{I}(\tau > t_0 - b)\Big) \\ &\geq \inf_{\xi\in D_a} P(b,\xi)\mathbb{P}_{x_0}(\tau_a > t_0 - b), \end{aligned} \tag{4.23}$$

provided that:

$$t_0 \geq b \geq C,\ t_0 - b \in \mathbb{Z}\ ; \quad a \geq C. \tag{4.24}$$

The corresponding lower estimate is slightly more subtle, because it is necessary to choose $\theta, a, b \geq C$ such that $F^{-}_{\theta}$ is sub-solution in $D(b,a)$, and such that the following additional condition is verified:

$$F^{-}(t,x) \leq 0\ ; \quad x \notin D_a,\ t \geq b. \tag{4.25}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

To achieve (4.25) we first choose $a \geq C_0 \geq 10$, so that $F^-$ is a subsolution in $D(C_0 - 1, C_0 - 1)$ as in the proposition. Then we choose $b \geq a^2$ and $b \geq C_0$ so that the lower estimate Section 3.5 scaled for $t = b$ holds. Then we finally choose $\theta \geq C_0$ large enough so that (4.25) holds. (The freedom to let $\theta \to \infty$ is exploited in an essential way at this point).

On the event $[\dot{\tau} < t_0 - b]$ we have

$$\dot{z}(\dot{\tau}) = \big(t_0 - \dot{\tau}, z(\dot{\tau})\big) \in (b, +\infty) \times \sim D_a \tag{4.26}$$

and therefore, (4.25) implies that

$$F^-\big(\dot{z}(\dot{\tau})\big)\mathbb{I}[\dot{\tau} < t_0 - b] \leq 0. \tag{4.27}$$

This together with (4.6) and the fact that

$$F^- \leq P \leq 1, \quad \dot{\tau} \leq \tau, \tag{4.28}$$

implies that

$$\begin{aligned} F^-(t_0, x_0) &\leq \mathbb{E}\Big(F^-\big(\dot{z}(\dot{\tau})\big)\Big) \\ &\leq \mathbb{E}\big[F^-\big(\dot{z}(\dot{\tau})\big)\mathbb{I}(\dot{\tau} \geq t_0 - b)\big] \\ &\leq \mathbb{P}[\dot{\tau} \geq t_0 - b] \leq \mathbb{P}_{x_0}[\tau \geq t_0 - b]\ ; \quad t_0 > b,\ x_0 \in D_a. \end{aligned} \tag{4.29}$$

By a simple change of notation (*cf.* (4.9), (4.10)) and $a$, $b$ as in (4.25), (4.23) and (4.29) give:

$$\begin{gathered} F^-(t + b, x_{-a}) \leq \mathbb{P}_x[\tau_0 \geq t]\ ; \\ F^+(t + b, x_{-a}) \geq \inf_{\xi \in D_a} P(b, \xi)\mathbb{P}_x(\tau_0 > t)\ ; \quad t = 0, 1, \dots,\ x \in D_a. \end{gathered} \tag{4.30}$$

(Observe that in both (4.23) and (4.29) the "extreme" inequality holds for any $x_0 \in \mathbb{R}^d$ for obvious reasons.)

The probability of life $P(t, x)$ on the other hand satisfies:

$$\frac{P(t,x)}{P(t+b,x)} \leq \left(1 + \frac{b}{t}\right)^C \leq 1 + \frac{Cb}{t}\ ; \quad x \in D,\ t > b, \tag{4.31}$$

$$\left|\frac{P(t,x)}{P(t,y)} - 1\right| \leq C\frac{|x-y|}{\delta(x)}\ ; \quad \delta(x) \geq 1,\ |x-y| \leq 1/2\delta(x), \tag{4.32}$$

$$P(t,x) \geq 1 - C\frac{t^2}{\delta(x)^4}\ ; \quad x \in D,\ t > 0. \tag{4.33}$$

(4.31) and (4.32) are immediate consequences of (0.14) and (2.3). To prove the estimate (4.33) it suffices to scale and to show that:

$$P(h,x) \geq 1 - Ch^2\ ; \quad \delta(x) = 1,\ 0 < h.$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

This is an immediate consequence of the differentiability of $P(t,x)$ in $t$ up to the smooth part of the boundary, $t = 0$, where $\frac{\partial P}{\partial t} = \Delta P = 0$ (*cf.* [27]).

If we observe that $P(t, x_{-a})$, $a > 0$, for fixed $t$ and $x$, is an increasingly function of $a$ (*cf.* (4.10)), we see that the first part of (4.30), together with (4.31), for an appropriate choice of $a$, $b$, (with $c < a, b < C$) gives:

$$P_\mu(t,x) \geq P(t,x)\left[1 - \frac{C}{\delta^\varepsilon(t,x)}\right]; \quad t, \delta(x) \geq C, 0 < \varepsilon < 1. \tag{4.34}$$

This is the lower estimate of the Main Theorem. Similarly for an appropriate, but fixed, choice of $a$ and $b$, the estimates (4.31), (4.32), (4.33) and the second part of (4.30) will show that for any $\eta > 0$ we have

$$P_\mu(t,x) \leq (1+\eta)P(t,x); \quad t, \delta(x) \geq C, \tag{4.35}$$

where $C$ also depends on $\eta$. This is a weak form of the upper estimate of the Main Theorem.

To prove the upper estimate of the Main Theorem in its full thrust we must optimize over $a$ and $b$. To illustrate the procedure we shall first choose $b$ appropriate but fixed, and then set $a \sim \delta(x)^{1/5}$. Combining then (4.31), (4.32), (4.33), and the second part of (4.30), we obtain

$$P_\mu(t,x) \leq P(t,x)\left[1 + \frac{C}{\delta(t,x)^\varepsilon} + \text{Error}\right],$$

where:

$$\text{Error} \leq \frac{Cb^2}{a^4} + \frac{Ca}{\delta(x)} - \frac{Cb}{t} \leq C\delta^{-4/5}(x). \tag{4.36}$$

(4.34) and (4.36) complete the proof of (0.11). We shall have to leave matters at that and the proof of (0.11) for $0 < \varepsilon < 1$ will have to wait for the second installment of this paper. Observe, however, that we can optimize over both *a and b* and use the precise version of the Proposition for which we have to set

$$1 \ll a \ll \delta(x); \quad 0 < b < 1; \quad a^{1-\varepsilon} b^{\frac{\varepsilon}{2}} \gg 1.$$

This will improve slightly the $\varepsilon$ to some $4/5 < \varepsilon_0 < 1$, but in this argument the support of $\mu$ has to be compact because of the precise version of the Proposition of Section 4.2.

# 5 Applications to Homogenization Theory

## 5.1 The Main Theorem

Let us consider an elliptic differential operator

$$Q = \frac{\partial}{\partial x_i} a_{ij}(x) \frac{\partial}{\partial x_j} \quad \text{(summation convention)},\ x \in \mathbb{R}^d, \tag{5.1}$$

$$\lambda^{-1}|\xi|^2 \leq a_{ij}\xi_i\xi_j \leq \lambda|\xi|^2; \quad \xi \in \mathbb{R}^d, \tag{5.2}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

with periodic coefficients $a_{ij} \in C^\alpha(\mathbb{R}^d)$, for some $\alpha > 0$ and $\lambda > 0$, *i.e.*, Hölder coefficients that satisfy

$$a_{ij}(x + e_k) = a_{ij}(x) \; ; \quad x \in \mathbb{R}^d, \; i, j, k = 1, \dots, d, \tag{5.3}$$

where $e_j = (0, 0, \dots, 0, 1, 0, \dots, 0)$ is the standard orthonormal basis of $\mathbb{R}^d$ (by a change of coordinates any periodic coefficients can be reduced to (5.3)).

I shall also consider $\chi_i \in C^{1+\alpha}(\mathbb{R}^d)$ $(i = 1, \dots, d)$ the correctors of $Q$ (*cf.* [10], [11]) and the corresponding homogenized operator

$$Q_0 = q_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \; ; \quad q_{ij} = \int_I [a_{ij}(x) - a_{ik} \partial_k \chi_j(x)] \, dx, \tag{5.4}$$

where $I$ is the period parallelopiped spanned by $e_1, \dots, e_d$.

We shall now consider $D \subset \mathbb{R}^d$ some Lipschitz domain as in (0.10), with Lipschitz constant $A > 0$, and the probability of life:

$$\begin{aligned} P_Q(t, x) &= \mathbb{P}_x[z(s) \in D, 0 < s < t], \\ P_0(t, x) &= \mathbb{P}_x[z_0(s) \in D, 0 < s < t], \end{aligned} \tag{5.5}$$

for the diffusion $(z(s), s > 0)$ (resp., $(z_0(s), s > 0)$) generated by $Q$ (resp. $Q_0$). The diffusion $z_0(s)$ is just Brownian motion after a change of coordinates. Observe that if $\alpha < 1$, $z(s)$ may not have continuous paths. To deal with this difficulty in what follows we shall make the qualitative assumption that $a_{ij} \in d^\infty$ but the constants will only depend on the $C^\alpha$ norms of these coefficients. We then have:

***Main Theorem: Homogenization*** *For all* $0 < \varepsilon < 1$ *there exists* $C = C(d, A, \varepsilon, \lambda)$ *such that:*

$$|P_Q(t, x) - P_0(t, x)| \le C \frac{P_0(t, x)}{(\mathrm{Min}[\delta(x), \sqrt{t}])^\varepsilon} \; ; \quad t, \delta(x) \ge C. \tag{5.6}$$

Just as for random walks, we can take $\varepsilon = 1$ when $D$ is convex, but I shall not prove this here.

### 5.2 The Second Correctors and the Super- (Sub-) Solution

Let us denote

$$L = Q - \frac{\partial}{\partial t} \; ; \quad L_0 = Q_0 - \frac{\partial}{\partial t}, \tag{5.7}$$

and let $u(t, x) \in C^\infty$, $(t, x) \in \mathbb{R}^{d+1}$, we have then

$$L\left(u - \chi_i \frac{\partial u}{\partial x_i} - \chi_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j}\right) = L_0 u + O[\nabla^3 u + \nabla^4 u + \partial_t \nabla u + \partial_t \nabla^2 u], \tag{5.8}$$

where:

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

(i) $\chi_i$ are the correctors and $\chi_{ij} \in C^{1+\alpha}(\mathbb{R}^d)$ are the second correctors (*cf.* [33], [34], [9]). These are periodic functions with periods (5.3).

(ii) $\nabla^k$ indicates $\frac{\partial^k}{\partial x_{i_1} \cdots \partial x_{i_k}}$ the $k$-th spacial gradient and $\partial_t$ indicates $\frac{\partial}{\partial t}$.

(iii) The "$O$" notation in (5.8) indicates a linear combination of the functions in $[\cdots]$ with coefficients that are bounded and periodic (provided that $\alpha \geq 1$).

The above formula is basic for us here. In [9] I have given a review of the aspect of homogenization theory that leads to the formula (5.8). The reader who is not an expert in the subject may find [9] useful.

