Quasistationarity and Conditioned Markov Processes

Laird Arnault Breyer

B.A., B.Sc. (Hons.)

A thesis presented to the
Department of Mathematics,
The University of Queensland,

in fulfilment of the requirements
for the degree of
Doctor of Philosophy.

June, 1997 (revised January, 1998).

Statement of originality

The work presented in this thesis is, to the best of my knowledge and belief, original and my own work, except as acknowledged in the text. The material has not been submitted, either in whole or part, for a degree at this or any other university.

Laird Arnault Breyer

Abstract

Many of the stochastic processes used in Applied Probability are fundamentally transient in nature. This occurs in population growth models such as branching processes, where the number of individuals can become zero in a finite time; in certain chemical reactions, where one of the reactants is eventually depleted; in teletraffic and queueing models, with the phenomenon of multiple stable states, and in interacting particle systems, where clustering can happen.

In all these models, there is a random time ζ which can be used to describe a qualitative change in the behaviour of the representing process. Frequently, this arises as follows: the evolving system is represented by a Markov process X_t and the change occurs when X_t first enters an absorbing part of the state space. The entry time can be finite, and very long. A typical example is an epidemic model, for which the process representing the number of infected individuals never leaves zero if once it reaches that state. Yet the time until this occurs may well be finite and is often estimated in millions of years. In the meantime, the process appears stationary. We call this phenomenon *quasistationarity*.

In realistic models of Applied Probability, the evolving state probabilities of the process are rarely explicitly available. The analysis of the model therefore centers around certain well-chosen functionals of the process. The models described above have all been investigated, among other approaches, using *quasistationary distributions*; see Pollett (1993) for a review.

The functionals which have proved to be of greatest interest in this context are variously known as *limiting conditional distributions*, or *Yaglom limits*. These are probability measures κ satisfying the relation

$$\lim_{t\to\infty} \mathbb{P}(X_t \in A \,|\, \zeta > t) = \kappa(A).$$

and can be viewed as a generalization of the concept of limiting distribution associated with the positive recurrent case. Variations on these limits are possible, leading in particular to the *doubly limiting conditional distribution*. It has been realized recently (Pollett (1988), Jacka and Roberts (1995)) that the best way of

investigating this other distribution is via a certain conditioned process Y, related to X by

$$\lim_{s\to\infty} \mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n \,|\, \zeta > s) = \mathbb{P}(Y_{t_1} \in A_1, \ldots, Y_{t_n} \in A_n).$$

I shall refer to this as simply the *conditioned process*. The two objects κ and Y are fundamental, in that many other functionals of X and ζ can be derived from them. They form the core of what I call the "quasistationary tool box".

In this thesis, I consider the tool box from a very general point of view: continuous time Markov processes on a general state space. I prove solidarity results for the exponential decay of the transition function; this depends on an assumption of ν-irreducibility, and generalizes the work of Kingman (1963) and Tweedie (1974a) in discrete state space and discrete time respectively. The λ-classification of the process is deduced, and some basic properties of λ-invariant functions and measures are proved. There is presently no systematic account of this type in the literature. I then describe the classical tool box theorems in cases where X is classified as positive Λ-recurrent, and prove a "dual" version of a test for this, due to Tweedie (1974b). A counterexample is given to show that the existence of Yaglom limits requires more conditions in general than for discrete state space. A new interpretation of the solidarity results in terms of branching Markov processes over X is also offered.

The last two chapters deal respectively with the conditioned process Y, and the Yaglom limits, in terms of parabolic Martin boundary theory. This is a new approach to these problems, and explains many of the conditions such as bounded jumps, compactly supported initial distributions, and the Strong Ratio Limit Property etc. that previous authors have used. Consider the graph (t, X_t) of the process X in spacetime. By restricting the state space E if necessary, it is convenient to think about ζ as the exit time of X from E. Conditioning X to get Y has the effect of making the graph of Y "exit" the set $(0, +\infty) \times E$ in the hyperplane $t = +\infty$. The process (t, Y_t) can therefore be viewed as a version of (t, X_t), Doob-conditioned to exit the Martin boundary on the hyperplane $t = +\infty$. A rigorous interpretation of this idea, which is developed in the thesis,

requires an identification of that boundary point (if any) on the hyperplane that is "chosen" by the conditioning procedure used to define Y. If no suitable boundary point can be found, the conditioning procedure fails and Y does not exist. For technical reasons, the "backward graph" $(-t, X_t)$ is used instead of (t, X_t).

In the last chapter, the parabolic Martin boundary techniques are used, once more, but now to investigate the existence of a Yaglom limit κ. This is achieved by viewing κ as an entrance law for X, the limit of a suitable sequence of entrance laws. By switching to a process $\widehat{X}$ in duality with X, the problem is first reduced to a conditioning problem for $(-t, \widehat{X}_t)$ very similar to that described in the previous chapter. Once this is solved, the result is interpreted in terms of the original process X, and used to get the Yaglom limit itself. In this last step, a tightness condition is used, and this is related, via a Ray-Knight compactification, to the concepts of *asymptotic remoteness* and *asymptotic proximity* due to Ferrari et al. (1995) and Pakes (1995) respectively.

Acknowledgements

I thank Phil Pollett, my supervisor, for a few very enjoyable years. Especially fun were the "Stochastic Coffee" mornings, which set the scene for much useful interaction with visitors. His help with English prose is much appreciated, too.

Thanks also go to Gareth Roberts, for his hospitality during my visit to Cambridge, and the many helpful conversations we had. The beginnings of the idea of spacetime conditioning developed in Chapter 4 can be traced back to one of the late afternoon sessions in his office, Earl Grey in one hand, chalk in the other.

I am grateful to Terry Lyons for convincing me of the importance of the Martin boundary, following a talk I gave at Imperial College. When I asked him where best to learn the techniques, his magical reply was simply: "Doob's papers".

I thank Terence Chan for the interest he took in my work, both here and during my visit at Heriot-Watt. His invitation to give a talk at Edinburgh was much appreciated.

While I was learning the basics of Markov processes, the innumerable conversations with my fellow Ph.D. student, Andrew Hart, helped set my ideas into place. Regrettably, the papers we wrote on truncation approximations for quasistationary distributions did not fit the theme of this thesis. See Hart (1997).

I thank Phil's visitors and collaborators of the last two years: Nigel Bean, Peter Taylor, Ilze Ziedins, Mark Bebbington and Colm O'Cinneide. The conversations on Mathematics and other matters held at "Stochastic Coffee", besides generating research ideas, provided a much needed perspective.

Thanks go also to the crowd of other Ph.D. students at the University of Queensland, for making the atmosphere in the department so unique. I should like to thank especially Katrina Hibberd, Peter Watson, Adrian Koerber, Barbara Maenhaut, Catherine Belward, Elliot Tonkes, Regina Basman, Barbara Remenyi, Jane Sexton, Yuri Anissimov, Adelle Howse, Brenton Gray, Sarah Zarai and Scott McCue.

During my stay in Cambridge, I was made to feel very welcome, and for that

I wish to thank Richard Gibbens, Violet Lo, Rua Murray, David Rose, Stephen Turner, Martin Baxter, Anthony Quas, Sarah Shea-Simonds and Susan Pitts.

Last but certainly not least, I thank my family for the encouragement and help they gave me over the years. I feel very lucky to have them, and this thesis is dedicated to them.

De cette irrcusable exprience, s'cria-t-il, j'ai dduit l'existence de *l'Absolu!* Une substance commune toutes les crations, modifie par une force unique, telle est la position nette et claire du problme offert par l'absolu et qui m'a sembl *cherchable.*

- Balthazar Clas

Honor de Balzac, *La Recherche de l'absolu.*

Contents

Chapter 1

Introduction

1.1 Quasistationary Tool Box

Markov processes are often used as models in Applied Probability, and the single most widely used result is the ergodic theorem. As a tool, it allows a simple asymptotic analysis of the transition probabilities, which are often unavailable explicitly. Its scope is limited however to recurrent processes.

Transient Markov processes occur in applications just as frequently, and their transition probabilities exhibit a wide range of qualitative behaviour. Consider such a process (X_t), evolving on some state space E (see Section 2.1 for the standard formalism and assumptions, and note especially the definition of $\langle \mu, f \rangle$ therein). At the most basic level, we have $\lim_{t\to\infty} \mathbb{P}(X_t \in A) = 0$ for a class of subsets A exhausting E. Beyond that, general results are scarce.

If the process leaves the state space a.s. in a finite time, it may nevertheless appear stationary over a long period of time. This occurs, for example, in epidemic models (see Nåsell (1995)). *Quasistationary distributions* (qsds) have been used to account for this type of behaviour (the phenomenon itself is known as *quasistationarity*). Their role is similar to that of stationary distributions in ergodic theory. However, there can be many distinct quasistationary distributions for the process.

Just as finding the stationary distribution of a recurrent Markov process is

only the beginning, rather than the end, of the analysis, so too are there many questions to be answered in a "quasistationary analysis", besides the existence and identification of qsds.

I call the body of results related to such an analysis the "quasistationary tool box". Besides procedures for identifying qsds, it also contains theorems which ensure the convergence of (conditional) transition probabilities. Like their counterparts in Markovian ergodic theory, these "tools" are used to gain an understanding of the Markov process which models the physical system.

The theme of the work before you is the quasistationary tool box as it applies to general Markov processes (continuous time and state space). Such an analysis has not been done before at this level of generality. Most work on quasistationary distributions (with some significant exceptions) has hitherto been done in the setting of countable state space Markov chains. Besides its unifying virtue, there is another, more compelling reason for the approach taken in this thesis; I shall discuss the main theorems of the quasistationary tool box from the point of view of spacetime processes and associated Martin boundaries. Even when the original state space is countable, this requires a more general formalism (spacetime is always continuous). The parabolic Martin boundary approach to quasistationarity is entirely new.

The plan of this introduction is as follows: I first discuss the main tool box components below, then give an extensive outline of the thesis in the next section. Some open questions and research ideas will appear at the end of each chapter.

In the typical setting in which quasistationarity is investigated, there is an absorbing state labelled 0 which is eventually entered from any starting position. Here what we mean is that the absorption time,

$$T_0 = \inf\{t > 0 : X_t = 0\},$$

is a.s. finite.

In applications, the process X_t might represent the size of a population, in which case T_0 can be interpreted as an extinction time. Thus we assume that extinction occurs in a finite, but random, time. Other possibilities include chemical

reactions, when T_0 represents the first time one of the reactants is depleted, or queueing models, with T_0 marking the transition from one stable state to another. For a specific example, take the simple epidemic model of Ridler-Rowe (1967). In this model, individuals within the population follow a cycle wherein they first become susceptible to the disease, then infected, and finally recover (or die). Figure 1 shows a simulation of the sample path of a two-dimensional process $X_t = (X_t^1, X_t^2)$, where X_t^1 is the number of susceptibles and X_t^2 the number of infected by time t. For this process, the X^1-axis is an absorbing set, and it can be shown that X_t must eventually enter it.

The important feature of this model from our point of view is the long time till extinction. An observer without the mathematical model before him might be forgiven for thinking that the process were stationary, on the basis of a very long (but not long enough!) time spent observing it. Other epidemic models exhibit the same phenomenon (see for example Nåsell (1995)). How should one model this quasistationarity quantitatively, given the law of the process X? This question is what the tool box is designed to address.

The first step towards a description of the phenomenon was taken by Yaglom (1947). He considered a branching process X_n in discrete time, with a.s. extinction. What he showed was the existence of a probability law κ on $\{1, 2, 3, \dots\}$ describing the number of individuals in the population at some distant time in the future, *given that extinction hasn't occurred in the meantime.*

$$\lim_{n\to\infty} \mathbb{P}(X_n = j \,|\, T_0 > n) = \kappa_j, \quad j = 1, 2, \dots. \tag{1.1}$$

If the extinction time can be shown to be a.s. very large, Yaglom's conditioning represents a valuable addition to the prior information about the process, given the validity of the model as a whole. In the epidemic example of Figure 1, the analogue of (1.1) can be computed; an approximate density of κ is shown in Figure 2. Efficient methods for computing the qsd have been discussed in Pollett and Stewart (1994).

Clearly, the law κ is only useful as an approximation of the medium to long term behaviour of X_t provided absorption occurs after a long time. How long

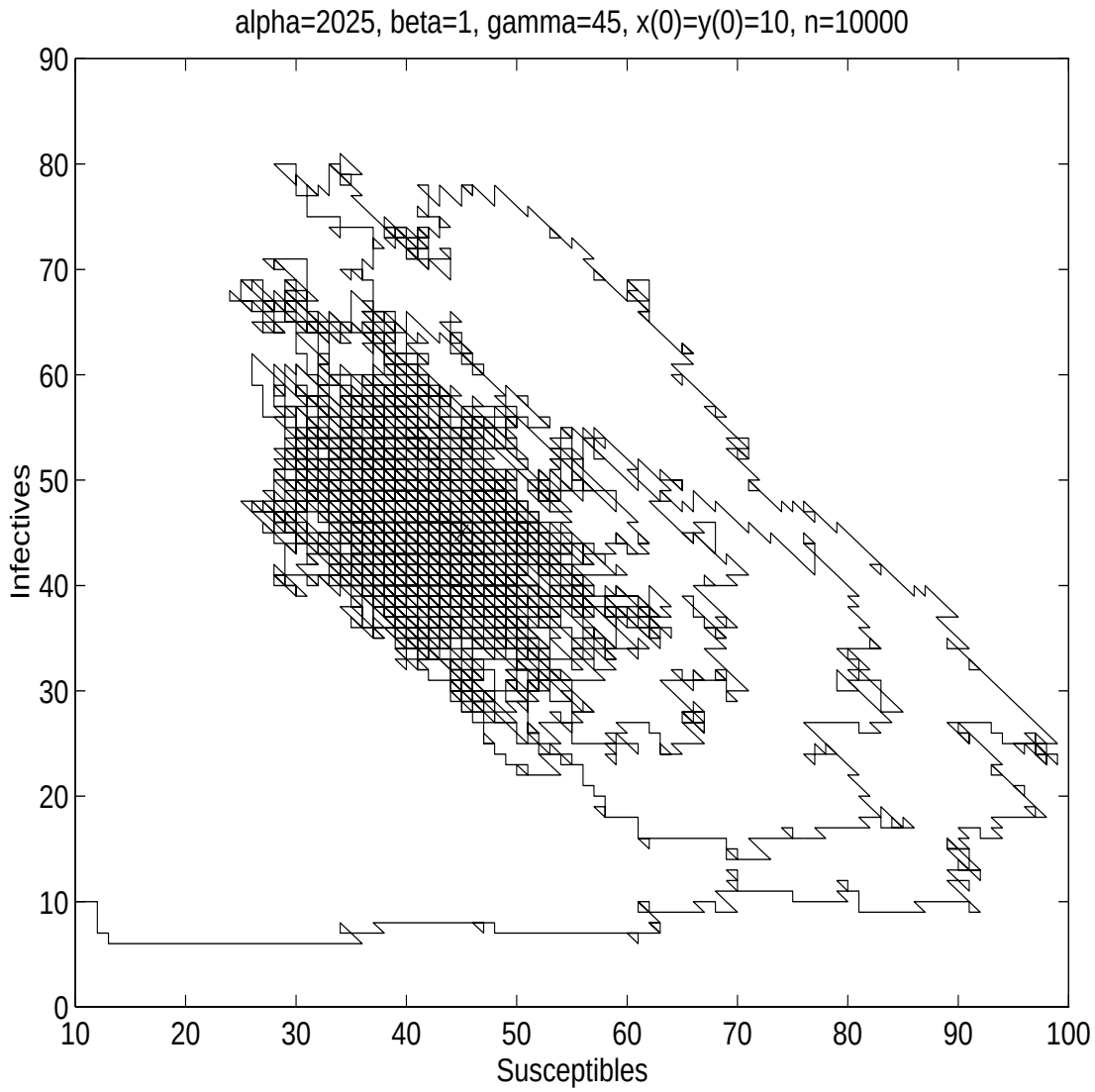

Figure 1. Simulation of an epidemic process (Ridler-Rowe (1967)) starting at $(10, 10)$.

X^1 = Number of susceptibles	$(X^1, X^1) \to (X^1 + 1, X^2)$ at rate α
X^2 = Number of infectives	$(X^1, X^2) \to (X^1 - 1, X^2 + 1)$ at rate $\beta X^1 X^2$
n = Number of transitions	$(X^1, X^2) \to (X^1, X^2 - 1)$ at rate γX^1

is enough could be, for a general process, the source of considerable argument. The simulation depicted in Figure 1 suggests that the process X naturally tends towards a statistical equilibrium, and only gets absorbed when a sufficiently large deviation of the sample path occurs. Chan (1997) has discussed this type of behaviour for density-dependent birth-death processes. In the case of contemporary epidemics models, the typical time till absorption can last (theoretically) for sev-

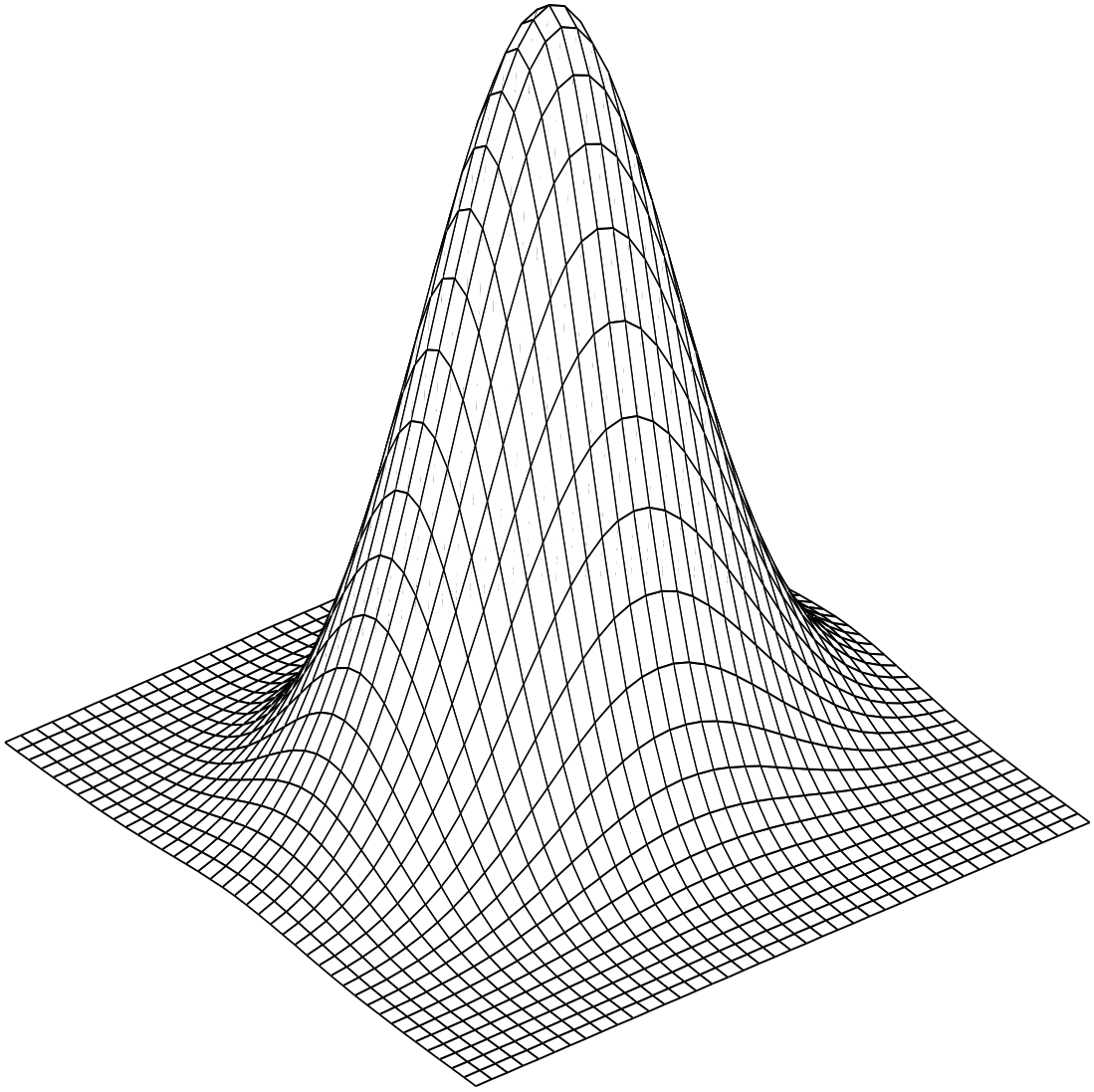

Figure 2. A quasistationary distribution.

eral times the age of the universe, and the measure κ offers a useful description of reality. In this work, I shall bypass this all-important question completely, concentrating instead on the *existence* of various limits such as (1.1) for a class of Markov processes. I shall refer to (1.1) as a *Yaglom limit*.

There is some debate in the literature on the appropriateness of this terminology. Other names for (1.1) include *quasistationary distribution* (we shall use this term for a different concept) and *limiting conditional distribution*, sometimes even ν-LCD, where ν is the distribution of X_0. This last term is favoured by Pakes (1995) on the grounds that Yaglom limits should refer only to branching processes. The terminology I shall use follows Ferrari et al. (1995). As we shall

see later, there is a close connection between Yaglom limits for general Markov processes and certain associated branching processes, and this is another reason why I have chosen this terminology.

The form of (1.1) is vaguely reminiscent of a limit theorem in the ergodic theory of Markov processes. Indeed, were the process X positive recurrent, with state space $E\backslash\{0\}$, then $T_0 = \infty$ a.s. , and (1.1) would state that the transition law of X converges to the (unique, stationary) limiting distribution. With this contrived analogy in mind, one can ask about the existence of a (unique, *quasistationary*) Yaglom limit, thereby studying a *Bayesian form* of ergodicity. Here a law μ is called *quasistationary* provided the process, started with μ, stays with μ given that it doesn't absorb at 0:

$$\mathbb{P}_\mu(X_t \in A \,|\, T_0 > t) = \mu(A), \quad A \subseteq E.$$

I shall also use the term "quasistationary distribution" (abbreviated qsd) to refer to such a probability measure. There can be many quasistationary laws (van Doorn (1991)), and a Yaglom limit is always quasistationary. However, there can be many Yaglom limits, depending on the entrance law of X. This is in stark contrast to the positive recurrent case alluded to earlier.

Another type of limit of interest in studying the quasistationary behaviour of a Markov process X is the double limit

$$\lim_{s\to\infty}\lim_{t\to\infty} \mathbb{P}(X_s \in A \,|\, T_0 > t + s) = \pi(A), \tag{1.2}$$

where π represents some probability measure on E and A is some subset of the state space, typically not containing 0. This is known sometimes as a type II limiting conditional distribution, whence the Yaglom limit is referred to as type I (Flashpohler (1974)). There will be little to say about the limit (1.2), much less about its use in applications. This is mainly due to the fact that (1.2) only ever appears when the process X is positive Λ_*-recurrent; see Theorem 3.15 of Chapter 3. Of more interest will be the question of whether we can define a new Markov process Y, by stipulating that its finite dimensional distributions satisfy

$$\mathbb{P}(Y_{t_1} \in A_1, \ldots, Y_{t_n} \in A_n) = \lim_{r\to\infty} \mathbb{P}(X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n \,|\, T_0 > r). \tag{1.3}$$

We will call the process Y a conditioned process. It describes the behaviour of X given that we *know* that absorption won't occur for a long time. It appears in various contexts, and was first considered by Darroch and Seneta (1967) in the context of quasitationary limit theorems. In relation to (1.2), note that π is simply the limiting law of the process Y, and this necessitates that Y be positive recurrent. The validity of (1.3) was conjectured by Pollett (1988).

Ratio limit theorems are a type of limit commonly investigated. Papangelou (1967) discusses these for Λ_*-recurrent processes (these processes are defined in Chapter 3, Section 3.2) in discrete time. The ratio of the absorption distributions from two different starting points,

$$\lim_{t\to\infty} \frac{\mathbb{P}_x(T_0 > t)}{\mathbb{P}_y(T_0 > t)} = \frac{\varphi(x)}{\varphi(y)}, \tag{1.4}$$

appears naturally when investigating quasistationarity (for the existence and properties of the function φ above, see Chapter 3, Theorem 3.15). We shall see in Chapter 4 that the limit (1.4) exists as a by-product of the existence of the process Y satisfying (1.3). Thus there is little need to analyse (1.4) on its own. If the Yaglom limit from x and y also exists, we get a form of the Strong Ratio Limit Property (compare with the classical form, Chapter 4, last paragraph of Section 4.1):

$$\lim_{t\to\infty} \frac{\mathbb{P}_x(f(X_t), \zeta > t)}{\mathbb{P}_y(g(X_t), \zeta > t)} = \lim_{t\to\infty} \frac{\mathbb{P}_x(f(X_t) \,|\, \zeta > t)\mathbb{P}_x(\zeta > t)}{\mathbb{P}_y(g(X_t) \,|\, \zeta > t)\mathbb{P}_y(\zeta > t)} = \frac{\langle \kappa, f\rangle \varphi(x)}{\langle \kappa, g\rangle \varphi(y)}.$$

I shall refer to the limits in (1.1) - (1.4) generically as *quasistationary limits*. Once these have been identified, a fundamental problem remains. How can these limits be computed? Here, one turns to equations involving the generator of the process (definition in Section 2.5). Much work has been done in this direction these last ten years. See, for example, papers by Pollett (1988) and Nair and Pollett (1993).

In this thesis, I shall always illustrate my results with reference to only two types of Markov processes. The first is that of diffusions on a subset of $\mathbb{R}^d$ (defined in Example 2.1). The second is that of continuous time Markov chains

on a countable state space (defined in Example 2.2). The literature related to quasistationary distributions is much greater in the case of Markov chains than diffusions. Pollett (1993) has given an overview, highlighting the applications to biological modelling. Pakes (1995) also has a good review in his introduction. In contrast, the literature relating to diffusions is much smaller, with Collet et al. (1995), Jacka and Roberts (1997), Pinsky (1985) being the main papers in the field so far. This discrepancy is of course due to the wider variety of Markov chain models (as opposed to diffusions) in use by investigators.

Nevertheless, the relations between diffusions and Markov chains are many, and I will affirm that a knowledge of quasistationary tool box theorems requires both cases to be properly understood. For example, the Feller-McKean chain, whose states are all instantaneous (see Rogers and Williams (1994)), may be viewed as a Brownian motion on a suitable compactification of the state space. This is a particular case of the Ray-Knight compactification theory, which turns the Markov chain jump from any instantaneous state into a continuous motion on the compactification.

Diffusion approximations of birth-death chains were used in the calculation of quasistationary distributions already by Seneta (1966); for this example, it is surprising that the corresponding tool box theorems were only proven much later (Collet et al. (1995), Jacka and Roberts (1997)).

Finally, potential-theoretic tools are much more accessible for diffusions than for Markov chains. The former therefore represent an ideal first choice for the application of potential theory towards the proof of tool box theorems. Here, I have in mind results in Chapters 4 and 5 of this thesis.

1.2 Outline of the Thesis

In this section, I shall outline in some detail the results contained in the thesis. My aim at this stage is to present an overview, with room for discussion and conjecture.

The model I shall use throughout is that of a (continuous time) Markov process X, with a.s. finite lifetime ζ, evolving on a metric state space E. This means that X leaves E for the first and last occasion at the random time ζ. The law of X when started in state $x \in E$ is written $\mathbb{P}_x$. There are some advantages of this formalism.

Since the exact manner of death is not prescribed, the random time ζ may equal T_0, the first hitting time of the state (or more generally, the set) 0. In this case, all equations and results in the thesis may be reinterpreted, replacing ζ by T_0, and the theory then applies to any Markov process, irrespective of its lifetime. This is the conventional form of presenting quasistationary limit theorems. For example, if X is a positive recurrent Markov process, and $\zeta = T_0$, then Yaglom limits can help in understanding the excursions away from 0.

A seemingly quite different class of limit theorems has begun to appear only recently (Schrijner (1996), Kijima et al. (1997) and Hart (1997)). Let X be a transient process, and suppose that we denote by L_0 the *last* time that the process ever visits state 0. The random variable L_0 is not a stopping time. Nevertheless, one can ask if the analogous Yaglom limit

$$\lim_{t\to\infty} \mathbb{P}_\nu(f(X_t) \,|\, L_0 > t) = \langle \kappa, f \rangle \tag{1.5}$$

exists. For example, X may be a birth death process on $\mathbb{Z}$, with positive drift. It is known that this process converges to $+\infty$ (in an infinite time). However, the time L_0 is certainly finite, and the existence of the limit (1.5) is not trivial. Pollett (personal communication), see also Hart (1997), has an example of such a process, which exhibits a form of quasistationarity. The process lingers for a very long time around a point near the state zero, and finally drifts off to $+\infty$, never to return.

It is possible to fit this type of study into the framework considered in this thesis. Indeed, it was shown by Meyer et al. (1972) that if $c(x) = \mathbb{P}_x(L_0 > 0)$,

then the process Y constructed by killing X at time L_0, namely

$$Y_t = \begin{cases} X_t & \text{if } L_0 > t, \\ \partial & \text{if } L_0 \leq t, \end{cases}$$

where ∂ is some cemetery state added to E, is under the probability measure $\mathbb{P}_x$, when $c(x) > 0$, a time-homogeneous honest Markov process with state space $\{x : c(x) > 0\} \cup \{\partial\}$, and semigroup (Q_t) related to the semigroup (P_t) of X by the formula $Q_t(x, dy) = P_t(x, dy)c(y)/c(x)$. Here the lifetime of the process Y can be viewed as the hitting time of the cemetery state ∂, as this state is absorbing. Thus to study (1.5), one has to formally replace X_t by Y_t, and L_0 by $\zeta = \zeta_Y$. More generally, one can consider any cooptional time L (Dellacherie and Meyer (1992)), and killing by way of multiplicative functionals.

Chapter 2 begins with some standard definitions and results about Markov processes. One aim here is to fix a notation to be used in the rest of the thesis. Noteworthy is the use of $\mathbb{P}$ instead of the usual $\mathbb{E}$ to symbolize expectations. This is designed to make it easier to keep track of the underlying probability measure, as this will frequently vary. Thus when the measure is $\mathbb{Q}$, not $\mathbb{P}$, the associated expectation is $\mathbb{Q}$, as opposed to the more awkward $\mathbb{E}_{\mathbb{Q}}$ used by some authors. I also describe three examples of processes: diffusions (generated by an elliptic differential operator), Markov chains (generated by a q-matrix), and spacetime processes $\overline{X}_r = (t - r, X_r)$, which correspond to the graph of X_t if the time axis is inverted. For this last type of process, X is typically either a diffusion or a Markov chain. $\overline{X}$ is needed in Chapter 4, and was first studied in depth by Doob (1955).

I shall then recall the definition of excessive functions (in Section 2.2, Chapter 2) of a Markov process, and outline their use in conditioning processes on the Martin boundary. This follows classical theory due to Doob (1957), Kunita and Watanabe (1965) and Meyer (1968). As these ideas are pivotal to Chapter 4, it pays to recall the theory for reference. Noteworthy here is the concept of a

harmonic function, that is a function h such that

$$\mathbb{P}_x(h(X_\sigma), \zeta > \sigma) = h(x), \quad x \in E,$$

where $\sigma = \inf\{t > 0 : X_t \notin K\}$ and K ranges over all compact subsets of E. When considering the spacetime process $\overline{X}$, a harmonic function (for $\overline{X}$) is also called *parabolic* (for X).

I finish the chapter with a characterization of harmonic functions in terms of the local martingale generator $\mathfrak{A}$. When X is a diffusion, $\mathfrak{A} = L$, whereas if X is a Markov chain, $\mathfrak{A}$ is the q-matrix. For the spacetime process $\overline{X}$, the generator is $\bar{\mathfrak{A}} = \mathfrak{A} - \partial/\partial t$.

By Theorem 2.9 of Chapter 2, a locally bounded function h is harmonic if and only if $\mathfrak{A}h = 0$. This is part of the folklore of Markov processes, and can be found routinely in textbooks for particular processes (see Bass (1995) for Brownian motion, Rogers and Williams (1987) for Markov chains, and Dynkin (1965) for Feller-Dynkin processes). Unfortunately, there does not appear to be a simple direct proof for general Markov processes (particularly the spacetime processes considered in Chapter 4) in the literature. I have therefore provided one (Theorem 2.9). By the previous remarks, a parabolic function is a solution to the Kolmogorov backward equation $(\partial/\partial t)h(t,x) = \mathfrak{A}h(t,x)$.

It should be emphasized that I only ever consider the *minimal* process X associated with the generator $\mathfrak{A}$. This is quite natural when the lifetime coincides with the first hitting time of 0.

Chapter 3 begins with a study of decay parameters and the existence of λ-excessive functions and measures. A λ-excessive function is by definition a function which is excessive for the semigroup $e^{-\lambda t}P_t$. Here, λ can be either positive or negative. A similar definition can be given for λ-excessive measures. If a function h is λ-harmonic for X, then the spacetime function $(t,x) \mapsto e^{\lambda t}h(x)$ is parabolic, and conversely.

Functions which are λ-harmonic arise naturally in the study of the conditioned process defined by (1.3). See for example the first generic study of (1.3) for

Markov chains, by Jacka and Roberts (1995). The case of λ-invariant measures is more ancient. It was noted by Vere-Jones (1969) that the probability measure κ arising in the Yaglom limit (1.1) must be λ-invariant for some $\lambda \leq 0$. Later, it was shown by Nair and Pollett (1993) that a probability measure μ is a quasistationary distribution according to (1.1) if and only if it is λ-invariant for some λ.