Just as in Section 4, we shall now consider the function

$$F^{\pm}(t,x) = P_0(t,x) \pm \theta w_\varepsilon(t,x), \tag{5.9}$$

where:

$$w_\varepsilon(t,x_0) = \int_0^t \int_D p^0_{t-s}(x,x_0) \frac{P_0(s,x)}{\delta_0^{2+\varepsilon}(x)}\, ds\, dx, \tag{5.10}$$

is a modification of (0.25), where $p^0(\cdot,\cdot)$ is the heat diffusion kernel of $(z_0(s)\ ;\ s>0)$ and, more to the point, where $\delta_0 \in C^\infty(D)$ is the "smoothing" of $\delta(x)$ that was used in Section 3.4.

Just as in Section 4 we have:

$$L_0 F^{\pm} = \mp\theta \frac{P_0(t,x)}{\delta_0^{2+\varepsilon}(x)}\ ; \quad t>0,\ x \in D. \tag{5.11}$$

We shall now apply the formula (5.8) and set

$$u = F^{\pm}\Phi^{\pm} = F^{\pm} - \chi_i \frac{\partial F^{\pm}}{\partial x_i} - \chi_{ij} \frac{\partial^2 F^{\pm}}{\partial x_i \partial x_j}, \tag{5.12}$$

so that:

$$L\Phi^{\pm} = L_0 F^{\pm} + O[\nabla^3 F^{\pm} + \cdots]. \tag{5.13}$$

We shall use the notation of Section 4 and we have:

***Proposition*** *There exists $C_0 = C_0(d, A, \lambda, \varepsilon)$ such that*

$$L\Phi^{+}(t,x) \leq 0 \quad (\text{resp.: } L\Phi^{-}(t,x) \geq 0)\ ; \quad \theta \geq C_0,\ t \geq C_0,\ \delta(x) \geq C_0, \tag{5.14}$$

*i.e., $\Phi^+$ (resp.: $\Phi^-$) is a super- (resp.: sub-) solution of $L$.*

The proof is identical to the one given in Section 4 and if anything simpler. It consists in verifying that in the required range the $O[\cdots]$ of (5.13) is absorbed by (5.11).

Concerning the "error term":

$$\Phi^{\pm}(t,x) - P_0(t,x) = \pm\theta w(t,x) + O[\nabla P_0 + \nabla^2 P_0 + \nabla w + \nabla^2 w], \tag{5.15}$$

with the same meaning of $O[\cdots]$ as before, the estimates of Sections 2 and 3 allow us to assert that:

$$(5.16) \qquad |\Phi^{\pm}(t,x) - P_0(t,x)| \le C\frac{P_0(t,x)}{\delta^{\varepsilon}(t,x)}\,; \quad t > C,\ \delta(x) > C,\ x \in D.$$

Similarly, from the analogue of the lower estimate Section 3.5 we have the one-sided estimate: For all $c > 0$ there exists $C > 0$ such that

$$(5.17) \qquad \Phi^{-}(t,x) - P_0(t,x) \le -C\theta\frac{P_0(t,x)}{\delta(x)^{\varepsilon}} + O[\cdots]\,; \quad \delta(x) \le c\sqrt{t},$$

where $O[\cdots]$ is as (5.15).

**Proof of (5.6)** We consider again the time space process

$$\big(t_0 - t, z(t)\big) \in \dot{\mathbb{R}}^d = \mathbb{R} \times \mathbb{R}^d\,; \quad t > 0,$$

which is now a diffusion, and we follow *verbatim* the analysis of Section 4 which now if anything, is simpler. It is only the analogues of (4.22), (4.28) and (4.25) that have to be verified again. For the analogues of (4.22) and (4.28) we use (5.16). This guarantees that

$$(5.18) \qquad 0 \le \Phi^{+}\,; \quad \Phi^{-} \le P_0\left(1 + \frac{C}{\delta(x)}\right) \text{ in } D(C,C),$$

for $C > 0$ large enough (with the notation of Section 4). The only difference then is an extra factor $1 + \frac{C}{a}$ that will appear in the analogue of the right hand side of (4.29). The analogue of (4.34) then follows by setting $a \sim \delta(x)$. For the analogue of (4.25) we use (5.17) and we have to control the $O[\cdots]$ of (5.17) at the exit point. But this here is even easier than in Section 4 because of the continuity of the path. This guarantees that at that exit point we have $\delta(x) = C$ (and we have *not* jumped over). The details will be left to the reader.

The rest of the proof finishes as in Section 4.

# 6 Applications to Translation Invariant Markov Chains

## 6.1 The General Setup

We shall consider the abelian group $\Gamma = \mathbb{R}^{d_1} \times \mathbb{Z}^{d_2}$ ($d = d_1 + d_2$) that acts on the space $X = B \times \Gamma$, where $B$ is some Borel space, and the action is

$$(6.1) \qquad X \ni x = (a,h) \to \gamma x \underset{\text{def}}{=} (a, h+\gamma) \underset{\text{def}}{=} x + \gamma \in X\,; \quad a \in B\,; \quad h,\gamma \in \Gamma.$$

We shall consider also $K(x,A) \ge 0$ ($x \in X$, $A \subset X$), submarkovian kernels and the corresponding submarkovian operators

$$(6.2) \qquad Kf(x) = \int K(x,dy)f(y),$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

that are $\Gamma$-invariant, *i.e.*,

$$K(x+\gamma, P+\gamma) = K(x, P)\ ; \quad x \in X,\ P \subset X. \tag{6.3}$$

The induced kernel on $B$ can then be defined:

$$K_B(a, A) = K(\tilde{a}, A \times \Gamma)\ ; \quad a \in B,\ A \subset B, \tag{6.4}$$

where $\tilde{a} \in X$ is arbitrary under the condition $\pi(\tilde{a}) = a$ and where $\pi\colon X \to B$ is the canonical projection.

A measure $\mu$ on $X$ is $\Gamma$-invariant if and only if $\mu = \mu_B \otimes d\gamma$ where $\mu_B$ is some measure on $B$ and $d\gamma$ is the Haar measure of $\Gamma$.

If $\mu_B$, as above, is $K_B$-invariant (*i.e.*, $K_B^* \mu_B = \mu_B$ for the adjoint operator) then $\mu$ is $K$-invariant. Observe (no use of this remark will be made) that if $dx \geq 0$ is a $K$-invariant measure so is the image of $dx$ by $\gamma$, $\overset{\vee}{\gamma}(dx)$. It follows that if we assume that $\sigma_\gamma = \frac{\overset{\vee}{\gamma}(dx)}{dx}$ satisfies $\sigma_\gamma, \sigma_\gamma^{-1} \in L^\infty(dx)$, uniformly in $\gamma$, then by the amenability of $\Gamma$ we can find some limit:

$$\sigma dx = \lim_F \frac{\sum_F \overset{\vee}{\gamma}(dx)}{|F|}\ ; \quad F \subset \Gamma \text{ finite}$$

that is both $\Gamma$- and $K$-invariant ($\sigma \in L^\infty$).

The above general setup occurs naturally in Lie group theory (*cf.* [13], [14]) but also in the analysis of Markov chains in a random environment (*cf.* [12]).

### 6.2 The Correctors

In the setup of the previous section, I shall consider $x_i\colon X \to \mathbb{R}$ ($i = 1, \ldots, d$):

$$x_i(x) = h_i \in \mathbb{R}\ ; \quad x = (a, h),\ h = (h_1, \ldots, h_d) \in \Gamma \subset \mathbb{R}^d. \tag{6.5}$$

It is, of course, clear that $x_i(x+\gamma) = x_i(x) + \gamma_i$, $\gamma = (\gamma_1 \cdots \gamma_d) \in \Gamma$, and therefore, if we assume that $K1 = 1$ in (6.2) (*i.e.*, $K$ is strictly Markovian), the new functions

$$(I - K)x_i = \theta_i\ ; \quad i = 1, \ldots, d, \tag{6.6}$$

are $\Gamma$-invariant and can thus be identified to functions on $B$. (We assume here that the integral that defines (6.6) is absolutely convergent.)

We shall say that the $\Gamma$-invariant functions $(\chi_1, \ldots, \chi_d) = \chi$ on $X$ (these can be identified to functions on $B$) are correctors if

$$(I - K)\chi_i = \theta_i\ ; \quad i = 1, \ldots, d. \tag{6.7}$$

The expression $(I - K)\chi_i = (I - K_B)\chi_i$ is not, *a priori,* meaningful, unless we make additional assumptions on the $\chi_i$'s, *e.g.* that they are bounded.

A natural context for which $(I - K)\chi_i$ can be defined is the case where

$$Kf(x) = \int_X K(x, y) f(y)\, dy\,; \quad f \geq 0, \tag{6.8}$$

is given by some doubly submarkovian kernel with respect to some $\Gamma$-invariant measure $dx = da \otimes d\gamma$, so that:

$$\int K(x, y)\, dy, \int K(y, x)\, dy \leq 1\,; \quad x \in X. \tag{6.9}$$

In that case $K_B$ contracts all the corresponding $L^p(B)$ norms ($1 \leq p \leq +\infty$) and if $\theta_i \in L^p$ then the correctors could be in $L^p$.

The above definition is equivalent to the fact that

$$y_i = x_i - \chi_i, \quad i = 1, \dots, d, \tag{6.10}$$

are harmonic functions on $X$ for the kernel (6.2), *i.e.*, that

$$\big( y_i(z_n) = x_i(z_n) - \chi_i(z_n)\,;\; n \geq 1 \big), \quad i = 1, \dots, d \tag{6.11}$$

is a martingale, for any starting probability, of the Markov process

$$z_n \in X\,; \quad n \geq 0, \tag{6.12}$$

that is generated by the kernel (6.2). This simply says that we can perform the change of variables on $X$

$$X \ni (a, h) \leftrightarrow \big(a, h - \chi(a)\big) \in X, \tag{6.13}$$

(alternatively, this amounts to choosing a new section $B' \subset X$ of the projection $\pi$) and then the new coordinate functions (6.5) are harmonic. In this context we sometimes say that we "chose harmonic coordinates".

## 6.3 The Fredholm Problem and the Existence of the Correctors

I shall place myself in the situation where a doubly submarkovian kernel exists with respect to $dx = da \otimes d\gamma$ as in (6.9). I shall also assume that $K$ in (6.2) gives rise to some compact operator $K_B$ in $L^2(B)$ so that $I - K_B$ is a Fredholm operator [30], [31]. The existence of the $\chi_i \in L^2(B)$ is then a consequence of the Fredholm alternative and $\chi_i$ exists if and only if

$$\theta_i \perp \operatorname{Ker}(I - K_B^*)\,; \quad i = 1, 2, \dots, d, \tag{6.14}$$

for the adjoint operator $K_B^*$.