A quasistationary distribution need not be equal to κ, and there may be many values of λ associated with qsds. However, the value of λ associated with κ (if the Yaglom limit exists) is always the smallest value associated with *any* qsd (this follows from a study of decay parameters, and results of Pollett (1988)). Indeed, if we start the process X with initial distribution μ, then its lifetime is exponential with mean $|\lambda|^{-1}$. Thus κ has the smallest absorption time. Just precisely why this is so is not entirely clear. The reason lies probably somewhere within Martin boundary theory. I shall come back to this remark later, when I discuss open problems after Chapter 4.

In the first section of Chapter 3, I shall prove some solidarity results (Theorem 3.3) for the kernels $V_\lambda f(x) = \mathbb{P}_x \int_0^\zeta e^{-\lambda t} f(X_t) dt$ associated with an irreducible Markov process; here $\lambda \leq 0$, otherwise (V_λ) would just be resolvent of X: there exists a parameter Λ_* (Kingman's decay parameter - see Kingman (1963)) and a negligible set $N \subset E$ such that $V_\lambda(x, dy)$ is transient for $\lambda > \Lambda_*$, and only takes the values $+\infty$ or 0 for $\lambda < \Lambda_*$. This forms the basis for the Λ_*-recurrence/transience classification of X to be described in Section 2. A proof of Theorem 2.9 for countable state processes is due to Vere-Jones (1962, discrete time), Kingman (1963, continuous time); for general state spaces it was shown by Tweedie (1974a, discrete time). None of these proofs is entirely satisfactory in continuous time with a general state space. See Chapter 3 for details.

Section 2 of Chapter 3 introduces another decay parameter, denoted Λ_1. It is characterized as the smallest (negative) λ for which $\mathbb{P}(\zeta > t) = o(e^{\lambda t})$ as $t \to \infty$. When the Yaglom limit exists, then κ is associated with Λ_1. Various characterizations and properties of Λ_1 are studied. Perhaps the most interesting fact stated here is that, in order to have a continuum of qsds (associated with different values

of λ), it is necessary to have a certain form of "asymptotic remoteness". This concept was used by Ferrari et al. (1995) to prove the existence of quasistationary distributions; it was first observed by Pakes (1995) that qsds also exist without this condition. Section 1 of Chapter 5 discusses these problems further.

Section 3 summarizes known results about positive Λ_*-recurrence. Processes which are of this type include Markov chains on a finite state space and uniformly elliptic diffusions on bounded open sets. The theorems here are mainly due to Darroch and Seneta (1967), and Tweedie (1974a), Tuominen and Tweedie (1979). Most useful for testing Λ_*-recurrence is Tweedie's Test (Proposition 3.11), which we "dualize" in Proposition 3.12. Theorem 3.15 states that the existence of *all* the limits (1.1)-(1.4) is equivalent to a strong form of positive Λ_*-recurrence, namely a simultaneous skeleton irreducibility together with a finite Λ_*-invariant measure. The simultaneous skeleton irreducibility does hold for Markov chains and diffusions, but we show by a counterexample that this need not be so for every Markov process. In particular, the counterexample fails the conditions of Theorem 3.15, even though it is positive Λ_*-recurrent and satisfies a natural (but weak) form of irreducibility.

In Theorem 3.15, the existence of the type II limit (1.2) is crucial. Indeed, there are many examples of processes X for which the limits (1.1), (1.3) and (1.4) exist, but X is not even Λ_*-recurrent. Consequently, (1.2) does not exist (it is identically zero).

Sections 4 and 5 of Chapter 3 are an attempt to interpret probabilistically the preceding material. Given X and γ, a branching Markov process B is constructed as follows: place n particles in E, each evolving independently according to the law of X (in particular, each has its own lifetime). At the jump times of an independent Poisson process with rate γ, choose a surviving particle at random and replace it with two independent particles placed at the same location. If γ is large enough, the expected number of particles alive in any one set will grow without bound; if γ is too small, the particles die faster than they can reproduce. The critical value is $\gamma = -\Lambda_*$ (Theorem 3.18). A quasistationary distribution

μ has an equally simple interpretation. If we let $Z_t(A)$ denote the number of particles alive within the set A at time t, then the measure μ is an equilibrium distribution for the measure valued process Z.

An extensive study of the process Z can be found in Asmussen and Hering (1983). Note however that they assume that X is positive Λ_*-recurrent, and even a little more. When B is supercritical, they showed that

$$\lim_{t\to\infty} \frac{Z_t(A)}{Z_t(B)} \overset{\text{a.s.}}{=} \frac{\kappa(A)}{\kappa(B)} \quad \text{on } \{Z_t(B) > 0 \ \forall t\},$$

where κ is the unique quasistationary distribution and Yaglom limit. It is also true that $\lim_{t\to\infty} \mathbb{Q}_x[Z_t(A)]/\mathbb{Q}_x[Z_t(B)] = \kappa(A)/\kappa(B)$, where $\mathbb{Q}_x$ is the law of B started with one particle at $x \in E$, regardless of the value of the branching rate. The Yaglom limit (1.1) can be viewed as a special case of this, occurring when the branching rate is zero. Kesten (1978) studied the process B in the case that X is a Brownian motion on $(0, +\infty)$ with constant negative drift; this process X is Λ_*-transient.

I shall now outline Chapter 4. The main problem considered here is whether or not (1.3) may be used to define a Markov process Y. If the finite dimensional distributions (1.3) do indeed converge, then we will say that the (process level) conditioning problem has a solution. Ideally, Y will be a time homogeneous Markov process; this is the case for finite state space Markov chains (Darroch and Seneta (1967)), birth death processes (Jacka and Roberts (1995), Jacka et al. (1997) and Schrijner and van Doorn (1997)), one dimensional diffusions (Collet et al. (1995), McKean (1963) and Jacka and Roberts (1997)), for example. At the other extreme, the conditioning problem does not have a solution; an example was found by Jacka and Roberts (1995), and I shall have more to say about this at the end of Chapter 4.

To understand the appearance of spacetime Martin boundaries (for definitions, see Sections 2.4 and 4.1), we need to re-express the right side of (1.3) using the

Markov property:

$$\mathbb{P}_x\big[X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n \big| \zeta > r\big] = \mathbb{P}_x\big[X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n, h_{-r}(t_n, X_{t_n})\big],$$

where $h_{-r}(t,y) = \mathbb{P}_y(\zeta > r - t)/\mathbb{P}_x(\zeta > r)$.

Doob (1955) was the first to show that $(t,y) \mapsto \mathbb{P}_y(\zeta > t)$ is parabolic. Indeed, it is easy to see that this function is in fact invariant for $\overline{X}_u = (t_0 - u, X_u)$ on the set $(0, +\infty) \times E$ (see Chapter 2). Consequently, the function $h_{-r}(t,y)$ is parabolic on $(-r, 0) \times E$, and moreover it is always normalized so that $h_{-r}(0,x) = 1$. A fundamental result in Martin boundary theory (see Chapter 2) now asserts the existence of a probability measure μ_{-r} representing the function h_{-r} on the boundary. By Lemma 4.5, this measure must not charge that part of the boundary "over" $(-r, 0) \times E$.

We can now study the weak limit points of the family (μ_{-r}) on the Martin compactification. Clearly, if $\mu_{-r_n} \Rightarrow \mu$, then μ only charges that part of the boundary directly "over" $\{-\infty\} \times E$. Corresponding to μ is a function $h(t,y) = \lim_{n\to\infty} h_{-r_n}(t,y)$ (Lemma 4.6); and this will be independent of the sequence $r_n \to +\infty$ if and only if μ is. Once the uniqueness of μ is established, the parabolic Harnack inequality is then used to prove the existence of a law $\mathbb{Q}$ solving (1.3), and consequently the existence of a Markov process Y (see Theorem 4.21, and compare with Jacka and Roberts (1995)).

Up to this point, the assumptions needed will be minimal: Assumption II (which includes a form of irreducibility) and the parabolic Harnack inequality. These assumptions certainly hold for diffusions and Markov chains (I give a proof of the parabolic Harnack inequality in the case of chains - I could not find one in the literature). It then remains only to prove that $\mu_{-r} \Rightarrow \mu$, or equivalently that the family (h_{-r}) has a unique limit point.

For this, two independent assumptions are needed: Assumption III (which requires the jumps of X to be bounded) guarantees the *existence* of a nonzero parabolic function with representing measure supported "over" $\{-\infty\} \times E$. Assumption IV is used to prove uniqueness, through the concept of a *cemetery*

neighbourhood. This is a subset N of E with the property that the sample path X_t spends the last segment of its life there. According to Lemma 4.14, any limit measure μ must be supported by the "projection" of N on $\{-\infty\} \times E$. The expectation is that smallness of N translates into few parabolic limit functions $h = \lim_{n\to\infty} h_{-r_n}$ (see Proposition 4.24 and Theorem 4.27). A range of examples is given to show that one can often deduce that there can be only one such function $h > 0$. It is worth pointing out here what goes wrong with the counterexample of Jacka and Roberts (1995) alluded to earlier. The cemetery neighbourhood consists of a single state, so by Theorem 4.27 there is at most one nonzero limit function h. The problem is that the process does not have bounded jumps, and there does not *exist* a suitable limiting parabolic function h. See the last example of Chapter 4 for details.

Chapter 5 builds in an essential way on the techniques of Chapter 4. We study the convergence, as $r \to \infty$, of the measures $\langle \nu_r, f\rangle = \mathbb{P}_\nu(f(X_r) \,|\, \zeta > r)$.

In Section 1, we obtain probabilistic conditions for the tightness of the family (ν_t). This is done in terms of an extended process X' which exists on a compactification of the original state space E. When restricted to E, this process coincides with X. Terms such as *asymptotic remoteness* and *asymptotic proximity* are defined using X'. These concepts were introduced by Ferrari et al. (1995) and Pakes (1995) respectively, but the interpretation using X' is new, and unifies them.

In Section 2, we return to the parabolic Martin boundary, and begin a study, not of (ν_t) itself, but of the associated entrance laws $\eta_t^r = \nu_r P_t$. Using a slight modification of Assumption II (which we call Assumption II(bis)), we define a dual process $\widehat{X}$, whose semigroup satisfies $\widehat{P}_t(x, dy) = p_t(y, x)m(dy)$ if $P_t(x, dy) = p_t(x, y)m(dy)$ is the semigroup of X. The entrance laws η_t^r have a density $\widehat{k}^r(t, y)$ with respect to the measure m, and this density is parabolic for $\widehat{X}$.

It is shown in the remainder of the chapter that the existence of the Yaglom limit is equivalent to the convergence, as $r \to \infty$, of the parabolic functions $(\widehat{k}^r)$. This is entirely analogous to the situation in Chapter 3, where the existence of

the conditioned process Y is shown to be equivalent to the convergence of the functions h_s as $s \to -\infty$. Once again, every limiting function $\widehat{k}(t, y)$ will be representable on the hyperplane $\{-\infty\} \times E$ of the parabolic Martin boundary associated with $\widehat{X}$ (Proposition 5.11); the uniqueness of the limit uses the same techniques as in Chapter 4, namely parabolic Harnack estimates and cemetery neighbourhoods. Finally, the existence of a nonzero limit function $\widehat{k}$ is shown by the tightness criterion of Section 1.

Chapter 2

Markov Processes and Martin Boundaries

This chapter introduces the notation and terminology used in the rest of this work. It introduces the formalism of Markov processes, and describes basic results of Martin boundary theory. Standard references for the claims below are Rogers and Williams (1987, 1994), Dellacherie and Meyer (1988, 1992), Chung (1982), Revuz and Yor (1991), and Sharpe (1988). See also Ancona (1990) for spacetime processes. The last section deals with local martingale generators, and uses these to characterize harmonic functions. Many different types of generators are used in Markov process theory, and while the characterization we give is familiar, it cannot be found in standard references.

2.1 Markov Processes

Let E be a locally compact metric space. The σ-algebra of Borel sets is denoted $\mathcal{E}$. A process X with state space E is simply a collection of random variables $(X_t : t \geq 0)$ with values in E. More formally, one defines a measurable space $(\Omega, \mathcal{F})$ and requires $X_t : \Omega \to E$ to be an $\mathcal{E}/\mathcal{F}$ measurable mapping for each $t \in \mathbb{R}_+ = \{t \geq 0\}$. Probability measures do not enter the framework yet.

Among suitable event spaces $(\Omega, \mathcal{F})$, we single out a particular one, the *space*

of (càdlàg) paths in E, which is defined as follows: Let Ω denote the set of curves (with lifetime) $\omega : [0, \zeta(\omega)) \to E$ which are right continuous and left limited in E, for all $t \in [0, \zeta(\omega))$. Here, $\zeta(\omega)$ is called the lifetime of the path ω. It is important that the left limit at the lifetime

$$X_{\zeta-}(\omega) = \lim_{t \uparrow \zeta(\omega)} \omega(t),$$

may not exist in the topology of E. There is more about this below. We emphasize that this means the left limit $\lim_{t \to \zeta-} \omega(t)$ may *not* exist (in the topology of E).

We define the *coordinate process* by $X_t(\omega) = \omega(t)$, and this will be a stochastic process on E provided we set $\mathcal{F} = \sigma(X_t, \zeta > t : t \geq 0)$, that is, $\mathcal{F}$ is the σ-algebra generated by all sets of the form $\{X_t \in A, \zeta > t\}$, where $A \subseteq \mathcal{E}$. Under this definition, the lifetime ζ is automatically a random variable on $(\Omega, \mathcal{F})$, with values in $[0, +\infty]$. The natural filtration of X is the collection $(\mathcal{F}_t^0)$, where $\mathcal{F}_t^0 = \sigma(X_s, \zeta > s : 0 \leq s \leq t)$, that is, $\mathcal{F}_t^0$ is the history of the process X up to and including time t.

Before going on, we need to describe some measure theoretic notation used throughout the rest of this work. Let μ be a measure, and f a function. Unless otherwise specified, we shall always assume that a function is measurable with respect to the structure defined on its domain. The integral of f with respect to μ is written variously as

$$\int f d\mu = \int f(x)\mu(dx) = \langle \mu, f \rangle = \mu(f).$$

If $P(x, dy)$ is a kernel, we shall write $Pf(x) = \int P(x, dy) f(y) = P(x, f)$ when it acts on functions, and $\mu P(dy) = \int \mu(dx) P(x, dy)$ when it acts on measures. Recall that the composition of *positive* kernels is associative.

Suppose now that $(P_t : t \geq 0)$ is a semigroup of positive, sub–Markovian kernels on E, that is $P_{t+s} = P_t P_s$ and $P_t(\cdot, E) \leq P_0(\cdot, E) = 1$ for all $t, s \geq 0$. We shall always assume that P_t maps Borel functions into Borel functions. If $\mathbb{P}$ is a measure on $(\Omega, \mathcal{F})$, we say that a process X (defined on Ω) is *Markovian under* $\mathbb{P}$ *with semigroup* (P_t) if, whenever f is a positive Borel function,

$$\mathbb{P}(f(X_t), \zeta > t | \mathcal{F}_s^0) = P_{t-s}(X_s, f) \quad \text{on } \{\zeta > s\}. \tag{2.1}$$

Note our convention: we use $\mathbb{P}$ (not $\mathbb{E}$) to denote expectation. We shall be considering the same coordinate process under many different measures $\mathbb{P}, \mathbb{Q}, \ldots$, and this notation makes it easier to keep track. The equality in (2.1) holds of course only a.s. $\mathbb{P}$, but we will usually not mention this further.

If, under $\mathbb{P}$, X_0 has law μ, we indicate this by writing $\mathbb{P}_\mu$ instead of $\mathbb{P}$. In particular, $\mathbb{P}_x$ is to be associated with $\mu = \epsilon_x$, the point mass at $x \in E$.

The resolvent of X is the family $(V_p : p \geq 0)$ of kernels given by

$$V_p(x, f) = \int_0^\infty e^{-pt} P_t(x, f) dt = \mathbb{P}_x \int_0^\zeta e^{-pt} f(X_t) dt. \tag{2.2}$$

Note the appearance of the lifetime ζ (let $s = 0$ in (2.1) to get the right hand side of (2.2)). Once ζ has occurred, the process does not belong to E anymore. When $p = 0$, we often simply write $V(x, dy) = V_0(x, dy)$, and call this the potential kernel of X. The process (in fact, any kernel V) is called *transient* if and only if there exists a function $g > 0$ such that $0 < V(\cdot, g) < +\infty$ on E. This does not necessarily imply that the lifetime ζ is finite, but it does mean that (X_t) spends a.s. a finite time in each subset B of E such that $V(\cdot, B) < +\infty$.

Recall that a stopping time relative to some arbitrary filtration $(\mathcal{G}_t) \supset (\mathcal{F}_t^0)$ is a random variable T such that $\{T \leq t\} \in \mathcal{G}_t$ for each $t \geq 0$. Examples are given by hitting times of Borel sets $B \subset E$:

$$T_B = \inf\{t > 0 : X_t \in B\},$$

but only if the filtration $(\mathcal{G}_t)$ contains all $\mathbb{P}$-null sets (see Rogers and Williams (1994)). To take care of all possible initial distributions, we write $(\mathcal{F}_t)$ to mean the smallest, right continuous filtration containing $(\mathcal{F}_t^0)$ and such that $\mathcal{F}_0$ contains all $\mathbb{P}_\mu$-null sets, as μ varies over the possible initial distributions. Later, we will be mostly interested in the first exit time from a set B, $\sigma_B = T_{B^c}$.

The process X is called Strong Markov under $\mathbb{P}$ if

$$\mathbb{P}(f(X_{T+t}), \zeta > T + t \,|\, \mathcal{F}_T) = P_t(X_T, f) \quad a.s. \text{ on } \{T < \infty\} \tag{2.3}$$

holds for all $(\mathcal{F}_t)$-stopping times.

We shall sometimes make use of the shift operator θ_t which is defined on path space Ω by the formula $(\theta_t\omega)(s) = \omega(t+s)$. In terms of it, the Strong Markov property becomes

$$\mathbb{P}(H \circ \theta_T \,|\, \mathcal{F}_T) = \mathbb{P}_{X_T}(H) \quad a.s. \text{ on } \{T < \infty\},$$

for any positive random variable H (which can be $\mathcal{F}_\infty$ measurable). A useful fact in this connection is that hitting times of Borel sets are terminal times, as is ζ. This means that $\zeta = \zeta \circ \theta_t + t$ on $\{\zeta > t\}$. Applying the Markov property gives the following calculation, which is of fundamental importance for Chapter 4, and illustrates the use of the shift operator:

$$\begin{aligned}\mathbb{P}_\nu(g(X_t), \zeta > t+s) &= \mathbb{P}_\nu(g(X_t), \zeta \circ \theta_t > s, \zeta > t)\\ &= \mathbb{P}_\nu(g(X_t)\mathbb{P}_\nu(\zeta \circ \theta_t > s \,|\, \mathcal{F}_t), \zeta > t)\\ &= \mathbb{P}_\nu(g(X_t)\mathbb{P}_{X_t}(\zeta > s), \zeta > t).\end{aligned}$$

We now give some examples of the Strong Markov processes we have in mind, fixing various elements of notation.

Example 2.1. Let E be a subdomain of $\mathbb{R}^d$. Thus E is open, and its boundary is denoted ∂E. We shall be interested in a diffusion (X_t) whose lifetime coincides with the first exit time from E. Thus $X_{\zeta-}$ does not belong to E. The process is specified uniquely as the minimal process whose generator is given on $C_K^\infty(E)$, the space of C^∞ functions with compact support strictly contained in E, by

$$Lf(x) = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\frac{\partial^2}{\partial x^i \partial x^j} f(x) + \sum_{j=1}^{d} b_j(x)\frac{\partial}{\partial x^j} f(x), \quad f \in C_K^\infty(E),$$

with Hölder continuous coefficients. We say that the diffusion X is uniformly elliptic if there exists a constant λ such that

$$\lambda^{-1}\|\theta\|^2 \le \langle \theta, a(x)\theta\rangle \le \lambda\|\theta\|^2, \quad \theta \in \mathbb{R}^d,$$

and $\|b(x)\| \le \lambda$. Thus the eigenvalues of the positive definite, symmetric matrix $a_{ij}(x)$ are bounded above and below (by λ and λ^{-1} respectively), uniformly on E.

Example 2.2. Let E be locally compact. A Strong Markov process X whose sample paths consist exclusively of right continuous step functions will be called a Markov chain on E. It is specified by a function $q(x)$ which describes how long the process waits in any given state,

$$\mathbb{P}_x(\sigma_x > t) = e^{-q(x)t}, \quad \sigma_x = \inf\{t > 0 : X_t \neq x\},$$

and a distribution for the location after the next jump:

$$\mathbb{P}_x(X_{\sigma_x} \in dy) = \pi(x, dy).$$

When the state space E is countable, this information is encapsulated in the q-matrix of the process,

$$q_{ij} = \begin{cases} -q(i) & \text{if } i = j, \\ q(i)\pi(i, \{j\}) & \text{if } i \neq j. \end{cases}$$

For this process, the left limit $X_{\zeta-}$ may or may not belong to E.

Example 2.3. If E is a state space with a Strong Markov process X defined on it, let $\overline{E} = (a, b) \times E$, where $-\infty \leq a < b \leq \infty$. We define a semigroup $(\bar{P}_r)$ of kernels on $\overline{E}$ by the formula

$$\bar{P}_r\big((t, x); (ds, dy)\big) = 1_{(a,b)}(s)\epsilon_{t-r}(ds)P_r(x, dy),$$

where $\epsilon_t(ds)$ is the point mass at $t \in \mathbb{R}$. The associated Strong Markov process on $\overline{E}$ is $\overline{X}_r = (L_r, X_r)$, where L_r is the process of translation at unit speed towards zero, killed once it hits the point a. The semigroup of L_r is just $Q_r(t, ds) = 1_{(a,b)}(s)\epsilon_{t-r}(ds)$. The process $\overline{X}$ dies once either $X_r \notin E$ or $L_r = a$, whichever comes first. Spacetime quantities will always be denoted by a bar, as in $\overline{x} = (t, x)$,

$$\bar{\zeta}(\omega) = \inf\{s > 0 : L_s(\omega) = a\} \wedge \zeta(\omega).$$

It is obvious that $\overline{\mathbb{P}}_{(t,x)}(z \leq t - a) = 1$, and the process $\overline{X}$ is always transient, although z can be infinite if $a = -\infty$ (in which case the process "runs off to

$-\infty$", although it will exit $\overline{E}$ before the end of the run if $\zeta < \infty$.). The resolvent is given by

$$\bar{V}_p f(\overline{x}) = \overline{\mathbb{P}}_{(t,x)} \int_0^z e^{-ps} f(\overline{X}_s) ds = \int_a^t e^{-p(t-s)} ds \int_E P_{t-s}(x, dy) f(s, y). \tag{2.4}$$

Sharpe (1988) can be consulted for a careful construction of $\overline{X}$. In particular, $\overline{X}$ can always be constructed on the same probability space Ω as X. We will always state clearly what the interval (a, b) is in any given situation.

2.2 Excessive Functions and h-Transforms

We fix a transient Strong Markov process X as in the previous section, and work with its canonical realization on path space. Many of the quantities we will deal with are constructed out of excessive functions, thus we collect here the relevant results about them. Much of this is culled from Dellacherie and Meyer (1987), but it may also be found in Chung (1982) among many others.

A Borel function $h : E \to [0, +\infty]$ is called excessive if it satisfies

$$P_t h \leq h, \quad \lim_{t \to 0} P_t h = h \quad \text{on } E.$$

Equivalently, this may also be stated in terms of the resolvent (V_p) by the requirement that $pV_p h \uparrow h$ as $p \to \infty$.

A prime example of an excessive function is the function $h = V(\cdot, g)$, where g is some arbitrary positive Borel function. Excessive functions of this form are sometimes known as potentials. One can ask if all excessive functions are of this form, and the answer is no. However, for any excessive h, we can set $g_n = n(h - P_{1/n} h)$, and then

$$h = \uparrow \lim_{n \to \infty} V(\cdot, g_n).$$

This result requires transience of X.

When h is excessive, the process $h(X_t) 1_{(\zeta > t)}$ is a positive supermartingale under any probability law $\mathbb{P}_\nu$ where $\langle \nu, h \rangle < +\infty$, though if h is not ν-integrable,

the process can have infinite expectation. If we do not have $\lim_{t\to 0} P_t f = f$, then the process is not right continuous.

One of the many uses of excessive functions is in testing for inaccessible sets. More precisely, if h is excessive, then $B = \{x : f(x) = +\infty\}$ is a polar set, which means $\mathbb{P}_x(T_B = +\infty) = 1$ for all $x \in B^c$. This fact is used in Chapter 3.

Various useful classifications of excessive functions exist. In this work, we are mainly interested in

Invariant functions: An excessive h is called invariant if

$$P_t h(x) = h(x), \quad x \in E.$$

Harmonic functions: An excessive h is called harmonic if

$$\mathbb{P}_x(h(X_{\sigma_K}), \zeta > \sigma_K) = h(x), \quad \text{for all compact } K \subset E.$$

As always, $\sigma_K = \inf\{t > 0 : X_t \notin K\}$ is the first exit time from K.

As we shall see, the distinction is mainly that $h(X_t)1_{(\zeta>t)}$ is a local martingale when h is harmonic, but a true martingale only if h is invariant. The analogous definition for the backward spacetime process $\overline{X}$ merits its own terminology:

Parabolic functions: A function $h(t,x)$ is called parabolic in $(a,b)\times E$ if it is harmonic for the process $\overline{X}$, killed upon exit from $(a,b)\times E$.

An important example of a parabolic function in $(0,\infty)\times E$ is given by the function

$$u_f(t,x) = \mathbb{P}_x(f(X_t), \zeta > t), \quad f \geq 0.$$

This function is in fact invariant for $\overline{X}$ (hence harmonic) since by the simple Markov property

$$\bar{P}_r u_f(t,x) = \mathbb{P}_x(\mathbb{P}_{X_r}(f(X_t), \zeta > t-r), \zeta > r) = \mathbb{P}_x(f(X_t), \zeta > t).$$

As pointed out by Doob (1957), the function u_f solves the Cauchy problem

$$\partial u/\partial t = \mathfrak{A}u, \quad t > 0, \quad u(0,x) = f(x).$$

Here $\mathfrak{A}$ is a suitable generator. See the end of this chapter.

We now introduce the setup of h-transforms, another concept originally due to Doob (1957). For the remainder of this section, h is a fixed excessive function. We put

$$E_h = \{x \in E : 0 < h(x) < \infty\},$$

which defines a Borel set in E. Associated with E_h is a semigroup (P_t^h), defined on E_h as follows:

$$P_t^h(x, f) = h^{-1}(x)P_t(x, fh), \quad x \in E_h.$$

Related quantities will be denoted by a superscript h in the obvious way. It is important that the process with semigroup (P_t^h) can be realized as a Strong Markov process on the same space Ω we started with. In fact, its path measure $\mathbb{P}_x^h$ is related to that of the original process by the formula

$$h(x)\mathbb{P}_x^h(\Lambda; \zeta > T) = \mathbb{P}_x(\Lambda; h(X_T)), \quad x \in E_h,$$

where T denotes a finite stopping time and $\Lambda \in \mathcal{F}_T$. In particular, we can deduce from the above that

$$\mathbb{P}_x^h(X_t \in E_h \text{ and } X_{t-} \in E_h \text{ for all } t < \zeta) = 1,$$

so that we do not need to worry about how (P_t^h) is defined on $E \backslash E_h$ (Meyer (1968)). Note that it is certainly possible that $X_{\zeta-} \notin E_h$.

2.3 Integral Representations

In this section, we recall the Choquet integral representation theorem for convex sets, and the consequent integral representation of excessive functions. The former will be used for various proofs in Chapter 4, while the latter is used in the proofs of Chapter 3.

Let C be a convex, metrizable, compact subset of a locally convex topological vector space. An extreme point of C is any point $x \in C$ which is not the midpoint

of any line segment entirely contained in C. We denote the set of extreme points of C by $\partial_e C$. According to Choquet, every point of C is a weighted sum of extreme points:

Choquet's integral representation theorem: states that, for each $x \in C$, there exists a probability measure μ_x concentrated on $\partial_e C$ such that

$$\lambda(x) = \int_{\partial_e C} \lambda(z)\mu_x(dz) \tag{2.5}$$

holds for every continuous linear map $\lambda : C \to \mathbb{R}$, and conversely (see Dellacherie and Meyer (1983)).

Now let r be a probability measure on the state space E of a transient Strong Markov process, and set

$$C_r = \{h : h \text{ is excessive and } \langle r, h\rangle = 1\}.$$

This set is convex, and its extreme points are minimal excessive functions.

Minimal excessive function: An excessive function h is minimal if, whenever $h = k + l$ where k and l are excessive, they are both proportional to h.

Suppose now that hypothesis (L) below holds:

Hypothesis (L): There exists a measure m on E such that the resolvent of X is absolutely continuous:

$$V_p(x, dy) = v_p(x, y)m(dy), \quad p \geq 0.$$

For the spacetime process $\overline{X}$, due to the explicit formula (2.4) for the resolvent, a sufficient condition for hypothesis (L) to hold is that the semigroup of X can be written $P_t(x, dy) = p_t(x, y)\eta(dy)$. Then $m(ds, dy) = ds \cdot \eta(dy)$.

We return to the general situation. When the set C_r is given the $L^1_{\text{loc}}(dm)$ topology, in which $h_n \to h$ if and only if $\int_K h_n dm \to \int_K h dm$ for all compact subsets K of E, it becomes compact and metrizable.

By concentrating attention on functionals of the type $\lambda(h) = pV_p(x,h)$ and using the fact that $pV_p h \uparrow h$ as $p \to \infty$ for all $h \in C_r$, it is possible (see Walsh (1976), p.148) to deduce Choquet's integral representation theorem for excessive functions: every function $h \in C_r$ can be written

$$h(x) = \int_{\partial_e C_r} k(x)\nu_h(dk), \quad x \in E, \tag{2.6}$$

for some probability measure ν_h concentrated on the minimal excessive functions (extreme points of C_r). This representation also has the added property that, whenever h is harmonic, the measure ν_h is actually concentrated on harmonic minimal functions. It is not true that invariance of h implies that ν_h is concentrated on invariant minimal functions. See Dellacherie and Meyer (1987), Meyer (1968) and Walsh (1976).

2.4 Martin Boundary

The integral representation of the previous section has a probabilistic counterpart, which we now describe. Hypothesis (L) is not necessary (see Jeulin (1978)), but we do assume it below, for simplicity.

Let $v(x,y)$ be the density of $V_0(x,dy)$ with respect to $m(dy)$, and let r be a probability measure on E such that $\int r(dx)v(x,y)$ is finite and nonzero. The Martin kernel is defined by

$$K(x,y) = \frac{v(x,y)}{\int r(dx)v(x,y)}.$$

There exists a compact set F containing E, known as the minimal Martin compactification of E, with the following property. For fixed x, $K(x,\cdot)$ can be uniquely extended to a continuous function on F, with values in $[0,+\infty]$, and such that $K(\cdot,y)$ is minimal excessive for each $y \in F$. This requires a change in the topology of E; the new topology is known as the Martin topology, and under it, the set E is dense in F. From the point of view of the study of the set C_r of the previous section, the meaning of these facts is that the set $\partial_e C_r$ can be

identified with some subset of F. Thus a measure on $\partial_e C_r$ can be viewed as a measure on F, and conversely. This is crucial for the next few statements.

The probabilistic interpretation of the Choquet integral representation of the previous section now states that, for an excessive function $h \in C_r$,

$$\mathbb{P}_x^h(X_{\zeta-} \text{ exists}) = 1, \quad x \in E_h \quad \text{if and only if } h = \int K(\cdot, y)\nu_h(dy), \tag{2.7}$$

for some probability measure ν_h which is concentrated on E if $X_{\zeta-}$ exists in the original topology, and concentrated on F if $X_{\zeta-}$ exists in the Martin topology. This probability measure ν_h is, up to a constant factor, the same as the measure ν_h of the previous section. Moreover, we have explicitly

$$h(x)\mathbb{P}_x^h(X_{\zeta-} \in A \text{ exists}) = \int_A K(x, y)\nu_h(dy), \quad x \in E_h, \quad A \subseteq F. \tag{2.8}$$

If $\mathbb{P}_x^h(\zeta = \infty) = 1$, as is the case for example if h is invariant, the representation (2.7) and (2.8) still holds if we replace the process X by a speeded up version $\widetilde{X}$. The function h then becomes excessive (but not invariant) for $\widetilde{X}$, and such that $X_{\zeta-}$ exists in F.

The preceding claims can be found with proofs in Rogers and Williams (1994), Meyer (1968) and Kunita and Watanabe (1965).

2.5 Local Martingale Generator

In this section, we take up a quick study of invariance versus harmonicity of excessive functions, in terms of a generator for the process X. It is part of the folklore of Markov process theory that a function h satisfying an equation $\mathfrak{A}h = 0$, where $\mathfrak{A}$ is a (suitable) generator, is harmonic. Stronger conditions are needed to prove invariance. This observation is the basis of Doob's work on Laplace's equation and the heat equation (1957, 1955) and is made rigorous for the restricted class of Feller-Dynkin processes in Dynkin (1965), where $\mathfrak{A}$ is Dynkin's characteristic operator.

A corresponding characterization for more general Markov processes seems hard to find in the literature, although the result is known for many specific

classes of processes. Thus for example with Markov chains, which are rarely Feller-Dynkin, the role of $\mathfrak{A}$ is taken by the q-matrix. We give now a formal characterization which is valid for any Strong Markov (right, Borel) process.

Definition 2.4. *A locally bounded Borel function f is said to belong to the domain of the local martingale generator $\mathfrak{A}$ if there exists a Borel function $g(x) = \mathfrak{A}f(x)$ such that the process*

$$M_t^f = f(X_t)1_{(\zeta>t)} - f(X_0) - \int_0^{t\wedge\zeta} \mathfrak{A}f(X_s)ds$$

is, for each $(\Omega, (\mathcal{F}_t), \mathbb{P}_x)$, a right continuous local martingale up to ζ in the following sense: there exists a sequence of stopping times $T_n \uparrow \zeta$ such that $M^f_{t\wedge T_n}$ is a $\mathbb{P}_x$ martingale for each $x \in E$.