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

To avoid uninteresting complications let us assume that $da$ is a probability measure on $B$, let us also assume that $K$ is doubly markovian and ergodic ($K_B^* 1 = K_B 1 = 1$; $K_B f = f$ or $K_B^* f = f \Rightarrow f = 1$). Then the condition (6.14) simply says that

$$\int_B \theta_i \, da = 0\,; \quad i = 1, \dots, d. \tag{6.15}$$

It will be convenient to write the kernel in terms of the coordinate functions as follows:

$$K(x, y) = K(\gamma - \lambda\,; a, b)\,; \quad x = (a, \gamma),\ y = (b, \lambda) \in X. \tag{6.16}$$

An elementary computation shows then that the conditions (6.15) are equivalent to

$$\int_\Gamma \int_B \int_B K(\gamma\,; a, b)\gamma \, d\gamma \, da \, db = 0. \tag{6.17}$$

The condition (6.17) is certainly verified if $K$ is symmetric (*i.e.*, $K(x, y) = K(y, x)$) with respect to $dx$ for then $K(-\gamma\,; a, b) = K(\gamma\,; b, a)$. At any rate, in the above situation if (6.17) holds, then the correctors $\chi$ exist in $L^2(B)$ and are uniquely determined up to an additive constant (by the ergodicity of $K$).

There are two examples that deserve special attention. First we could assume that $B$ is a finite set. In that case $L^2(B) = L^\infty(B)$ and the correctors are bounded. Assume next that $B$ is some compact $C^\infty$ manifold, that $\Gamma = \mathbb{R}^d$, and that the Markov chain (6.12) can be identified with $z_n = z(n)$ $(n = 0, 1, \dots)$, where

$$z(t) \in X\,; \quad t > 0, \tag{6.18}$$

is the diffusion generated by $\mathcal{A}$ some $\Gamma$-invariant second order subelliptic (say elliptic if you are not quite sure of the terminology!) differential operator. The function $y(\cdot) = x(\cdot) - \chi(\cdot)$ $(x(\cdot) = (x_1(\cdot), \dots, x_d(\cdot))$ is then $\mathcal{A}$-harmonic. If $\mathcal{A}x = \theta$ then $\theta \in C^\infty(B)$ and we can find $\chi \in L^2(B)$ such that $\mathcal{A}\chi = \theta$. The subellipticity then guarantees that $\chi \in C^\infty(B)$ as long as the Fredholm alternative (6.14) is verified. This is certainly the case if $\mathcal{A}$ is formally self adjoint with respect to some $C^\infty$ non-vanishing $\Gamma$-invariant measure $dx$ on $X$. The symmetry is by no means essential here but I shall not elaborate.

Another typical situation that deserves special attention is the case of a $C^\infty$ manifold $\tilde{M}$, and a covering map $\pi\colon \tilde{M} \to M$, where we assume that $M$ is compact, and where we assume that the deck transformation group $\Gamma$ of $\pi$ is $\Gamma \cong \mathbb{Z}^d$. We can then consider $dm$ some smooth non-vanishing measure on $M$, and $\mathcal{A}$, some second order sub elliptic operator on $M$, and assume that $\mathcal{A}1 = \mathcal{A}^*1 = 0$, where $\mathcal{A}^*$ is the formal adjoint of $\mathcal{A}$. $\mathcal{A}$ and $dm$ induce then a differential operator and a measure on $\tilde{M}$ that are invariant by the action of the group $\Gamma$. We obtain thus a $\Gamma$-invariant diffusion on $\tilde{M}$ where the formalism of Section 6.1 applies. The correctors exist by the Fredholm alternative ($\mathcal{A}^* f = 0$ if and only if $f =$ constant, by the maximum principle) and they are $C^\infty$ by subellipticity.

## 6.4 The Limit Operator and the Main Theorem

We shall place ourselves here in the setup of a Markovian kernel (6.2) that admits first correctors $\chi \equiv 0$. This, as we pointed out in the previous section, can be achieved by the change of coordinates (6.13), as long as correctors exist at all. What is assumed, in other words, is that the coordinate functions (6.5) are already harmonic. We shall also assume that $K$ admits a bi-Markovian kernel $K(x, y)$ with respect to $dx = da \otimes d\gamma$ (both integrals in (6.9) are $= 1$ and $da \in \mathbb{P}(B)$). In other words, the measure $dx$ (resp.: $da$) is $K$- (resp.: $K_B$-) invariant.

We shall denote by $\mathcal{F}_n$ the field generated by $(z_0, \dots, z_n \in X)$ (*cf.* (6.12)) and denote by:

$$X_n = x(z_n) \in \mathbb{R}^d \,; \quad X_n = X_0 + d_1 + \cdots + d_n, \tag{6.19}$$

the vector-valued martingale of the coordinate functions (6.5). We have then:

$$\mathbb{E}(\langle d_n, \xi\rangle^2 // \mathcal{F}_{n-1}) = [(K - I)\langle x, \xi\rangle^2](z_{n-1}) = \Phi_{ij}(z_{n-1})\xi_i\xi_j \,; \quad \xi \in \mathbb{R}^d, \tag{6.20}$$

and if $z_n = \big(a_n, x(z_n)\big) \in X$ $(n \geq 0)$ one easily verifies that

$$\Phi_{ij}(z_{n-1}) = \Phi_{ij}(a_{n-1}) \,; \quad i, j = 1, \dots, d. \tag{6.21}$$

It follows, in particular, that if we take as starting probability for the process (6.12) any measure $\tilde{da} \in \mathbb{P}(X)$ that has the property that $\overset{\vee}{\pi}\tilde{da} = da$ then

$$\mathbb{E}_{\tilde{da}}\langle d_n, \xi\rangle^2 = q_{ij}\xi_i\xi_j \,; \quad q_{ij} = \int_B \Phi_{ij}(a)da \,; \quad i, j = 1, \dots, d. \tag{6.22}$$

This is the symbol of a differential operator on $\mathbb{R}^d$:

$$Q_0 = q_{ij}\frac{\partial^2}{\partial x_i \partial x_j}, \tag{6.23}$$

which is called the limit operator.

Under appropriate "ellipticity" conditions on $K$ (*cf.* [12]: (Ell)) the operator (6.23) is non-singular, *i.e.*, the matrix $(q_{ij}) \gg 0$ is positive definite. It is clear from the definition that for this it suffices to demand for instance, that (*cf.* (6.16)):

$$K(\gamma \,; a, b) \geq \varepsilon_0 > 0 \,; \quad a, b \in B, \; |\gamma| = |\gamma_1| + \cdots + |\gamma_d| \leq 10, \tag{Ell}$$

for some $\varepsilon_0 > 0$. If that is the case, for "all practical purposes", the Brownian motion generated by (6.23) "approximates" the martingale (6.19) (*cf.* [12]). I do not intend to go into this here; but I should point out that for such an approximation we also need to impose decay conditions at infinity on the kernel. The kernels that are considered in [12] satisfy the finite span condition

$$K(\gamma \,; a, b) = 0 \,; \quad a, b \in B, \; \gamma \in \Gamma, \; |\gamma| \geq R, \tag{FS}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

for some $R > 0$. Weaker moment conditions would, in fact, suffice

$$(KM_E) \qquad \int_\Gamma K(\gamma\,;a,b)(1+|\gamma|)^E\,d\gamma \le M_E\,; \quad a,b\in B,$$

or even average conditions

$$(KM_E^*) \qquad \int_B\int_B\int_\Gamma K(\gamma\,;a,b)(1+|\gamma|)^E\,d\gamma\,da\,db \le M_E,$$

for $E > 0$ large enough.

To state our Main Theorem in the present setting we shall need to impose also another condition. I shall assume first of all that there exist $\psi_{ij}(a)$ functions on $B$ such that

$$(K_B - I)\psi_{ij}(a) = \Phi_{ij}(a) - q_{ij}. \tag{6.24}$$

In the case when $K_B$ is a compact operator on $L^2(B)$, by the Fredholm alternative, we see that such functions $\psi_{ij} \in L^2(B)$ exist under natural conditions. To conform with the terminology of homogenization theory we shall introduce the following terminology and say (quite generally, *i.e.*, not just in the case where $\chi \equiv 0$) that $K$ admits first correctors $\chi_i$ and second correctors $\chi_{ij}$, which are functions on $B$ with average $\int_B x = 0$, and that $Q_0$ (6.23) is the limit operator if the following holds: For any second order polynomial in the coordinates $u$ we have

$$\begin{gathered}(K-I)(u-\chi_i\partial_i u-\chi_{ij}\partial_i\partial_j u) = Q_0 u\,;\\ u(x_1,\dots,x_d) = a + a_i x_i + a_{ij}x_i x_j, a, a_i, a_{ij}\in\mathbb{R}.\end{gathered} \tag{6.25}$$

With this terminology, in the above case, we have:

$$\chi_i = 0, \quad (1+\delta_{ij})\chi_{ij} = \psi_{ij}. \tag{6.26}$$

The additional condition that we have to be imposed for the Main Theorem to hold is that the first and second correctors are bounded. In the above context, *i.e.*, (6.26) this says:

$$\chi_{ij} \in L^\infty(B). \tag{6.27}$$

As we already point out at the end of Section 6.3, this is going to be the case if $B$ is finite or if the Markov chain is induced on the manifold $B \times \mathbb{R}^d$ by some subelliptic operator.

Let us now go back to the general situation and assume that $X = B\times\Gamma$ and $K$ are as in Section 6.1 (6.2). The assumption that I shall make is that $K$ admits *bounded* first and second correctors, *i.e.*, that (6.25) holds for an appropriate choice of bounded $\Gamma$-invariant functions $\chi_i$, $\chi_{ij}$ on $X$ and a symmetric positive definite matrix $Q_0 = (q_{ij})$.

I shall then consider $D$ some Lipschitz domain in $\mathbb{R}^d$ as in (0.10) where $d_1+d_2 = d$, $\Gamma \subset \mathbb{R}^d$, and I will identify $X \subset B \times \mathbb{R}^d$. The Markov chain induced by $K$ will be

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

denoted by $\big(z_n = (a_n, X_n) \in B \times \mathbb{R}^d \ ; \ n \geq 1\big)$ (*cf.* (6.19)), and I shall define the probability of life in $D$

(6.28) $$P(t,x) = \mathbb{P}_x[X_n \in D\ ; n = 1,\dots,t].$$

The corresponding probability of life $P_0(t,x)$ (*cf.* (0.3)) for the Brownian motion generated by $Q_0$ will also be considered.

The additional moment condition $(KM_E)$ or $(KM_E^*)$ will be imposed on $K$ for some appropriate $E$ large enough $E \geq B_0$: How large $B_0$ will have to be, and whether $KM_E$ or $KM_E^*$ or simply (FS), are imposed is something that I will stay "vague" at this point so as to avoid becoming cumbersome. The necessary precisions will be given in the course of the proof. (The reader who is unhappy with this attitude could assume the (FS) condition throughout.)

***The Main Theorem* (Random Environment)** *Let $K, D$ be as above. Then for all $0 < \varepsilon < 1$ there exists $C = C(d, A, K, \varepsilon)$ such that:*

(6.29) $$|P(t,x) - P_0(t,x)| \leq C \frac{P_0(t,x)}{\delta(t,x)^\varepsilon}\ ; \quad t, \delta(x) \geq C.$$

If $D$ is convex we can even set $\varepsilon = 1$ in (6.29), but no proof of this fact will be offered here.