In this work, $\mathfrak{A}$ is the only type of generator we deal with, so we will usually call it simply the generator of X. It is an extension of the Hille-Yosida infinitesimal generator of (P_t), viewed as an operator semigroup acting on a suitable space of bounded functions. The "operator" $\mathfrak{A}$ is generally multivalued, since the function $g = \mathfrak{A}f$ can be arbitrary on sets which are visited by the process for a time set of zero Lebesgue measure.

Before proceeding, we give a few examples of well known generators.

Example 2.5. Suppose that X is the minimal diffusion on a domain $E \subset \mathbb{R}^d$ associated with an elliptic differential operator L. If f is a C^2 function with compact support strictly contained in E, it is a consequence of Ito's formula that the process

$$M_t^f = f(X_t)1_{(\zeta>t)} - f(X_0) - \int_0^{t\wedge\zeta} Lf(X_s)ds$$

is a continuous martingale under each $\mathbb{P}_x$. Now let $K_n \subset E$ be compact such that $K_n \uparrow E$, and set $T_n = \sigma_{K_n}$. We have $T_n \uparrow \zeta$, and $M^f_{t\wedge T_n}$ is a martingale for any C^2 function f on E, again by Ito's formula. Thus $\mathfrak{A}f = Lf$ for all $f \in C^2(E)$.

Example 2.6. Let X be a Markov chain on $E = \{1, 2, 3, \dots\}$ with stable q-matrix (q_{ij}), that is $q_{ii} > -\infty$. The generator $\mathfrak{A}$ coincides with the q-matrix

(see Rogers and Williams (1987)):

$$\mathfrak{A}f(i) = \sum_{i=1}^{\infty} q_{ij} f(j), \quad i \in E,$$

and its domain includes all functions $f : E \to \mathbb{R}$ such that $\mathfrak{A}f$ is finite. The sequence $T_n \uparrow \zeta$ can be taken as $T_n = \sigma(K_n)$, where the K_n are finite sets such that $K_n \uparrow E$.

Example 2.7. The generator of the spacetime process $\overline{X}_r = (L_r, X_r)$ is given by

$$\bar{\mathfrak{A}}f(t,x) = \mathfrak{A}f(t,x) - \frac{\partial}{\partial t} f(t,x).$$

Its domain includes all functions f such that $f(\cdot, x)$ is differentiable and $f(t, \cdot)$ belongs to the domain of $\mathfrak{A}$.

To characterize harmonic functions as the solutions to $\mathfrak{A}h = 0$, we use the following technical lemma. In plain language, it signifies that the process cannot leave the state space by a jump when the lifetime is predictable.

Lemma 2.8. *Let $T_n \uparrow \zeta$ be a sequence of stopping times. If $\zeta < \infty$, then for every $x \in E$ and compact set $K \subset E$,*

$$\mathbb{P}_x(\sigma_K = \zeta > T_n \ \forall n) = 0.$$

Proof. If $\zeta(\omega) > T_n(\omega)$ for each n, then $X_{\zeta-}(\omega) \notin E$; this is seen as follows. The stopping time defined by

$$R(\omega) = \begin{cases} \zeta(\omega) & \text{on } \{\zeta > T_n \ \forall n\}, \\ \infty & \text{otherwise,} \end{cases}$$

is predictable. If X_{R-} belonged to E on $\{R < \infty\}$, we would have (see Rogers and Williams (1994))

$$1 = (P_0 1_E)(X_{R-}) 1_{(R<\infty)} = \mathbb{P}_x(1_E(X_R), R < \infty | \mathcal{F}_{R-}),$$

which is absurd, the right side being zero since $X_R \notin E$ on $\{R < \infty\}$. We now finish the proof by noting that, on $\{\sigma_K = \zeta > T_n \ \forall n\}$, we would have

$$1 = 1_K(X_{\sigma_K-}) \leq 1_E(X_{\zeta-}) = 0,$$

and this is a contradiction. □

Theorem 2.9. *A function $h \geq 0$ is harmonic if and only if it satisfies*

$$\mathfrak{A}h = 0 \text{ in } E.$$

Proof. If h is harmonic, it is excessive and the process $h(X_t)1_{(\zeta>t)}$ is a right continuous supermartingale. Because the function is harmonic, we also have by optional stopping

$$h(x) = \mathbb{P}_x(h(X_\sigma), \zeta > \sigma) \leq \mathbb{P}_x(h(X_{t\wedge\sigma}), \zeta > t \wedge \sigma) \leq h(x),$$

whenever σ is the first exit time of a compact set. The process $h(X_{t\wedge\sigma})1_{(\zeta>t\wedge\sigma)}$ is thus a martingale. Taking a sequence of compact sets K_n increasing to E, we find that $M^h_{t\wedge\sigma_n}$ is a martingale with $\mathfrak{A}h = 0$ and $\sigma_n = \sigma(K_n) \uparrow \zeta$. Conversely, suppose that $\mathfrak{A}h = 0$ in E, and let T_n be a localizing sequence. Take a compact set K and put $\sigma = \sigma(K)$. By the martingale stopping theorem,

$$\begin{aligned} h(x) &= \mathbb{P}_x(h(X_{\sigma\wedge T_n}), \zeta > \sigma \wedge T_n) \\ &= \mathbb{P}_x(h(X_\sigma), \zeta > \sigma, \sigma \leq T_n) + \mathbb{P}_x(h(X_{T_n}), \zeta > T_n, \sigma > T_n). \end{aligned}$$

On $\{\sigma > T_n\}$, the random variable $h(X_{T_n})$ is bounded, so

$$\begin{aligned} \overline{\lim}_{n\to\infty} \mathbb{P}_x(h(X_{T_n}), \zeta > T_n, \sigma > T_n) &\leq \overline{\lim}_{n\to\infty} \|h\|_K \cdot \mathbb{P}_x(\zeta > T_n, \sigma > T_n) \\ &= \|h\|_K \, \mathbb{P}_x(\sigma_K = \zeta > T_n \ \forall n), \end{aligned}$$

which equals zero by the technical lemma above. Now it suffices to apply the monotone convergence theorem (since h is positive) to get

$$h(x) = \mathbb{P}_x(h(X_\sigma), \zeta > \sigma).$$

It remains to check that h is excessive. By Fatou's lemma and the assumed right continuity,

$$\underline{\lim}_{t\to 0} P_t h(x) = \underline{\lim}_{t\to 0} \mathbb{P}_x(h(X_t), \zeta > t) \geq \mathbb{P}_x(h(X_0), \zeta > 0) = h(x),$$

and since $h(X_t)1_{(\zeta>t)}$ is a positive local martingale, it is also a supermartingale, which means $P_t h \leq h$. □

It is worth pointing out explicitly that

Corollary 2.10. *A function h is parabolic for X in $(a,b) \times E$ if and only if*

$$\frac{\partial}{\partial t} h(t,x) = \mathfrak{A} h(t,x) \quad \text{in } (a,b) \times E.$$

As already pointed out by Doob (1955), if $u(x)$ satisfies an equation $\mathfrak{A}u(x) = \lambda u(x)$, $\lambda \in \mathbb{R}$, the function $v(t,x) = e^{\lambda t} u(x)$ is parabolic. Thus there is no need to develop a sophisticated theory of eigenfunctions of $\mathfrak{A}$; rather, their properties can be deduced from the behaviour of the backward spacetime process $\overline{X}$.

Chapter 3

Solidarity Results and Λ_*-Recurrence

This chapter is about the 'λ-classification' theory of an irreducible Markov process. The body of results described here has a long history. In the first section, we extend to general state space some basic solidarity results for the resolvent. The first results of this type appeared in Vere-Jones (1962) and Kingman (1963), who dealt with Markov chains on countable state spaces. Tweedie (1974a) proved the general state space version for discrete time, and used this as a basis for the continuous time extension (Tuominen and Tweedie (1979)). This required simultaneous irreducibility of skeleton chains (defined in Section 3). The proof given here does not rely on discrete time results, and dispenses entirely with simultaneous irreducibility. Next comes a study, in a very general setting, of decay parameters, λ-excessive functions and measures. A counterexample is given in Section 3, to show that positive Λ_*-recurrence with bounded Λ_*-invariant functions and measures is not sufficient for the existence of a Yaglom limit. This contradicts a well known countable state space result, stated as Theorem 3.15 below. The last two sections deal with a branching Markov process B constructed over X, in terms of which the solidarity results are interpreted.

3.1 Solidarity Results

The purpose of this section is to describe some qualitative features of certain kernels $(V_\lambda : \lambda \leq 0)$ defined below. The success of this undertaking will depend on good irreducibility assumptions, so we provide a suitable definition below.

Assumption I: There exists a nontrivial σ-finite (irreducibility) measure ν such that

$$\nu(A) > 0 \text{ implies } \mathbb{P}_x \int_0^\infty 1_A(X_s)ds > 0 \text{ for all } x \in E.$$

In other words, we assume that X is ν-irreducible. This concept goes back at least to Orey (1971). See the references therein.

There are other, stronger forms of irreducibility for Markov processes, and we shall use one such in the next chapter, but for the present, the above assumption is all we require.

Irreducibility measures are certainly not unique, nor are they equivalent. Some charge more sets than others. However, there always exists a maximal irreducibility measure η, characterized by the properties that

(i) η is an irreducibility measure, and

(ii) $\eta \gg \nu$ for every irreducibility measure ν.

In particular, property (ii) implies that η is unique only up to measure equivalence. For a proof of this assertion, see the discussion in Tweedie (1974a) or Nummelin (1985), who work in discrete time. The continuous time result is proved by switching to the resolvent chain with one-step probabilities $P(x, dy) = V_1(x, dy)$.

Our goal in the present section is to establish solidarity results for the "operators" V_λ defined by the formula

$$V_\lambda f(x) = \mathbb{P}_x \int_0^\zeta e^{-\lambda t} f(X_t)dt, \quad x \in E, \lambda \in \mathbb{R}. \tag{3.1}$$

This is used later in Section 3 to study some basic properties of the set of λ-excessive functions and measures. Note that if $\lambda \geq 0$, (V_λ) is just the resolvent of the process.

We are interested in whether the quantity in (3.1) is finite or infinite, simultaneously for all $x \in E$ and positive functions f. Clearly, this will be the case only if the behaviour of the process when started in x can be linked to its behaviour when started somewhere else, which is why we need Assumption I.

Besides being useful for proving Theorem 3.3, the following technical lemma shows that $(V_\lambda : \lambda \leq 0)$ is, as one would expect, an extension of the resolvent $(V_p : p \geq 0)$.

Lemma 3.1. *For every $\lambda \in \mathbb{R}$, the formula* (3.1) *defines a kernel $V_\lambda(x, dy)$, that is, a function $V_\lambda : E \times \mathcal{E} \to [0, \infty]$ such that*

1. *for each $x \in E$, $V_\lambda(x, \cdot)$ is a (not necessarily σ-finite) measure,*

2. *for each $A \in \mathcal{E}$, $V_\lambda(\cdot, A)$ is a $\mathcal{E}$ measurable function with values in $[0, \infty]$.*

Moreover, the extended resolvent equation holds:

$$V_\lambda f = V_\mu f + (\mu - \lambda) V_\lambda V_\mu f, \quad V\mu V\lambda f = V_\lambda V_\mu f, \tag{3.2}$$

for every $f \geq 0$ and $-\infty < \lambda < \mu < +\infty$, both sides being possibly infinite.

Proof. For each n, we define a kernel $V_\lambda^n : E \times \mathcal{E} \to [0, \infty)$ by

$$V_\lambda^n f(x) = \mathbb{P}_x \int_0^n e^{-\lambda s} f(X_s) 1_{(\zeta > s)} ds, \quad f \in b\mathcal{E}.$$

Letting $n \to \infty$, we see that V^λ is the limit of an increasing sequence of kernels on $(E, \mathcal{E})$, thus is itself a kernel on $(E, \mathcal{E})$. For the second assertion, let f be a positive bounded Borel function. Integration by parts gives, for $T = \zeta \wedge n$ and $f \geq 0$,

$$\begin{aligned}
&\int_0^T e^{-\lambda t} f(X_t) dt \\
&\qquad = \int_0^T e^{-\mu t} f(X_t) dt + (\mu - \lambda) \int_0^T e^{-(\lambda - \mu)t} \int_t^T e^{-\mu s} f(X_s) ds dt \\
&\qquad = \int_0^T e^{-\mu t} f(X_t) dt + (\mu - \lambda) \int_0^T e^{-(\mu - \lambda)t} \int_t^T e^{-\lambda s} f(X_s) ds dt.
\end{aligned}$$

Letting n tend to infinity and using monotone convergence, we see that the above also holds when $T = \zeta$, both sides being possibly infinite. Now take expectations on both sides, and note that

$$V_\lambda f(x) = e^{\lambda t}\mathbb{P}_x\big[\int_t^\zeta e^{-\lambda s} f(X_s)ds \,|\, \mathcal{F}_t\big] \quad \text{on } \{\zeta > t\},$$

by the Markov property, whether this is finite or not. We immediately get (3.2). □

We now show that, for each $\lambda \in \mathbb{R}$, either the kernel V_λ only takes the values 0 and $+\infty$, or it is transient when restricted to some subset $E_\lambda \subset E$. Here, transient means that there exists a function g such that $0 < V_\lambda g < +\infty$ on E_λ. Moreover, when starting in E_λ, the process never enters $E\backslash E_\lambda$ in a finite time, so that this part of the state space can be ignored.

As mentioned before, such a result has been known for a long time for specific processes. Vere-Jones (1962) discusses discrete space and time, Kingman (1962) allows continuous time, while Tweedie (1974a) treats discrete time and general state space. Moreover, Tweedie shows that the exceptional set $E\backslash E_\lambda$, which only appears when the state space is non-denumerable, is not charged by the irreducibility measure η. In discrete time, this is enough to guarantee that $E\backslash E_\lambda$ is not visited again once the process enters E_λ. However, this is not so in continuous time (and space), and a slightly more involved argument is necessary. This is the main reason for stating and proving Theorem 3.3 below.

While, according to that theorem, the set $E\backslash E_\lambda$ can be ignored, it is not in general empty. The following example illustrates the situation.

Example 3.2. Let the state space consist of the unit circle S^1 in $\mathbb{C}$ together with the origin $\{0\}$. The process X, when starting at $\{0\}$, waits for an exponentially distributed random time with mean one before jumping to the point 1 of S^1. Once there, it moves counterclockwise with constant speed v, while being killed at rate $\beta > 1$. The irreducibility measure for this process is the Lebesgue

surface measure η on S^1.

$$\mathbb{P}_x \int_0^\zeta e^{-\lambda s} f(X_s) ds = \begin{cases} \int_0^\infty e^{-(\lambda+1)t} \int_0^t e^{(1-\beta)s} f(e^{ivs}) ds dt & \\ \qquad + \int_0^\infty e^{-(\lambda+1)t} f(0) dt & \text{if } x = \{0\}, \\ \int_0^\infty e^{-(\lambda+\beta)t} f(e^{i(\theta+vt)}) dt & \text{if } x = e^{i\theta} \in S^1. \end{cases}$$

Here, we must take $E_\lambda = S^1$ in case $1 < -\lambda < \beta$, as a simple computation shows. The process never hits the origin from E_λ.

Note that in the previous example, any point $x \in S^1$ has zero η measure, but the process can (and sometimes does) still hit the set $\{x\}$ in a finite time. Thus it is not enough to prove that $E \backslash E_\lambda$ has zero η measure to dispense with it.

We are now ready to prove Theorem 3.3. The essential new feature in the proof we present, as compared with the proofs of the corresponding result for discrete time or state space (see Vere-Jones (1962), Tweedie (1974a)), is that the set A^r of regular points for A defined below need not be empty. This can occur only in the jointly continuous setting (assuming we restrict ourselves to *stable* chains in the continuous time, discrete space setting).

Theorem 3.3. *Suppose Assumption I holds. For every $\lambda \in \mathbb{R}$, one and only one of the following holds:*

(i) $V_\lambda f \equiv 0$ or $V_\lambda f \equiv +\infty$ for each $f \geq 0$.

(ii) The kernel $V_\lambda(x, dy)$ is transient when restricted to some set $E_\lambda \subset E$ such that $E \backslash E_\lambda$ is polar: $0 < V_\lambda g < \infty$ on E_λ for some function g satisfying $0 < g < \infty$, and

$$\mathbb{P}_x\big(X_t \in E_\lambda \quad \forall t \in [0, \zeta)\big) = 1, \quad x \in E_\lambda.$$

Proof. We loose no generality in restricting $\lambda \leq 0$, as we have $V_\lambda(\cdot, E) \leq \lambda^{-1}$ if $\lambda > 0$. Suppose that $V_\lambda(x_0, B_0) < \infty$ for some accessible $B_0 \subset E$ and some $x_0 \in E$. The strictly positive function $f(x) = V_\lambda(x, B_0)$ is excessive for the process X, since we have

$$P_t f(x) = e^{\lambda t} \mathbb{P}_x\big[\int_t^\zeta e^{-\lambda s} 1_{B_0}(X_s) ds; \zeta > t\big] \leq f(x),$$

from which we deduce $f(x) = \lim_{t\to 0} P_t f(x)$. Now let $A = \{f = \infty\}$. By Lemma 3.1, this is a Borel set. Denoting by T_A its first hitting time (which is therefore an $(\mathcal{F}_t)$ stopping time), we have $X(T_A) \notin A$ in general. But $X(T_A) \in A \cup A^r$, where $A^r = \{x : \mathbb{P}_x(T_A = 0) = 1\}$. Because f is excessive, it follows that (e.g. Chung (1982))

$$\inf_A f \leq f(x) \leq \sup_A f, \quad x \in A^r,$$

and hence $f(X(T_A)) = \infty$ a.s. on $\{T_A < \infty\}$; combining these facts with the Strong Markov property, we get, for any $x \in \{f < \infty\}$:

$$\begin{aligned}
\infty > V_\lambda(x, B_0) &= \mathbb{P}_x\big[\int_0^\zeta e^{-\lambda s} 1_{B_0}(X_s) ds\big] \\
&\geq \mathbb{P}_x\big[T_A < \zeta; e^{+\lambda T_A} \int_{T_A}^\zeta e^{-\lambda s} 1_{B_0}(X_s) ds\big] \\
&= \mathbb{P}_x\big[T_A < \zeta; \mathbb{P}_{X(T_A)} \int_0^\zeta e^{-\lambda s} 1_{B_0}(X_s) ds\big] \\
&= \mathbb{P}_x(T_A < \zeta; f(X_{T_A})),
\end{aligned}$$

and this clearly implies $\mathbb{P}_x(T_A < \zeta) = 0$.

We now set $E_\lambda = \{f < \infty\}$, and exhibit a strictly positive function g such that $0 < V_\lambda g < \infty$ on E_λ. By the resolvent equation of Lemma (3.1), if we take $\mu = 0$, then $g = V_0(\cdot, B_0) < \infty$ on E_λ, and

$$\begin{aligned}
\infty > V_\lambda(x, B_0) &= V_0(x, B_0) - \lambda(V_\lambda V_0 1_{B_0})(x) \\
&\geq V_\lambda g(x) \geq V_0 g(x) > 0,
\end{aligned}$$

for all $x \in E_\lambda$ as required. □

Note also that, since $E \backslash E_\lambda$ is polar, we have $V_0(\cdot, E\backslash E_\lambda) = 0$ on E_λ, and hence $\eta(E\backslash E_\lambda) = 0$ by irreducibility.

3.2 λ-Excessive Functions and Measures

By Theorem (3.3), the following parameter is well defined:

$$\Lambda_* = \inf\{\lambda : V_\lambda \text{ is transient when restricted to } E_\lambda \neq \emptyset\}.$$

Now set

$$E_* = \begin{cases} E_{\Lambda_*} & \text{if } E_{\Lambda_*} \neq \emptyset, \\ \bigcap_n E_{\Lambda_*-1/n} & \text{otherwise.} \end{cases}$$

Clearly, the set $E \backslash E_*$ is polar when starting from E_* and so we can, and will henceforth, restrict the process X to the state space $(E_*, \mathcal{E} \cap E_*)$. To avoid complicated notation, we denote this new state space again by $(E, \mathcal{E})$. Now we can say that, for each λ, the (restricted) kernel V_λ is either transient (certainly when $\Lambda_* < \lambda$), or else takes only the values 0 and $+\infty$ (certainly if $\lambda < \Lambda_*$). We therefore make the following classification:

Λ_*-recurrence/transience: If $V_{\Lambda_*}(x, dy)$ only takes the values 0 and $+\infty$, we say that the process X is Λ_*-recurrent. Otherwise, the kernel $V_{\Lambda_*}(x, dy)$ is transient, and we call the process Λ_*-transient.

We shall look at Λ_*-recurrent processes more closely in the next section.

Example 3.4. If X is a Markov chain on a countable state space E, the irreducibility (Assumption I) reduces to $p_t(x, y) > 0$ for all $x, y \in E$. Kingman (1963) showed that the integrals

$$\int_0^\infty e^{-\lambda t} p_t(x, y) dt, \quad x, y \in E,$$

either all converge or diverge simultaneously for a given value of λ. The critical value Λ_* is thus Kingman's decay parameter for the transition probabilities.

In many ways, a more important parameter for quasistationary limit theorems is the following:

$$\Lambda_1 = \inf\{\lambda \leq 0 : V_\lambda(\cdot, 1) < \infty\}.$$

It is obvious that $\Lambda_* \leq \Lambda_1$, but equality may not hold. The following alternative characterizations of Λ_1 are sometimes useful.

Lemma 3.5. *The following alternative characterizations of* Λ_1 *exist:*

(i) $\Lambda_1 = -\sup\{\epsilon > 0 : \mathbb{P}_x(e^{\epsilon\zeta}, \zeta < \infty) < +\infty\}$.

(ii) $\Lambda_1 = \inf\{\lambda \leq 0 : \mathbb{P}_x(\zeta > t) = o(e^{\lambda t}) \text{ as } t \to \infty\}$.

(iii) $\Lambda_1 = \inf\{\lambda \leq 0 : \text{ a } \lambda\text{-excessive probability exists}\}$.

(iv) $\Lambda_1 = \inf\{\lambda \leq 0 : \text{ some } \lambda\text{-excessive function } f \text{ satisfies } 1 \leq f < \infty\}$.

Proof. Statements (i) and (ii) follow immediately upon using the identities

$$\begin{aligned} V_\lambda(x, E) &= \int_0^\infty e^{-\lambda t} \mathbb{P}_x(\zeta > t) dt \\ &= (-\lambda)^{-1} \left(\mathbb{P}_x(e^{(-\lambda)\zeta}) - 1 \right). \end{aligned}$$

For statement (iii), supposing that $0 \geq \lambda > \Lambda_1$, the measure

$$\mu(\cdot) = V_\lambda(x, \cdot)/V_\lambda(x, E) \tag{3.3}$$

is a λ-excessive probability; conversely, if a λ-excessive probability μ exists, then by Fubini's theorem, for $\epsilon > 0$,

$$\begin{aligned} \int V_{\lambda+\epsilon}(x, E)\mu(dx) &= \int_0^\infty e^{-(\lambda+\epsilon)t} \langle \mu, P_t 1 \rangle dt \\ &\leq \int_0^\infty e^{-\epsilon t} \mu(B) dt = \epsilon^{-1} < \infty, \end{aligned}$$

which implies (since $\mu >> \eta$) that $V_{\lambda+\epsilon} 1 < \infty$ a.e., and hence $\lambda \geq \Lambda_1$. For statement (iv), if $0 \geq \lambda > \Lambda_1$, then by (i), the function

$$f(x) = \mathbb{E}_x(e^{-\lambda\zeta}, \zeta > 0) \tag{3.4}$$

is excessive, and clearly $f(x) \geq \mathbb{P}_x(\zeta > 0) = 1$ on E by the Blumenthal zero-one law. Conversely, let f be λ-excessive and satisfy $1 \leq f < \infty$ on E, then

$$\begin{aligned} V_{\lambda+\epsilon}(x, E) &= \int_0^\infty e^{-(\lambda+\epsilon)t} P_t(x, 1) dt \\ &\leq \int_0^\infty e^{-(\lambda+\epsilon)t} P_t(x, f) dt \\ &\leq f(x)/\epsilon < \infty, \end{aligned}$$

and hence $\lambda \geq \Lambda_1$. □

The parameters Λ_* and Λ_1 need not be equal: suppose again that X is a Markov chain on discrete state space, with transition matrix $p_t(x, y)$. Kingman (1963) proved

$$\Lambda_* = \lim_{t\to\infty} \frac{1}{t} \log p_t(x, x), \quad x \in E,$$

whereas it is shown in Jacka and Roberts (1995) that

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{P}_x(\zeta > t) = \Lambda_1,$$

provided this limit exists. When it does, they construct a Markov chain for which $\Lambda_* < \Lambda_1$.

In the more general situation of this chapter, we can say that

$$\overline{\lim}_{t\to\infty} \frac{1}{t} \log \mathbb{P}_x(\zeta > t) \le \Lambda_1, \quad x \in E. \tag{3.5}$$

This follows from Lemma 3.5 by noting that, for every $\epsilon > 0$, there exists a function $f_\epsilon \ge 1$ which is $(\Lambda_1 + \epsilon)$-excessive. Using this function in the inequalities

$$\begin{aligned} \frac{1}{t} \log \mathbb{P}_x(\zeta > t) &\le \frac{1}{t} \log \mathbb{P}_x(f_\epsilon(X_t), \zeta > t) \\ &\le \frac{1}{t} \log\big(f_\epsilon(x) e^{(\Lambda_1+\epsilon)t}\big) \\ &= \Lambda_1 + \epsilon + \frac{1}{t} \log f_\epsilon(x), \end{aligned}$$

we get (3.5) upon first letting $t \to \infty$, and then $\epsilon \to 0$. A corresponding lower bound requires further assumptions. We can show that

$$\underline{\lim}_{t\to\infty} \frac{1}{t} \log \mathbb{P}_x(\zeta > t) \ge \lambda, \quad x \in E, \tag{3.6}$$

provided a bounded λ-invariant function φ exists. For to get (3.6), it suffices to let $t \to \infty$ in the inequality

$$\begin{aligned} \frac{1}{t} \log \mathbb{P}_x(\zeta > t) &\ge \frac{1}{t} \log \mathbb{P}_x(\varphi(X_t), \zeta > t) \\ &= \frac{1}{t} \log\big(\varphi(x) e^{\lambda t}\big) \\ &= \lambda + \frac{1}{t} \log \varphi(x). \end{aligned}$$

On combining (3.5) and (3.6), we see that bounded λ-invariant functions can only exist if $\lambda \leq \Lambda_1$.

Part (i) of Lemma 3.1 in fact shows that $\Lambda_1 < 0$ can occur only if the lifetime of X has all finite moments:

$$\mathbb{P}_x[\zeta^k] \leq \sum_{n=0}^{\infty} (\epsilon^n/n!)\mathbb{P}_x(\zeta^n) = \mathbb{P}_x[e^{\epsilon\zeta}] < \infty, \quad k \geq 0.$$

The relationship between the parameter Λ_1 and exit times can be taken somewhat further (recall that σ_A denotes the first exit time from a set A):

Proposition 3.6. *Suppose that $A \subseteq B$. If $\inf_{x\in A} \mathbb{P}_x(\sigma_A > 0) > 0$, then also*

$$\inf_{x\in A} \mathbb{P}_x[\sigma_A] \leq \frac{1}{|\Lambda_1|}. \tag{3.7}$$

If there exists either a bounded λ-invariant function or λ-invariant probability measure, then

$$\frac{1}{|\lambda|} \leq \sup_{x\in B} \mathbb{P}_x[\zeta].$$

Proof. (i) If $\Lambda_1 = 0$, there is nothing to prove. Otherwise, take $\epsilon > 0$ such that $\Lambda_1 + \epsilon < 0$, and let $f \geq 1$ be $(\Lambda_1 + \epsilon)$-excessive. For any $x \in A$, we have

$$\begin{aligned}
(\inf_{y\in A} f(y)) \cdot \mathbb{P}_x(\sigma_A) &\leq \mathbb{P}_x \int_0^{\sigma_A} f(X_s)ds \\
&\leq \mathbb{P}_x \int_0^{\zeta} f(X_s)ds \\
&\leq f(x) \cdot \int_0^{\infty} e^{(\Lambda_1+\epsilon)s} ds \\
&\leq |\Lambda_1 + \epsilon|^{-1} f(x),
\end{aligned}$$

from which we get, upon taking the infimum over $x \in A$,

$$\inf_{x\in A} \mathbb{P}_x(\sigma_A) \leq |\Lambda_1 + \epsilon|^{-1},$$

and it remains only to let ϵ tend to zero.

(ii) Take μ to be λ-invariant ($\lambda < 0$) and such that $\mu(1) < \infty$. Then

$$|\lambda|^{-1} \langle \mu, 1\rangle = \langle \mu V_0, 1\rangle = \mathbb{P}_\mu(\zeta) \leq \langle \mu, 1\rangle \cdot \sup_{x\in E} \mathbb{E}_x(\zeta).$$

Otherwise, take f to be λ-invariant and such that $\sup_{x\in E} f(x) = 1$. Then

$$f(x) = |\lambda|\, V_0 f(x) \leq |\lambda|\, V_0 1(x) = |\lambda| \mathbb{P}_x[\zeta],$$

and it remains only to take the supremum over x on both sides of the inequality. □

We now give some examples to which the bounds in Proposition 3.6 apply.

Example 3.7. Suppose that X is an elliptic diffusion on a connected, bounded domain $E \subset \mathbb{R}^d$ with smooth boundary. It is well known (see Pinsky (1995)) that there exists a unique bounded positive λ-invariant function, known as the ground state of the differential operator $L = \mathfrak{A}$. The number λ is the principal eigenvalue of L, and the process is in fact positive λ-recurrent (see next section). Thus $\Lambda_* = \Lambda_1 = \lambda$ here. The first bound in Proposition (3.6) is uninteresting, since $\inf_{x\in A} \mathbb{P}_x[\sigma_A] = 0$ for all nice sets A. However, the second inequality provides a lower bound on the modulus of the principal eigenvalue,

$$|\Lambda| \geq \left(\sup_{x\in E} \mathbb{P}_x(\zeta) \right)^{-1}.$$

Example 3.8. Suppose that X is a Markov chain with a countable state space $E = \{1, 2, 3, \dots\}$. Taking $A = \{x\}$ for each $x \in E$ in turn, we find that

$$\begin{aligned} |\Lambda_1| &\leq \inf_{x\in E} \left(\mathbb{P}_x[\sigma_x] \right)^{-1} \\ &= \inf_{x\in E} q(x), \end{aligned}$$

where $q(x) = -q(x, x)$ is the diagonal element of the q-matrix, i.e. $\mathbb{P}_x(\sigma_x > t) = e^{-q(x)t}$. Thus Proposition (3.6) generalizes a well known bound (see Kingman (1963)). To illustrate the second inequality in Proposition (3.6), suppose that the state space E is finite and irreducible, in which case the expectation $\mathbb{P}_x[\zeta]$ is uniformly bounded over $x \in E$. It is known (Seneta and Vere-Jones (1966)) that the process X is Λ_*-recurrent. Alternatively, we can note that the semigroup (P_t), being representable by a finite matrix, is a compact operator on the set of all

functions on E. Its spectrum thus consists only of eigenvalues, and the Perron-Frobenius theorem guarantees that the largest eigenvalue λ has an associated positive eigenfunction/eigenmeasure pair. Since their sum is finite, Tweedie's test (see next section) implies that $\lambda = \Lambda_* = \Lambda_1$ and the process is Λ_*-recurrent. We again get a lower bound

$$|\Lambda| \geq \left(\sup_{x\in E} \mathbb{P}_x(\zeta)\right)^{-1}.$$

For a finite, irreducible state space, the function $x \mapsto \mathbb{P}_x[\zeta]$ is the solution to the matrix equation

$$\sum_{y\in E} q(x,y)\mathbb{P}_y[\zeta] = -1, \quad x \in E.$$

For example, if we assume that X can leave the state space from anywhere within E, and writing $v(x) = \mathbb{P}_x[\zeta]$, $v = \max_{x\in E} v(x)$, we have

$$v(x) \leq q(x)^{-1}\left(1 + v\sum_{y\neq x} q(x,y)\right),$$

which after rearranging gives

$$v \leq \frac{\max_{x\in E} q(x)^{-1}}{1 - \max_{x\in E}\sum_{y\neq x} q(x,y)/q(x)},$$

and hence

$$|\Lambda| \geq \frac{\min_{x\in E}\left(1 - \sum_{y\neq x\in E} q(x,y)/q(x)\right)}{\max_{x\in E} q(x)^{-1}}.$$

If $\Lambda_1 < 0$, there may be many λ-invariant probability measures; however, these can occur only for $\Lambda_1 \leq \lambda < 0$. If a 0-invariant probability μ were to exist, then because $\mathbb{P}_\mu(\zeta > t) = \mu(1) = 1$, we would have $\mathbb{P}_x(\zeta > t) = 1$ a.e., meaning that the constant function 1 is 0-invariant. As a result, we would have $\Lambda_1 = 0$, a contradiction.

Both examples given fall into the category of Λ_1-recurrent processes, which are treated in more detail in the next section. We end this section with a small result on the spectral structure of Λ_1-transient processes.