## 6.5 The Proof of the Main Theorem: The Finite Span Case

I shall first give the proof of the Main Theorem in Section 6.4 under the finite span condition (FS) of Section 6.4. All the key ideas are already there. For the proof as in Sections 4 and 5 we shall consider

(6.30) $$F^\pm = P_0(t,x) \pm \theta w_\varepsilon(t,x)\ ; \quad t > 0,\ x \in D,\ 0 < \varepsilon < 1,\ \theta > 0,$$

where the $w_\varepsilon$ is as in (5.10) and is constructed from the Brownian motion generated by $Q_0$. We set also

(6.31) $$\Phi^\pm = F^\pm - \chi_i \frac{\partial F^\pm}{\partial x_i} - \chi_{ij} \frac{\partial^2 F^\pm}{\partial x_i \partial x_j},$$

where $\chi_i, \chi_{ij}$ are the correctors of $K$. We shall consider then $P_0, F^\pm, \Phi^\pm$, as functions in $\dot{D} = \mathbb{R} \times D \subset \dot{\mathbb{R}}^d$ (*cf.* Section 4 and [1]), and consider as in Section 4 the discrete parabolic operator

(6.32) $$L_K = K - I - \delta_t\ ; \quad \delta_t f = f(t+1,\cdot) - f(t,\cdot)$$

on $\dot{\mathbb{R}}^d$. In terms of $L_K$ the formula (6.25) can be rewritten:

(6.33) $$L_K(u - \chi_i \partial_i u - \chi_{ij} \partial_i \partial_j u) = L_0 u,$$

where now $u = a + bt + a_i x_i + a_{ij} x_i x_j$ (*i.e.*, it is also a linear function of $t$) and where $L_0 = Q_0 - \frac{\partial}{\partial t}$.

***Proposition*** *For all $\varepsilon > 0$ there exists $C = C(d, A, K, \varepsilon)$ such that*

(6.34) $$L_K \Phi^+ \leq 0\ ; \quad L_K \Phi^- \geq 0,$$

*in $(t,x) \in D(C,C)$ (notation of Section 4) for all $\theta > C$.*

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

The exact dependence of $C$, as far as $K$ is concerned, is in terms of $R > 0$ of (FS), $\|\chi_i\|_\infty$, $\|\chi_{ij}\|_\infty$ of Section 6.3, and $\lambda, \varepsilon_0 > 0$ the ellipticity constant of (Ell) and (6.23) $(\lambda^{-1}I \le Q_0 \le \lambda I)$.

The proof now, after all that has been done, is straight forward. We have:

$$L_0F^{\pm} = \mp\theta\frac{P(t,x)}{\delta^{2+\varepsilon}(x)}\,; \quad 0 < \varepsilon < 1,\ t > 0,\ x \in D, \tag{6.35}$$

and therefore by Taylor's theorem we have

$$L_K\Phi^{\pm}(t,x) = \mp\theta\frac{P(t,x)}{\delta^{2+\varepsilon}(x)} + O[\cdots], \tag{6.36}$$

where the $\cdots$ under the big $O$ sign is the sup in the set $[t, t+1] \times B_R(x)$ (provided that $\delta(x) > R$) of the following ("third" or "fourth" order) gradients:

$$\nabla^3, \quad \partial_t^2, \quad \partial_t\nabla, \quad \partial_t\nabla^2, \tag{6.37}$$

which are applied to the two functions:

$$P(t,x), \quad w(t,x). \tag{6.38}$$

These are exactly the type of estimates that had to be carried out in the proof of the first proposition of Section 4, and we shall not repeat the proofs here.

Once we have (6.34) we can finish the proof of the Main Theorem exactly as in Section 4,5. For the upper estimate of the Main Theorem instead of (4.22) we use the positivity of $\Phi^+$ in $D(C,C)$, for $C$ large enough. For the lower estimate we need the analogues of (4.25) and (4.28). This is done exactly as in the end of Section 5 or in Section 4. We no longer have the continuity of the path as in Section 5, but what we know instead (because of (FS) with some $R > 0$), is that if $\tau$ is the first exit time of $z(t)$ from $D_a$ (notation of (4.8)), then $z(\tau) \in D_{a-R}$. This, together with the analogue of (5.17), suffices for our estimates.

### 6.6 Proof of the Main Theorem: Finite Moment Case

To obtain the Main Theorem of Section 6.3, with the finite span condition (FS) replaced by the more general $KM_E$, or even $KM_E^*$, we can proceed as in Section 4 and truncate the kernel $K = K_R + N_R$, where $K_R$ (this corresponds to $\mu_R$ in (4.15)) has finite span $R$, and where the error term can be estimated in the appropriate norm $KM_E$ or $KM_E^*$ and satisfies $\|N_R\| = O(R^{-B})$ for some large $B > 0$. This construction is elementary, but the necessity to keep track of the covariance, of the second correctors, *etc.*, makes it rather tedious. In this final section I will outline a different approach, that although more sophisticated, and perhaps less natural, is in fact simpler to carry out even in the case of Section 4.

With schematic notation ($\nabla$ indicates here linear combinations of first and second derivatives) we can combine (6.30) and (6.31) and we can write

$$\Phi^{\pm} = P \pm \theta w + \nabla P + \theta\nabla w. \tag{6.39}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

If $a_0 \geq 1$ is large enough we have (*cf.* Section 3)

$$P \geq 10^{10}|\nabla P| \; ; \quad w \geq C_1 \frac{P}{\delta^\varepsilon} \; ; \quad |\nabla w| \leq C_2 \frac{P}{\delta} \; ;$$
$$t, x \in \dot{\mathbb{R}}^d, \; t \geq 1, \; \delta(x) \geq a_0, \; \sqrt{t} \geq \delta(x),$$

for some $C_1, C_2 > 0$. It follows that we can choose $a_1 \geq a_0$, such that for all $a_2 \geq a_1 + 10^{10}$, we can choose $\theta_0 > 0$ such that

$$\begin{aligned} &P \geq 10^{10}|\nabla P| \; ; \quad \theta w \geq 10^{10} P \; ; \quad \theta w \geq 10^{10}\theta|\nabla w| \; ; \\ &a_2 \geq \delta(x) \geq a_1, \; \sqrt{t} \geq a_2, \; \theta \geq \theta_0. \end{aligned} \tag{6.40}$$

We shall now truncate

$$\Phi_1^\pm = \Phi^\pm \chi[\delta(x) \geq a_1],$$

and then use the convolution to define:

$$G^\pm(t, x) = \Phi_1^\pm * \varphi(t, x) = \int \Phi_1^\pm(t, x - y)\varphi(y)\, dy,$$

for an appropriate molifier $0 \leq \varphi \in C_0^\infty(\mathbb{R}^d)$; supp $\varphi \subset B_a(1)$, $0 < a \ll 1$, $\int \varphi = 1$.

The same construction can be done of course directly on $F^\pm$ as in (4.11) (*i.e.*, where the third and fourth term of (6.39) are replaced by zero and $\Phi^\pm = F^\pm$), and for simplicity, I shall place myself in that case. We have then

$$LG^\pm(t, x) = \mp\theta \left[ \frac{P}{\delta^{2+\varepsilon}} * \varphi \right](t, x) \; ; \quad t > 0, \; \delta(x) > a,$$

as in (4.13). The error term

$$P(t, x) - G^\pm(t, x) = E(t, x),$$

can be controlled by the estimates of Section 3 and routine calculations. We have:

$$|E(t, x)| \leq C \frac{P(t, x)}{\delta^\varepsilon(t, x)} \; ; \quad t \geq C, \quad \delta(x) \geq C. \tag{6.41}$$

Furthermore

$$\pm G^\pm > 0 \; ; \quad \sqrt{t} \geq a_2, \; \delta(x) \leq a_2. \tag{6.42}$$

The advantage of $G^\pm$ over $F^\pm$ is that it is smooth in ($x \in \mathbb{R}^d$, $t > 0$) and we can use the Taylor's theorem with the global integral remainder:

$$\begin{aligned} f(x + y) &= \sum_{|\alpha|<k} \partial^\alpha f(x) y^\alpha/\alpha! \\ &\quad + k \sum_{|\alpha|=k} \int_0^1 (1 - t)^{k-1} \partial^\alpha f(x + ty) y^\alpha/\alpha!\, dt, \quad x, y \in \mathbb{R}^d, \end{aligned} \tag{6.43}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

(*cf.* [31]). Using this we can express the convolution with a measure

$$\int f(x+y)\,d\mu(y)\,dy = \sum_{|\alpha|\le 2} \frac{\partial^\alpha f(x)}{\alpha!}\int y^\alpha\,d\mu(y) \tag{6.44}$$
$$+\,3\sum_{|\alpha|=3}\int_0^1\int (1-t)^2\partial^\alpha f(x+ty)y^\alpha\,d\mu(y)\,dt.$$

The moment condition $(M_B)$ of Section 0 is just what is needed to control the error term in (6.44) in the context of Section 4. To estimate that error term we truncate again $\mu\chi_R = \mu\chi[|x| < R]$ ($\mu = \mu\chi_R + O(R^{-B})$ with $R$ as (4.21)) and use Section 2 and 3 and the Lemma of Section 4.2. But the point is that now we do *not* have to worry about the center or the covarience of $\mu\chi_R$.

As an illustration we can use the above estimates to deduce that the mixed differential-difference operator satisfies:

$$\mp\left[(\mu-\delta)-\frac{\partial}{\partial t}\right]G^{\pm}\ge 0\,; \quad \delta(x), \sqrt{t}\ge a_2, \tag{6.45}$$

where $\mu$ is as in (0.6), $(M_B)$, and $a_2$ is as in (6.42).

The estimates (6.41), (6.42) and (6.45) suffice to give the version of our Main Theorem for the continuous time Markov process (*i.e.*, the Markov process induced by the semi-group $\exp\bigl(-t(\delta-\mu)\bigr)$. For the discrete time process of our Main Theorem of Section 0 in (6.45) we must replace $\frac{\partial}{\partial t}$ by $\delta_t$ as in (4.3) and prove that $\mp L_\mu G^{\pm}\ge 0$. To cope with this, Taylor's theorem (6.44), has to be used now in the $\dot{\mathbb{R}}^d$ space. The details will be left to the reader.

The modifications that are needed to treat the Main Theorem in Section 6.3 under the general condition $KM_E^*$ (where the third and fourth term of $\Phi^{\pm}$ in (6.39) are not necessarily zero) are straight forward and will also be left to the interested reader.

# 7 Sampling

## 7.1 The Motivation

Let $s_j = X_1 + X_2 + \cdots + X_j \in \mathbb{R}^d$ be a standard random walk in $\mathbb{R}^d$, *i.e.*, $X_j$ are independent identically distributed random variables. Let $A\subset\mathbb{R}^d$ be some fixed set, or more generally $A_1, A_2, \dots, A_j, \dots \subset \mathbb{R}^d$ a sequence of sets, and let

$$J = \{j = 1, 2, \dots\,; X_j\in A_j\}\subset\mathbb{Z}^+, \tag{7.1}$$

this is a random set, and we can define

$$Q(n,x) = \mathbb{P}_x[s_j\in D\,;\, j\in J, 1\le j\le n] \tag{7.2}$$

where $D\subset\mathbb{R}^d$ is as before a Lipschitz domain. The problem of having to find the asymptotic behaviour of the above "sampled" probability of life appeared in my work

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

in Lie groups (*cf.* [7]). It is this, together with a number of ramifications, that has motivated the work in this paper.