In various papers (see van Doorn (1991), Pakes (1995), Martinez and San Martin (1994)), it has been observed that certain processes X can have a continuum of quasistationary distributions. More precisely, these studies have shown the existence of λ-invariant probability measures μ_λ for every $\lambda \in [\Lambda_1, 0)$. By using Proposition (3.6), we see that this necessitates a weak form of "asymptotic remoteness", a concept discussed at length in Pakes (1995), and which we shall define in Chapter 5:

Corollary 3.9. *If a continuum $(\mu_\lambda : \lambda \in [\Lambda_1, 0))$ of λ-invariant probabilities exists, we have necessarily*

$$\sup_{x \in E} \mathbb{P}_x[\zeta] = +\infty. \tag{3.8}$$

The condition (3.8) is not sufficient for the existence of a continuum $(\mu_\lambda : \lambda \in [\Lambda_1, 0))$, as an example in Ferrari et al. (1995) shows. Indeed, the condition (3.8) is compatible with Λ_1-recurrence. One may wonder whether (3.8) *together* with knowledge that X is Λ_1-transient *is* sufficient.

3.3 Λ_*-Recurrence

In this section, we group together various well known results about Λ_1-recurrent processes, in a version appropriate to the general Markov processes we study. For the most part, these results are due to Tweedie (1974a,b) and Tuominen and Tweedie (1979). We also give a counterexample, to show that positive Λ_1-recurrence in general state space is not sufficient for the existence of a Yaglom limit. This is in stark contrast to the case of countable state spaces, or even merely discrete time.

For the remainder of this section, we fix a process X, which is assumed Λ_*-recurrent. Recall that, for any $\Lambda \in \mathbb{R}$, Λ-recurrence means that $V_\Lambda f \equiv 0$ or ∞ for any $f \geq 0$ and not just $V_\Lambda 1 \equiv \infty$, and similarly Λ-transience means that $V_\Lambda g < \infty$ for some $g > 0$ and not necessarily $V_\Lambda 1 < \infty$. We shall in fact often assume that X is Λ_1-recurrent, which means that we also have $\Lambda_* = \Lambda_1$.

The results below are due to Tweedie (1974a,b,c) and Tuominen and Tweedie (1979) in a series of papers devoted to the Λ-classification theory of discrete time and continuous time Markov processes on general state space. We begin with the probabilistic form of the well known Perron-Frobenius theorem on the spectrum of positive matrices.

Theorem 3.10 (Tuominen and Tweedie (1979)). *If X is Λ_*-recurrent and satisfies Assumption I (ν-irreducibility), there exists a unique strictly positive, finite Λ_*-excessive function (resp. measure), and this is in fact Λ_*-invariant. There are no other λ-invariant functions (resp. measures) if $\lambda \in (\Lambda_*, 0)$.*

This result does not preclude the existence of λ-harmonic functions (resp. measures) if $\lambda \geq \Lambda_*$. If the Λ_*-invariant measure whose existence is claimed above is bounded, we also have $\Lambda_* = \Lambda_1$.

We introduce the following standard terminology:

Positive/null Λ_*-recurrence: If X is Λ_*-recurrent, let μ (resp. f) be the unique nontrivial measure (resp. function) satisfying $\mu P_t = e^{\Lambda_* t}\mu$ (resp. $P_t f = e^{\Lambda_* t} f$); X is called positive if $\langle \mu, f \rangle < +\infty$, and null otherwise.

Tweedie (1974) gave a useful condition for testing positive Λ_*-recurrence. His result was originally proved for discrete time Markov chains, but also applies to continuous time processes via the resolvent chain with one-step transition function $pV_p(x, dy)$, $p > 0$, and a Laplace transform inversion.

Proposition 3.11 (Tweedie's Test (Tweedie (1974b))). *Let $\lambda \leq 0$, and suppose that some nontrivial measure μ (resp. finite positive function f) satisfies $\mu P_t \leq e^{\lambda t}\mu$ (resp. $P_t f \geq e^{\lambda t} f$). Then $\langle \mu, f \rangle < +\infty$ if and only if $\lambda = \Lambda_*$ and X is positive Λ_*-recurrent.*

This result can be "dualized" as follows:

Proposition 3.12 (Tweedie's Test (dual version)). *Let $\lambda \leq 0$, and suppose that some nontrivial measure μ (resp. finite positive function f) satisfies $\mu P_t \geq$*

$e^{\lambda t}\mu$ (resp. $P_t f \leq e^{\lambda t} f$). Then $\langle \mu, f\rangle < +\infty$ if and only if $\lambda = \Lambda_*$ and X is positive Λ_*-recurrent.

Proof. We may assume without loss of generality that $\zeta < \infty$ a.s. , for otherwise we can always take $f = 1$ and $\lambda = 0$, and the test reduces to checking standard positive recurrence. Consider the time reversal $\hat{X}_t = X_{\zeta - t}$. It is well known (Chung and Walsh (1970), Dellacherie and Meyer (1992)) that under $\mathbb{P}_\mu$, the process $\widehat{X}$ has a semigroup $\widehat{P}_t$ in duality with P_t:

$$\int P_t h(y) \cdot k(y) \mu V_0(dy) = \int h(y) \cdot \widehat{P}_t k(y) \mu V_0(dy), \quad h, k \geq 0. \tag{3.9}$$

Multiplying both sides of this equation by e^{-pt} and integrating over t gives the equation

$$\int V_p(y, h) \cdot k(y) \mu V_0(dy) = \int h(y) \cdot \widehat{V}_p(y, k) \mu V_0(dy), \quad h, k \geq 0, \quad p \in \mathbb{R},$$

and this shows immediately that X and $\widehat{X}$ have the same Λ_*-recurrence/transience classification under $\mathbb{P}_\mu$. Moreover, the assumption $\mu P_t \geq e^{\lambda t}\mu$ implies $\mu V_0 \gg \mu$, and hence there exists a Radon-Nikodym derivative $\widehat{f}$ such that $\mu(dy) = \widehat{f}(y)\mu V_0(dy)$. Using (3.9) we see that $\widehat{P}_t \widehat{f} \geq e^{\lambda t}\widehat{f}$, except perhaps on a μV_0-null set. Similarly, if we let $\widehat{\mu}(dy) = f(y)\mu V_0(dy)$, we find that $\widehat{\mu}\widehat{P}_t \leq e^{\lambda t}\widehat{\mu}$. Using Tweedie's test with $\widehat{\mu}$, $\widehat{f}$ shows that $\widehat{X}$ is positive Λ_*-recurrent (and consequently so is X) if and only if $\langle \widehat{\mu}, \widehat{f}\rangle = \langle \mu, f\rangle < +\infty$. □

The dual version of Tweedie's Test is not just provided for reasons of symmetry. In practice, the original test is typically used as follows:

(i) Find a λ-excessive measure μ and function f such that $\langle \mu, f\rangle < +\infty$.

(ii) Check that f is λ-invariant.

By using the dual version of the test, (ii) is replaced by

(ii') Check that μ is λ-invariant.

This may well be easier, especially since the methods for checking (ii') have been developed extensively. See Pollett (1988, 1995), Hart and Pollett (1997).

Example 3.13. Let X be a uniformly elliptic diffusion on a bounded open subset $E \subset \mathbb{R}^d$ with smooth boundary. The following facts may all be found in Pinsky (1995): The potential operator $V_0(x, dy) = v_0(x, y)dy$ is compact on the space of bounded continuous functions which vanish on the boundary of E. It follows by the Krein-Rutman theorem (which is the analogue, for positive operators, of the Perron-Frobenius theorem) that there exist strictly positive, continuous functions $\varphi(x)$, $\varphi^*(x)$ and a number λ satisfying $V_p\varphi(x) = -\lambda\varphi(x)$, $V_0^*\varphi^*(x) = -\lambda\varphi^*(x)$, and $\varphi(x) = \varphi^*(x) = 0$ on ∂E. Here $V_0^*(x, dy) = v_0(y, x)dy$. Using the resolvent equation gives $\varphi(x) = (p-\lambda)V_p\varphi(x)$, and inverting the Laplace transform produces the equation $P_t\varphi(x) = e^{\lambda t}\varphi(x)$. Entirely analogously, one gets $\int \varphi^*(x)P_t(x, dy) = e^{\lambda t}\varphi^*(y)dy$. Tweedie's Test now applies to $\mu(dx) = \varphi^*(x)dx$ and $f(x) = \varphi(x)$. Since the eigenfunctions are continuous and the closure of E is compact,

$$\langle \mu, f \rangle = \int \varphi^*(x)\varphi(x)dx < +\infty,$$

and the process X is Λ_*-recurrent. The number $\lambda = \Lambda_*$ is known as the principal eigenvalue of the generator L on E, with Dirichlet boundary conditions.

Example 3.14. Let X be a Markov chain on a finite state space. The fact that this process is Λ_*-positive recurrent was shown by Seneta and Vere-Jones (1967). Alternatively, the exact same method as in the previous example can be used, *mutatis mutandis*.

We shall now state the most general known results relating to Yaglom limits and Λ_*-recurrent processes. Before doing so, we state a stronger form of irreducibility than Assumption I.

Simultaneous skeleton irreducibility: For some nontrivial irreducibility measure ν,

$$\nu(A) > 0 \text{ implies } \sum_{n=0}^{\infty} \mathbb{P}_x(X_{n\delta} \in A) > 0, \quad \delta > 0.$$

This form of irreducibility implies Assumption I, but is strictly stronger; it allows the application of discrete time methods in continuous time situations. The theorem below is no longer true if X is not simultaneously skeleton irreducible. A counterexample will be discussed below.

Theorem 3.15 (Tuominen and Tweedie (1979)). *Suppose that X is simultaneously skeleton irreducible. Then X is positive Λ_*-recurrent with a bounded Λ_*-invariant measure κ and Λ_*-invariant function φ if and only if the following limits all exist a.e. with respect to η (the maximal irreducibility measure):*

(i) $\lim_{t\to\infty} \mathbb{P}_x(f(X_t) \,|\, \zeta > t) = \langle f, \kappa\rangle / \langle \kappa, 1\rangle$, x *a.e.* , $f \in L^\infty(\kappa)$,

(ii) $\lim_{s\to\infty} \lim_{t\to\infty} \mathbb{P}_x(f(X_s) \,|\, \zeta > t+s) = \langle \kappa, \varphi f\rangle / \langle \kappa, \varphi\rangle$, x *a.e.* , $f \in L^\infty(\kappa)$,

(iii) $\lim_{t\to\infty} \mathbb{P}_x(\zeta > t)/\mathbb{P}_y(\zeta > t) = \varphi(x)/\varphi(y)$, x, y *a.e.*

The source of the equivalence between positive Λ_*-recurrence with bounded κ and the existence of limits (i), (ii), (iii) in Theorem (3.15) is mainly the limit (ii), known as the doubly limiting conditional distribution. There are many examples of Λ_*-transient processes for which the limits (i) and (iii) exist, but the double limit (ii) must tend to zero.

Indeed, as we shall see in the next chapter, it is often possible to define a new Markov process Y with law $\mathbb{Q}_x$ and infinite lifetime such that

$$\mathbb{Q}_x(f(Y_t)) = \lim_{s\to\infty} \mathbb{P}_x(f(X_t) \,|\, \zeta > s).$$

The limit (ii) in Theorem (3.15) exists as stated (defining a probability measure $\pi(dx)$ proportional to $\varphi(x)\kappa(dx)$ only if Y is positive recurrent. Since under $\mathbb{Q}_x$, (Y_t) has transition semigroup $Q_t(x, dy) = e^{-\Lambda_* t} P_t(x, dy)\varphi(y)/\varphi(x)$, this process can be recurrent if and only if (X_t) is, under $\mathbb{P}_x$, Λ_*-recurrent.

The simultaneous irreducibility condition cannot be dropped entirely, as the following counterexample shows.

Example 3.16. Let $E = S^1$, and consider the Markov process X defined in the following way. Starting in $x \in E$, the process performs a counterclockwise

motion with unit speed, until the lapse of an exponential time with mean ρ, after which it dies. The transition semigroup of X is given by

$$P_t(e^{i\theta}, f) = e^{-\rho t} f(e^{i(\theta+t)}), \quad e^{i\theta} \in S^1,$$

and it is easy to see that X is Λ_*-recurrent with $\Lambda_* = -\rho$. Moreover, the unique Λ_*-invariant function is $f = 1$, while the Λ_*-invariant measure κ is Lebesgue surface measure on S^1. Thus the process is positive Λ-recurrent, by Tweedie's Test. It is easy to check that the Yaglom limits

$$\lim_{t\to\infty} \mathbb{P}_x(g(X_t) \,|\, \zeta > t), \quad x \in E,$$

do not exist for any $x \in E$, even though $\langle \kappa, 1\rangle = 1$. However, the process can still be conditioned to live for ever, and the resulting process Y is ordinary counterclockwise motion on S^1. Clearly, the simultaneous skeleton irreducibility fails for the process X, although Assumption I holds (with η being Lebesgue surface measure).

This example sits in contrast to the situation encountered with Markov chains on countable state spaces, where positive Λ_*-recurrence together with $\langle \kappa, 1\rangle < +\infty$ always implies, not only the success of process level conditioning, but also the existence of Yaglom limits from all possible initial distributions. The process X has a periodic transition semigroup $P_t(x, dy)$, something a Markov chain on a countable state space can never have.

3.4 Branching Markov Processes

The contents of this section can be found in standard references such as Ikeda, Nagasawa and Watanabe (1968,1969) or Asmussen and Hering (1983), to which we refer for proofs of all the statements below. Our only aim is to introduce the concepts and notation we shall use in the following section.

We start with our transient Strong Markov process $X = (\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}_x, X_t, \zeta)$ evolving on the state space $(E, \mathcal{E})$. A branching Markov process B over X can be constructed as follows. Suppose we place a particle in E and let it evolve according

to the law of X. At the same time, we take an independent, exponentially distributed time τ with mean γ^{-1}, called the first branching time. When τ occurs, we replace instantaneously the particle (provided it hasn't died already) with a random number of identical particles, all placed at the current location of the original particle. These new particles now evolve independently and according to the law of X, each carrying its own independent branching time, upon the occurrence of which a new batch is born, and so forth.

The process (B_t), representing the number and locations of all particles in existence, is called a branching Markov process over X. It is a Strong Markov process itself (Ikeda, Nagasawa, Watanabe (1968,1969)) and is described formally below. For the interpretation of the solidarity results, we will be interested also in another process, denoted by Z, whose significance is that it counts for each subset $A \subset E$ the number of particles in A at time t. This is achieved by a measure-valued process $Z_t(dy)$.

We now introduce the formal framework. For each $n \geq 1$, let $E^{(n)}$ denote the unordered product of n copies of E, with corresponding Borel σ-algebra. We set

$$\widehat{E} = \bigcup_{n=0}^{\infty} E^{(n)},$$

with $E^{(0)} = \{\emptyset\}$. Here, a point $\widehat{x} = (x_1, \ldots, x_n) \in E^{(n)}$ represents the joint location of n particles, while $\emptyset$ stands for the nonexistence of particles. A σ-algebra $\widehat{\mathcal{E}}$ is generated by the unordered sets of the form

$$\widehat{A} = A_1 \times A_2 \times \cdots \times A_n = A_2 \times \cdots \times A_n \times A_1 = \text{etc.}, \quad A_i \in E, n \geq 0.$$

Let (B_t) denote the coordinate process on the space $\widehat{\Omega}$ of $\widehat{E}$-valued paths, with corresponding natural filtration $(\mathcal{G}_t{}^0)$, and lifetime $\widehat{\zeta} = \inf\{t > 0 : B_t \notin \widehat{E}\}$. We say that a family $(\mathbb{Q}_{\widehat{x}} : \widehat{x} \in \widehat{E})$ of probability measures turns B into a branching Markov process if $B = (\widehat{\Omega}, B_t, \mathcal{G}_t, \mathbb{Q}_{\widehat{x}}, \widehat{\zeta})$ is a Strong Markov process and if the following branching property holds: For any function $f(x)$ on E, define $\hat{f}$ to be

a function on $\hat{E}$ by

$$\widehat{f}(\widehat{x}) = \begin{cases} 1 & \text{if } \widehat{x} = \emptyset, \\ \prod_{k=1}^{n} f(x_k) & \text{if } \widehat{x} = (x_1, \ldots, x_k), \end{cases}$$

for every function f on E such that $|f| \leq 1$, we require

$$\mathbb{Q}_{\widehat{x}}(\widehat{f}(B_t), \widehat{\zeta} > t) = \left(\mathbb{Q}_x(\widehat{f}(B_t), \widehat{\zeta} > t)\right)^{\widehat{}}, \tag{3.10}$$

for all $t > 0$ and functions f with $|f| \leq 1$. This property characterizes branching Markov processes, and in fact, the σ-algebra $\widehat{\mathcal{E}}$ is generated by functions of the form $\widehat{f}$, where f ranges over all functions in the unit ball. By the branching property (3.10), it is enough to construct the measures $(\mathbb{Q}_x : x \in E)$ to get a realization of B.

Suppose now that $\pi(x, d\widehat{y})$ is a Markovian kernel on $E \times \widehat{\mathcal{E}}$ which we will, moreover, assume to be of the form

$$\pi(x, d\widehat{y}) = \sum_{n=2}^{\infty} p_n \epsilon_{\underbrace{(x, \cdots, x)}_{n}}(d\widehat{y}).$$

When branching occurs at time τ, the branching particle X_t is replaced by a population distributed according to $\pi(X_{\tau-}, \cdot)$. The form of π prescribes that X should be replaced by n particles with probability p_n. If the particles behave identically to the Markov process $X = (\Omega, \mathcal{F}_t, X_t, \mathbb{P}_x, \zeta)$, we can solve for the semigroup of the process B, $Q_t(\widehat{x}, g) = \mathbb{Q}_{\widehat{x}}(g(B_t), \widehat{\zeta} > t)$, by writing

$$\begin{aligned} \mathbb{Q}_x(g(B_t), \widehat{\zeta} > t) &= \mathbb{Q}_x(g(B_t), \widehat{\zeta} > t, \tau > t) \\ &\quad + \int_0^t \int_E \mathbb{Q}_x(\tau \in ds, \widehat{\zeta} > s, B_{\tau-} \in dy) \int \pi(y, d\widehat{z}) \mathbb{Q}_{\widehat{z}}(g(B_{t-s}), \widehat{\zeta} > t - s), \end{aligned}$$

which transforms into

$$\begin{aligned} \mathbb{Q}_x(g(B_t), \widehat{\zeta} > t) &= e^{-\gamma t} \mathbb{P}_x(g(X_t), \zeta > t) \\ &\quad + \sum_{n=2}^{\infty} p_n \int_0^t \gamma e^{-\gamma s} \mathbb{P}_x\left[\left(\mathbb{Q}_{X_s}(g(B_{t-s}), \widehat{\zeta} > t - s)\right)^n, \zeta > s\right] ds, \end{aligned} \tag{3.11}$$

provided $g = \widehat{f}$ for some f and τ is distributed exponentially, with mean γ^{-1}. We call γ the branching rate. When Q_x satisfies (3.11), we say that B is a branching Markov process over X, with branching rate γ.

The semigroup (Q_t) which solves (3.11) is unique provided the corresponding lifetime $\widehat{\zeta}$ is infinite. This occurs in our setting if the mean number of offspring is finite:

$$\widehat{\zeta} = +\infty \quad a.s. \text{ if } \sum_{n=2}^{\infty} n \cdot p_n < \infty.$$

We shall always assume this to be the case. It does not preclude the process dying out in the sense of being trapped in $\emptyset$.

3.5 Solidarity Results Interpreted

Suppose that f is a bounded real-valued function on E, and set

$$g(\widehat{x}) = g_f(\widehat{x}) = \begin{cases} 0 & \text{if } \widehat{x} = \emptyset, \\ \sum_{k=1}^{n} f(x_k) & \text{if } \widehat{x} = (x_1, \ldots, x_k) \in E^{(n)}. \end{cases}$$

For each $t \geq 0$, we define a random measure $Z_t(dy)$ on E by the prescription

$$\langle f, Z_t(\widehat{\omega}) \rangle = g_f(B_t(\widehat{\omega})).$$

Upon taking f as the indicator function of a set $A \subset E$, we see that $Z_t(A)$ counts the number of particles alive and belonging to A at time t.

Now let $\theta > 0$ be some real number, and notice that

$$e^{-\theta \langle Z_t, f \rangle} = e^{g_{-\theta f}(B_t)} = \widehat{e^{-\theta f}}(B_t).$$

Inserting this function into (3.11), we find (assuming $\widehat{\zeta} = +\infty$ from now on)

$$\begin{aligned} (3.12) \quad \mathbb{Q}_x[e^{-\theta \langle Z_t, f \rangle}] &= e^{-\gamma t} \mathbb{P}_x[e^{-\theta f(X_t)}, \zeta > t] \\ &\quad + \sum_{n=2}^{\infty} p_n \int_0^t \gamma e^{-\gamma s} \mathbb{P}_x \left[\left(\mathbb{Q}_{X_s}[e^{-\theta \langle Z_{t-s}, f \rangle}] \right)^n, \zeta > s \right] ds. \end{aligned}$$

For $f \geq 0$, we now differentiate (from one side) this equation with respect to the variable θ, and let $\theta \downarrow 0$. The result is an equation for the first moment $v(t,x) = \mathbb{Q}_x Z_t(f)$ (see Ikeda, Nagasawa and Watanabe (1968,1969) or Asmussen and Hering (1983)):

$$v(t,x) = e^{-\gamma t} P_t(x,f) + \sum_{n=2}^{\infty} np_n \int_0^t \gamma e^{-\gamma s} P_s(x, v(t-s,\cdot))ds.$$

It is easy to verify that the above equation is solved by the function $v(t,x) = e^{a\gamma t} P_t(x,f)$ with $a = \sum_{n=2}^{\infty} np_n$. The solution is known to be unique, so we have in fact

Proposition 3.17. *For each bounded function f,*

$$\mathbb{Q}_x Z_t(f) = e^{(c-1)\gamma t} P_t(x,f), \quad c = \sum_{n=2}^{\infty} np_n.$$

The semigroup $e^{(c-1)\gamma t} P_t$ which we have obtained is known as the moment semigroup of B.

The interpretation of the solidarity results is now obvious. Suppose that only two particles are born at a time (or more generally that the mean number of offspring is 2, if we include the possibility $p_0 > 0$ of no offspring being created).

Theorem 3.18. *Suppose $c = 2$, and Assumption I holds. If $\lambda > \Lambda_*$ (or $\lambda \geq \Lambda_*$ if X is Λ_*-transient) then the branching Markov process B with rate $\gamma = -\lambda$ satisfies*

$$\mathbb{Q}_x(B_t \to \emptyset) = 1.$$

If $\lambda < \Lambda_*$, then the expected number of particles in any one set grows without bound or is a.s. identically zero for all times.

Proof. Suppose first that the kernel $V_\lambda(x,dy)$ is transient. There exists a sequence of sets $E_k \uparrow E$ such that $V_\lambda(x,E_k) < \infty$. Thus by the Proposition (3.17),

$$\int_0^\infty \mathbb{Q}_x Z_t(E_k)dt = V_\lambda(x,E_k) < \infty,$$

which implies that $\mathbb{Q}_x Z_t(E_k) \to 0$ as $t \to \infty$. But $Z_t(E_k)$ can only take integer values, hence $\mathbb{Q}_x(Z_t(E_k) > 0) \leq \mathbb{Q}_x(Z_t(E_k))$ and we have a.s. $Z_t(E_k) = 0$ for t sufficiently large; the set E_k being arbitrary, we must have $\mathbb{Q}_x(B_t \to \emptyset) = 1$. When $\lambda < \Lambda_*$, let A be any measurable subset of E. Then $V_\lambda(x, A) = 0$ or $+\infty$ according as the function $t \mapsto e^{-\lambda t} P_t(x, A) = \mathbb{Q}_x Z_t(A)$ is either identically zero, or grows without bound. Indeed, if in the latter case $e^{-\lambda t} P_t(x, A)$ were bounded, multiplying by $e^{-\epsilon t}$ $(\epsilon > 0)$ and integrating would show that $\lambda + \epsilon \geq \Lambda_*$, so that in view of the arbitrariness of ϵ we must have $\lambda = \Lambda_*$. □

Thus the branching process B is critical, subcritical or supercritical depending on whether $\lambda = \Lambda_*$, $\lambda > \Lambda_*$ or $\lambda \leq \Lambda_*$, respectively.

The λ-invariant measures, functions, etc. have equally simple interpretations, namely:

(i) A measure μ is λ-invariant for X if and only if μ is an invariant measure for the counting process Z associated with B when the branching rate is $-\lambda$. If μ is a quasistationary distribution, then it is an equilibrium distribution for Z.

(ii) A function φ is λ-invariant for X if and only if the process $\langle Z_t, \varphi \rangle$ is a positive martingale under each $\mathbb{Q}_x$, $x \in E$. Indeed, this follows by the branching property (see Asmussen and Hering (1983), p.154) and

$$\mathbb{Q}_x[Z_t(\varphi) \,|\, \mathcal{G}_s] = \langle Z_s, \mathbb{Q}_\cdot[Z_{t-s}(\varphi)] \rangle = \langle Z_s, e^{-\Lambda_*(t-s)} P_{t-s}(\cdot, \varphi) \rangle = \langle Z_s, \varphi \rangle.$$

(iii) A Yaglom limit of the type

$$\lim_{t \to \infty} \mathbb{P}_\nu(f(X_t) \,|\, \zeta > t) = \langle \kappa, f \rangle \tag{3.13}$$

exists if and only if, for every branching rate $\lambda < 0$,

$$\lim_{t \to \infty} \mathbb{Q}_\nu[Z_t(f)] / \mathbb{Q}_\nu[Z_t(g)] = \langle \kappa, f \rangle / \langle \kappa, g \rangle \tag{3.14}$$

holds for all bounded functions f, g with $\langle \kappa, g \rangle \neq 0$. Thus we may interpret (3.13) as the continuation of (3.14) in the degenerate case when the

branching rate is zero. We also mention that Asmussen and Hering (1983) have investigated the existence of the limit

$$\lim_{t\to\infty} \langle Z_t, f\rangle / \langle Z_t, g\rangle \stackrel{\text{a.s.}}{=} \langle \kappa, f\rangle / \langle \kappa, g\rangle, \tag{3.15}$$

in the case that X is positive Λ_*-recurrent, under slightly stronger assumptions than those of Theorem (3.15). These assumptions are satisfied by finite Markov chains, or uniformly elliptic diffusions on bounded domains.

3.6 Open questions

We have seen that the quasistationary tool box is well understood for positive Λ_*-recurrent processes, and in view of the counterexample after Theorem 3.15, at least for Markov chains and diffusions. The examples given in Section 3 suggest that such processes will have in some sense small state spaces, with no room to wander about without returning often to regions already visited. This line of argument certainly accounts for the cases of finite Markov chains and uniformly elliptic diffusions on bounded domains; in the latter, dropping the assumed uniformity allows the process to "slow down" arbitrarily, effectively making the state space much bigger than it is.

In both examples, one can show that the resolvent operators are compact on spaces of continuous functions, after which standard spectral theory (Krein-Rutman and Perron-Frobenius theorems) gives the Λ_*-invariant measure and function. If the process is symmetric, then the same procedure will work when the resolvent generates a Hilbert-Schmidt operator. This gives some examples of positive Λ_*-recurrent processes on unbounded domains. For example, this is often the case when X is the Brownian motion, killed upon exiting a set of finite Lebesgue measure (this set need not be bounded). The semigroup $p_t(x,y)dy$ has an eigenfunction expansion

$$p_t(x,y) = \sum_{k=0}^{\infty} e^{\lambda_k t}\varphi_k(x)\varphi_k(y),$$

and the function $\varphi_0(x)$ is the unique (positive) λ_0-invariant function.

An open problem here is to find some other set of conditions, which are easy to check and will ensure Λ_*-recurrence. One promising direction may be to look for Doeblin-type conditions (see Meyn and Tweedie (1993), Jacka and Roberts (1995)).

In another direction, the interpretation of the solidarity results in terms of branching Markov processes over X offers some exciting prospects. As mentioned before, the relationships between Λ_*-recurence and Yaglom limits have been extensively studied by Asmussen and Hering (1983). I am not aware of similar results in the case when X is Λ_*-transient. Does the existence of a Yaglom limit for X imply the existence of limits (3.15)? Secondly, the branching process B need not have a constant growth rate γ; it may be state-dependent: $\gamma = \gamma(x)$. It seems reasonable that the decay theory and Λ_*-classification should extend without major changes. Can the analogous theory in this chapter be used to shed some light on gauge theorems? See Chung (1995) and Sturm (1991).

Chapter 4

Conditioned Processes and Harnack's Inequality

This chapter and the next present a new approach to the main quasistationary tool box theorems, by linking them with the parabolic Martin boundary. Here, we deal with the conditioned process. After a description of the Martin boundary of $\overline{X}$ intended to complement the facts laid out in Chapter 2, the major assumptions are progressively introduced: the parabolic Harnack inequality is shown to hold for Markov chains and diffusions (the latter is well known), the set of parabolic functions is shown to be closed if the process has bounded jumps, and cemetery neighbourhoods are introduced. These are then used to characterize a part of the boundary. The main result is Theorem 4.21. It combines all these ideas into a sufficient condition for the success of process level quasistationary conditioning.

4.1 Parabolic Martin Boundary

We now make an assumption which will stay in effect throughout the chapter.

Assumption II: The process X has a semigroup given by

$$P_t(x, dy) = p_t(x, y)m(dy),$$

with a strictly positive, jointly continuous transition density (in all three variables). The measure m is assumed finite on compact sets.

Assumption II is quite strong. There are two reasons which justify it here. Firstly, we will be discussing tool box theorems foremost in the Martin topology. This topology is usually distinct from the original topology of E, especially when nothing is assumed about X (see Jeulin (1978) or Rogers and Williams (1994), where a Martin topology is constructed as the Ray-Knight compactification of the time reversal of X; examples in Getoor (1971) show that the Ray-Knight topology need not be comparable to the original topology). Our aim is to prove theorems which are applicable in the original topology, and it is here that Assumption II becomes useful. Secondly, the two classes of processes which are of greatest interest in this thesis, diffusions and Markov chains, typically *do* satisfy Assumption II.

Thus the assumption is relatively harmless. The strict positivity requirement represents a form of irreducibility. It clearly implies Assumption I, but is stronger than it.

Recall that $\overline{X}_r = (L_r, X_r)$ denotes the backward spacetime process on $(a, b] \times E$. We shall be mainly interested in the case $(a, b] = (-\infty, 0]$, and we put $\overline{E} = (-\infty, 0] \times E$. By Assumption II, $\overline{X}$ has an absolutely continuous resolvent (see (2.4)) in Chapter 2). In particular, the potential $\overline{V}_0$ has a density with respect to the measure $\eta(ds, dy) = ds \otimes m(dy)$ given by

$$v_0((t, x); (s, y)) = p_{t-s}(x, y) 1_{(-\infty, t)}(s).$$

Now let $r(ds, dy)$ be a normalizing probability measure on $\overline{E}$, and define the Martin kernel in terms of it (see (2.4) in Chapter 2). We are mostly interested in the case that $r(ds, dy) = \epsilon_0(ds)\nu(dy)$ where ν is a probability measure on E such that $(t, y) \mapsto \int \nu(dx) p_t(x, y)$ is continuous. The (spacetime) Martin kernel now takes the form

$$\overline{K}\Big((t, x); (s, y)\Big) = \frac{1_{(-\infty, t)}(s) p_{t-s}(x, y)}{\int \nu(dz) p_{-s}(z, y)}, \quad t, s < 0, \quad x, y \in E.$$

The Martin topology is induced by a metric

$$d(\overline{y},\overline{z}) = \int 1 \wedge \left|\overline{K}(\overline{x},\overline{y}) - \overline{K}(\overline{x},\overline{z})\right| \cdot f(\overline{x})\eta(d\overline{x}),$$

where $f(\overline{x})\eta(d\overline{x})$ is any probability measure equivalent to η. We denote by $\overline{F}$ the completion of $\overline{E}$ by d. It is well known (e.g. Doob (1984), p.197, or Meyer (1968)) that $\overline{F}$ is compact and that $\overline{K}(\overline{x},\cdot)$ has a continuous extension to $\overline{F}$, for each $\overline{x} \in \overline{E}$. The continuity of $\overline{K}(\overline{x},\cdot)$ in the original topology, which comes from the continuity of $p_t(x,y)$, is used to show that the original and Martin topologies coincide on $\overline{E}$.

The Martin boundary is $\partial\overline{E} = \overline{F}\backslash\overline{E}$. For each $\overline{y} \in \overline{F}$, the function $\overline{K}(\cdot,\overline{y})$ (defined on $\overline{E}$) is excessive, but not always minimal. We shall be mainly interested in that part of the boundary, which we shall denote by $\overline{F}_{-\infty}$, consisting of points $\overline{y} = \lim_{n\to\infty}(s_n, y_n) \in \overline{E}$ with the property that $\lim_{n\to\infty} s_n = -\infty$. These points have previously been studied in conjunction with the so called ratio limit property (see Pakes (1995), Kesten (1995) and references therein). Indeed, if the sequence (y_n) is constant and equal to $y \in E$, and ν is the probability mass concentrated at a single point $\nu(dz) = \epsilon_{x_0}(dz)$, then

$$\overline{K}\Big((t,x);\overline{y}\Big) = \lim_{r_n\to\infty} \frac{p_{r_n+t}(x,y)}{p_{r_n}(x_0,y)}, \quad r_n = -s_n.$$

We recognize the spacetime function $\overline{K}(\cdot,\overline{y})$ as a possible limit point of the ratio $(t,x) \mapsto p_{t+r}(x,y)/p_r(x_0,y)$.

The Strong Ratio Limit Property (SRLP) states that there exist functions f, g and a constant λ such that

$$\lim_{t\to\infty} \frac{p_{t+s}(x,y)}{p_t(x_0,y_0)} = e^{\lambda s}\frac{f(x)g(y)}{f(x_0)g(y_0)}.$$

Its use in proving quasistationary limit theorems is well established; see Anderson (1991). In terms of Martin boundary theory, it implies that all sequences of the form $\overline{y}_n = (s_n, y)$ with $s_n \to -\infty$ converge to the same boundary point $\overline{y}$, whose associated spacetime excessive (actually parabolic) function is

$$\overline{K}\Big((t,x);\overline{y}\Big) = e^{\lambda t}f(x), \quad \langle\nu, f\rangle = 1.$$

This function also appears when we condition X to have infinite lifetime. This will be discussed in detail below, where we shall derive, under suitable conditions, a weakened form of the SRLP. However, this theory requires various conditions, which must be dealt with first.

4.2 Parabolic Harnack Inequality

In this section, we discuss the following assumption.

Parabolic Harnack Inequality: Let $s > a > 0$ be positive real numbers, and K be compact in E. For each $\delta > 0$ there exists a constant $C > 0$ such that

$$\sup_{r\in[0,a]} \sup_{x\in K} u(t+r,x) \leq C \cdot \inf_{y\in K} u(t+s,y), \quad t > \delta,$$

holds for all parabolic functions $u(t,x)$ in $(0,\infty) \times E$.

This inequality will be essential in proving many of the results below. Note that K is strictly contained in E.

Example 4.1. Suppose that X is a uniformly elliptic diffusion on a (possibly unbounded) domain $E \subset \mathbb{R}^d$ (see Chapter 2). It was first shown by Moser (1964) that the parabolic Harnack inequality holds, in the form

$$\frac{u(t,x)}{u(t+r,y)} \leq \exp A\left(\frac{\|x-y\|^2}{r} + \frac{r}{d^2} + 1\right),$$

with constants $A > 0$ and $0 < d < t$ depending only on K (and the generator L). See Friedman (1964), and also Ancona (1991).

Example 4.2. Let X be an irreducible Markov chain with a countable state space. Here, a set is compact if and only if it is finite. Let K be such a set. Since the state space E is taken to be irreducible, for any $x, y \in K$, there always exists a finite chain of states $x_1, \ldots, x_n$ such that X can jump from x to x_1, from x_1 to x_2, ... , $x_n \to y$. We construct a new set $\tilde{K}$ from K by adding all these states to K, for any combination of states $x, y \in K$. The set $\tilde{K}$ need not be uniquely determined, but it can always be taken finite, and thus compact, since there are

only finitely many ordered pairs (x, y) with $x, y \in K$. The usefulness of $\tilde{K}$ stems from the fact that X can get from $x \in K$ to $y \in K$ without ever leaving $\tilde{K}$.

Proposition 4.3. *The parabolic Harnack inequality holds with*

$$C(K, \delta, a) = \left(\inf_{r \in [0,a]} \inf_{x,y \in K} \mathbb{P}_x(X_{s-r} = y, \sigma_{\tilde{K}} > s - r) \right)^{-1},$$

where $\sigma_{\tilde{K}} = \inf\{t > 0 : X_t \notin \tilde{K}\}$ *and* $\tilde{K}$ *is constructed from* K *as above.*

Proof. Let $u(t, x)$ be parabolic, and choose K compact (finite) with $x, y \in K$. Fix $s, t > 0$, and let T denote the first exit time of $\overline{X}$ from the set $[\delta, \infty) \times \tilde{K}$. Under the probability measure $\overline{\mathbb{P}}_{(t,x)}$, the stopping time T coincides a.s. with the first exit time from the compact set $[\delta, t] \times \tilde{K}$ if $t > \delta$. Moreover, on the event $\{T > r\}$, we have $T = T \circ \bar{\theta}_r + r$, since T is a terminal time for $\overline{X}$ (as is z). We also have $\overline{X}_T = \overline{X}_T \circ \overline{\theta}_r$ on $\{T > r\}$. Since u is parabolic, we can now write

$$\begin{aligned}
u(t+s, x) &= \overline{\mathbb{P}}_{(t+s,x)}(u(\overline{X}_T), z > T) \\
&\geq \overline{\mathbb{P}}_{(t+s,x)}\Big(u(\overline{X}_T), z > T > s - r, \overline{X}_{s-r} = (t + r, y) \Big) \\
&= \overline{\mathbb{P}}_{(t+s,x)}\Big(\overline{\mathbb{P}}\big[u(\overline{X}_T \circ \bar{\theta}_{s-r}), z \circ \bar{\theta}_{s-r} > T \circ \bar{\theta}_{s-r} \,|\, \overline{\mathcal{F}}_{s-r} \big], \\
&\qquad\qquad T > s - r, \overline{X}_{s-r} = (t + r, y) \Big) \\
&= \overline{\mathbb{P}}_{(t+s,x)}\Big(\overline{\mathbb{P}}_{\overline{X}_{s-r}}\big[u(\overline{X}_T), z > T \big], T > s - r, \overline{X}_{s-r} = (t + r, y) \Big) \\
&= \overline{\mathbb{P}}_{(t+s,x)}(T > s - r, X_{s-r} = y) \overline{\mathbb{P}}_{(t+r,y)}(u(X_T), z > T) \\
&= \mathbb{P}_x(\sigma_{\tilde{K}} > s - r, X_{s-r} = y) u(t + r, y),
\end{aligned}$$

where we have used the fact that X cannot have left $\tilde{K}$ by time s if $T > s$, and conversely. Since u is arbitrary, this concludes the proof. □

Example 4.4. The parabolic Harnack inequality generally fails for Markov chains on an uncountable state space. For one, it is easy to see that the previous proof fails, since in general $\mathbb{P}_x(X_s = y) = 0$. Two, Markov chains in uncountable state spaces do not satisfy Hypothesis (L). By a theorem of Walsh and Winkler (1981), this hypothesis is equivalent to the condition that the state space is the

union of at most a countable number of sets B with the property that $\mathbb{P}_x(\sigma_B > 0) = 1$ for all $x \in B$. If X is a Markov chain, every singleton set $B = \{x\}$ has this property.

4.3 Conditioned Processes and the Martin Boundary

In this section, we examine the connections between Doob's conditioning of processes to hit particular boundary points and the standard quasistationary procedure of conditioning processes not to hit certain parts of the state space. As explained in the introduction, all such problems can be reduced to conditioning X so as to have an infinite lifetime.

For reasons best left to practitioners, it is interesting to define, if possible, a law $\mathbb{Q}_\nu$ on the canonical path space Ω of X with the property

$$\mathbb{Q}_\nu(d\omega)|_{\mathcal{F}_t} = \lim_{r\to\infty} \mathbb{P}_\nu(d\omega \,|\, \zeta > r)|_{\mathcal{F}_t}. \tag{4.1}$$

As we shall see, under such a law, if it exists, the coordinate process X is again Markovian. This problem is trivial if $\mathbb{P}_\nu(\zeta = \infty) > 0$, so we will assume throughout the chapter that $\mathbb{P}_\nu(\zeta = \infty) = 0$ for all initial distributions ν on E.

For the record, we note that if $h(x) = \mathbb{P}_x(\zeta = \infty) > 0$ on E, the law $\mathbb{Q}_x$ is simply $\mathbb{P}^h_x$, the law of the h-transform associated with the invariant function h. There is no reason to expect X to be, under $\mathbb{Q}_\nu$ defined by (4.1), a time homogeneous Markov process, but as we shall see, this happens frequently.

We now return to a consideration of (4.1). The measure $\mathbb{P}_\nu(d\omega \,|\, \zeta > r)$ is obviously absolutely continuous with respect to $\mathbb{P}_\nu(d\omega)$. When restricted to the σ-algebra $\mathcal{F}_t$, it is easy to compute the corresponding Radon-Nikodym derivative. The Markov property gives

$$\mathbb{P}_\nu(d\omega \,|\, \zeta > r)|_{\mathcal{F}_t} = \frac{\mathbb{P}_{X_t(\omega)}(\zeta > r - t)}{\mathbb{P}_\nu(\zeta > r)} \cdot \mathbb{P}_\nu(d\omega)|_{\mathcal{F}_t}, \quad t > 0.$$

A study of the limit (4.1) therefore corresponds to a study, as $s \to -\infty$, of the functions

$$h_s(t,x) = \frac{\mathbb{P}_x(\zeta > -s+t)}{\mathbb{P}_\nu(\zeta > -s)}, \quad x \in E, s < t < 0.$$

This was already pointed out by Jacka and Roberts (1995), who showed that the pointwise convergence of $h_s(t,x)$, as $s \to -\infty$, is both necessary and sufficient for vague convergence of $\mathbb{P}_\nu(d\omega \,|\, \zeta > r)$ to $\mathbb{Q}_\nu(d\omega)$, provided X is a Markov chain on a countable state space, or a process with regular conditional probabilities (Jacka and Roberts (1997)). We shall derive conditions which ensure this convergence. These are related to some conditions of Kesten (1995) in countable time and space, but the methods of proof are quite different here, and applicable to general state spaces.

From now on, unless otherwise specified, the variables s and t will range over negative real values only. Each function $h_s(\overline{x})$ is parabolic (indeed, invariant for $\overline{X}$) on $(s,\infty) \times E$. It is reasonable to expect that the limit should be parabolic in $\mathbb{R} \times E$, and a fortiori in $(-\infty, 0] \times E$, hence representable by a probability measure on $\partial\overline{E}$. Once this measure is identified, we shall have a corresponding law $\mathbb{Q}_\nu$ satisfying (4.1). As far as possible, this programme will be implemented.

We begin by identifying the location of the Martin representing measure for h_s. This requires the function h_s to be spacetime excessive in all of $(-\infty, 0) \times E$. Accordingly, we shall extend it by setting

$$h_s(t,x) = 1/\mathbb{P}_\nu(\zeta > -s) \quad \text{for } (t,x) \in (-\infty, s) \times E.$$

The extended function is continuous, since $\lim_{t\uparrow 0} \mathbb{P}_x(\zeta > t) = 1$ for all $x \in E$. Using the fact that any constant spacetime function is excessive for $\overline{X}$ on $(-\infty, 0) \times E$, it is straightforward to check that the extended function h_s is excessive on $(-\infty, 0) \times E$. If $\mathfrak{A}1 = 0$, then the function h_s is parabolic on all of $(-\infty, 0) \times E$. From now on, h_s is assumed to be defined everywhere.

Lemma 4.5. *For every $s < 0$, there exists a probability measure μ_s on $\overline{F}$ such*

that

$$h_s(t,x) = \int \overline{K}\Big((t,x);\overline{y}\Big)\mu_s(d\overline{y});$$

This measure is supported by the set $\overline{F}_s$ *of those points* $\overline{y} = \lim_{n\to\infty}(s_n, y_n)$ *satisfying* $\limsup_{n\to\infty} s_n \le s$.

Proof. Since $\langle \nu, h_s(0,\cdot)\rangle = 1$, the integral representation presents no problems. By the results of Chapter 2, we know that the representing measure satisfies

$$\overline{\mathbb{P}}^{h_s}_{\overline{x}}(\overline{X}_{\zeta-} \in A) = \int_A \overline{K}(\overline{x},\overline{y})\mu_s(d\overline{y}), \quad A \subset F.$$

We shall show that the left hand side is zero if $A \subset F\backslash\overline{F}_s$, for any $\overline{x} \in (-\infty, 0]\times E$. The fact that $\overline{K}((t,x);\cdot)$ is strictly positive on $(-\infty, t)\times E$ then gives the result. By the definition of h_s, for any $r > 0$

$$\begin{aligned}\overline{\mathbb{P}}^{h_{-r}}_{\overline{x}}(z > r) &= h_{-r}(t,x)^{-1}\mathbb{P}_x(h_{-r}(t-r, X_r), \zeta > r)\\ &= \mathbb{P}_x(\mathbb{P}_{X_r}(\zeta > t), \zeta > r)/\mathbb{P}_x(\zeta > t+r)\\ &= 1.\end{aligned}$$

Thus regardless of the initial state $\overline{x} \in (-\infty, 0]\times E$, the process has, under $\mathbb{P}^{h_{-r}}_x$, a lifetime which is longer than r. In particular, $X_{\zeta-}$ must belong to $\overline{F}_{-r}$, and the proof is complete. □

The previous result tells us that any weak limit point μ of the family of measures (μ_s) has support in $\overline{F}_{-\infty} = \cap_{s<0}\overline{F}_s$. Correspondingly, the spacetime excessive function h represented by μ is a limit point of (h_s).

Lemma 4.6. *The family of functions* (h_s) *is compact: every subfamily contains a subsequence such that*

$$\lim_{n\to\infty} h_{s_n}(\overline{x}) = h(\overline{x}) \text{ on } (-\infty, 0]\times E.$$

The function h *is an excessive function (possibly zero) with a representing measure* μ *supported in* $\overline{F}_{-\infty} = \cap_s \overline{F}_s$, *and the convergence occurs boundedly on compact subsets of* $(-\infty, 0]\times E$.

Proof. Suppose first that D is a compact subset of $(-\infty, 0) \times E$. Without loss of generality, we shall assume that $D = [a, b] \times G$, where $\nu(G) > 0$. By the Parabolic Harnack Inequality, since $(t, x) \mapsto p_t(x, y)$ is parabolic in $(0, \infty) \times E$, there exists, if $s < \delta < 0$, a constant C such that

$$\begin{aligned}
\sup_{\overline{x} \in D} \overline{K}(\overline{x}, (s, y)) &= \sup_{(t,x) \in [a,b] \times G} \frac{1_{(-\infty, t)}(s) p_{t-s}(x, y)}{\int \nu(dw) p_{-s}(w, y)} \\
&\leq C \cdot \inf_{z \in G} \frac{p_{-s}(z, y)}{\int \nu(dw) p_{-s}(w, y)} \\
&\leq \left(C/\nu(G) \right) \frac{\int_G \nu(dz) p_{-s}(z, y)}{\int \nu(dw) p_{-s}(w, y)} \\
&\leq C/\nu(G).
\end{aligned}$$

Thus the map $\overline{y} \mapsto \overline{K}(\overline{x}, \overline{y})$ is continuous, and uniformly bounded on $\overline{F}_s$ for $s < \delta < 0$ and $\overline{x} \in D$. Now $\overline{F}$ is compact, so the probability measures (μ_s) have a weak limit point μ, i.e. $\mu_{s_n} \Rightarrow \mu$. Thus if $\overline{x} = (t, x)$ with $t < 0$,

$$\begin{aligned}
\lim_{n \to \infty} h_{s_n}(\overline{x}) &= \lim_{n \to \infty} \int \overline{K}(\overline{x}, \overline{y}) \mu_{s_n}(d\overline{y}) \\
&= \int \overline{K}(\overline{x}, \overline{y}) \mu(d\overline{y}) \\
&= h(\overline{x}),
\end{aligned}$$

and the convergence is uniformly bounded on D. It remains only to show that $\lim_{n \to \infty} h_{s_n}(0, x) = h(0, x)$. Since $t \mapsto h_s(t, x)$ is decreasing for each s, the same is true of $t \mapsto h(t, x)$ on $(-\infty, 0)$. Then the limit $h(0, x) := \lim_{t \to 0} h(t, x)$ exists, and clearly $h(0, x) \geq \overline{\lim}_{s_n \to -\infty} h_{s_n}(0, x)$. On the other hand, the function

$$g(t, x) = \underline{\lim}_{s_n \to -\infty} h_{s_n}(t, x), \quad (t, x) \in (-\infty, 0] \times E,$$

satisfies $\overline{P}_r g \leq g$ and it is bounded below by the excessive function $\tilde{g} = \lim_{r \to 0} \overline{P}_r g$ on $\overline{E}$; we now have

$$\begin{aligned}
h(0, x) = \lim_{p \to \infty} p(\overline{V}_p h)(0, x) &= \lim_{p \to \infty} p(\overline{V}_p g)(0, x) \\
&= g(0, x) = \lim_{s_n \to -\infty} h_{s_n}(0, x),
\end{aligned}$$

since h and g agree η-almost everywhere. $\square$

Importantly, it is possible for the limit function (which we as yet know only to be excessive) to be identically zero. Indeed, whereas each function h_s satisfies $\langle h_s(0,\cdot),\nu\rangle = 1$, Fatou's lemma shows only that $\langle h(0,\cdot),\nu\rangle \leq 1$.

If ν is compactly supported, then we must have $\langle h(0,\cdot),\nu\rangle = 1$, by bounded convergence; it follows that h is nonzero. However, suppose that ν is a quasistationary distribution for X (or more generally, it satisfies $\nu \geq C\mu$ where $C > 0$ and μ is a quasistationary distribution). Thus there exists $\lambda \geq \Lambda_1$ such that $\mathbb{P}_\nu(\zeta > t) = e^{\lambda t}$; then

$$h_s(t,x) = e^{\lambda s}\mathbb{P}_x(\zeta > -s+t),$$

and this tends to zero as $s \to -\infty$ whenever X is λ-transient (see Chapter 3). Thus $h(t,x) = 0$ in this case, and indeed also when X is λ-null recurrent, by the same argument. As a direct consequence, the conditioning procedure given by (4.1) fails to give a well-defined process. The only case when the limit $h(t,x)$ is nonzero above is when X is λ-positive recurrent, and in that case the initial distribution does not usually matter.

From now on, we shall always assume that ν is compactly supported (and possibly a point mass ϵ_x), so that any limit h automatically satisfies $\langle h(0,\cdot),\nu\rangle = 1$, and is therefore nonzero, and in fact strictly positive by Assumption II.

4.4 Closedness of Parabolic Functions

The next step consists of proving that any limit h given by Lemma 4.6 is in fact parabolic in $(-\infty,0]\times E$. This is true for diffusions, where the parabolic Harnack inequality implies that any family of parabolic functions has a convergent subsequence to another parabolic function, see Ancona (1990) or Doob (1984). Moreover, the convergence is then uniform on compact subsets of $(-\infty,0)\times E$. For Markov chains in general, this result is not so clear. Indeed, the parabolic functions in $(a,b]\times E$ are the positive solutions to the equation

$$u(t,x) = u(b,x) + \int_0^t \sum_{y\in E} q(x,y)u(b-s,y)ds. \tag{4.2}$$

If $h_s(t,x)$ solves (4.2), the same holds for h if and only if one may interchange limits and integrals on the right. A simple condition which guarantees this is that, for each $x \in E$, the measure $\{y\} \mapsto q_{xy}$ be supported by only a finite number of points y. This means probabilistically that there is at most a finite number of destinations for each jump of X.

Assumption III: For each compact set $K \subset E$, there exists another compact set $K' \subset E$ such that

$$\mathbb{P}_x(X_{\sigma_K} \in K', \zeta > \sigma_K) = 1, \quad x \in K,$$

where $\sigma_K = \inf\{t > 0 : X_t \notin K\}$ is the first exit time from K.

As remarked above, this assumption holds for Markov chains with bounded jumps, for we can take

$$K' = \{y : q_{xy} \neq 0, x \in K\}.$$

The assumption also holds for diffusion processes, for we can take $K' = K$ on account of the continuity of the sample paths. Note that the jump to the cemetery state which occurs at the end of the life of the process is allowed to be arbitrarily large.

Proposition 4.7. *Suppose that Assumption III holds. If (h_n) is a sequence of parabolic functions which converges boundedly on compacts to a function h, then h is itself parabolic.*

Proof. Let D denote a compact subset of $(-\infty, 0] \times E$, and put $\sigma = \inf\{r > 0 : \overline{X}_r \notin D\}$. Since h_n is parabolic,

$$h_n(\overline{x}) = \overline{\mathbb{P}}_{\overline{x}}(h_n(\overline{X}_\sigma), z > \sigma).$$

Moreover, our Assumption III implies that $\overline{X}_\sigma$ belongs to some compact set D' a.s. $\overline{\mathbb{P}}_{\overline{x}}$. Thus $h_n(\overline{X}_\sigma)$ is uniformly bounded a.s. , and using the bounded convergence theorem we can let $n \to \infty$ on both sides of the equation. Thus h is itself parabolic. □

In view of Lemma 4.6, we immediately get the following corollary.

Corollary 4.8. *If Assumption III holds, then every limit point of* (h_s) *is parabolic in* $(-\infty, 0] \times E$.

Proof. For each $a < 0$, the function h_s is parabolic in $(a, 0) \times E$ whenever $s < a$. By Lemma 4.6 and the previous proposition (which we apply to the process $\overline{X}$, killed when leaving $(a, 0) \times E$), any limit point of (h_s) must be parabolic in $(a, 0) \times E$, and hence in $(-\infty, 0) \times E$ since a is arbitrary. □

It is worth pointing out that Assumption III implies, via the above corollary, the existence of at least one non-trivial minimal parabolic boundary point belonging to $\overline{F}_{-\infty}$. This is because any non-trivial limit function h may be represented by Choquet's theorem as a measure on the spacetime harmonic minimal functions, and we saw that this measure charges only $\overline{F}_{-\infty}$.

There exists a way of dropping the bounded jump condition and still keep the conclusion of Corollary 4.8 intact. We assume the existence of a spacetime excessive function $g(t, x)$ such that $g(t, x) \geq h_s(t, x)$ for all s. Since $g(\overline{X}_{\sigma(K)})$ is by definition integrable for each compact K, the bound $h_s(\overline{X}_{\sigma(K)}) \leq g(\overline{X}_{\sigma(K)})$ and Lebesgue's dominated convergence theorem can be used instead of the bounded convergence theorem in Proposition 4.7. This replacement condition seems however much more difficult to check, and we shall not delve any further into the existence of such a function g.

4.5 Characterization of Limit Functions

It was shown in the previous sections that, under suitable technical conditions, there exist nonzero parabolic functions $h(t, x)$ in $(-\infty, 0] \times E$ such that

$$\lim_{s_n \to -\infty} \frac{\mathbb{P}_x(\zeta > -s_n + t)}{\mathbb{P}_\nu(\zeta > -s_n)} = h(t, x), \quad (t, x) \in (-\infty, 0] \times E. \tag{4.3}$$

This result would be much more useful if the subsequential limit could be replaced by a full limit, that is $\lim_{s \to -\infty} h_s(t, x) = h(t, x)$. In that case, one could

deduce as in Jacka and Roberts (1995) that $h(t,x) = e^{\lambda t}g(x)$ for some constant $\lambda \leq 0$.

Indeed, it suffices to note that, for $t < 0$,

$$h_{s+t}(0,x) = h_s(t,x)\mathbb{P}_\nu(\zeta > -s \,|\, \zeta > -s+t), \tag{4.4}$$

so that writing

$$L_\nu(t) = \lim_{s\to-\infty} \mathbb{P}_\nu(\zeta > -s \,|\, \zeta > -s+t),$$

which exists since the other two terms in (4.4) have a limit as $s \to \infty$, one gets

$$h(0,x) = h(t,x)L_\nu(t),$$

and therefore the decreasing function L_ν satisfies $L_\nu(a+b) = L_\nu(a)L_\nu(b)$, and hence $L_\nu(t) = ae^{-\lambda t}$, $\lambda \leq 0$.

Returning to the case at hand, since the subsequential limit function h of (4.3) is represented by a measure concentrated on the minimal part of $\overline{F}_{-\infty} = \cap_s \overline{F}_s$, the reasoning above suggests that we should look for conditions under which

$$\overline{K}\Big((t,x);\overline{y}\Big) = e^{\lambda t}g(x), \quad \overline{y} \in \overline{F}_{-\infty} \text{ minimal.} \tag{4.5}$$

We shall begin with a definition. Recall that we have assumed that $\zeta < \infty$ a.s.

Definition 4.9. *A subset $N \subset E$ is called a cemetery neighbourhood if*

$$\lim_{t\to\zeta} 1_N(X_t) = 1 \quad a.s.$$

for all starting points $x \in E$.

There always exists at least one cemetery neighbourhood, namely $N = E$. The definition states that the process spends the last segment of its lifetime in N.

Example 4.10. Suppose X is a Brownian motion killed upon leaving the unit ball $E = \{x : \|x\|^2 < 1\}$. On account of the continuity of sample paths, any annulus $N = \{x : \epsilon < \|x\|^2 < 1\}$ is a cemetery neighbourhood.

Example 4.11. Let B be a Brownian motion on $E = \{x : \|x\|^2 < 1\}$ as in the previous example, and let $c(x) > 0$ be a bounded function on E. Define a Markov

process X by killing B according to the additive functional $A_t = \int_0^t c(B_s)ds$. The semigroup of X is given by the formula

$$\mathbb{P}_x(f(X_t), \zeta > t) = \mathbb{P}_x(f(B_t)e^{-\int_0^t c(B_s)ds}, T_{\partial E} > t), \quad x \in E,$$

here $T_{\partial E} = \inf\{t > 0 : |B_t| = 1\}$. If U is any nonempty open subset of E, the probability that X_t belongs to U at its moment of death is strictly positive. Hence the only cemetery neighbourhood is $N = E$.

Example 4.12. Let Y be a Markov chain on $E = \{0, 1, 2, 3, \dots\}$ and suppose that Y gets absorbed in 0 in a finite time. Define X as the Markov chain on $E = \{1, 2, 3, \dots\}$ which is constructed by killing Y at the first hitting time of state 0. A cemetery neighbourhood is given by the set of states from which the process Y can directly jump to zero, namely

$$N = \{y > 0 : q_{y0} > 0\};$$

here (q_{ij}) is the q-matrix of Y. The set we defined is clearly the smallest possible cemetery neighbourhood.

Example 4.13. Let X be an explosive pure birth process on the countable set $E = \{1, 2, 3, \dots\}$. The lifetime of X coincides with the explosion time, that is

$$\zeta = \inf\{t > 0 : |X_t| = \infty\}.$$

A typical cemetery neighbourhood is given by $N = \{n, n+1, n+2, \dots\}$. There exists no 'smallest' such set.

The importance of cemetery neighbourhoods for our study stems from the following fact. We denote by $\overline{N}$ the closure, in the Martin topology, of the cylinder set $(-\infty, 0] \times N$.

Lemma 4.14. *If h is a limit point of (h_s), then a representing measure for h is concentrated on $\overline{N} \cap \overline{F}_{-\infty}$.*

Proof. Let μ be a representing probability measure for h; there exists a sequence (s_n) such that $\mu_{s_n} \Rightarrow \mu$, where μ_s is a representing measure for h_s. Consider the

event $\{\overline{X}_{z-} \in \overline{N}, z < \infty\}$; a rephrasing of Definition 4.9 in terms of the spacetime process gives

$$\overline{\mathbb{P}}_{\overline{x}}(\overline{X}_{z-} \in \overline{N}, z < \infty) = 1, \quad \overline{x} \in (-\infty, 0] \times E.$$

From the proof of Lemma 4.5, we known that $\overline{\mathbb{P}}^{h_s}_{\overline{x}}(z \geq -s) = 1$; we also have $\overline{\mathbb{P}}^{h_s}_{\overline{x}}(z < \infty) = 1$, since the constancy of h_s on $(-\infty, s) \times E$ implies that

$$\begin{aligned}
\lim_{u\to\infty} \overline{\mathbb{P}}^{h_s}_{\overline{x}}(z > u) &= \lim_{u\to\infty} \mathbb{P}_x(h_s(t-u, X_u), \zeta > u)/h_s(t,x) \\
&= \lim_{u\to\infty} c\mathbb{P}_x(\zeta > u) \\
&= 0,
\end{aligned}$$

where $c^{-1} = h_s(t,x)\mathbb{P}_\nu(\zeta > -s)$. Hence for $0 < r < -s$ we have

$$\begin{aligned}
\overline{\mathbb{P}}^{h_s}_{\overline{x}}(\overline{X}_{z-} \in \overline{N}, z < \infty) &= \overline{\mathbb{P}}^{h_s}_{\overline{x}}(\overline{X}_{z-} \in \overline{N}, r < z < \infty) \\
&= \overline{\mathbb{P}}_{\overline{x}}(h_s(\overline{X}_r), \overline{X}_{z-} \in \overline{N}, r < z < \infty)/h_s(\overline{x}) \\
&= \overline{\mathbb{P}}_{\overline{x}}(h_s(\overline{X}_r), z > r)/h_s(\overline{x}) \\
&= 1.
\end{aligned}$$

But by (2.7) of Chapter 2, this means that the representing measure μ_s is concentrated on $\overline{N}$ (in fact $\overline{N} \cap \overline{F}_s$). Indeed, if A is any set such that $A \cap \overline{N} = \emptyset$,

$$0 = \overline{\mathbb{P}}_{\overline{x}}(\overline{X}_{z-} \in A) = \int_A \overline{K}(\overline{x}, \overline{y})\mu_s(d\overline{y}),$$

and since $\overline{x}$ was arbitrary, this follows as in the proof of Lemma 4.5. Now by weak convergence, since $\overline{N}$ is closed,

$$\mu(\overline{N}) \geq \overline{\lim}_{s_n\to\infty} \mu_{s_n}(\overline{N}) = 1,$$

and this completes the proof. □

The last result means that, in a study of h we can restrict attention to those minimal parabolic functions $\overline{K}(\cdot, \overline{y})$ arising out of sequences $\overline{y} = \lim_{n\to\infty}(s_n, y_n)$ such that $y_n \in N$ for all n.

We now introduce the assumption which will allow us to prove that (4.5) holds.

Assumption IV: There exists a cemetery neighbourhood N with the following property: for each $r > 0$, there exist $C(r), T(x) > 0$ such that

$$p_t(x, y) \leq C(r) p_{t+r}(x, z), \quad t > T(x), \quad x \in E, \quad y, z \in N.$$

Bearing in mind that $(t, x) \mapsto p_t(x, y)$ is parabolic in $(0, \infty) \times E$, Assumption IV is a kind of "reverse" Harnack inequality. More will be said about this later.

The following result is an adaptation in our context of a result due to Koranyi and Taylor (1985). See Ancona (1990) and also Lyons and Sullivan (1984).

Proposition 4.15. *Suppose that Assumption IV holds. Every minimal parabolic function $\overline{K}(\cdot, \overline{y})$ with $\overline{y} \in \overline{N} \cap \overline{F}_{-\infty}$ is of the form*

$$\overline{K}\Big((t, x); \overline{y}\Big) = e^{\lambda t} g(x),$$

where g is a minimal positive solution to the equation $\mathfrak{A}g = \lambda g$ in E.

Proof. By the assumption, if $s < t \leq 0 < r$, $s < -T(x)$,

$$\begin{aligned} \overline{K}\Big((t - r, x); (s, y)\Big) &= \frac{p_{t-r-s}(x, y)}{\int \nu(dx) p_{-s}(x, y)} \\ &\leq C(r) \cdot \frac{p_{t-s}(x, y)}{\int \nu(dx) p_{-s}(x, y)} \\ &= C(r) \overline{K}\Big((t, x); (s, y)\Big). \end{aligned}$$

Suppose now that $w(\overline{x}) = \lim_{n \to \infty} \overline{K}(\overline{x}, \overline{y}_n)$ is a minimal parabolic function corresponding to a sequence (s_n, y_n) satisfying $\lim_n s_n = -\infty$ and $y_n \in N$ for all n. For any $r > 0$, the function $w_r(t, x) = w(t - r, x)$ is also parabolic in $(-\infty, 0] \times E$, and the above computation shows that $w_r(t, x) = w(t - r, x) \leq C(r) w(t, x)$. By minimality of w (see the definition in Chapter 2, Section 2.3), this means there exists a constant $L(r)$ such that $w(t, x) = L(r) w(t - r, x)$. Now L satisfies $L(a + b) = L(a) L(b)$ and $L(0) = 1$. Moreover, L is continuous since $t \mapsto w(t, x)$ is. Hence $L(t) = e^{\lambda t}$ for some constant $\lambda \in \mathbb{R}$, and then $w(t, x) = e^{\lambda t} w(0, x)$. Finally, since w is parabolic, it satisfies the equation

$$\frac{\partial}{\partial t} w(t, x) = \mathfrak{A} w(t, x) \quad \text{in } (-\infty, 0] \times E,$$

and hence the function $g(x) = w(0, g)$ is an eigenfunction of $\mathfrak{A}$ with eigenvalue λ. For the minimality, put $g = k + l$ where both k and l are positive eigenfunctions with eigenvalue λ. The functions $k'(t, x) = e^{\lambda t}k(x)$ and $l'(t, x) = e^{\lambda t}l(x)$ are both parabolic, and dominated by w; hence they are constant multiples of w, and multiplying by $e^{-\lambda t}$ shows that k and l are both multiples of g. □

The above result does not require the full force of Assumption IV, but merely that $p_t(x, y) \leq C(r) \cdot p_{t+r}(x, y)$ for all $y \in N$. We now give some examples of processes satisfying Assumption IV.

Example 4.16. Let X be a uniformly elliptic diffusion (bounded coefficients) on a bounded open set with regular boundary; the transition density $p_t(x, y)$ (with respect to Lebesgue measure) is the fundamental solution of the parabolic operator $L - \partial/\partial t$, where $L = \mathfrak{A}$ is the generator of X. The regularity of the boundary ensures that $\lim_{x\to\partial E} p_t(x, y) = 0$. Extend L to a uniformly elliptic operator in all of $\mathbb{R}^d$. This is always possible when the coefficients of L are sufficiently smooth; it suffices to take continuous functions $\tilde{a}_{ij}(x)$ (with $\tilde{a}_{ij}(x) = \tilde{a}_{ji}(x)$), $\tilde{b}_j(x)$ which agree with $a_{ij}(x)$, $b_j(x)$ in E and have the same bounds above and below as these. Then set

$$\tilde{L}f(x) = \frac{1}{2}\sum_{i,j=1}^{d} \tilde{a}_{ij}(x)D_{ij}f(x) + \sum_{j=1}^{d} \tilde{b}_j(x)D_jf(x).$$

Now if

$$\tilde{p}_t(x, y) = \begin{cases} p_t(x, y) & \text{if } x, y \in E, \\ 0 & \text{otherwise,} \end{cases}$$

then the function $(t, x) \mapsto \tilde{p}_t(x, y)$ is parabolic on $(0, \infty) \times \mathbb{R}^d$ for $\tilde{L}$, and the function $(t, y) \mapsto \tilde{p}_t(x, y)$ is parabolic for the adjoint $\tilde{L}^*$ of $\tilde{L}$:

$$\tilde{L}^*f(x) = \frac{1}{2}\sum_{i,j=1}^{d} D_{ij}(\tilde{a}_{ij}f)(x) - \sum_{j=1}^{d} D_j(\tilde{b}_jf)(x)$$

Assuming that the parabolic Harnack inequality holds for this adjoint, there then exists a constant C such that

$$\tilde{p}_t(x, y) \leq C \cdot \tilde{p}_{t+r}(x, z), \quad y, z \in E, \quad t > \delta,$$

independently of x. Thus Assumption IV holds for this process.