We can define also the following stopping time

$$T = \inf[\, j = 1, \dots ; X_j \in A], \tag{7.3}$$

and iterate:

$$T_1 = T\,; \quad T_{n+1} = T_n + T \circ \theta_{T_n},\ n \geq 1, \tag{7.4}$$

where $\theta_j(\omega_1, \omega_2, \dots) = (\omega_{j+1}, \omega_{j+2}, \dots)$ is the usual shift operation on the path space $\omega = (s_1, s_2, \dots)$ of the Markov chain (*cf.* [35]). We can consider then the random walk $\Sigma_j = s_{T_j}$ $(j = 1, 2, \dots)$.

The probability of life $P(n, x)$ of this new random walk in $D$ is closely related with the above $Q(n, x)$ and it is easy to prove that (if $A_1 = A_2 = \cdots = A$) then $Q(n, x) \simeq P(n, x)$ in a sense that can be made precise.

It is in view of these types of problems that I will carry out in this section a number of formal, but not always trivial, calculations.

One should observe that it is this sampling that forces us to consider measures of unbounded support (: $\underline{Y} \notin L^\infty$ in (7.8) in general, even when $\underline{X} \in L^\infty$ in (7.5)). The final conclusion of this section is summarized at the end of Section 7.5, and it has applications in Lie group theory. The reader who is not interested in these applications could skip this section.

## 7.2 Random Walks

Let $X_1, \dots, X_n, \dots \in \mathbb{R}^d$ be identically distributed, independent, random variables centered at $\overline{X}$ and covariance $I$, *i.e.*,

$$\mathbb{E}|X|^2 \leq +\infty, \quad \mathbb{E}(\langle X_j - \overline{X}\,; \xi\rangle^2 = |\xi|^2),\ \xi \in \mathbb{R}^d. \tag{7.5}$$

I shall denote by $\mathcal{F}_j$ the $\sigma$-field generated by $X_1, \dots, X_j$ and I shall consider some stopping time $T \geq 1$ (with respect to $\mathcal{F}_0 = (\varnothing, \Omega) \subset \mathcal{F}_1 \subset \mathcal{F}_2 \cdots$). I shall also define the following sequence of stopping times

$$T_{n+1} = T_n + T \circ \theta_{T_n},\ n \geq 1\,; \quad T_1 = T, \tag{7.6}$$

and consider the random walks

$$s_j = X_1 + \cdots + X_j \in \mathbb{R}^d\,; \quad \Sigma_j = s_{T_j} = Y_1 + \cdots + Y_j, \tag{7.7}$$

where clearly $Y_1, \dots$ are independent random variables identically distributed with

$$Y = s_T = \sum_{j=1}^{\infty} X_j [T \geq j]. \tag{7.8}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

The following calculations are easy and well known:

$$\mathbb{E}Y = \Sigma\mathbb{E}[\mathbb{E}(X_j//j \leq T)\mathbb{I}(j \leq T)], \tag{7.9}$$

and since $[j \leq T] \in \mathcal{F}_{j-1}$ this gives

$$\mathbb{E}Y = \overline{X} \cdot \sum_{j=1}^{\infty} \mathbb{P}(j \leq T), \tag{7.10}$$

provided that the series on the right hand side converges. If in particular $T$ is time homogeneous and satisfies:

$$T = j + T \circ \theta_j \text{ on } T \geq j, \tag{7.11}$$

then (7.10) gives:

$$\mathbb{E}Y = \frac{\overline{X}}{1 - \mathbb{P}(T > 1)}. \tag{7.12}$$

A similar calculation can be carried out for the covariance: Assume (w.l.o.g.) that $\overline{X} = 0$. Then:

$$\mathbb{E}\langle Y, \xi\rangle^2 = \sum_{j=1}^{\infty} \mathbb{E}[\mathbb{E}(\langle X_j, \xi\rangle^2 // \mathcal{F}_{j-1})\mathbb{I}(j \leq T)] = |\xi|^2 \sum_{j\geq 1} \mathbb{P}(j \leq T). \tag{7.13}$$

What is therefore obvious, but important, is that the covariance of the new walk $\Sigma_j$ is proportional to the covariance of $s_j$. This point will be generalized to a more general setup in the next section.

### 7.3 Markov Chains

Let $(z_n \in X\ ;\ n = 0, 1, \dots)$ be some Markov chain and $T = 0, 1, 2, \dots$ some stopping time. The following considerations and facts are easy exercises in the terminology and notation of Markov chains (*cf.* [35]). Let $\hat{X} = X \cup \infty$ be the augmented state space where we can define the corresponding killed chain $(\hat{z}_n \in \hat{X}\ ;\ n \geq 0)$ by

$$\hat{z}_n = \begin{cases} z_n \in X, & n < T \\ \infty, & n \geq T. \end{cases} \tag{7.14}$$

Similarly if $0 = T_0 \leq T_1 \leq \cdots$ is a sequence of stopping times we can define $(\tilde{z}_n = z_{T_n} \in X, n \geq 0)$ which is a new Markov chain.

Let us assume that $(z_n \in X)$ is time homogeneous and $K(x, dy) = \mathbb{P}(z_1 \in dy // z_0 = x)$ is the transition kernel. Then $(\hat{z}_n \in \hat{X})$ is time homogeneous if $T$ satisfies

$$T = j + T \circ \theta_j \text{ on } [T \geq j]. \tag{7.15}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

The transition kernel of $(\hat{z}_n \in \hat{X})$ is then

$$\hat{K}(x, dy) = \mathbb{P}_x[z_1 \in dy\ ;\ T > 1] = K(x, dy)\chi_A(x, y)\ ; \quad x, y \in X, \tag{7.16}$$

where $\chi_A$ is the characteristic function of the set $A \subset X \times X$ defined by

$$(z_0, z_1) \in A \Leftrightarrow T > 1. \tag{7.17}$$

The point is that $[T > j]$ is measurable with respect to $\mathcal{F}_j$, the $\sigma$-field generated by $z_0, \dots, z_j$.

Similarly the chain $(\tilde{z}_n \in X)$ is time homogeneous if the stopping times $T_1 \le T_2 \le \cdots$ are defined inductively

$$T_{n+1} = T_n + T_1 \circ \theta_{T_n}\ ; \quad T_1 = T. \tag{7.18}$$

The transition kernel of $(\tilde{z}_n \in X)$ is then

$$\tilde{K}(x, dy) = \mathbb{P}_x(\tilde{z}_1 \in dy) = \sum_{j=0}^{\infty} \mathbb{P}_x[z_j \in dy\ ;\ T = j]. \tag{7.19}$$

Clearly, if $j \ge 1$, $[z_j \in dy\ ;\ T = j] = [z_1 \in dy\ ;\ T = 1] \circ \theta_{j-1}$ on $[T \ge j]$ by (7.15), and since $[T \ge j] \in \mathcal{F}_{j-1}$, we have by the Markov property:

(7.20)
$$\begin{aligned}\mathbb{P}_x[z_j \in dy\ ;\ T = j] &= \mathbb{P}_x\big([z_1 \in dy\ ;\ T = 1] \circ \theta_{j-1}\mathbb{I}(T \ge j)\big)\\ &= \mathbb{E}_x[\mathbb{I}(T \ge j)\mathbb{P}_{z_{j-1}}(z_1 \in dy\ ;\ T = 1)]\\ &= \mathbb{E}_x\big[\mathbb{I}(T \ge j)\big(K(z_{j-1}, dy) - \hat{K}(z_{j-1}, dy)\big)\big]\ ; \quad j \ge 1,\end{aligned}$$

because $[T \ge j] \cap [T \ne 1] = [T \ge j] \cap [T > 1]$. The right hand side of (7.20) is

$$\int_X \big(K(z, dy) - \hat{K}(z, dy)\big)\,\mathbb{P}_x[\hat{z}_{j-1} \in dz]\ ; \quad j \ge 1. \tag{7.21}$$

If we use the usual notation for the composition of submarkovian kernels

$$K_1 \circ K_2(x, dy) = \int_X K_1(x, dz)K_2(z, dy), \tag{7.22}$$

(7.20) becomes:

$$(\hat{K})^{j-1} \circ (K - \hat{K})(x, dy)\ ; \quad j \ge 1, \tag{7.23}$$

and we obtain, at least formally:

$$\tilde{K}(x, dy) = \mathbb{P}_x[T = 0]\delta_x(dy) + \hat{R} \circ (K - \hat{K})(x, dy), \tag{7.24}$$

$$\hat{R} = I + \hat{K} + \hat{K}^2 + \cdots. \tag{7.25}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

We have

$$\tilde{K} = \hat{R} \circ (K - \hat{K}) + \mathbb{P}_x[T = 0]I. \tag{7.26}$$

If we assume that $T \geq 1$ we obtain

$$I - \tilde{K} = \hat{R}(I - \hat{K}) - \hat{R}(K - \hat{K}) = \hat{R}(I - K), \tag{7.27}$$

and

$$K1 = 1 \Rightarrow \tilde{K}1 = 1,$$

at least formally. It follows that $\tilde{K}$ in (7.26) is a Markovian kernel as long as $K$ is Markovian, provided that the following series converges

$$\hat{R}(x, X) < +\infty, \quad x \in X. \tag{7.28}$$

(7.28) holds if and only if:

$$\sum_{j \geq 1} \mathbb{P}_x(T \geq j) < +\infty, \tag{7.29}$$

and, because of (7.15), this is the case if $\mathbb{P}_x(T = +\infty) \neq 1$.