The above example can be modified for diffusions on unbounded domains E, provided a sufficiently small cemetery neighbourhood exists. The following example illustrates the procedure.

Example 4.17. Let X be Brownian motion on $E = (0, \infty)$, killed upon first hitting zero. Its generator is $(1/2)d^2/dx^2$ on $C_K^2((0,\infty))$. The set $(0,1)$ is a cemetery neighbourhood, and the transition function of X is,

$$p_t(x,y) = \sqrt{2/\pi t}\exp\left(-\frac{x^2+y^2}{2t}\right)\sinh(xy/t), \quad x,y,t>0.$$

Since

$$\frac{p_t(x,y)}{p_{t+r}(x,y)} = \sqrt{\frac{t+r}{t}}\cdot\frac{\sinh(xy/t)}{\sinh(xy/t+r)},$$

we can take $C(r) = 4$, provided we take

$$T(x,r) = r \vee \inf\left\{t>0 : \frac{\sinh(x/t)}{\sinh(x/t+r)} \le 2\right\}.$$

The previous example can be generalized as follows:

Example 4.18. Let X be an elliptic diffusion process in an unbounded domain $E \subset \mathbb{R}^d$, and suppose that the lifetime of X coincides with the first hitting time of some compact set $K \subset E$. Any bounded open set D containing K is a cemetery neighbourhood. We shall assume that the generator is uniformly elliptic in some such set. If the boundary of K is sufficiently regular, and the coefficients of $L = \mathfrak{A}$ are sufficiently smooth, we can again argue in terms of the adjoint of L, as in the first example, that Assumption IV must hold.

We give one further example, within the realm of Markov chains.

Example 4.19. Let X be an irreducible Markov chain on a countable state space, and suppose that a finite cemetery neighbourhood N exists. Assumption IV must hold, since if $p_t(x,y)$ is the density of the transition function with respect to counting measure,

$$p_{t+r}(x,y) \ge p_t(x,z)p_r(z,y), \quad x \in E, \quad y,z \in N,$$

and we can take $C(r)^{-1} = \min_{z,y \in N} p_r(z,y) > 0$ by irreducibility. This argument also shows, by taking $y = z$, that the conclusion of Proposition 4.15 holds as soon as there exists a (not necessarily finite) cemetery neighbourhood N satisfying

$$\sup_{y \in N} |q(y)| < +\infty.$$

In particular, this is always true when the q-matrix is bounded.

Alternatively, Assumption IV may be seen as follows: let $\xi(dy)$ be an excessive measure on E with density $\xi(y)$ with respect to counting measure, that is

$$\sum_{x \in E} \xi(x) p_t(x,y) \leq \xi(y).$$

Such a measure always exists; then

$$\widehat{p}_t(x,y) = p_t(y,x)\xi(y)/\xi(x)$$

is the transition density of the minimal Markov chain $\widehat{X}$ with q-matrix $\widehat{q}(x,y) = q(y,x)\xi(y)/\xi(x)$. For each $y \in E$,

$$\frac{\partial}{\partial t}\widehat{p}_t(x,y) = \sum_{z \in E} \widehat{q}(x,z)\widehat{p}_t(z,y),$$

that is $(t,x) \mapsto p_t(x,y)$ is parabolic for $\widehat{X}$. By the parabolic Harnack inequality on the compact set N, $\widehat{p}_t(x,y) \leq C(r)\widehat{p}_{t+r}(z,y)$ for all $x, z \in N$ and fixed $y \in E$. It follows that

$$p_t(y,x) \leq C'(r) p_{t+r}(y,z), \qquad y \in E, x, z \in N,$$

where $C'(r) = \max_{x,z \in N} C(r)\xi(x)/\xi(z)$.

We remark that the above assumption, that a finite cemetery neighbourhood exists, also implies that the decay parameters Λ_* and Λ_1 coincide; this was shown in Jacka and Roberts (1995), Lemma 3.3.5.

We end this section by summarizing the results of our study of (h_s). In the next section, we will describe the implications for the conditioning procedure (4.1).

Proposition 4.20. *Suppose that Assumptions II, III, IV hold together with the parabolic Harnack inequality. If ν is a compactly supported probability measure on E, then the limit points of (h_s) are all of the form*

$$\lim_{r_n\to\infty} \frac{\mathbb{P}_x(\zeta > r_n + t)}{\mathbb{P}_\nu(\zeta > r_n)} = \int e^{\lambda t} g_\lambda(x)\mu(d\lambda),$$

where μ is a probability measure concentrated on $[\Lambda_*, 0]$ and $g_\lambda > 0$ satisfies $P_t g_\lambda = e^{\lambda t} g_\lambda$ (and hence $\mathfrak{A} g_\lambda = \lambda g_\lambda$) with $\langle \nu, g_\lambda\rangle = 1$.

Proof. The integral representation comes from Lemma 4.14, Proposition 4.15 and the results of Chapter 2, Section 2.3. The fact that μ must be concentrated on $[\Lambda_*, 0]$ is due to the fact that any limit point h of (h_s) is a decreasing function of the time variable (so that $\lambda \le 0$), as well as the results of Chapter 3 (which imply $\lambda \ge \Lambda_*$). For the normalization of g_λ, we first have by Lemma 4.6 (see also the discussion thereafter) and the dominated convergence theorem that $1 = \langle \nu, h(0,\cdot)\rangle$. Then, since $t \mapsto h(t,\cdot)$ is decreasing,

$$1 = \langle \nu, h(0,\cdot)\rangle \le \langle \nu, h(t,\cdot)\rangle = \int e^{\lambda t}\langle \nu, g_\lambda\rangle \mu(d\lambda).$$

Then by monotone convergence,

$$1 \le \int \langle \nu, g_\lambda\rangle \mu(d\lambda),$$

whereas Fatou's lemma shows that

$$\langle \nu, g_\lambda\rangle \le \lim_{n\to\infty} \langle \nu, \overline{K}(0,\cdot;\overline{y}_n)\rangle = 1,$$

where $(\overline{y}_n)$ is any sequence converging to the boundary point which represents the function $e^{\lambda t} g_\lambda(x)$. It follows that $\langle \nu, g_\lambda\rangle = 1$ a.e. with respect to μ. Finally, it is clear that $h^\lambda(t,x) = e^{\lambda t} g_\lambda(x)$ is spacetime invariant, or equivalently that $z = \infty$ a.s. under $\overline{\mathbb{P}}^{h^\lambda}$. Otherwise we would have $\overline{\mathbb{P}}^{h^\lambda}_{\overline{x}}(\zeta < \infty) > 0$, and therefore $\overline{X}_{\zeta-} \notin \overline{F}_s$ on $\{z \le t - s\}$ with positive probability. Then the representing measure for h^λ would not be concentrated on $\overline{F}_{-\infty}$, contradicting Lemma 4.5. Now

$$e^{\lambda t} g_\lambda(x) = \overline{\mathbb{P}}_{(t,x)}(h^\lambda(\overline{X}_r), z > r) = \mathbb{P}_x(e^{\lambda(t-r)} g_\lambda(X_r), \zeta > r),$$

and dividing by $e^{\lambda t}$ gives $P_r(x, g_\lambda) = e^{\lambda r} g_\lambda(x)$. □

For every probability measure μ on $\overline{F}_{-\infty} \cap \overline{N}$, the invariant function $(t,x) \mapsto \int e^{\lambda t} g_\lambda(x) \mu(d\lambda)$ represents a different way of exiting the set $\overline{N}$ via $\overline{F}_{-\infty}$. To make the limit in Proposition 4.20 independent of the sequence $r_n \to \infty$, we could show that $\overline{F}_{-\infty} \cap \overline{N}$ consists of a single minimal point. This will be done in the next section, where we also return to a discussion of the conditioning procedure (4.1).

4.6 Single Exit Conditions

In this section, we use the result of Proposition 4.20 to study the conditioning problem (4.1). We shall say that this problem has a sequential solution if any family (r_n), $r_n \to \infty$, has a subsequence $r_{n(k)}$ such that

$$\lim_{k\to\infty} \int H(\omega) d\mathbb{P}_\nu(\omega \,|\, \zeta > r_{n(k)}) = \int H(\omega) d\mathbb{Q}_\nu(\omega)$$

holds for some probability law $\mathbb{Q}_\nu$ on Ω and all bounded $\mathcal{F}_t$ measurable random variables, $t > 0$. Thus we are saying that the laws $\mathbb{P}_\nu(d\omega \,|\, \zeta > r)$ are sequentially compact in the space of all probability measures. If the limit law $\mathbb{Q}_\nu$ is independent of the particular chosen sequence, then we will say that the problem (4.1) has a (full) solution.

The main result of this section is the following theorem, which tells us that the problem (4.1) has a solution if the set $\overline{F}_{-\infty} \cap \overline{N}$ consists of a single minimal point. The remainder of this section will be a discussion of this condition, with examples relating it to existing theories.

Theorem 4.21. *Suppose that Assumptions II, III and IV hold, together with the parabolic Harnack inequality. If ν is a probability measure with compact support in E, and $\overline{F}_{-\infty} \cap \overline{N} = \{\overline{z}\}$ consists of a single minimal point, with corresponding invariant parabolic function*

$$\overline{K}(t,x;\overline{z}) = e^{\lambda t} g(x), \quad \langle \nu, g \rangle = 1,$$

then the conditioning problem (4.1) *has a solution given by*

$$\mathbb{Q}_\nu(d\omega) = e^{-\lambda t} g(X_t(\omega)) \cdot \mathbb{P}_\nu(d\omega) \quad \text{on } \mathcal{F}_t, \quad t \geq 0.$$

In particular, the coordinate process X is, under $\mathbb{Q}_\nu$, a time homogeneous Markov process with semigroup $Q_t(x,dy) = e^{-\lambda t}P_t(x,dy)g(y)/g(x)$ and initial distribution $g(x)\nu(dx)$.

Some comments are in order before we start the proof.

It is possible to characterize the eigenvalue λ by the formula

$$\lambda = \lim_{r\to\infty} \frac{1}{t} \log \mathbb{P}_\nu(\zeta > r) = \Lambda_1.$$

Indeed, this follows from the fact that (Proposition 4.20 and bounded convergence)

$$\lim_{r\to\infty} \frac{\mathbb{P}_\nu(\zeta > t+r)}{\mathbb{P}_\nu(\zeta > r)} = e^{\lambda t},$$

and the formula

$$\frac{1}{n}\log \mathbb{P}_\nu(\zeta > n) = \frac{1}{n}\sum_{k=0}^{n-1} \log\Big(\mathbb{P}_\nu(\zeta > k+1)/\mathbb{P}_\nu(\zeta > k)\Big);$$

see Jacka and Roberts (1997) or Collet et al. (1995).

We need to keep in mind that the set $\overline{N}$ is that referred to in Assumption IV. We shall list below a number of sufficient conditions for $\overline{F}_{-\infty}\cap\overline{N}$ to be a singleton. For example, it suffices that there exists a symmetrizing measure for the transition semigroup. Of course, Λ_*-recurrence is another sufficient condition.

Even if $\overline{F}_{-\infty}\cap\overline{N}$ consists of more than one point, the other assumptions of the theorem guarantee that the conditioning problem (4.1) always has a sequential solution. This improves on existing results, notably Jacka and Roberts (1995), who do not discuss what happens if (h_s) does not converge.

In view of the spacetime picture we have been working with, it is not surprising that the law $\mathbb{Q}_\nu$ is that of an h-transform of the spacetime process $\overline{X}$ with respect to the invariant function $h(t,x) = e^{\lambda t}g(x)$. We are simply conditioning $\overline{X}$ to hit that part of the Martin boundary which we denoted $\overline{F}_{-\infty}\cap\overline{N}$. In light of this, the interplay between X and the geometry of N and E encapsulated by Assumption IV play a major role in the conditioning problem, as was already noted by Kesten (1995) in his discussion of Yaglom limits. One can therefore

expect that sequential solutions are the best one can hope for in general, with full solutions representing rather extreme cases. This may explain that, when X is Λ-transient, most "successful" conditional limit theorems have been proved for effectively one-dimensional processes (birth-death processes or stochastically monotone processes are examples).

The time homogeneity of X under the law $\mathbb{Q}_\nu$, claimed in the above theorem, is largely due to Assumption IV which implies (4.5). If $\overline{F}_{-\infty} \cap \overline{N}$ does not consist of a single point, a typical sequential limit law ought to give rise to a time inhomogeneous Markov process, whose transition law will be

$$(4.6) \quad P(t,x;s,dy) = h(t,x)^{-1} p_{s-t}(x,y) h(s,y) m(dy), \quad x,y \in E, \quad s > t > 0,$$

where

$$h(t,x) = \int e^{\lambda t} g_\lambda(x) \mu(d\lambda).$$

The infinite lifetime of the conditioned process (equivalent to the spacetime invariance of $h(t,x)$) is intuitively obvious, but it should be borne in mind that the conditioning procedure does not yield a well defined Markov process for *arbitrary* initial distributions ν (see discussion after Lemma (4.6)); the integrability of g with respect to ν is crucial. Our choice of dealing only with compactly supported probability measures ν is dictated by the parabolic Harnack inequality, which underlies many of our proofs.

In the proof of Theorem 4.21 below, we will actually show that any sequential limit law $\mathbb{Q}_\nu$ gives rise to a Markov process with transition function (4.6). When μ is concentrated on a single point, the function h reduces to $\overline{K}(\cdot,\overline{z})$ as required.

Proof of Theorem 4.21. Let H denote any bounded $\mathcal{F}_t$ measurable random variable, $t \geq 0$, and suppose that (r_n) gives rise to a limit point h of (h_s) as in the previous section. Then if $\sigma_K = \inf\{t > 0 : X_t \notin K\}$ denotes the first exit time from a compact set $K \subset E$, the bounded convergence theorem gives

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}_\nu(H, \sigma_K > t \,|\, \zeta > r_n) &= \lim_{n\to\infty} \mathbb{P}_\nu\left(H, \sigma_K > t, \frac{\mathbb{P}_{X_t}(\zeta > r_n - t)}{\mathbb{P}_\nu(\zeta > r_n)}\right) \\ &= \mathbb{P}_\nu(H, \sigma_K > t, h(-t, X_t)). \end{aligned}$$

We now dispense with the set $\{\sigma_K > t\}$ above. Taking $H = 1$ and remembering that $\langle g_\lambda, \nu\rangle = 1$ gives

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}_\nu(\sigma_K \le t \,|\, \zeta > r_n) &= 1 - \lim_{n\to\infty} \mathbb{P}_\nu(\sigma_K > t \,|\, \zeta > r_n) \\
&= 1 - \int \mathbb{P}_\nu(\sigma_K > t, e^{-\lambda t} g_\lambda(X_t))\mu(d\lambda) \\
&= \int \mu(d\lambda)\nu(dx)\Big(g_\lambda(x) - \mathbb{P}_x(e^{-\lambda t} g_\lambda(X_t), \sigma_K > t)\Big) \\
&= \int \mu(d\lambda)\nu(dx)\mathbb{P}_x(e^{-\lambda t} g_\lambda(X_t), \zeta > t, \sigma_K \le t) \\
&= \mathbb{Q}_\nu(\sigma_K \le t),
\end{aligned}$$

where we have used the λ-invariance of g_λ. Now

$$\begin{aligned}
&\overline{\lim}_{n\to\infty} \big|\mathbb{P}_\nu(H \,|\, \zeta > r_n) - \mathbb{Q}_\nu(H)\big| \\
&\qquad \le \overline{\lim}_{n\to\infty} |\mathbb{P}_\nu(H, \sigma_K > t \,|\, \zeta > r_n) - \mathbb{Q}_\nu(H, \sigma_K > t)| \\
&\qquad\qquad + \overline{\lim}_{n\to\infty} 2\,\|H\| \Big(\mathbb{P}_\nu(\sigma_K \le t \,|\, \zeta > r_n) + \mathbb{Q}_\nu(\sigma_K \le t)\Big) \\
&\qquad = 4\,\|H\|\,\mathbb{Q}_\nu(\sigma_K \le t),
\end{aligned}$$

and, since $\sigma_K \uparrow \zeta$ as $K \uparrow E$ and $\zeta = \infty$ a.s. under $\mathbb{Q}_\nu$, the right hand side can be made arbitrarily small by choosing K arbitrarily large, the result follows. □

We now present some examples illustrating the scope of the 'single exit' assumption $\overline{F}_{-\infty} \cap \overline{N} = \{\overline{z}\}$. Throughout, we will suppose that the other conditions of Theorem 4.21 are met.

Suppose first that the decay parameter Λ_* of the Markov process X is zero. Since the interesting minimal points are represented by eigenfunctions of $\mathfrak{A}$ with eigenvalue $\lambda \in [\Lambda_*, 0] = \{0\}$ here, the points of $\overline{F}_{-\infty}$ correspond to the solutions g of the equation $\mathfrak{A}g = 0$ in E which also satisfy $P_t g = g$. This is illustrated in the next example.

Example 4.22. Let X denote the one dimensional Brownian motion on $E = (0, \infty)$, killed upon first hitting zero. We saw earlier that this process satisfies all the other assumptions of Theorem 4.21. Since its generator is given by $\mathfrak{A} = (1/2)d^2/dx^2$, the solutions to $\mathfrak{A}g = 0$ are given by $g(x) = a + bx$, for

some positive constants a and b. If $b = 0$, g is a multiple of the excessive function 1, which is not invariant for X since the process has finite lifetime under the h-transformed law $\mathbb{P}^1_x = \mathbb{P}_x$. The Martin compactification of X is well known to be $[0, +\infty]$, and the excessive function 1 corresponds here to the point 0. If $a = 0$, then g is a multiple of the excessive function x. Now this function is invariant, since the corresponding h-transform (with $h(x) = x$) is a three dimensional Bessel process with generator

$$\mathfrak{A}^h f(x) = h(x)^{-1}\mathfrak{A}(hf)(x) = \frac{1}{2}\frac{d^2}{dx^2}f(x) + \frac{1}{x}\frac{d}{dx}f(x),$$

and it is well known that this process is nonexplosive. Thus we see that $\overline{F}_{-\infty} \cap \overline{N}$ consists of only the function $g(x) = x/\int x\nu(dx)$. In terms of the Martin boundary of X, the excessive function x corresponds to the point $+\infty$. Clearly, the assumption $\Lambda_* = 0$ always implies that we can identify $\overline{F}_{-\infty}$ with a subset of the Martin boundary of the original process X, as occurs here. In particular, the conditioned process must converge to a point on the Martin boundary of X; here it is the point $+\infty$. The interpretation of the three dimensional Bessel process as a conditioned Brownian motion has a long history. See papers by McKean (1963), Williams (1974), Pitman (1975) and references therein. We end this example with some further remarks on the parabolic Martin boundary, taken from Doob (1984, p. 375). For every $\tau < 0$, set

$$K_0(t, x; \tau) = \begin{cases} \frac{x}{\sqrt{2\pi(t-\tau)^3}} \exp\left(-\frac{x^2}{2(t-\tau)}\right) & \text{if } t > \tau, \\ 0 & \text{if } t \le \tau. \end{cases}$$

For each $\gamma \le 0$, set

$$K_1(t, x; \gamma) = \begin{cases} \sinh(-\gamma x)\exp\left(\frac{\gamma^2 t}{2}\right) & \text{if } \gamma < 0, \\ x & \text{if } \gamma = 0. \end{cases}$$

Without changing notation, we will assume that the functions K_0 and K_1 are properly normalized, that is $\langle \nu, K_0(0, \cdot; \tau)\rangle = \langle \nu, K_1(0, \cdot; \gamma)\rangle = 1$. The Martin sequences are as follows: if $\overline{y}_n \to (\tau, 0)$, then $\lim_n \overline{K}(\cdot, \overline{y}_n) = K_0(\cdot; \tau)$; if $s_n \to -\infty$

and $y_n/s_n \to \gamma \leq 0$, then $\lim_n \overline{K}(\cdot, \overline{y}_n) = K_1(\cdot, \gamma)$, and if either $y_n \to +\infty$ with $y_n/(1+|s_n|) \to +\infty$ or else $s_n \to 0$ with no restriction on y_n, then $\lim_n \overline{K}(\cdot, \overline{y}_n) = 0$. Every positive parabolic function u with $\underline{\lim}_{s\to 0} u(s,x) < +\infty$ then has the Martin representation

$$u(t,x) = \int K_0(t,x;\tau)\mu_0(d\tau) + \int K_1(t,x;\gamma)\mu_1(d\gamma).$$

For the function which interests us, namely $h_s(t,x)$, the quickest way to get this representation explicitly is to use the Bachelier-Levy formula

$$\mathbb{P}_x(\zeta > r) = \int_r^\infty \frac{x}{\sqrt{2\pi u^3}} \exp\bigl(-x^2/2u\bigr) du,$$

and make the change of variable $\tau = r - u$. One then finds that

$$\begin{aligned} \frac{\mathbb{P}_x(\zeta > t-s)}{\mathbb{P}_\nu(\zeta > -s)} &= \int_{-\infty}^s \frac{K_0(t,x;\tau)}{\langle \nu, K_0(0,\cdot;\tau)\rangle} \left(\int \nu(dz) \frac{z}{\sqrt{-2\pi\tau^3}} \exp(z^2/2\tau) d\tau \right) \\ &= \int \overline{K}\Bigl(t,x;(\tau,0)\Bigr) 1_{(-\infty,s]}(\tau)\mu_s(d\tau), \end{aligned}$$

and a representing probability measure μ_s is concentrated on $\overline{F}_s$ as predicted. The set $\overline{F}_{-\infty}$ here consists of the half-line $\gamma \leq 0$, where each point γ is identified with the function $K_1(\cdot,\gamma)$. A cemetery neighbourhood is given by the set $(0,1) \subset E$. Now the points belonging to $\overline{F}_{-\infty} \cap \overline{N}$ must be arrived at through sequences (s_n, y_n) such that $y_n < 1$ for all n. In view of the characterization of Martin sequences above, every such sequence must give the function $K_1(t,x;0) = x$ (up to normalization).

Example 4.23. Suppose that X is Λ_*-recurrent. As described in Chapter 3, there exists a unique function $g > 0$ on E such that $P_t g = e^{\lambda t} g$ holds for some $\lambda \leq 0$. This function is associated with the eigenvalue $\lambda = \Lambda_*$. The assumptions of Theorem 4.21 are satisfied for both Markov chains on finite state space and uniformly elliptic diffusions on bounded domains, see Chapter 3. In those cases, the set $\overline{F}_{-\infty}$ consists of a single minimal boundary point $\overline{z}$, which satisfies $\overline{K}(t,x;\overline{z}) = e^{\Lambda_* t} g(x)$. This may be seen as follows. For each fixed $y \in E$, the Λ_*-recurrence implies that we can recover g by the formula

$g(x) = \lim_{t\to\infty} e^{-\Lambda_* t} p_t(x,y)$; see Chapter 3. Now

$$\lim_{s\to-\infty} \overline{K}(t,x;(s,y)) = \lim_{s\to-\infty} \frac{e^{\Lambda_* s} p_{t-s}(x,y)}{e^{\Lambda_* s}\int \nu(dz) p_{-s}(z,y)} = e^{\Lambda_* t} g(x)/\langle \nu, g\rangle,$$

which means that all sequences (s_n, y) with $s_n \to -\infty$ tend to $\overline{z}$ in the Martin topology. On its own, this is not enough to show that $\overline{F}_{-\infty} = \{\overline{z}\}$, for we haven't considered the sequences (s_n, y_n) where y_n may vary. However, according to Proposition 4.24 below, this is not necessary, since both classes of processes considered here satisfy Assumption IV with $N = E$.

Proposition 4.24. *If Assumption IV holds, then the following three statements are equivalent:*

(i) For some $y \in N$,

$$\underline{\lim}_{s\to-\infty} \overline{K}\big(\overline{x},(s,y)\big) > 0$$

holds for all $\overline{x}$ in a set of positive η-measure,

(ii) The set $\overline{F}_{-\infty} \cap \overline{N}$ contains a single nontrivial, minimal parabolic boundary point $\overline{z}$,

(iii) For every sequence $\overline{y}_n = (s_n, y_n)$ such that $s_n \to -\infty$ and $y_n \in N$ for all n,

$$\lim_{n\to\infty} \overline{K}(\overline{x}, \overline{y}_n) = \overline{K}(\overline{x}, \overline{z}) = e^{\lambda t} g(x).$$

Proof. By Assumption IV, if $\epsilon > 0$ is fixed, we have the inequality

$$\begin{aligned}\overline{K}\Big((t,x);(s,y)\Big) &= \frac{p_{t-s}(x,y)}{\int \nu(dw) p_{-s}(w,y)} \\ &\le C(\epsilon)^{-2} \frac{p_{t-s+\epsilon}(x,z)}{\int \nu(dw) p_{-s-\epsilon}(w,z)} \\ &= C(\epsilon)^{-2} \overline{K}\Big((t+2\epsilon,x);(s+\epsilon,z)\Big),\end{aligned}$$

for all $y, z \in N$ and $t-s, -s > T(x,\nu)$. Now let $(s_n+\epsilon, z_n) \to \overline{z}$, where $\overline{z}$ is any nonzero minimal parabolic boundary point belonging to $\overline{F}_{-\infty} \cap \overline{N}$. The function $k(\overline{x}) = \underline{\lim}_{s\to-\infty} \overline{K}(\overline{x},(s,y))$ satisfies $\overline{P}_r k \le k$ by Fatou's lemma. It is thus

bounded below by the excessive function $\tilde{k} = \lim_{r\to 0} \overline{P}_r k$, from which it differs at most on a set of η-measure zero; the minimality of $\overline{z}$ together with the bound $k(t-2\epsilon, x) \le C(\epsilon)^{-2}\overline{K}((t,x),\overline{z})$ implies that $\tilde{k}(t-2\epsilon,x)$ is a constant multiple of $\overline{K}((t,x),\overline{z})$. Since $\tilde{k}$ is necessarily parabolic, we have $\tilde{k}(t,x) = e^{c\epsilon}k(t-2\epsilon,x)$ for some constant c, when $\tilde{k}$ is also a multiple of $\overline{K}(\cdot,\overline{z})$. If $\overline{z}'$ is another minimal parabolic boundary point, the same argument shows that $\overline{K}(\cdot,\overline{z}')$ is a constant multiple of $\tilde{k}$, and hence of $\overline{K}(\cdot,\overline{z})$. Thus there exists at most one mimimal point in $\overline{F}_{-\infty} \cap \overline{N}$, and hence all sequences converge to it. Conversely, if there are two distinct minimal points $\overline{z}, \overline{z}' \in \overline{F}_{-\infty} \cap \overline{N}$, then $\tilde{k}(\cdot)$ must be proportional to both $\overline{K}(\cdot,\overline{z})$ and $\overline{K}(\cdot,\overline{z}')$, and hence identically zero. □

If X is Brownian motion and E is the unit ball in $\mathbb{R}^d$, then this process is positive Λ_*-recurrent, as explained in the last chapter. The parameter $\Lambda_* = \Lambda_1$ is the principal eigenvalue of $\frac{1}{2}\Delta$ with Dirichlet conditions on ∂E, and therefore is nonzero here. Unlike the earlier example, the parabolic Martin boundary $F_{-\infty}$ is not identifiable with the ordinary Martin boundary of X, which is well known to be ∂E here. In particular, the conditioned process does not converge to any point on ∂E (as is also to be expected by the rotational symmetry of X).

Example 4.25. Let X be a uniformly elliptic diffusion on an open set $E \subset \mathbb{R}^d$, not necessarily bounded. Assume that the generator is in divergence form,

$$Lf(x) = \sum_{i,j=1}^{d} \frac{\partial}{\partial x^i}\left(a_{ij}(x)\frac{\partial f}{\partial x^j}\right)(x), \quad f \in C^2(E).$$

Bass and Burdzy (1992) have shown that, if the set E is given locally by the graph of an L^p function with $p > d-1$, then the following parabolic boundary Harnack principle holds: for every $u > 0$, there exists $C(u)$ such that

$$\frac{p_a(y,x)}{p_a(z,x)} \ge C(u) \cdot \frac{p_b(y,v)}{p_b(z,v)}, \quad a, b \ge u, \tag{4.7}$$

for all $v, x, y, z \in E$. See their paper and references therein for a precise definition of L^p domains, and for other conditions ensuring the validity of (4.7).

Suppose now that a slightly weaker form of (4.7) holds, namely for v, x merely belonging to N. Integrating both sides of (4.7) over $y \in E$ with respect to the

measure $\int \nu(dw)p_r(w,\cdot)m(\cdot)$ and inverting gives

$$\frac{p_a(z,x)}{\int \nu(dw)p_{a+r}(w,x)} \le C(u)\cdot \frac{p_b(z,v)}{\int \nu(dw)p_{b+r}(w,v)}, \quad a,b \ge u, \quad x,v \in N, \quad z \in E.$$

Changing variables according to $t = -r$, $-s' = a + r$, $-s' = b + r$ we find $\overline{K}\big((t,z);(s',x)\big) \le C(u)\overline{K}\big((t,z);(s,v)\big)$, and this is enough to guarantee that $\overline{F}_{-\infty}\cap\overline{N}$ consists of a single minimal boundary point. Indeed, let (s_n, z_n) converge to some minimal $\overline{z} \in F_{-\infty} \cap \overline{N}$ (which exists by Assumption III), and let (s'_n, y_n) converge to any parabolic point $\overline{y}$ in $F_{-\infty}$ with the correct normalization. The inequality ensures that

$$e^{\lambda' t}g'(z) = \overline{K}((t,z),\overline{y}) \le C(u)\cdot \overline{K}((t,z),\overline{z}) = e^{\lambda t}g(z),$$

and the minimality of $\overline{z}$ ensures that these two parabolic functions are proportional. Thus there exists a constant c such that $e^{\lambda' t}g'(z) = c\cdot e^{\lambda t}g(z)$. Integrating both sides with respect to $\nu(dz)$ gives $c = 1$, and hence $\overline{y} = \overline{z}$. Since $\overline{y}$ was arbitrary, the set $F_{-\infty}$ must consist of a single minimal point with the properties required by Theorem 4.21.

In our next example, we return to Markov chains, and expand on Proposition 4.24.

Example 4.26. Suppose that X is a Markov chain on a countable state space E, whose transition function is symmetric with respect to the measure m; here m is the measure used in Assumption II. The symmetry of X means that $p_t(x,y) = p_t(y,x)$ for the transition density associated with m by Assumption II.

In dealing with Markov chains, it is usual to work with respect to counting measure, so that m has the representation $m(A) = \sum_{x\in A} m(x)$. Let us denote by $p_{xy}(t)$ the transition density of $P_t(x,dy)$ with respect to counting measure. We then have the formula $p_t(x,y) = p_{xy}(t)/m(y)$, so that the symmetry requirement becomes the familiar formula

$$m(x)p_{xy}(t) = m(y)p_{yx}(t), \quad x,y \in E.$$

Kendall (1959) showed the existence, for each $x, y \in E$, of finite signed measures $\mu(x, y; d\lambda)$ on $(-\infty, 0]$ such that

$$p_t(x, y) = \int_{-\infty}^{0} e^{\lambda t} \mu(x, y, d\lambda). \tag{4.8}$$

It is also known that the measure $\mu(x, x; d\lambda)$ is positive for $x \in E$, and that, for each $y \in E$, the signed measure $\mu(x, y; d\lambda)$ is absolutely continuous with respect to the measure $\mu(y, y; d\lambda)$, independently of $x \in E$. Let us now fix once and for all a state $y \in N$, and denote the bounded measure $\mu(y, y; d\lambda)$ simply by $\mu(d\lambda)$. Since

$$\int_0^\infty e^{\epsilon t} p_t(y, y) dt = \int_{-\infty}^{0} \left(\int_0^\infty e^{(\lambda+\epsilon)t} dt \right) \mu(d\lambda),$$

it is obvious that μ is supported in the set $(-\infty, \Lambda_*]$. Then we can write

$$\overline{K}\Big((t, y); (s, y)\Big) = \frac{\int_{-\infty}^{\Lambda_*} e^{\lambda(t-s)} \mu(d\lambda)}{\int_{-\infty}^{\Lambda_*} e^{-\lambda s} \mu(d\lambda)}, \tag{4.9}$$

if the normalizing measure ν is taken as the point mass at y.

This representation is all we need to prove that the hypotheses of Proposition 4.24 hold.