If $K^*$ is the dual kernel (*i.e.*, an operator on the space of measures on $X$) then we say that $K$ is bi-markovian and uniquely ergodic if there exists a unique positive measure $dx$ on $X$ such that

(7.30)
$$(I - K^*)(d\mu) = 0, \quad d\mu \text{ a positive measure} \Leftrightarrow d\mu \text{ is a scalar multiple of } dx.$$

Assume that this is the case. Then from the formula (7.27) we see that the measure $d\mu$ on $X$ is invariant by $\tilde{K}$, if and only if

$$(I - \tilde{K}^*)(d\mu) = (I - K^*)\big(\hat{R}^*(d\mu)\big) = 0, \tag{7.31}$$

*i.e.*, if and only if:

$$\hat{R}^*(d\mu) = \text{scalar multiple of } dx. \tag{7.32}$$

### 7.4 Harmonic Functions

Let $(z_n \in X \; ; \; n \geq 0)$ be a Markov chain as in Section 7.3 that is assumed to be time homogeneous, and let $y(x) \in \mathbb{R}^d$ $(x \in X)$ be some harmonic function $(y(x) = \int_X K(x, dz) y(z))$. We shall consider then the martingale and the martingale differences:

(7.33)
$$y_0 = y(z_0) \; ; \quad y_j = y(z_j) = y_0 + d_1 + \cdots + d_j \; ; \quad \mathbb{E}(d_j // \mathcal{F}_{j-1}) = 0 \; ; \quad j \geq 1.$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

We shall also consider the covarience (here we assume that $K$ is strictly Markovian):

$$\begin{aligned}\mathbb{E}_x\langle d_1,\xi\rangle^2 &= \int (K-I)(x,dz)\langle y(z),\xi\rangle^2 \\ &= [(K-I)\langle y(\cdot),\xi\rangle^2](x) = \Phi(x) = \Phi_{ij}(x)\xi_i\xi_j\ ; \quad \xi\in\mathbb{R}^d.\end{aligned} \tag{7.34}$$

If we stop the process at $T$ where $T = 0,1,\dots$ is a stopping time as in Section 7.3 (7.15) we see that

$$\mathbb{E}_x\langle y_T - y_0,\xi\rangle^2 = \mathbb{E}_x\sum_{\nu=1}^{\infty}\big(\langle d_\nu,\xi\rangle^2\mathbb{I}(\nu\le T)\big), \tag{7.35}$$

$$\begin{aligned}\mathbb{E}_x[\langle d_\nu,\xi\rangle^2\mathbb{I}(\nu\le T)] &= \mathbb{E}_x[\mathbb{I}(\nu\le T)\mathbb{E}(\langle d_\nu,\xi\rangle^2//\mathcal{F}_{\nu-1}] \\ &= \mathbb{E}_x[\mathbb{I}(\nu\le T)\mathbb{E}_{z_{\nu-1}}\langle d_1,\xi\rangle^2] \\ &= \int \mathbb{P}_x(\hat z_{\nu-1}\in dz)\mathbb{E}_z(\langle d_1,\xi\rangle^2) \\ &= (\hat K^{\nu-1}\Phi)(x) \\ &= \Big(\big(\hat K^{\nu-1}\circ(K-I)\big)\langle y(\cdot),\xi\rangle^2\Big)(x)\ ; \quad \nu\ge 1.\end{aligned} \tag{7.36}$$

This gives, at least formally:

$$\mathbb{E}_x\langle y_T-y_0,\xi\rangle^2 = [\hat R\circ(K-I)\langle y(\cdot),\xi\rangle^2](x) = [(\tilde K - I)\langle y(\cdot),\xi\rangle^2](x), \tag{7.37}$$

provided that $T\ge 1$. This is a correct formula if (7.29) holds and if the integrals in (7.37) converge absolutely. In that case $(\tilde y_n = y(\tilde z_n)\in\mathbb{R}^d, n\ge 0)$ is a martingale and $\tilde K$ is strictly Markovian and we have:

$$\tilde y_n = y_0 + \tilde d_1 + \cdots\ ; \qquad \mathbb{E}[\langle \tilde d_\nu,\xi\rangle^2//\tilde z_0,\dots,\tilde z_{\nu-1}] = \mathbb{E}_{\tilde z_{\nu-1}}(\langle y_T-y_0,\xi\rangle^2). \tag{7.38}$$

What is interesting is to consider a starting probability $\mu$ for $(z_n\in X, n\ge 0)$ for which

$$\mathbb{E}_\mu\langle d_\nu,\xi\rangle^2 \text{ is independent of } \nu. \tag{7.39}$$

The reason is that in that case we can formulate interesting limit theorems for the martingale. It is of interest therefore to find conditions under which the property (7.39) passes from $(y_n)$ to $(\tilde y_n)$. This will be done in the next section in the context of the $\Gamma$-invariant chains of Section 6.

## 7.5 $\Gamma$-Invariant Chains

In this section I shall specialize $X = B\times\Gamma$ where $B$ and $\Gamma$ are as in Section 6.1 and I shall assume that $K(x,dy)$ in (6.2) is Markovian and $\Gamma$-invariant. I shall also

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

assume that the stopping time $T$ is $\Gamma$-invariant, in the following sense: $\Gamma$ induces an obvious action, by the translation $x \mapsto x + \gamma$ *cf.* (6.1), on the path space ($\omega \in \Omega$) of ($z_n \in X$ ; $n \geq 1$). We shall assume that $T\big(\gamma(\omega)\big) = T\omega$.

I shall assume that $z_n = (a_n, y_n) \in B \times \Gamma$ is expressed in harmonic coordinates, *i.e.*, I shall assume that the $\Gamma$-invariant chain admits correctors $\chi$, and by a change of coordinates, if necessary, I shall assume that these correctors $\chi \equiv 0$. With the notation

$$y_n = y_0 + d_1 + \cdots + d_n, \tag{7.40}$$

we then have (*cf.* (6.20), (6.21)):

$$\mathbb{E}(\langle d_n, \xi\rangle^2 // \mathcal{F}_{n-1}) = \Phi_{ij}(a_{n-1})\xi_i\xi_j = \Phi(a_{n-1}) \; ; \quad \xi \in \mathbb{R}^d. \tag{7.41}$$

Finally if $dx = da \otimes d\gamma$ is an invariant measure such that $da \in \mathbb{P}(B)$, as in Section 6.3, the symbol of the limit operator is

$$Q_0 = \int_B \Phi(a)\, da. \tag{7.42}$$

Let $K(\gamma \; ; a, b)$ ($\gamma \in \Gamma$ ; $a, b \in B$) be the kernel of ($z_n$) with respect to $da \otimes d\gamma$ (with the notation (6.16)). Then if the stopping time $T$ is $\Gamma$-invariant, the two kernels $\hat{K}$ and $\tilde{K}$ of Section 2, are also $\Gamma$-invariant and

$$\hat{K}(\gamma \; ; a, b) = K(\gamma \; ; a, b)\chi_{a,b}(\gamma), \tag{7.43}$$

where $\chi_{a,b}(\cdot)$ is the characteristic function of a set $A_{a,b} \subset \Gamma$ (*cf.* (7.16)).

The new chain ($\tilde{z}_n \in X$) is $\Gamma$-invariant, but unless

$$\hat{R}_B(da) = \text{scalar multiple of } da, \tag{7.44}$$

the measure $da \otimes d\gamma$ is not invariant for $\tilde{K}$. If $K_B$ is uniquely ergodic in the sense of (7.30) $\tilde{K}$ may not even be bi-markovian with respect to any measure on $X$ (*cf.* (7.31)). Independently of this difficulty, however, we have (with the notation $(a, y) \in B \times \Gamma = X$ for the coordinates):

$$(K - I)\langle y, \xi\rangle^2 = \Phi_{ij}(a)\xi_i\xi_j \; ; \quad \xi \in \mathbb{R}^d, \tag{7.45}$$

$$(\tilde{K} - I)\langle y, \xi\rangle^2 = \hat{R}(K - I)\langle y, \xi\rangle^2 = \hat{R}\Phi_{ij}(a)\xi_i\xi_j. \tag{7.46}$$

If, as in (6.24), the $\chi_{ij}$'s are such that

$$(K_B - I)\chi_{ij} = \Phi_{ij}(a) - q_{ij}, \tag{7.47}$$

then

$$(\tilde{K} - I)\chi_{ij} = \hat{R}(K - I)\chi_{ij} = \hat{R}\Phi_{ij}(a) - \lambda q_{ij}, \tag{7.48}$$

where $\lambda = \int_B \hat{R}_B(da)$.

In particular the covarience of the martingale $(\tilde{y}_n)$ with respect to $\tilde{da}$ (notation of (6.22)) is proportional to the covarience of the original martingale $(y_n)$, and the new martingale $(\tilde{y}_n)$ has the same second correctors as the original one, in the sense (6.26). This holds despite the fact that the new chain $\tilde{K}$ is not open to the same treatment as the original chain $K$ (unless (7.44) holds). The conclusion is that the probability of life of $(\tilde{z}_n)$ in some Lipschitz domain satisfies the same asymptotic behaviour as that of $(z_n)$ and it is given by the Main Theorem of Section 6.3. Results like this are vital for the study of the heat kernel on a Lie group (*cf.* [7], [13], [14]). These results are very complicated to obtain without the machinery developed in the last three sections (*cf.* [7], [4], [5], [9]).

***Remark*** In view of the truncation of the kernels that was proposed in Section 6.6, it is of interest to be able to find sets $A_{a,b} \subset \Gamma$ $(a, b \in B)$ that satisfy:

$$A_{a,b} \subset [\gamma \in \Gamma \,;\, |\gamma| \leq R] \,; \quad \hat{K}_B(a, b) = \lambda K_B(a, b),$$

with $R > 0$, and $0 < \lambda < 1$ independent of $a, b \in B$. It is easy to see that this is possible under the condition $K_E$ of Section 6.6 provided that we impose in addition some smoothness conditions on the kernel: *e.g.* with the notation of [13, Section 2.2] it suffices to demand that each measure $\mu_{h,k}$ $(h, k \in K)$ is continuous. We can then pre-assign for $\lambda$ any value $0 < \lambda < 1$. Furthermore, if in addition we assume that $K_B$ is symmetric, we can demand that $\hat{K}_B$ is symmetric.

These considerations however are rather "esoteric" and of very little general interest. We shall not elaborate.

# 8 The Kernel Estimates

## 8.1 The Upper Estimate (0.15)

We shall consider $\mu$ a measure as in (0.5) that has the additional property that

(8.1) $$\operatorname{supp} \mu \subset \mathbb{Z}^d \cap [|x| \leq \delta],$$

for some $\delta > 0$. This gives rise to a "pure lattice" random walk (*cf.* [5, Sections 0.7, 0.8]) and the kernel $p_n^\mu(x, y)$ (*cf.* (0.2)) is now defined with respect to the counting (Haar-) measure $dx$ of $\mathbb{Z}^d$.

I shall use the notation and the over all treatment of the Doob $h$-processes for discrete stopped Markov chains that I described in [5, Section 3]. The globally Lipschitz domain $D$ is as in (0.10), (1.1) and I shall denote:

(8.2) $$D_R = D + (R, 0, 0, \dots, 0) \,; \quad \chi_R(\cdot) = \text{char. function of } D_R \,; \quad R \in \mathbb{R}.$$

(8.3) $$K_R(x, y) = \chi_R(x)\chi_R(y)\mu(x - y) \,; \quad x, y \in \mathbb{Z}^d$$

(abusively: $\mu(x) = \mu(\{x\})$ ; $x \in \mathbb{Z}^d$). Clearly we have:

(8.4) $$p_n^\mu(x, y) = (K_0)^n(x, y) = (K_0 \circ \cdots \circ K_0)(x, y) \,; \quad x, y \in \mathbb{Z}^d,$$

and for the corresponding kernel associated with the domain $D_R$ we have

$$p_n^R(x,y) = p_n^\mu\big(x-(R,0,\cdots), y-(R,0,\cdots)\big) = (K_R)^n(x,y) \\ (= K_R \circ K_R \circ \cdots \circ K_R)\,; \quad x,y \in \mathbb{Z}^d. \tag{8.5}$$

We shall now define

$$\phi(T,x) = T^{-d/4} P_\mu(T,x)\,; \quad T \geq T_0,\ \delta(x) \geq R_0. \tag{8.6}$$

If $T_0$, $R_0$ are large enough we have $\phi > 0$. This fact (which follows, among other things, from our Main Theorem) is a consequence of the local central limit theorem. For every fixed $T$, large enough, if $R_0$ is large enough, the function $\phi(T,\cdot)$ is $K_R$-superharmonic in the range $D_R$ for $R > R_0$ (*cf.* [5, Section 3.2]). To see this we can use the operator in (4.3) and the fact that $P_\mu(T,x)$ is decreasing in $T$ for all fixed $x \in D_{R_0}$.