Theorem 4.27. *Suppose that the transition function $p_t(x, y)$ is symmetric. Then the set $\overline{F}_{-\infty} \cap \overline{N}$ is a singleton.*

Proof. Consider the probability measures

$$\gamma_s(d\lambda) = \frac{e^{-\lambda s} \mu(d\lambda)}{\int_{-\infty}^{\Lambda_*} e^{-\theta s} \mu(d\theta)}.$$

Using the bound $\int_{-\infty}^{\Lambda_*} e^{-\lambda s} \mu(d\lambda) \geq e^{-(\Lambda_* - \epsilon)s} \mu((\Lambda_* - \epsilon, \Lambda_*])$, it follows that

$$\gamma_s((-\infty, \Lambda_* - \epsilon]) \leq \mu((\Lambda_* - \epsilon, \Lambda_*])^{-1} e^{(\Lambda_* - \epsilon)s} \int_{-\infty}^{\Lambda_* - \epsilon} e^{-\lambda s} \mu(d\lambda),$$

and this tends to zero as $s \to -\infty$. Consequently, the measures γ_s are tight, and converge weakly to the point mass at Λ_*. If the normalizing measure is the point mass at y, it follows that

$$\begin{aligned} \lim_{s \to -\infty} \overline{K}\Big((t, y); (s, y)\Big) &= \lim_{s \to -\infty} \int e^{\lambda t} \gamma_s(d\lambda) \\ &= e^{\Lambda_* t} > 0. \end{aligned}$$

Since the set $(-\infty, 0) \times \{y\}$ has positive η-measure, it follows from Proposition 4.24 that $\overline{K}(\cdot, \overline{y}_n) \to \overline{K}(\cdot, \overline{z})$ as $\overline{y}_n \to \overline{F}_{-\infty}$ while remaining in $\overline{N}$. This convergence is clearly unaffected if we integrate both sides with respect to some compactly (finitely) supported measure ν, that is we shall have $\langle \nu, \overline{K}((0, \cdot), \overline{y}_n)\rangle \to \langle \nu, \overline{K}((0, \cdot), \overline{z})\rangle$. Finally, we can remove the assumption that $\overline{K}((0, y), \overline{y}) = 1$, by writing

$$\begin{aligned}\lim_{n\to\infty} \frac{\overline{K}(\overline{x}, \overline{y}_n)}{\langle \nu, \overline{K}(0, \cdot; \overline{y}_n)\rangle} &= \lim_{n\to\infty} \frac{\overline{K}(\overline{x}, \overline{y}_n)/\overline{K}(0, y; \overline{y}_n)}{\langle \nu, \overline{K}(0, \cdot; \overline{y}_n)\rangle/\overline{K}(0, y; \overline{y}_n)} \\ &= \frac{\overline{K}(\overline{x}, \overline{z})}{\langle \nu, \overline{K}(0, \cdot; \overline{z})\rangle}.\end{aligned}$$

Applying Proposition 4.24 gives the result. □

Example 4.28. Let $0 < d < +\infty$, and denote by E the set of strictly positive integer-valued bounded measures on the set $C = \{1, \ldots, d\}$. Thus a typical measure in E assigns an integer $\mu(\{k\})$, which we shall interpret as counting the number of individuals in the k-th group or colony. The set E is a metric space under total variation distance, that is $\mathrm{dist}(\mu, \nu) = \sum_{k\in C} |\mu(\{k\}) - \nu(\{k\})|$, and is clearly denumerable.

An open migration process on E is a Markov chain X_t with the following dynamics (see Kelly (1979)): from any state $\mu \in E$, the process may jump to neighbouring states $\mu + \epsilon_k$ (interpreted as a birth in the k-th colony), $\mu - \epsilon_j$ if $\mu(\{j\}) > 0$ (interpreted as a death in the j-th colony), or $\mu - \epsilon_j + \epsilon_k$ (interpreted as the movement of some individual from colony j to colony k). Clearly such a Markov chain has compactly supported jumps. The lifetime coincides with the first time X_t exits E:

$$\zeta = \inf\{t > 0 : X_t(C) = 0 \text{ or } X_t(C) = +\infty\}.$$

The following q-matrix specifies the above behaviour:

$$\begin{aligned} q(\mu, \mu + \epsilon_k) &= \alpha_k \psi_k(\mu(\{k\})), \\ q(\mu, \mu - \epsilon_j) &= \beta_j \phi_j(\mu(\{j\})), \\ q(\mu, \mu - \epsilon_j + \epsilon_k) &= \gamma_{jk} \phi_j(\mu(\{j\})), \\ q(\mu, \mu) &= -\sum_{j,k \in C} q(\mu, \mu + \epsilon_k) + q(\mu, \mu - \epsilon_j) + q(\mu, \mu - \epsilon_j + \epsilon_k). \end{aligned}$$

Here (α_k), (β_j) and (γ_{jk}) are positive numbers and (ψ_k), (ϕ_j) are positive functions. We shall suppose that they have been chosen so that X is irreducible, i.e. every colony is accessible from any other, possibly via a chain of colonies, and that the birth and death parameters allow for unbounded growth (resp. complete death) of the total population.

Suppose now that there exist strictly positive numbers $(a_1, \ldots, a_d)$ satisfying the equations

$$a_j \gamma_{jk} = a_k \gamma_{kj}, \quad a_j \beta_j = \alpha_j, \quad j, k \in C.$$

The detailed balance equations $m(\mu)q(\mu, \nu) = m(\nu)q(\nu, \mu)$ on E have a solution (Kelly (1979))

$$m(\mu) = \prod_{j=1}^{d} \left(a_j^{\mu(\{j\})} \prod_{r=1}^{\mu(\{j\})} \frac{\psi_j(r-1)}{\phi_j(r)} \right).$$

and this defines a symmetrizing measure $m(d\mu)$, provided each $m(\mu)$ is finite. We can ensure that $\mathbb{P}_\mu(X_t(C) = 0 \text{ eventually}) = 1$ for any initial state (population configuration) μ, with a suitable choice of the parameters. In that case, a cemetery neighbourhood is given by the set $N = \{\epsilon_k : k \in C\}$, whose cardinality is at most d. It follows that Assumption IV holds and according to Theorem 4.27, the conditioning problem has a solution.

In the proof of Theorem 4.27, all that we have really used is the fact that the function $f(t) = p_t(y, y)$ is the Laplace transform of some finite positive measure. This is guaranteed by Bernstein's theorem on completely monotonic functions (see Widder (1941) whenever the function f satisfies the following conditions:

$$f(0+) < \infty, \quad (-1)^k \frac{d^k}{dt^k} f(t) \geq 0 \quad k = 0, 1, 2, \ldots$$

For one-dimensional diffusions, it was shown by McKean (1956) that the representation (4.8) is valid. Indeed, in a follow-up paper (McKean (1963)), the conditioning problem (4.1) was shown to have a solution. For other symmetric processes, one would use the spectral representation to get a similar result.

Kijima (1993) showed that a representation of $p_t(y,y)$ as the Laplace transform of a bounded measure also holds for Markov chains which are skip-free to the left on $E = \{1, 2, 3 \dots\}$. Thus for such processes too, the conditioning problem has a solution.

Consider now the following example, due to Jacka and Roberts (1995), of a process for which the conditioning problem *does not* have a solution.

Example 4.29. Let X be the Markov chain on $E = \{1, 2, 3, \dots\}$ with nonconservative q-matrix given by

$$q(x,y) = \begin{bmatrix} -1 & 2^{-2} & 2^{-3} & 2^{-4} & \cdots \\ 2^{-2} & -2^{-2} & & & \\ 2^{-3} & & -2^{-3} & & \\ 2^{-4} & & & -2^{-4} & \\ \vdots & & & & \ddots \end{bmatrix}.$$

When started in state 1, X waits for an exponential time with mean 1 before either jumping to state k with probability 2^{-k+1} or getting killed with probability $1/2$. In state $k > 1$, it first waits for an exponential time with mean 2^{k+1} and then jumps back to state 1. Thus the process is irreducible, and the smallest cemetery neighbourhood is given by $N = \{1\}$. In particular, Assumptions II, IV and the parabolic Harnack inequality are satisfied. Moreover, X is clearly symmetric, so that $\overline{F}_{-\infty} \cap \overline{N}$ consists of at most one point. Nevertheless, the quasistationary conditioning procedure does not work. Indeed, suppose that $g(x)$ is a positive solution to the equation $Qg = \lambda g$. We must therefore solve the system

$$\begin{aligned} -g(1) + \sum_{k=2}^{\infty} 2^{-k} g(k-1) &= \lambda g(1), \\ 2^{-k} g(1) - 2^{-k} g(k-1) &= \lambda g(k-1), \quad k \geq 2. \end{aligned}$$

Clearly the only solution is $g(x) = 0$. Now Proposition 4.15 guarantees that the minimal parabolic functions associated with points of $\overline{F}_{-\infty} \cap \overline{N}$ are of the form $e^{\lambda t} g(x)$; thus there are no nonzero minimal parabolic points in $\overline{F}_{-\infty} \cap \overline{N}$ for this process, and in particular the sequence (h_s) cannot converge. The source of this failure lies in the fact that Assumption III, which requires that the jumps of X be bounded, does not hold. It is this assumption which guarantees that there exists at least one parabolic function in $\overline{F}_{-\infty}$. Here, the process can jump into each and every state from state 1.

We end this chapter with one last remark. Jacka and Roberts (1995) gave a number of equivalent conditions for (4.1), among them the existence of a Yaglom limit

$$\lim_{r \to \infty} \mathbb{P}_x(f(X_r) \,|\, \zeta > r) = \langle \kappa, f \rangle,$$

where κ is a probability measure, independent of the initial position $x \in E$. When this holds, it suffices to take $f(x) = \mathbb{P}_x(\zeta > t)$ to find that

$$\lim_{s \to -\infty} \frac{\mathbb{P}_x(\zeta > t - s)}{\mathbb{P}_\nu(\zeta > -s)} = e^{\lambda t} g(x),$$

with $e^{\lambda t} g(x)$ invariant, and consequently the conditioning problem (4.1) has a solution.

This also means that the existence of some λ-invariant function is necessary for the existence of a Yaglom limit. The converse is not true. Recalling the last example above, since the q-matrix does not have any strictly positive eigenfunctions, it cannot have a Yaglom limit either.

In the next chapter, we shall discuss some necessary conditions for the existence of the Yaglom limit.

4.7 Open Questions

With reference to Section 4, are there simple conditions which can be used to drop the bounded jump Assumption III? In the last part of that section, it is stated that one could look for an excessive function g which majorizes (h_s). This

function, if it exists, will almost never be bounded, unless the process X turns out positive Λ_*-recurrent (by an argument involving Tweedie's Test).

We have restricted our attention to initial distributions ν which are compactly supported. This was due mainly to the local nature of the parabolic Harnack inequality. If this inequality were to hold on all of E, the results would apply to any initial distribution. The counterexample at the end of Section 3 shows that this can only occur for positive Λ_*-recurrent processes. However, one could still ask for a characterization of the initial distributions which would allow the conditioning procedure to work.

Another technical problem associated with Chapter 4 is the following: is the combination of Assumptions II, III and IV (together with the parabolic Harnack inequality) sufficient for success of the process level conditioning? The result stated in Proposition 4.24 offers strong evidence for this; however, I was only able to make progress under the further assumption that X has a symmetric transition density (Theorem 4.27). It seems to me that much weaker conditions, if any, should suffice.

Finally, there is the following open question: why does the conditioning procedure 4.1 automatically lead to a Λ_1-invariant function? Here the emphasis is on the location of the parameter Λ_1, near or at the bottom of the spectrum. A heuristic explanation can be given in the light of the results of this chapter. As was shown, the conditioned process Y may be identified with the process $\overline{X}$, Doob-conditioned to exit the Martin boundary in the window projected onto $\{-\infty\} \times E$ by a cemetery neighbourhood N. This set is in some sense closest to the cemetery state. When the conditioning procedure works, let $e^{\lambda t} g(x)$ be the parabolic function associated with the window. Since g is excessive for the original process X, we can consider the transformed semigroup $P_t^g(x, dy) = P_t(x, dy)g(y)/g(x)$. Under $\mathbb{P}_x^g$, the lifetime of X has mean $|\lambda|^{-1}$, and this is the shortest possible when compared with other λ-invariant functions. A challenging question here would be to ask if an underlying minimization principle operates. This would offer a very powerful method for checking the validity of (4.1). Indeed, the conditioning

problem might then be reduced to checking for the existence of some function g solving the problem

$$\text{Minimize } \{\lambda : \ \mathfrak{A}g = \lambda g, \quad g > 0 \text{ in } E\}.$$

Chapter 5

Compactifications and Yaglom Limits

The present chapter continues the study of the quasistationary tool box theorems via boundary theory. We are interested here in the existence of Yaglom limits. As was explained in the last chapter, it was shown by Jacka and Roberts (1995) that the existence of a Yaglom limit implies the success of the quasistationary conditioning procedure (4.1). We shall now ask the converse question. Specifically, suppose that

$$\lim_{r\to\infty} \mathbb{P}_\nu(H_t \,|\, \zeta > r) = \mathbb{P}_\nu(H_t, e^{\Lambda t} g(X_t)), \quad H_t \in \mathcal{F}_t; \tag{5.1}$$

does it necessarily follow that there exists a probability measure κ on E such that

$$\lim_{r\to\infty} \mathbb{P}_\nu(f(X_r) \,|\, \zeta > r) = \langle \kappa, f \rangle \ ? \tag{5.2}$$

As we will see, the answer is negative in general, but we shall identify some sufficient conditions for the implication to hold. We fix an initial distribution ν on E. In the previous chapter, we found conditions that guarantee the validity of the following assumption, which we will now state formally:

Assumption V: There exists $\Lambda \leq 0$ such that

$$\lim_{r\to\infty} \mathbb{P}_\nu(\zeta > t + r \,|\, \zeta > r) = e^{\Lambda t}, \quad t \geq 0.$$

In particular, this occurs when (5.1) holds. We will also need Assumption II of Chapter 3, whose continuity hypothesis guarantees that $x \mapsto P_t f(x)$ is bounded and continuous whenever f is bounded with compact support in E.

5.1 Compactifications and Tightness

In this section, we begin a study of the family of probability measures (ν_t) defined by

$$\langle \nu_t, f \rangle = \mathbb{P}_\nu(f(X_t) \,|\, \zeta > t).$$

These measures are defined on E, and their convergence to some probability κ is precisely the meaning of (5.2). What we shall look for are conditions guaranteeing $\nu_t \Rightarrow \kappa$ in the sense of weak convergence of probability measures. Accordingly, we use the standard two step procedure. First, we show tightness of the family (ν_t), and then in the next section, we shall study the possible limit points of (ν_t).

According to Prokhorov's criterion, tightness of probability measures always occurs when these are defined on a common compact set. Since all the probabilities ν_t are defined on E, the family (ν_t) is tight on any compactification F of E, and the possible limit points are consequently all probability measures on F, which may or may not charge E. To find out where the probability mass ends up as $t \to \infty$, we require a compactification F with good probabilistic properties.

One requirement is that the topology induced by F on E be at least as fine as the original topology. This is to prevent a relatively awkward class of test functions along which the Yaglom limit (5.2) would exist. The stated requirement on the topology induced by F on E guarantees that the Yaglom limit, if it exists at all, occurs along all originally continuous functions, without reference to the topology of F. This is important for there will in general exist many different compactifications (with correspondingly different topologies on the boundary $F \backslash E$) which satisfy our conditions.

To get a probabilistic understanding of tightness of (ν_t), we require the existence of some Markov process Y on F, whose restriction to E coincides with X.

This process should exhibit a continuous variation of the law of ζ, as a function of the initial position. According to the result below, all these conditions can be met using a suitable Ray-Knight compactification of E. This is intended as an existence result only. In applications, one would use any compactification F satisfying (i), (ii) and (iii) below.

Proposition 5.1. *Suppose Assumption II holds. There exists a compactification F' of E with the following properties:*

(i) Every compactly supported function on E which was originally continuous remains continuous on F';

(ii) There exists on F' a Strong Markov process (X'_t) extending (X_t): for every probability measure μ on F', there exists a probability measure $\mathbb{Q}_\mu$ on F'-path space such that

$$\mathbb{Q}_\mu(f_0(X'_{t_0})f_1(X'_{t_1})\cdots f_n(X'_{t_n})) = \mathbb{P}_\mu(f_0(X_{t_0})f_1(X_{t_1})\cdots f_n(X_{t_n})), \tag{5.3}$$

whenever μ is concentrated on E and $f_0, \ldots, f_n$ are positive functions on E.

(iii) The lifetime ζ' of X' has a Laplace transform which is continuous in x: equivalently, the function

$$x \mapsto U_p(x, F') = \int_0^\infty e^{-pt}\mathbb{Q}_x(\zeta' > t)dt \quad \textit{is continuous on } F'.$$

The items listed are standard properties of the Ray-Knight compactification procedure, as can be found in Getoor (1975), Dellacherie and Meyer (1987), Sharpe (1988) or Rogers and Williams (1994). This procedure is usually performed on initially Markovian processes, so we outline below the trivial changes required.

Proof. We start by turning X into an honest (conservative) process by adding a cemetery state ∂, isolated from E. Thus ζ coincides with the first hitting time of ∂. Now X is Markovian on $E \cup \{\partial\}$, and we let F be any Ray-Knight compactification of $E \cup \{\partial\}$. Since ∂ was isolated from E, it remains isolated in

F. Thus $F' = F\backslash\{\partial\}$ is again compact. By Assumption II, the resolvent (V_p) of X maps the set of uniformly continuous functions (in the original topology) into itself, and consequently part (i) follows (Getoor (1975)). On F, there exists a unique Markovian resolvent (U_p) with the following properties:

(a) $U_p f = V_p f$ on $E \cup \{\partial\}$, whenever $f = 0$ on $F\backslash(E \cup \{\partial\})$,

(b) $U_p : C_b(F) \to C_b(F)$, where $C_b(F)$ is the set of bounded continuous functions on F.

Associated with (U_p) is a Strong Markov process (Y_t) with semigroup (Q_t) on F satisfying (ii) with X' replaced by Y. In particular, if $T_\partial = \inf\{t > 0 : Y_t = \partial\}$ is the first time that Y hits ∂, we have

$$\mathbb{Q}_x(T_\partial > t) = \mathbb{P}_x(\zeta > t), \quad x \in E, \quad t \geq 0.$$

Thus we can construct X' by killing Y upon first hitting ∂, and (5.3) will hold. Clearly the semigroup of X' is just (Q_t) restricted to the set F'. Meanwhile, since the point ∂ is absorbing for Y, we have the identity $\mathbb{Q}_x(\zeta' > t) = \mathbb{Q}_x(Y_t \in F\backslash\{\partial\}) = Q_t(x, F')$. Now $1_{F'}$ is continuous on F because ∂ is isolated, so by (b), the function $V_p(x, F') = \int_0^\infty e^{-pt}\mathbb{Q}_x(\zeta' > t)dt$ is continuous on F, and a fortiori on F'. □

To distinguish the original from the extended process, we will keep the notation introduced above to state and prove the next result.

Theorem 5.2. *We suppose that Assumptions II and V hold. If ξ is any weak limit point of (ν_t) on F', then its mass is always concentrated on the set*

$$H = \{x \in F' : \mathbb{Q}_x(\zeta' > 0) = 1\}.$$

If $\Lambda < 0$, then ξ is also concentrated on

$$G = \{x \in F' : \mathbb{Q}_x(\zeta' < \infty) = 1\},$$

while if $\Lambda = 0$, the mass of ξ is entirely contained in

$$G' = \{x \in F' : \mathbb{Q}_x(\zeta' = +\infty) = 1\}.$$

Proof. Let ξ be a weak limit point of (ν_t), so there exists some sequence $t_n \to \infty$ such that $\nu_{t_n} \Rightarrow \xi$ weakly on F'. Since the function $U_p(x, F')$ is continuous on F', we have

$$\begin{aligned}\int_0^\infty e^{-ps}\mathbb{Q}_\xi(\zeta' > s)ds &= \int_0^\infty e^{-ps}\langle \xi, \mathbb{Q}_\cdot(\zeta' > s)\rangle ds \\ &= \langle \xi, U_p(\cdot, F')\rangle \\ &= \lim_{n\to\infty}\langle \nu_{t_n}, U_p(\cdot, F')\rangle.\end{aligned}$$

Since each probability measure ν_t is concentrated on E, we have by property (ii) of Proposition 5.1

$$\begin{aligned}\lim_{n\to\infty}\langle \nu_{t_n}, U_p(\cdot, F')\rangle &= \lim_{n\to\infty}\langle \nu_{t_n}, V_p(\cdot, E)\rangle \\ &= \lim_{n\to\infty}\int_0^\infty e^{-ps}\langle \nu_{t_n}, P_s(\cdot, E)\rangle ds \\ &= \lim_{n\to\infty}\int_0^\infty e^{-ps}\mathbb{P}_\nu(\zeta > t_n + s \,|\, \zeta > t_n)ds \\ &= \int_0^\infty e^{-ps}e^{\Lambda s}ds,\end{aligned}$$

and using Assumption V as well as the bounded convergence theorem. By the uniqueness of Laplace transforms, since both the functions $s \mapsto \mathbb{Q}_\xi(\zeta' > s)$ and $s \mapsto e^{\Lambda s}$ are right continuous, we find

$$\mathbb{Q}_\xi(\zeta' > s) = e^{\Lambda s}.$$

Now let $s \to 0$; we see that $\mathbb{Q}_\xi(\zeta' > 0) = 1$, and since $\mathbb{Q}_x(\zeta > 0) \leq 1$, the probability measure ξ must be concentrated on H. If we let $s \to \infty$ instead, we have two cases to consider. If $\Lambda < 0$, then by bounded convergence $\mathbb{Q}_\xi(\zeta' = \infty) = 0$, and hence ξ charges only G. If $\Lambda = 0$, then $\mathbb{Q}_\xi(\zeta' = \infty) = 1$ so that (once again, since $\mathbb{Q}_x(\zeta' = \infty) \leq 1$) ξ can only charge G'. □

In the next section, we shall proceed to an identification of the structure of the possible limit points ξ.

According to Theorem 5.2, there can never exist a Yaglom limit (5.2) (where κ is a probability measure) in the case $\Lambda = 0$, since any possible limit measure

ξ is then concentrated on that part of the boundary $F'\backslash E$ from which X' takes forever to die.

It is possible to classify some of the boundary points $x \in F'\backslash E$ as follows:

Definition 5.3. *(i) A point x is called asymptotically remote (from the cemetery state) if it forms a trap: $\mathbb{Q}_x(X'_t = x \quad \forall t) = 1$. This property is denoted (AR).*

(ii) A point x is called asymptotically proximate if it is regular for $\{\partial\}$: $\mathbb{Q}_x(\zeta' = 0) = \mathbb{Q}_x(T_\partial = 0) = 1$. This property is denoted (AP).

The terminology for (i) is due to Ferrari et al. (1995) and for (ii) it is due to Pakes (1995). The Blumenthal zero one law implies that every state $x \in F$ satisfies $\mathbb{Q}_x(\zeta' = 0) = 0$ or 1. Some of the remaining boundary points may be holding points, where X' waits an exponentially distributed time before jumping. If holding points do not exist on the boundary $F'\backslash E$, then each boundary point is either (AP), (AR), or else the sample path of X' must immediately hit E a.s. (after which it stays there until death); it is not clear whether it is worth classifying those types of points in a similar manner as above.

The value of the (AR)/(AP) classification is that it appears naturally in the conclusions of Theorem 5.2. Specifically, it is worth mentioning the following corollary:

Corollary 5.4. *Under the hypotheses of Theorem 5.2, suppose that each boundary point in $F'\backslash E$ is either (AP) or (AR). Then the measures (ν_t) are tight if and only if $\Lambda < 0$.*

Note that, if $\Lambda = 0$, (ν_t) cannot be tight regardless of how the boundary points in $F'\backslash E$ are classified. It is also worth stating explicitly here that tightness refers to the original topology, not the Ray-Knight topology where tightness is automatic.

Example 5.5. Let (R_t) be a uniformly elliptic diffusion on $\mathbb{R}^d$, and suppose we construct X by killing R upon first exit from the unit disc $E = \{x : \|x\| < 1\}$.

The Euclidean boundary points of E are well known to be regular for E^c, that is $\lim_{x\to\partial E} \mathbb{P}_x(\zeta > t) = 0$ for each $t > 0$. Set $F' = E \cup \partial E$, and construct X' by killing R upon first exit from F'. The process X' clearly extends X, and it is easy to see that its semigroup (hence its resolvent) maps $C_b(F')$ into itself. Clearly, every boundary point is (AP) here. By Theorem 5.2, the (ν_t) are tight.

Example 5.6. Let X be one-dimensional Brownian motion, killed upon first hitting zero. The state space is $E = (0, +\infty)$, and we shall take F' as the Martin boundary $[0, +\infty]$. Here also, the semigroup of X maps $C_b(F')$ into itself, as can be easily checked. Thus there exists a Feller-Dynkin process on F' extending X. The boundary point 0 is (AP), while the boundary point $+\infty$ is (AR). The measures (ν_t) are not tight since $\Lambda = 0$.

Example 5.7. Let X be a birth and death chain on $\mathbb{Z}$ with constant birth and death parameters, killed upon first hitting 0. We shall take $E = \{1, 2, 3, \dots\}$. This is the analogue of the previous example. The boundary point $+\infty$ is still (AR), but now the boundary point 0 is not needed. This gives a case where all boundary points are (AR).

Example 5.8. Let X be a Markov chain on $\{1, 2, 3, \dots\}$ whose behaviour may be described as follows: when started in $x \geq 1$, it may jump up to $x + 1$, or jump catastrophically back to state 1, or disappear from the state space. Such a process was called a pure birth process with catastrophes by Pakes (1995). In that paper, he gave conditions under which the point $+\infty$ is (AP), (AR), or neither. When this last possibility occurs, the point $+\infty$ can be a holding point (where the process waits for an exponentially distributed amount of time before jumping), or else it might split up into several distinct boundary points (this occurs when $\mathbb{P}_x(\zeta > t)$ oscillates as $x \to +\infty$).

5.2 Characterization of Limit Measures

In this section, we return to the Martin boundary methodology of the last chapter, and show how to deduce from it the existence (or not) of a Yaglom limit. This

point of view appears at first to be quite independent of that in the last section, but connections will surface in Section 3. At present, we need only keep in mind that Yaglom limits exist under more stringent conditions than the conditioned processes of Chapter 4.

To motivate the steps below, we restate the following fact, proved for Markov chains by Vere-Jones (1969). Suppose a Yaglom limit (5.2) exists with κ being a probability measure on E. Taking $f(x) = \mathbb{P}_x(g(X_t), \zeta > t)$, it follows from Assumption V that

$$\begin{aligned}\langle \kappa P_t, g\rangle &= \lim_{r\to\infty} \mathbb{P}_\nu(f(X_r) \,|\, \zeta > r) \\ &= \lim_{r\to\infty} \mathbb{P}_\nu(g(X_{r+t}) \,|\, \zeta > r+t)\mathbb{P}_\nu(\zeta > r+t \,|\, \zeta > r) \\ &= e^{\Lambda t}\langle \kappa, g\rangle.\end{aligned}$$

Thus κ is Λ-invariant. From a slightly different point of view, it defines an entrance law

$$\eta_t(dy) = e^{\Lambda t}\kappa(dy), \quad t \in \mathbb{R}.$$

Meanwhile, $\mathbb{P}_\nu(f(X_r) \,|\, \zeta > r) = \langle \nu_r P_t, g\rangle$, and so for each $r > 0$ we have another entrance law $\eta_t^r(dy) = \nu_r P_t(dy)$. When the Yaglom limit exists, the measures $(\eta_t^r : r > 0)$ converge to η_t, for each $t > 0$, as $r \to \infty$. In fact, we have

Lemma 5.9. *Under Assumption V, the Yaglom limit* (5.2) *exists if and only if* $\lim_{r\to\infty} \eta_t^r = \eta_t$ *for all* $t > 0$.

Proof. If the Yaglom limit exists, then the entrance laws converge as explained in the paragraph above. Conversely, choose $t > 0$ and g bounded; then

$$\begin{aligned}\lim_{r\to\infty} \langle \nu_r, g\rangle &= \lim_{r\to\infty} \langle \eta_t^{r-t}, g\rangle / \langle \eta_t^{r-t}, 1\rangle \\ &= \big(e^{\Lambda t}\langle \kappa, g\rangle\big) / \big(e^{\Lambda t}\langle \kappa, 1\rangle\big),\end{aligned}$$

so that the Yaglom limit exists. □

Accordingly, the remainder of this section comprises a study of these entrance laws. We shall be using similar assumptions to those of Chapter 4, but for a process $\widehat{X}$ in duality with X. We begin by stating

Assumption II(bis): Assumption II holds with m being some excessive measure.

When Assumption II holds, that is $P_t(x, dy) = p_t(x, y)m(dy)$ where m is any σ-finite measure, we can always construct the excessive measure

$$\widetilde{m}(dy) = \int_0^\infty p_t(a, y)dt \cdot m(dy),$$

where $a \in E$ is fixed (this requires transience of X, otherwise $\widetilde{m}$ may not be σ-finite). Since clearly $\widetilde{m} \ll m$, we also have $P_t(x, dy) = \widetilde{p}_t(x, y)\widetilde{m}(dy)$. The density $\widetilde{p}_t(x, y)$ will be continuous if $y \mapsto \int_0^\infty p_t(a, y)dt$ is continuous.

When Assumption II(bis) holds, we can define a dual semigroup $(\widehat{P}_t)$ by the formula

$$\widehat{P}_t(x, dy) = \widehat{p}_t(x, y)m(dy), \quad \widehat{p}_t(x, y) = p_t(y, x).$$

The associated Markov process will be denoted $\widehat{X}$. We also write $\widehat{\overline{X}}$ for the corresponding backward spacetime process, $\widehat{\overline{X}} = (L_r, \widehat{X}_r)$.

Consider now the entrance law

$$\eta_t^r(dy) = \nu_r P_t(dy) = \frac{\int \nu(da)p_{r+t}(a, y)}{\mathbb{P}_\nu(\zeta > r)} \cdot m(dy).$$

We shall be interested in the function

$$\widehat{k}^r(t, y) = \int \nu(da)\widehat{p}_{r+t}(y, a)/\mathbb{P}_\nu(\zeta > r);$$

it is obvious that $\eta_t^r(dy) = \widehat{k}^r(t, y)m(dy)$.

Lemma 5.10. *For each $r > 0$, the function $\widehat{k}^r$ is parabolic for $\widehat{X}$ on $(-r, \infty) \times E$, and satisfies*

$$\int m(dy)\widehat{k}^r(0, y) = 1.$$

Proof. We shall in fact show spacetime invariance with respect to $\widehat{\overline{X}}$, killed on

exiting $(-r,0)\times E$: for $(t,x)\in(-r,\infty)\times E$, $u<t+r$,

$$\begin{aligned}\widehat{\mathbb{P}}_{(t,x)}(\widehat{k}^r(\widehat{X}_u),\widehat{\zeta}>u) &= \widehat{P}_u(x;\widehat{k}^r(t-u,\cdot))\\ &= \int \widehat{p}_u(x,z)m(dz)\widehat{k}^r(t-u,z)\\ &= \int m(dz)\widehat{p}_u(x,z)\int \widehat{p}_{r+t-u}(z,a)\nu(da)/\mathbb{P}_\nu(\zeta>r)\\ &= \int \widehat{p}_{r+t}(x,a)\nu(da)/\mathbb{P}_\nu(\zeta>r)\\ &= \widehat{k}^r(t,x).\end{aligned}$$

Finally,

$$\begin{aligned}\int m(dy)\widehat{k}^r(0,y) &= \frac{\int\int m(dy)\widehat{p}_r(y,a)\nu(da)}{\mathbb{P}_\nu(\zeta>r)}\\ &= \frac{\int\int \nu(da)p_r(a,y)m(dy)}{\mathbb{P}_\nu(\zeta>r)} = 1,\end{aligned}$$

which is the claimed normalization. □

Formally, Lemma 5.10 has the effect of reducing the question of the convergence of the entrance laws (η^r) to that of the normalized parabolic functions $(\widehat{k}^r)$ for $\widehat{X}$, a very similar problem to that treated in the last chapter. There are some significant differences, however.

Unlike the case of Chapter 4, the functions $(\widehat{k}^r)$ are normalized with respect to the measure m, and this is generally a σ-finite measure, rather than a compactly supported probability measure ν. Recalling the comments made after Lemma 4.6, a limit function $k(t,y)=\lim_{r_n\to\infty}\widehat{k}^{r_n}(t,y)$ might well be zero, in which case also $\eta_t^{r_n}\Rightarrow 0$, and the Yaglom limit doesn't exist (as a probability measure). The tightness conditions of Corollary 5.4 guarantee that this does not happen.

Another point of note concerns the problem of extending the domain of the function $\widehat{k}^r(t,y)$ to the whole of $(-\infty,0)\times E$. We are aiming for an integral representation on the Martin boundary of $(-\infty,0)\times E$ as in the last chapter, so it is necessary that $\widehat{k}^r$ be the restriction to $(-r,0)\times E$ of some excessive function for $\widehat{\overline{X}}$. In the last chapter, we took h_s to be a positive constant on $(-\infty,s]\times E$, chosen so as to make the extended function parabolic when $\mathfrak{A}1=0$, but this is

not natural here. For example, if the measure ν satisfies $\nu(da) = g(a)m(da)$, then we generally have $\lim_{t\to -r} \widehat{k}^r(t,y) = g(y)$. Since the density function g can be arbitrary, the existence of a parabolic continuation of $\widehat{k}^r$ into $(-\infty, -r] \times E$ poses some problems, and will depend strongly on the measure ν. Another example concerns the case when X is Brownian motion on the unit disk, the transition density of which (relative to Lebesgue measure) satisfies

$$\lim_{t\downarrow 0} p_t(x,y) = \begin{cases} \infty & \text{if } x = y, \\ 0 & \text{if } x \neq y. \end{cases}$$

If $\nu(dz)$ is a point mass $\epsilon_a(dz)$, it follows that $\lim_{t\downarrow -r} \widehat{k}^r(t,a) = \infty$.

It is obvious that the manner in which we extend $\widehat{k}^r$ cannot affect the limit points $\widehat{k}(t,y) = \lim_{r_n\to\infty} \widehat{k}^{r_n}(t,y)$. Nevertheless, it can have a drastic effect on our ability to identify the limit in question.

In cases when the initial probability measure ν is compactly supported, a useful extension is to take $\widehat{k}^r(t,x) = 0$ on $(-\infty, -r] \times E$. It is easy to check that this definition turns $\widehat{k}^r$ into an excessive function for $\widehat{\overline{X}}$ on the whole of $(-\infty, 0) \times E$. As the extension does not affect the normalization, there exists a corresponding integral representation on the Martin compactification of $\widehat{\overline{X}}$. The Martin kernel used therein must be normalized with respect to the measure $r(ds, dy) = \epsilon_0(ds)m(dy)$, and hence

$$\widehat{\overline{K}}\Big((t,x);(s,y)\Big) = \frac{1_{(-\infty,t]}(s)\widehat{p}_{t-s}(x,y)}{1_{(-\infty,0]}(s)\int m(dz)\widehat{p}_{-s}(z,y)} = \frac{1_{(-\infty,t]}(s)p_{t-s}(y,x)}{\mathbb{P}_y(\zeta > -s)}.$$

Comparing this last expression with the definition of $\widehat{k}^r$, we find an explicit form for a representing measure:

Proposition 5.11. *For each $r > 0$, the function $\widehat{k}^r$ is represented on the parabolic Martin compactification by the probability measure ξ^r, concentrated on the set $\{-r\} \times supp(\nu)$, and given by*

$$\xi^r(ds, dy) = \epsilon_{-r}(ds)\mathbb{P}_y(\zeta > -s)\nu(dy)/\mathbb{P}_\nu(\zeta > -s).$$

Thus

$$\widehat{k}^r(t,x) = \int \widehat{\overline{K}}\Big((t,x);\overline{y}\Big)\xi^r(d\overline{y}).$$

Proof. We check simply that the representation works with ξ^r as given:

$$\begin{aligned}\int \widehat{\overline{K}}\Big((t,x);(s,y)\Big)\xi^r(ds,dy) &= \int_{\overline{E}} \frac{1_{(-\infty,t]}(s)p_{t-s}(x,y)}{\mathbb{P}_y(\zeta > -s)}\epsilon_{-r}(ds)\frac{\mathbb{P}_y(\zeta > -s)\nu(dy)}{\mathbb{P}_\nu(\zeta > -s)} \\ &= 1_{(-\infty,t]}(-r)\frac{\int p_{t+r}(x,y)\nu(dy)}{\mathbb{P}_\nu(\zeta > r)} \\ &= \widehat{k}^r(t,x). \quad \square\end{aligned}$$

Let $\widehat{\overline{D}}$ denote the closure, in the Martin topology of $\widehat{\overline{X}}$, of the set $(-\infty,0)\times \mathrm{supp}(\nu)$. Since ξ^r is concentrated on $\widehat{\overline{F}}_{-r} \cap \widehat{\overline{D}}$ by Proposition 5.11, every weak limit point of (ξ^r) is a probability measure on $\widehat{\overline{F}}_{-\infty} \cap \widehat{\overline{D}}$.

In cases when ν charges all of the state space E, all we can deduce from Proposition 5.11 is that the weak limit point ξ is concentrated on $\widehat{\overline{F}}_{-\infty}$. This set usually consists of a large number of points, and by choosing the initial distribution ν appropriately, ξ can often be made to charge any given one, as the next example shows. Compare it with the example following the proof of Theorem 4.21.

Example 5.12. Let $X_t = B_t - \alpha t$, where B is a one-dimensional Brownian motion, and X is killed upon first leaving $(0,\infty)$. The drift is towards zero, i.e. $\alpha \geq 0$. It is well known (Revuz and Yor (1991)) that this process is related to Brownian motion $(0,\infty)$ (that is, the case $\alpha = 0$) by a Girsanov transformation. Accordingly, the semigroup (Q_t) of X is related to the semigroup (P_t) of killed Brownian motion by the formula

$$Q_t(x,dy) = \frac{1}{e^{-\alpha x}}P_t(x,dy)e^{-\alpha^2 t/2}e^{-\alpha y}, t > 0, x > 0.$$

If we set $g(t,x) = e^{\alpha^2 t/2}e^{-\alpha x}$, this function is parabolic: $\frac{1}{2}\partial^2 g/\partial x^2 = \partial g/\partial t$.

In terms of the spacetime process, $\overline{X}_r = (L_r, X_r) = (L_r, B_r - \alpha r)$ has semigroup $\overline{P}^g_r(\overline{x},d\overline{y}) = \overline{P}_r(\overline{x},d\overline{y})g(\overline{y})/g(\overline{x})$, where $(\overline{P}_r)$ is the semigroup of $\overline{B}_r = (L_r, B_r)$. Then h is excessive for $\overline{X}$ if and only if $h = k/g$, where k is excessive for $\overline{B}$. The Martin kernel $\overline{K}^g$ of $\overline{X}$ is related to that of $\overline{B}$ via the formula

$$\overline{K}^g(\overline{x},\overline{y}) = \frac{\langle r, g\rangle}{g(\overline{x})}\overline{K}(\overline{x},\overline{y}),$$

where $r(ds, dy) = \epsilon_0(ds)\nu(dy)$ is the normalizing measure. Using the example after the proof of Theorem 4.21, the parabolic Martin boundary of X has minimal functions as follows: (we set $\nu_\alpha = \int e^{-\alpha z}\nu(dz) = \langle r, g\rangle$)

$$\overline{K}^g(\overline{x}, (\tau, 0)) = \begin{cases} \nu_\alpha e^{-\alpha^2\tau/2} \frac{x}{\sqrt{2\pi(t-\tau)^3}} \exp\left(-\frac{(x-\alpha(t-\tau))^2}{2(t-\tau)}\right) & \text{if } t > \tau, \\ 0 & \text{if } t \le \tau, \end{cases}$$

and for any real number $\gamma \le 0$, we also get another minimal function by the formula (where $c := (\gamma^2 - \alpha^2)/2$):

$$\overline{K}^g(\overline{x}, (-\infty, \gamma)) = \begin{cases} \nu_\alpha e^{ct} e^{\alpha x} \sinh(\sqrt{|\alpha^2 + 2c|} \cdot x) & \text{if } c = (\gamma^2 - \alpha^2)/2 > -\alpha^2/2, \\ \nu_\alpha x e^{ct} e^{\alpha x} & \text{if } c = -\alpha^2/2. \end{cases}$$

Only the boundary points $(-\infty, \gamma)$ belong to $\overline{F}_{-\infty}$. The measure $m(dx) = e^{2\alpha x}dx$ is symmetrizing, so Assumption II(bis) applies, and we can define the process $\widehat{X}$ as described above. Since the transition density with respect to m satisfies $p_t(x, y) = p_t(y, x)$, we find immediately that $\widehat{\overline{F}}_{-\infty}$ consists of the minimal functions

$$\widehat{\overline{K}}^g(\overline{x}, (-\infty, \gamma)) = \begin{cases} \nu_\alpha e^{ct} e^{-\alpha x} \sinh(\sqrt{|\alpha^2 + 2c|} \cdot x) & \text{if } c > -\alpha^2/2, \\ \nu_\alpha e^{ct} x e^{-\alpha x} & \text{if } c = -\alpha^2/2. \end{cases}$$

This agrees with results of Martinez and San Martin (1994). The quasistationary distribution with eigenvalue c is given (up to normalization) by

$$\mu_c(dx) = \begin{cases} e^{-\alpha x} \sinh(\sqrt{|\alpha^2 + 2c|} \cdot x)dx & \text{if } -\alpha^2/2 < c < 0, \\ x e^{-\alpha x} dx & \text{if } c = -\alpha^2/2. \end{cases}$$

If we choose the initial distribution ν to be μ_c, then the function $\widehat{k}^r(t, x)$ equals $e^{ct} d\mu_c/dx$ for $t > -r$, and a natural way to extend this into $(-\infty, -r] \times E$ is to set $\widehat{k}^r(t, x) = e^{ct} d\mu_c/dx$ there. Independently of $r > 0$, the $\widehat{k}^r$-process is Doob-conditioned to hit the boundary point $\gamma(c)$, and hence so is the limiting process as $r \to \infty$. In terms of Yaglom limits, this is the familiar result that

$$\lim_{t\to\infty} \mathbb{P}_{\mu_c}(f(X_t) \,|\, \zeta > t) = \langle \mu_c, f\rangle.$$

The point γ is associated with any Martin sequence with $\lim_{n\to\infty}(s_n, y_n/s_n) = (-\infty, \gamma)$, hence all sequences such that (y_n) is bounded are associated with the single point $\gamma = 0$, that is, the eigenfunction with eigenvalue $c = -\alpha^2/2$. If we take a compactly supported initial distribution ν with support D, the set $\widehat{\overline{F}}_{-\infty} \cap \widehat{\overline{D}}$ therefore consists of this single eigenfunction, and therefore $\nu_t \Rightarrow \mu_{-\alpha^2/2}$, provided the (ν_t) are tight (see Theorem 5.15).

The last example shows that the limit points $\widehat{k}$ of the family $(\widehat{k}^r)$ can range over the full boundary subset $\widehat{\overline{F}}_{-\infty}$, in cases when the initial distribution charges all of E. For such measures, the extension we chose for $\widehat{k}^r$ is not entirely satisfactory, and a different choice is needed. In view of the last example, the problem of extending $\widehat{k}^r$ to the whole of $(-\infty, 0) \times E$ is essentially equivalent to the domain of attraction problem for limiting conditional distributions (Pakes (1995)): given a QSD μ, for which initial distributions ν does the Yaglom limit, started with ν, converge to μ?

The explicit calculation of the measure ξ^r given by Proposition 5.11 is also interesting from the point of view of Chapter 4. If the process level conditioning (5.1) works, then the function $h_s(0, \cdot)$ must converge as $s \to \infty$. If this convergence takes place in $L^1(d\nu)$, then we must have $\xi^r \Rightarrow \xi$ for some probability measure ξ on $\widehat{\overline{F}}_{-\infty}$, and hence the functions $\widehat{k}^r$ must converge as well. However, we do not have a guarantee that the latter limit is nonzero.

Returning to the case that ν is supported by a compact set D, we can identify the boundary points in $\widehat{\overline{F}}_{-\infty} \cap \widehat{\overline{D}}$ as in Chapter 4, by letting D take the role of a good cemetery neighbourhood.

Proposition 5.13. *If D is a compact set and the parabolic Harnack inequality holds for $\widehat{X}$, then each nonzero minimal excessive function for $\widehat{\overline{X}}$ which belongs to $\widehat{\overline{F}}_{-\infty} \cap \widehat{\overline{D}}$ is of the form $\widehat{k}(t, x) = e^{\lambda t}\widehat{g}(x)$, where $\widehat{g}$ solves $\widehat{\mathfrak{A}}g = \lambda\widehat{g}$ in E. Moreover, the following two statements are equivalent:*

(i) For some $y \in D$,

$$\underline{\lim}_{s\to-\infty} \widehat{\overline{K}}\big(\overline{x}, (s, y)\big) > 0$$

holds for all $\overline{x}$ in a set of positive η-measure,

(ii) The set $\widehat{\overline{F}}_{-\infty} \cap \widehat{\overline{D}}$ contains a single nontrivial, minimal parabolic boundary point $\overline{z}$.

Proof. Since for each $\epsilon > 0$,

$$\begin{aligned}\widehat{\overline{K}}\Big((t,x);(s,y)\Big) &= p_{t-s}(y,x)/\mathbb{P}_y(\zeta > -s) \\ &= \frac{\mathbb{P}_y(\zeta > -s+\epsilon)}{\mathbb{P}_y(\zeta > -s)}\Big(p_{t-s}(y,x)/\mathbb{P}_y(\zeta > -s+\epsilon)\Big) \\ &\leq \widehat{\overline{K}}\Big((t-\epsilon,x);(s-\epsilon,y)\Big),\end{aligned}$$

the same proofs as those of Propositions 4.15 and 4.24 can be used, interchanging N and D. □

According to the above proposition, (i) or (ii) imply that $\lim_{r\to\infty} \widehat{k}^r(t,y) = e^{\lambda t} g(y)$ for some positive function g. Moreover, Assumption V obviously follows. The function g is not necessarily integrable with respect to m. This is exemplified by the case where X is Brownian motion killed upon leaving $(0,\infty)$, that is, the example after Proposition 5.11 with zero drift. Since the decay parameter Λ_1 equals zero, the measures (ν_t) are not tight, and in fact $g(x) = x$, which is not Lebesgue integrable.

5.3 Yaglom Limits

One drawback of Proposition 5.13 is that it implies convergence of the functions $\widehat{k}^r(t,x)$ for $t < 0$. To get the Yaglom limit (5.2), we require convergence for $t \geq 0$. Nevertheless, Proposition 5.13 essentially implies that the Yaglom limit, if it exists, must be $\kappa = g \cdot m$. This is the counterpart to Proposition 4.20 in Chapter 4. Before proving the counterpart to Theorem 4.21, we need a condition which guarantees that $\widehat{\overline{F}}_{-\infty} \cap \widehat{\overline{D}}$ contains an integrable eigenfunction g; a similar problem was solved in Chapter 4, where it was necessary to prove that the limit of (h_s) is strictly positive. This was achieved by restricting the initial distribution

to compactly supported ones. We shall show that the function g is positive by using the tightness conditions of Section 1.

Proposition 5.14. *Let X be a Markov chain with compactly supported initial distribution ν, and suppose that the family (ν_t) is tight. Then statement (ii) of Proposition 5.13 holds, and hence* $\lim_{r\to\infty} \widehat{k}^r(t,x) = e^{\Lambda t} g(x)$.

Proof. Without loss of generality, suppose that $\nu(dx) = \epsilon_{y_0}(dx)$ is the point mass at $y_0 \in E$. For fixed $(t,x) \in (-\infty, 0) \times E$, we calculate

$$\begin{aligned}\widehat{\overline{K}}\Big((t,x);(s,y_0)\Big) &= p_{t-s}(y_0,x)/\mathbb{P}_{y_0}(\zeta > -s) \\ &\geq p_{t-s}(y_0,x)/\mathbb{P}_{y_0}(\zeta > t-s) \\ &= \langle \nu_{t-s}, 1_{\{x\}} \rangle.\end{aligned}$$

Let $\nu_\infty = \underline{\lim}_{s\to-\infty} \nu_{t-s}$. An elementary calculation shows that $\nu_t P_r \leq \nu_{t+r}$ for $r > 0$, so by Fatou's lemma

$$\begin{aligned}\int \nu_\infty(dz) p_r(z,x) &\leq \underline{\lim}_{s\to-\infty} \sum \nu_{t-s}(\{z\}) p_r(z,x) \\ &\leq \underline{\lim}_{s\to-\infty} \nu_{t+r-s}(\{x\}) = \nu_\infty(\{x\}),\end{aligned}$$

and hence $\nu_\infty(\{x\}) > 0$, because $p_r(z,x) > 0$ by irreducibility, and $\nu_\infty(E) = 1$ by tightness. We have shown that statement (i) of Proposition 5.13 holds, and hence so does statement (ii). □

The existence of the Yaglom limit follows:

Theorem 5.15. *Let Assumption II(bis) hold, and let $\widehat{X}$ satisfy Assumption III and the parabolic Harnack inequality. If the family (ν_t) is tight and the conclusion of Proposition 5.14 holds, then the Yaglom limit* (5.2) *exists and is a probability measure: for every bounded measurable f,*

$$\lim_{r\to\infty} \mathbb{P}_\nu(f(X_t) \,|\, \zeta > t) = \int f(x) g(x) m(dx),$$

where g is strictly positive, Λ-invariant for $\widehat{X}$ (with $\Lambda < 0$), and such that $\int g(x) m(dx) = 1$.

Proof. By definition, since

$$\widehat{k}^r(-1,z) = \widehat{k}^{r-1}(0,z)/\mathbb{P}_\nu(\zeta > r \,|\, \zeta > r-1),$$

we have by Assumption V that $\lim_{r\to\infty} \widehat{k}^r(0,z) = g(z)$, and this convergence is bounded on compact subsets of E, by Proposition 5.13. Since $\int \widehat{k}^r(0,z)m(dz) = 1$, we shall have $\int g(z)m(dz) = 1$ if the convergence also occurs in $L^1(dm)$. This follows from the tightness of (ν_t): take $\epsilon > 0$ arbitrary, and choose D compact such that $\nu_r(D^c) < \epsilon$ for all r. There exists a constant C such that $\sup_r \widehat{k}^r(0,z) < C$ for all $z \in D$, hence we can use bounded convergence (since m is finite on compacts) to deduce

$$\begin{aligned}\lim_{r\to\infty} \int \widehat{k}^r(0,z)m(dz) &= \lim_{r\to\infty} \int_D \widehat{k}^r(0,z)m(dz) + \lim_{r\to\infty} \int_{D^c} \widehat{k}^r(0,z)m(dz) \\ &\leq \int_D g(z)m(dz) + \epsilon.\end{aligned}$$

Letting $\epsilon \downarrow 0$ (and consequently $D \uparrow E$) gives

$$1 = \lim_{r\to\infty} \int \widehat{k}^r(0,z)m(dz) \leq \int g(z)m(dz),$$

while Fatou's lemma implies $\int g(z)m(dz) \leq 1$. Using Scheffé's lemma, we deduce that $\widehat{k}^r(0,\cdot)$ converges to g in $L^1(dm)$. Hence for any bounded measurable function f,

$$\begin{aligned}\lim_{r\to\infty} \mathbb{P}_\nu(f(X_t) \,|\, \zeta > r) &= \lim_{r\to\infty} \int f(z)\widehat{k}^r(0,z)m(dz) \\ &= \int f(z)g(z)m(dz) \\ &= \langle \kappa, f \rangle,\end{aligned}$$

where κ is the probability measure $g \cdot m$. □

Example 5.16. Let X be an open migration process with a symmetric transition function (relative to some symmetrizing measure m) - see the example after Theorem 4.27. This process clearly satisfies Assumption II(bis), as the symmetrizing measure, being invariant, is also excessive. As seen in Chapter 4, the conditioned process exists if a finite cemetery neighborhood can be found.

Thus Assumption V holds. Assuming that $\Lambda < 0$ and that the F'-boundary points are all either (AP) or (AR), the measures (ν_t) must be tight. By Proposition 5.14 and Theorem 5.15, the Yaglom limit exists if the initial distribution ν is compactly supported.

The corresponding theory in discrete time is more complicated. Kesten (1995) gave an example of a process with uniformly bounded jumps on $E = \mathbb{Z}$ for which the measures (ν_t) are tight, but for which the Yaglom limit does not exist. Consequently, Proposition 5.14 must fail in this case. Complications arise, in the discrete time setting, because the transition function $p_t(x, y)$ need not be strictly positive on $E \times E$, even though the process may be irreducible. This happens in Kesten's example, where no matter what the time t is, $p_t(x, y) = 0$ if $|x - y| > t$. The proof of Proposition 5.14 breaks down, and consequently the conclusion of Proposition 5.13 fails. Indeed, Kesten has shown that there are at least two distinct subsequential limits for (ν_t), both of these quasistationary distributions. Thus the window $\widehat{\overline{F}}_{-\infty} \cap \widehat{\overline{D}}$ must consist of at least two distinct minimal points in this case.

5.4 Open Questions

The methods of this chapter are very closely related to those of the last chapter, and many of the questions posed there have their counterpart here:

How can we weaken the assumption that ν be compactly supported? As observed in Section 2, the limit points $\widehat{k}$ need not in general be representable by a measure on $\widehat{\overline{F}}_{-\infty} \cap \widehat{\overline{D}}$. When ν charges all of E, suppose that for each r, we can find a strictly positive parabolic function g^r on $(-\infty, 0) \times E$ which extends $\widehat{k}^r$. If the lifetime of the g^r-transform of $\widehat{\overline{X}}$ is finite, then the method of Chapter 4 can be used to show that g^r is representable by a probability measure ξ^r concentrated on $\widehat{\overline{N}}$, where N is a cemetery neighbourhood. Thus in this case, we can proceed exactly as in Chapter 4.

When ν is compactly supported, and the associated Yaglom limit κ exists,

the latter is a quasistationary distribution whose parameter Λ is minimal. The lifetime of the process X, started in κ, has the smallest mean of all possible quasistationary distributions. Why this is so is still a mystery, but it seems clear that this fact is closely linked to the corresponding problem for the conditioned process; see the open questions at the end of Chapter 4.

Finally, I shall mention one further problem, of great importance to practitioners. Once a Yaglom limit is shown to exist, it is desirable to have estimates on the convergence rate. This is necessary for the following reason: if the measures (ν_t) converge too slowly, the process is likely to die before quasistationarity manifests itself. Few results in this direction are known. The interested reader can start with the last section of van Doorn (1991), which identifies the spectral gap between the first and second eigenvalues as an important ingredient, but an actual "test for quasistationarity" is far from being a reality yet. Chan (1997) discusses this issue via the theory of Large Deviations.

Bibliography

[1] Ancona, A. (1990). *Théorie du potentiel sur les graphes et les variétés.* Lecture Notes in Mathematics 1427. Springer-Verlag, New York.

[2] Anderson, W.J. (1991). *Continuous-Time Markov Chains: An Applications-Oriented Approach.* Springer-Verlag, New York.

[3] Asmussen, S. and Hering, H. (1983). *Branching processes.* Birkhäuser, Boston.

[4] Bass, R.F. (1995). *Probabilistic Techniques in Analysis.* Springer-Verlag, New York.

[5] Bass, R.F. and Burdzy, K. (1992). Lifetimes of conditioned diffusions. *Probab. Theory and Rel. Fields* **91**, 405-444.

[6] Chan, T. (1997) Large deviations and quasi-stationarity for density-dependent birth-death processes. Preprint.

[7] Chung, K.L. (1982). *Lectures from Markov Processes to Brownian Motion.* Springer-Verlag, New York.

[8] Chung, K.L. and Walsh, J.B. (1970). To reverse a Markov process. *Acta Math.* **123**, 225-251.

[9] Chung, K.L. and Zhao, Z.X. (1995). *From Brownian Motion to Schrödinger's Equation.* Springer-Verlag, New York.

[10] Collet, P., Martinez, S. and San Martin, J. (1995). Asymptotic laws for one-dimensional diffusions conditioned to nonabsorption. *Ann. Probab.* **23** (3), 1300-1314.

[11] Darroch, J.N. and Seneta, E. (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Probab.* **4**, 192-196.

[12] Dellacherie, C. and Meyer, P.A. (1987). *Probabilités et Potentiel D. Théorie du potentiel associée à une résolvante. Théorie des processus de Markov*. Hermann, Paris.

[13] Dellacherie, C. and Meyer, P.A. (1992). *Probabilités et Potentiel E. Processus de Markov (fin). Compléments de calcul stochastique*. Hermann, Paris.

[14] Doob, J.L. (1955). A probability approach to the heat equation. *Trans. Amer. Math. Soc.* **80**, 216-280.

[15] Doob, J.L. (1957). Conditional Brownian motion and the boundary limits of harmonic functions. *Bull. Soc. Math. France* **85**, 431-458.

[16] Doob, J.L. (1984). *Classical Potential Theory and Its Probabilistic Counterpart*. Springer-Verlag, New York.

[17] van Doorn, E.A. (1991). Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. *Adv. Appl. Probab.* **23**, 683-700.

[18] Dynkin, E.B. (1965). *Markov Processes, Vol. I and II*. Springer-Verlag, New York.

[19] Dynkin, E.B. (1982). *Markov Processes and Related Problems of Analysis*. London Math. Soc. Lecture Notes Series 54. Cambridge University Press.

[20] Ferrari, P.A., Kesten, H., Martinez, S. and Picco, P. (1995). Existence of quasi-stationary distributions. A renewal dynamic approach. *Ann. Probab.* **23**, 501-521.

[21] Flaspohler, D.C. (1974). Quasi-stationary distributions for absorbing continuous-time denumerable Markov chains. *Ann. Inst. Statist. Math.* **26**, 351-356.

[22] Friedman, A. (1964). *Partial Differential Equations of Parabolic Type.* Prentice-Hall, Englewood Cliffs, N.J.

[23] Getoor, R.K. (1975) *Markov processes: Ray processes and right processes.* LNM 440. Springer-Verlag.

[24] Glover, J. (1986). Quasistationary distributions, eigenmeasures, and eigenfunctions of Markov processes, in *Seminar on Stochastic Processes 1984*, 71-98. Birkhäuser, Boston.

[25] Hart, A.G. (1997). *Quasistationary Distributions for Continuous-Time Markov Chains*. Ph.D. Thesis. The University of Queensland.

[26] Hart, A.G. and Pollett, P.K. (1996). Direct analytical methods for determining quasistationary distributions for continuous-time Markov chains. In C.C. Heyde, Yu.V. Prohorov, R. Pyke, and S.T. Rachev, editors, *Athens Conference on Applied Probability and Time Series Analysis, Volume I: Applied Probability, In Honour of J.M. Gani*, Lecture Notes in Statistics 114. Springer-Verlag, New York, 116-126.

[27] Ikeda, N., Nagasawa, M. and Watanabe, S. (1968). Branching Markov processes I. *J. Math. Kyoto Univ.* **8**, 233-278.

[28] Ikeda, N., Nagasawa, M. and Watanabe, S. (1969). Branching Markov processes III. *J. Math. Kyoto Univ.* **9**, 95-160.

[29] Jacka, S.D. and Roberts, G.O. (1994). Weak convergence of conditioned birth and death processes. *J. Appl. Probab.* **31**, 90-100.

[30] Jacka, S.D. and Roberts, G.O. (1995). Weak convergence of conditioned processes on a countable state space. *J. Appl. Probab.* **32**, 902-916.

[31] Jacka, S.D., Pollett, P.K. and Roberts, G.O. (1997). Non-explosivity of limits of conditioned birth and death processes. *J. Appl. Probab.* **34**, 35-45.

[32] Jacka, S.D. and Roberts, G.O. (1997) Conditional one-dimensional diffusions. Preprint.

[33] Jeulin, T. (1978). Compactification de Martin d'un processus droit. *Z. Wahrscheinlichkeitstheorie* **42**, 229-260.

[34] Kelly, F.P. (1979). *Reversibility and Stochastic Networks*. John Wiley and Sons, New York.

[35] Kendall, D.G. (1959). Unitary dilations of one-parameter semigroups of Markov transition operators, and the corresponding integral representations for Markov processes with a countable infinity of states. *Proc. London Math. Soc. (3)*, 417-431.

[36] Kesten, H. (1978). Branching Brownian motion with absorption. *Stoch. Proc. Appl.* **7**, 9-47.

[37] Kesten, H. (1995). A ratio limit theorem for (sub) Markov chains on $\{1, 2, 3, \dots\}$ with bounded jumps. *Adv. Appl. Probab.* **27**, 652-691.

[38] Kijima, M., Nair, M.G., Pollett, P.K. and van Doorn, E. (1997). Limiting conditional distributions for birth-death processes. *Adv. Appl. Probab.* **29**, 185-204.

[39] Kijima, M. (1993). Quasi-limiting distributions of Markov chains that are skip-free to the left in continuous-time. *J. Appl. Probab.* **30**, 509-517.

[40] Kingman, J.F.C. (1963). The exponential decay of Markov transition probabilities. *Proc. London Math. Soc.* **13**, 337-358.

[41] Koranyi, A. and Taylor, J.C. (1985). Minimal solutions of the heat equation and uniqueness of the positive Cauchy problem on homogeneous spaces. *Proc. Amer. Math. Soc.* **94** (2), 273-278.

[42] Kunita, H. and Watanabe, T. (1965). Markov processes and Martin boundaries I. *Illinois J. of Math.* **9**, 485-526.

[43] Lyons, T. and Sullivan, D. (1984). Function theory, random paths and covering spaces. *J. Differential Geometry* **19**, 299-323.

[44] Martinez, S. and San Martin, J. (1994). Quasi-stationary distributions for a Brownian motion with drift and associated limit laws. *J. Appl. Probab.* **31**, 911-920.

[45] McKean, H.P. Jr. (1956). Elementary solutions for certain parabolic partial differential equations. *Trans. Amer. Math. Soc.* **82**, 519-548.

[46] McKean, H.P. Jr. (1963). Excursions of a non-singular diffusion. *Z. Wahrscheinlichkeitstheorie* **1**, 230-239.

[47] Meyer, P.A. (1968). *Processus de Markov: la frontière de Martin.* Lecture Notes in Mathematics 77. Springer-Verlag, New York.

[48] Meyer, P.A., Smythe, R.T. and Walsh, J.B. (1972). Birth and death of Markov processes. *Proc. Sixth Berkeley Symp. Math. Stat. Prob. 1970/71* Vol. III. Berkeley, Univ. of California, 295-305.

[49] Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability.* Springer-Verlag, New York.

[50] Moser, J. (1964). A Harnack inequality for parabolic differential equations. *Comm. Pure Appl. Math.* **17**.

[51] Nair, N.G. and Pollett, P.K. (1993). On the relationship between μ-invariant measures and quasistationary distributions for continuous-time Markov chains. *Adv. Appl. Probab.* **25**,82-102.

[52] Nåsell, I. (1995). The quasi-stationary distribution of the closed endemic SIS model. *Adv. Appl. Probab.* **28**, 895-932.

[53] Nummelin, E. (1984). *On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states.* Cambridge Tracts in Mathematics, 83. Cambridge University Press.

[54] Orey, S. (1971). *Lecture Notes on Limit Theorems of Markov Chain Transition Probabilities.* Van Nostrand, London.

[55] Pakes, A.G. (1993). Explosive Markov branching processes: entrance laws and limiting behaviour. *Adv. Appl. Probab.* **25**, 737-756.

[56] Pakes, A.G. (1995). Quasistationary laws for Markov processes: examples of an always proximate absorbing state. *Adv. Appl. Probab.* **27**, 120-145.

[57] Papangelou, F. (1967). Strong ratio limits, R-recurrence and mixing properties of discrete parameter Markov processes. *Z. Wahrscheinlichkeitstheorie* **8**, 259-297.

[58] Pinsky, R.G. (1985). On the convergence of diffusion processes conditioned to remain in a bounded region for large time to limiting positive recurrent diffusion processes. *Ann. Probab.* **13** (2), 363-378.

[59] Pinsky, R.G. (1995). *Positive Harmonic Functions and Diffusion - An Integrated Analytic and Probabilistic Approach.* Cambridge University Press.

[60] Pitman, J.W. (1975). One-dimensional Brownian motion and the three-dimensional Bessel process. *Adv. Appl. Probab.* **7** (3), 511-526.

[61] Pollett, P.K. (1995). Modelling the long-term behaviour of evanescent ecological systems. *Ecological Modelling* **86**, 135-139.

[62] Pollett, P.K. (1988). Reversibility, invariance and μ-invariance. *Adv. Appl. Probab.* **20**, 600-621.

[63] Pollett, P.K. (1993). Recent advances in the theory and application of quasistationary distributions. *Stochastic Models in Engineering, Technology and Management*, S. Osaki and D.N.P. Murthy (eds), World Scientific, Singapore, 477-486.

[64] Pollett, P.K. (1995). The determination of quasistationary distributions directly from the transition rates of an absorbing Markov chain. *Math. Computer Modelling* **22**, 279-287.

[65] Pollett, P.K. and Stewart, D.E. (1994). An efficient procedure for computing quasi-stationary distributions of Markov chains with sparse transition structure. *Adv. Appl. Probab.* **26**, 68-79.

[66] Revuz, D. and Yor, M. (1991). *Continuous Martingales and Brownian Motion*. Springer-Verlag, New York.

[67] Ridler-Rowe, C.J. (1967). On a stochastic model of an epidemic. *J. Appl. Probab.* **4**, 19-33.

[68] Rogers, L.C.G. and Williams, D. (1987). *Diffusions, Markov processes, and Martingales, Vol. II*. Wiley, New York.

[69] Rogers, L.C.G. and Williams, D. (1994). *Diffusions, Markov processes, and Martingales, Vol. I*. Wiley, New York.

[70] Schrijner, P. (1996). *Quasi-Stationarity of Discrete-Time Markov Chains.* PhD Thesis. University of Twente.

[71] Schrijner, P. and van Doorn, E. (1997). Weak convergence of conditioned birth-deat processes in discrete time. *J. Appl. Probab.* **34** (1), 46-53.

[72] Seneta, E. (1966). Quasi-stationary distributions and time-reversion in genetics. *J. Roy. Statist. Soc.*, Ser B **28**, 253-277.

[73] Seneta, E. and Vere-Jones, D. (1966). On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states. *J. Appl. Probab.* **3**, 403-434.

[74] Sharpe, M. (1988). *General Theory of Markov Processes.* Pure and Applied Mathematics 133. Academic Press, Boston.

[75] Sturm, K.T. (1991). Gauge theorems for resolvents with application to Markov processes. *Probab. Theory and Rel. Fields* **89**, 387-406.

[76] Tuominen, P. and Tweedie, R.L. (1979). Exponential decay and ergodicity of general Markov processes and their discrete skeletons. *Adv. Appl. Probab.* **11**, 784-803.

[77] Tweedie, R.L. (1974a). R-theory for Markov chains on a general state space I: Solidarity properties and R-recurrent chains. *Ann. Probab.* **2** (5), 840-864.

[78] Tweedie, R.L. (1974b). R-theory for Markov chains on a general state space II: r-subinvariant measures for r-transient chains. *Ann. Probab.* **2** (5), 865-878.

[79] Tweedie, R.L. (1974c). Quasistationary distributions for Markov chains on a general state space. *J. Appl. Probab.* **11**, 726-741.

[80] Vere-Jones, D. (1962). Geometric ergodicity in denumerable Markov chains. *Quart. J. Math. Oxford* **13**, 7-28.

[81] Vere-Jones, D. (1969). Some limit theorems for evanescent processes. *Austral. J. Statist.* **11**, 67-78.

[82] Walsh, J. (1977). The cofine topology revisited. *Proceedings of Symposia in Pure Mathematics* **31**, 131-151.

[83] Walsh, J. and Winkler, W. (1981). Absolute continuity and the fine topology, in *Seminar on Stochastic Processes 1981*, 151-157. Birkhäuser, Boston.

[84] Widder, D.V. (1941). *The Laplace Transform*. Princeton University Press.

[85] Williams, D. (1974). Path decomposition and continuity of local time for one-dimensional diffusions, I *Proc. London Math. Soc. (3)* **28**, 738-768.

[86] Yaglom, A.M. (1947). Certain limit theorems of the theory of branching stochastic processes (in Russian). *Dokl. Akad. Nauk. SSSR* **56**, 795-798.