The $h$-transformed kernel that will be considered (for $R$ large enough) is:

$$K(n\,;x,y) = K_T(n\,;x,y) = \frac{K_R^n(x,y)}{\phi(T,x)\phi(T,y)}\,; \quad n = 1,2,\ldots,\ x,y \in D_R, \tag{8.7}$$

which this is the kernel of the $h$-transformed process, with respect to the measure (*cf.* [6]):

$$d\mu_T(x) = \phi^2(T,x)dx\,; \quad x \in D_R, \tag{8.8}$$

of the process induced by $K_R$ in $D_R$. The kernel (8.7) is bi-submarkovian with respect to (8.8) (*i.e.*, $K_T$ and $K_T^*$ are both submarkovian). In the considerations that follow it is important to bear in mind that $K_R$ is a decreasing function of $R$. As a result it is not important to keep track of the $R$ in the definition (8.7) and in (8.10) below. At the end the $R$ will be chosen large enough but *fixed*. [Analogous definitions can be made for the differential operator (0.16)

$$K(t\,;x,y) = \frac{T^{d/2} p_t(x,y)}{P(T,x)\,P^*(T,y)}\,; \quad d\mu(x) = T^{-d/2} P(T,x) P^*(T,x) dx, \tag{8.9}$$

with the notation of (8.21).]

Together with the kernel (8.7) I shall also consider the following kernel ([6], [20], [5, Section 4]):

$$K_s(n\,;x,y) = e^{s(\varphi(x)-\varphi(y))} K_T(n\,;x,y), \tag{8.10}$$

where $\varphi(x)$, $x \in \mathbb{Z}^d$ satisfies the following Lipschitz condition:

$$|\varphi(x) - \varphi(y)| \leq |x-y|\,; \quad x,y \in \mathbb{Z}^d. \tag{8.11}$$

What is now needed are the following two estimates:

$$K_T(n\,;x,y) \leq C\left(\frac{T}{n}\right)^C; \quad T,n \geq n_1,\ x,y \in D_R,\ R \geq R_1, \tag{8.12}$$

$$\|K_s(n\,;\cdots)\|_{2\to 2} \leq C\exp(Cns^2\delta^2 + C)\,; \quad n \geq n_1,\ |s\delta| \leq C, \tag{8.13}$$

where $n_1$, $R_1$ are large enough and $C > 0$ is independent of $T$, $n$, $s$, $x$, $y$. In (8.13) $\|\ \|_{2\to 2}$ indicates the $L^2(D_R) \to L^2(D_R)$ norm of the kernel $K_s(n\,;x,y)$ (equivalently: of $e^{s(\varphi(x)-\varphi(y))}\ K_R^n(x,y)$) with respect to $d\mu_T(x)$ (8.9) (equivalently: with respect to $dx$).

**Proof of (8.12) (*cf.* [7, p. 651])** We have:

$$\begin{aligned} p_{3n}^{\mu}(x,y) &\leq \int_D\int_D p_n^{\mu}(x,z)p_n^{\mu}(z,u)p_n^{\mu}(u,y)\,dz\,du \\ &\leq Cn^{-d/2}P_{\mu}(n,x)P_{\mu^*}(n,y)\,; \quad x,y\in D,\ n\geq 0 \end{aligned} \tag{8.14}$$

(where $d\mu^*(x) = d\mu(-x)$ is the adjoint measure), simply because the central factor in the integrant of (8.14) can be estimated by

$$p_n(x,y) \leq \mu^n(x-y) \leq Cn^{-d/2}\,; \quad n \geq 1,\ x,y \in D. \tag{8.15}$$

We now use the Main Theorem (in an essential way) to compare $P_{\mu}(\cdot,\cdot)$ with the brownian $P(\cdot,\cdot)$ and then use (0.20). This gives the estimate (8.12).

**Proof of (8.13) [5, Section 4]** We shall prove (8.13) in the special case where $\varphi = \langle a,x\rangle$ ($a \in \mathbb{R}^d$, $|a| = 1$). This will be good enough for our purposes since these linear functions "give distance" in $\mathbb{Z}^d$ (and, in any case, the general result for $\varphi$ as (8.11) follows from this very easily). A moment's reflection shows that it suffices to show that the $L^\infty \to L^\infty$ norm of the product operator:

(8.16) [Multiplication by $e^{-s\varphi(\cdot)}$] $\circ$ [convolution by $\mu$] $\circ$ [Multiplication by $e^{s\varphi(\cdot)}$]

satisfies

$$\|\ \|_{\infty\to\infty} \leq \exp(Cs^2)\,; \quad s \in \mathbb{R}, \tag{8.17}$$

where the $\delta$ is now absorbed in the $C$. This also holds in fact as long as $\mu$ has a Gaussian decay $\mu(x) \leq C\exp(-C|x|^2)$. From (8.17), by taking adjoints and by interpolation, (8.13) follows.

The proof of (8.17) for $|s| \geq 1$ is a trivial consequence of (8.11) and the decay of $\mu$. For $|s| \leq 1$ we use the estimate

$$\exp\bigl(s(x-y)\bigr) = 1 + s(x-y) + O[s^2|x-y|^2e^{s|x-y|}],$$

and the fact that $\mu$ is a centered probability measure.

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

Another way to prove (8.13) is to use the classical estimate $\mu^{*n}(x) \leq Cn^{-d/2}\exp(-\frac{|x|^2}{Cn})$ *cf.* [28] and to integrate it against $\exp(s|x|)$.

To deduce the upper estimate (0.15) from (8.12), (8.13) and (8.17) I shall follow closely [20]. Indeed, by the interpolation of [20, Section 2] we easily deduce from (8.12), (8.13) and (8.17)

$$\|K_s(n\,;\cdots)\|_{p\to q} < C\left(\frac{T}{n}\right)^{C(1/p-1/q)}\exp\left(Cqs^2n+\frac{C}{q}\right)\,; \tag{8.18}$$
$$T, n \geq C\,; \quad 2 \leq p < q < +\infty.$$

From this we can estimate

$$\|K_s(n\,;\cdots)\|_{2\to\infty} \leq \prod_{i=1}^{N}\|K_s(n_i\,;\cdots)\|_{p_i\to p_{i+1}}, \tag{8.19}$$

where $2 = p_1 < p_2 < \cdots < p_{N+1} = +\infty$ and $n_1 + n_2 + \cdots + n_N = n$ are chosen appropriately (*cf.* [20, Section 2]). The estimate

$$\|K_s(T\,;\cdot,\cdot)\|_{2\to\infty} \leq \exp(Cs^2T + C)\,; \quad T \geq C,\ s \in \mathbb{R}, \tag{8.20}$$

follows, where now $C > 0$ also depends on $\delta$ (8.1). The analogous estimate is also valid for the adjoint operator $K^*$ and therefore we obtain the same estimate (8.20) for $\|K_s\|_{1\to 2}$. Combining the two, we obtain the estimate for $\|K_s\|_{1\to\infty}$. Finally, by the usual optimization (*cf.* [6], [20]) over $\varphi$ (in (8.11)) and $s \in \mathbb{R}$, we obtain a proof of (0.15).

An identical proof can be given for absolutely continuous random walks (*cf.* [5, Section 0.7]). This means that instead of (8.1) we can assume that $d\mu(x) = f(x)dx$ with $f \in L^\infty$ and:

$$\operatorname{supp} f \subset [|x| \leq \delta].$$

The modifications needed for the proof of (0.18) starting from (8.9) are obvious.

## 8.2 The Lower Kernel Estimate

I shall consider $p_t(x,y) = p_t^{\mathcal{A}}(x,y)$ with $\mathcal{A}$ as in (0.16) and I shall prove

$$p_t(x,y) \geq Ct^{-d/2}P(t,x)P^*(t,y)\exp\left(-\frac{|x-y|^2}{Ct}\right)\,; \quad t > 0,\ x, y \in D, \tag{8.21}$$

where $P = P_{\mathcal{A}}$, $P^* = P_{\mathcal{A}^*}$, is as in (0.17) and $D$ is as in (0.10), and where $C$ only depends on $d$, $A$, and $\lambda$ of (0.19). It is, of course, enough to prove (8.21) for $t = 1$ and then scale. (8.21) shows that (0.18) is sharp.

This is very easy and is done in two steps. First if $z \in D$, $\delta(z) \geq C$, $t \in [c, C]$ we have

$$p_t^{\mathcal{A}}(z,z) \sim P(t,z) \sim P^*(t,z) \sim C. \tag{8.22}$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

This together with the parabolic Harnack boundary principle and (0.20) implies that we can find $a, b \in D$ such that

$$\delta(a), \delta(b) \geq 1 \; ; \quad |a - x|, |b - y| \leq 1, \tag{8.23}$$

and

$$\begin{gathered} p_t(x, a') \geq CP(1, x) \; ; \quad p_t(b', y) \geq CP^*(1, y) \; ; \\ a', b' \in D, \; |a' - a| \leq 1/2, \; |b' - b| \leq 1/2, \; t \in [c, C]. \end{gathered} \tag{8.24}$$

The second step is standard: We "link" $a = a_1, a_2 \ldots, a_N = b$ by a sequence of points:

$$a_j \in D, \; \delta(a_j) \geq 1/4, \; |a_{j+1} - a_j| \leq N^{-1/2} \; ; \quad j = 1, 2, \ldots, \tag{8.25}$$

and use the interior parabolic Harnack estimate:

$$p_{1+\frac{j+1}{N}}(x, a_{j+1}) \geq c p_{1+\frac{j}{N}}(x, a_j), \quad j = 1, 2, \ldots . \tag{8.26}$$

This gives:

$$p_3(x, b') \geq c^N p_1(1, a) \; ; \quad |b' - b| \leq 1/4. \tag{8.27}$$

(8.24) and (8.27), together with the semigroup property:

$$p_4(x, y) = \int p_3(x, b') p_1(b', y) \, db', \tag{8.28}$$

gives:

$$p_4(x, y) \geq c^N P(1, x) P^*(1, y). \tag{8.29}$$

And since it is possible to choose

$$\sqrt{N} \sim |x - y|,$$

we can scale (8.29) and we have (8.21) because of (0.20). What is essential in the above argument is not (0.20) but (8.22) and the argument clearly generalizes to time-dependent operators as in (2.4).

There is a way to avoid the parabolic Harnack boundary principle in the first step of the above proof. What we use instead is the upper estimate (0.18) and we deduce, that if $a \geq a_0$ is large enough, and if $0 < \eta < \eta_0$ is small enough, then:

$$\int p_t(x, y) [|x - y| \leq a\sqrt{t} \; ; \delta(y) \geq \eta\sqrt{t}] \, dy \geq 1/2 P(t, x) \; ; \quad t > 0, \; x \in D.$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

This approach is more involved but it has the advantage that it easily adapts to give the corresponding lower estimate for the kernel of a random walk: There exists $c > 0$ such that

$$p_n^\mu(x,y) \geq cP_\mu(n,x)P_\mu(n,y)\exp\left(-\frac{|x-y|^2}{cn}\right); \tag{8.30}$$
$$x, y \in D,\ \delta(x), \delta(y), n \geq c^{-1},\ |x-y| \leq cn.$$

For this, of course, additional technical, but obvious, conditions (*e.g.* the general position of [5, Section 0.7]) have to be imposed on $\mu$. We shall not elaborate but the reader who is willing to write the proof of (8.30) out for himself will have no difficulty to find the appropriate conditions on $\mu$ (no lower estimate can hold for instance for periodic walks that live in strictly smaller sublattice *e.g.* $\mu = 1/2(\delta_{-2}+\delta_2)$). The above upper and lower estimates easily generalize to non- (space or time) homogeneous random walks (*cf.* [29], [5] for the appropriate definitions).

In proving the lower estimate (8.30) the reader should observe that what replaces the scaled interior Harnack estimate that is used in the proof of (8.21), are the estimates $\mu^{*n}(x) \geq cn^{-d/2}$, $|x| \leq c\sqrt{n}$, and $\mu^{*n}(x) \leq Cn^{-d/2}\exp(-\frac{|x|^2}{cn})$ which follows from the Edgeworth expansion [28]. This easy approach through the Edgeworth expansion however only works for space and time homogeneous random walks—in the non-space homogeneous case the estimates in [37] have to be used.

### 8.3 Harnack Estimates: Further Results

It has become a real fashion in the subject, these days, to ascribe to a Markov chain various "abstract" properties *e.g.* the Harnack property or the upper Gaussian estimate (0.15), or the lower estimate (8.30), *etc.*, and then to study these properties for their own sake. I should add that one thing that seem to sell well, and for which there is now a flourishing (and promising) production is the study of the interconnection of these properties between themselves; *i.e.*, which property is implied by others. The next few lines could be considered to be my modest contribution in that direction.

Let $u(t,x) \geq 0$, $t > 0$, $x \in D$ be a positive parabolic function, $(\partial_t - \mathcal{A})u = 0$ where $D$ is as in (0.10) and $\mathcal{A}$ is as in (0.18), and let us assume that $u(t,x) = 0$ for $t > 0$, $x \in \partial D$. Then such a function admits an integral representation in terms of $U_y(t,x) = p_t^{\mathcal{A}}(x,y)$:

$$u = \int_D U_y u(0,y)\,dy\,; \qquad t > 0,\ x \in D. \tag{8.31}$$

This is easy to see, at least as long as we know in advance growth estimates for $u(\cdot,\cdot)$. The functions $U_y$, $y \in D$ satisfy on the other hand the upper and lower estimates (0.18), (8.21), (8.32):

(8.32)
$$\pi_{c_1} \leq U_y \leq \pi_{c_2}\,; \qquad \pi_C(t\,;x,y) = Ct^{-d/2}P_{\mathcal{A}^*}(t,x)P_{\mathcal{A}}(t,y)\exp\left(-\frac{|x-y|^2}{Ct}\right);$$
$$t > 0,\ x, y \in D.$$

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

From (8.31), (8.32) and (0.20) we can immediately deduce the following backward Harnack inequality of [3]:

For $u(t,x) \geq 0$ as above there exists $C(d, A, D)$ such that:

$$\frac{u(t_1,x)}{u(t_2,x)}, \frac{u(t_2,x)}{u(t_1,x)} \leq C \exp\bigl(C(1/t_1 + t_2 - t_1)\bigr) \;; \quad 0 < t_1 < t_2,\ x \in D,$$

provided that $D$ is a *bounded* Lipschitz domain. Such estimates are obviously false in unbounded domains *e.g.* "the Gaussian in $\mathbb{R}^d$".

What is more, the above can be adapted to time homogeneous operators in non-divergence form [38]. This last point presents a real interest (*i.e.*, it is not a question whether one "property" implies another). From these backwards Harnack estimates, we can obtain nontrivial consequences in Fatou type theorems and such like [16], [3]. Furthermore this type of analysis applies to more general domains and not just cylinders (*cf.* [16, 3.1.3], [25]):

$$\dot{D} = [(x_1, x', t) \in \dot{\mathbb{R}}^d \;; x_1 \in \mathbb{R}, x' \in \mathbb{R}^{d-1}, t > 0, x_1 > \varphi(x', t)], \tag{8.33}$$

where $\varphi$ satisfies the following anisotropic but scale invariant (*cf.* (1.12)) Lipschitz condition:

$$|\varphi(x',t) - \varphi(y',s)| \leq A[|x' - y'| + |t - s|^{1/2}] \;; \quad x', y' \in \mathbb{R}^{d-1},\ t, s > 0. \tag{8.34}$$

The same parabolic boundary principle can be proved by the same method in the context of discrete potential theory, *i.e.*, for $u(n,x) \geq 0$ such that $u(n,x) = 0$ ($x \notin D$, $n \geq 0$) and $L_\mu u = 0$ ($x \in D$; $n \geq 0$) where $L_\mu$ is as in (4.3). This result seems to be new even in the harmonic case (*cf.* [29]), and it has a number of obvious but not uninteresting consequences. These are similar to the consequences of the classical boundary Harnack principle for the standard Laplacian $\Delta$. A typical such application is the existence and the uniqueness of the reduite (*cf.* [8]), *i.e.*, of a unique (up to scalar multiple) positive ($\not\equiv 0$) harmonic function in $D$ that vanishes on the boundary. Analogous results, under appropriate conditions, hold therefore for a random walk. These results are a straightforward consequence of what has been said, but the precise statements and the details deserve to be written out. I hope to be able to develop the issues that were outlined in this section in a future publication.

## References

[1] J. L. Doob, *Classical Potential Theory and its probabilistic counterpart.* Springer-Verlag, New York, 1984.

[2] M. Fukushima, *Dirichlet Forms and Markov Processes.* North-Holland, Amsterdam-New York, 1980.

[3] E. Fabes, N. Garofalo and S. Salsa, *A backward Harnack inequality and Fatou theorem for nonnegative solutions of parabolic equations.* Illinois J. Math. **30**(1986), 536–565.

[4] N. Th. Varopoulos, *Potential theory in conical domains (II).* Math. Proc. Camb. Phil. Soc. **129**(2000), 301–319.

[5] ———, *Potential theory in conical domains.* Math. Proc. Camb. Phil. Soc. **125**(1999), 335–384.

[6] E. B. Davies, *Gaussian upper bounds for the heat kernels of some second-order operators on Riemannian manifolds.* J. Funct. Anal. **80**(1988), 16–32.

https://doi.org/10.4153/CJM-2001-041-6 Published online by Cambridge University Press

[7] N. Th. Varopoulos, *Diffusion on Lie groups. III.* Canad. J. Math. (3) **48**(1996), 641–672.
[8] A. Ancona, *Principe de Harnack à la frontière et théorème de Fatou pour un opérateur elliptique dans un domaine lipschitzien.* Ann. Inst. Fourier (4) **28**(1978), 162–213.
[9] N. Th. Varopoulos, *Potential theory in conical domains (III).* To appear.
[10] A. Bensoussan, J. L. Lions and G. Paparicolaou, *Asymptotic analysis for periodic structures.* North-Holland, 1978.
[11] V. V. Jikov, S. M. Kozlov and O. A. Oleinik, *Homogenization of Differential Operators and Integral Functionals.* Springer-Verlag, Berlin, 1994.
[12] S. M. Kozlov, *The method of averaging and walks in inhomogeneous environment.* Uspekhi Math. Nauk. (2) **40**(1985), 61–120.
[13] N. Th. Varopoulos, *Analysis on Lie groups.* Rev. Mat. Iberoamericana (3) **12**(1996), 791–917.
[14] S. Mustapha and N. Th. Varopoulos, *Analysis on $p$-adic Lie groups.* To appear.
[15] L. Carleson, *On the existence of boundary values for harmonic functions in several variables.* Ark. Mat. **4**(1962), 339–393.
[16] C. E. Kenig, *Harmonic Analysis Techniques for Second Order Elliptic Boundary Value Problems.* CBMS Regional Conferences Series in Math. **83**, Amer. Math. Soc., 1994.
[17] B. Dahlberg, *Estimates of harmonic measure.* Arch. Rational Mech. Anal. **65**(1977), 275–288.
[18] S. Salsa, *Some properties of nonnegative solutions of parabolic differential operators.* Ann. Mat. Pura. Appl. **128**(1981), 193-206.
[19] J. L. Doob, *Stochastic Processes.* John Wiley, New York, 1953.
[20] W. Hebisch and L. Saloff-Coste, *Gaussian estimates for Markov chains and random walks on groups.* Ann. Probab. (2) **21**(1993), 673–709.
[21] G. Stampacchia, *Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus.* Ann. Inst. Fourier Grenoble **15**(1965), 189–258.
[22] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes.* Springer-Verlag, Berlin-New York, 1979.
[23] R. Kaufman and J. M. Wu, *Singular parabolic measures.* Compositio Math. (2) **40**(1980), 243–250.
[24] K. Ito and H. P. McKean Jr., *Diffusion Processes and their sample paths.* Springer-Verlag, 1974.
[25] Y. Heurteaux, *Solutions positives et mesure harmonique pour des opérateurs paraboliques dans des ouverts"lipschitziens".* Ann. Inst. Fourier (3) **41**(1991), 601–649.
[26] G. F. Lawler and T. W. Polaski, *Harnack inequalities and difference estimates for random walks with infinite range.* J. Theoret. Probab. **6**(1993), 781–802.
[27] A. Friedman, *Partial Differential Equations of Parabolic Type.* Prentice Hall, Englewood Cliffs, NJ, 1964.
[28] W. Feller, *An introduction to Probability Theory, Vols. I & II.* John Wiley & Sons.
[29] G. F. Lawler, *Estimates for differences and Harnack inequality for difference operators coming from random walks with symmetric, spatially inhomogeneous, increments.* Proc. London Math. Soc. (3) **63**(1991), 552–568.
[30] N. Dunford and J. T. Schwartz, *Linear operators.* Interscience.
[31] L. Hörmander, *The Analysis of Linear Partial Differential Operators, Vols. I & II.*
[32] N. Th. Varopoulos, *Marches aléatoires et diffusions dans les domaines lipschitziens.* C. R. Acad. Sci. Paris Sér. I Math. **330**(2000), 317–320.
[33] S. M. Kozlov, Math. U.S.S.R. Sbornik (4) **35**(1979), 481–498.
[34] M. Avellaneda and F. H. Lin, *Compactness methods in the theory of homogenization. II. Equations in nondivergence form.* Comm. Pure Appl. Math. **42**(1989), 139–172.
[35] R. M. Blumenthal and R. K. Getoor, *Markov Processes and Potential Theory.* Academic Press, New York-London, 1968.
[36] D. G. Aronson, *Bounds for the fundamental solution of a parabolic equation.* Bull. Amer. Math. Soc. **73**(1967), 890–896.
[37] H. G. Kao, Tridinger N.S. Duke Math. J. **91**(1998), 587–607.
[38] N. Krylov and M. Safonov, Math. USSR Izv. **16**(1981), 151–164.

*Institut Universitaire de France*
*Université Paris VI*
*Département de mathématiques*
*4, place Jussieu*
*75005 Paris*
*France*