Importance sampling techniques for the multidimensional ruin problem and general first passage problems

Importance Sampling Techniques for the Multidimensional Ruin Problem and General First Passage Problems

Jeffrey F. Collamore

EURANDOM*

Abstract

Let $\{(X_n, S_n) : n = 0, 1, \dots\}$ be a Markov additive process, where $\{X_n\}$ is a Markov chain on a general state space and S_n is an additive component on $\mathbb{R}^d$. We consider $\mathbf{P}\{S_n \in A/\epsilon, \text{ some } n\}$ as $\epsilon \to 0$, where $A \subset \mathbb{R}^d$ is open and the mean drift of $\{S_n\}$ is away from A. Our main objective is to study the *simulation* of $\mathbf{P}\{S_n \in A/\epsilon, \text{ some } n\}$ using the Monte Carlo technique of importance sampling. If the set A is convex, then we establish: (i) the precise dependence (as $\epsilon \to 0$) of the estimator variance on the choice of the simulation distribution; (ii) the existence of a *unique* simulation distribution which is efficient and optimal in the asymptotic sense of Siegmund (1976). We then extend our techniques to the case where A is not convex. Our results lead to positive conclusions which complement the multidimensional counterexamples of Glasserman and Wang (1997).

1 Introduction

There has been much recent interest in simulation techniques for estimating rare event probabilities, or more precisely, the numerical computation of $\mathbf{P}(C_\epsilon)$ for small ϵ when $\mathbf{P}(C_\epsilon) \to 0$ as $\epsilon \to 0$. Generally, such probabilities cannot be computed using direct Monte Carlo techniques, because the relative error associated with samples averages of $\mathbf{1}_{C_\epsilon}$ is

$$\frac{\sqrt{\text{Var}(\mathbf{1}_{C_\epsilon})}}{\mathbf{E}(\mathbf{1}_{C_\epsilon})} \to \infty \quad \text{as} \quad \epsilon \to 0, \quad \text{where } \mathbf{1}_{C_\epsilon} = \text{indicator function on } C_\epsilon.$$

In this article, we will study rare event simulation in the context of the following multidimensional boundary crossing problem: Let $S_1, S_2, \dots$ be a sequence of random variables in $\mathbb{R}^d$, and consider the hitting probability of a region $A \subset \mathbb{R}^d$ by $\{S_n\}$, namely,

$$\mathbf{P}\Big\{S_n \in \frac{A}{\epsilon}, \text{ some } n\Big\} = \mathbf{P}\{T^\epsilon(A) < \infty\} \quad \text{as} \quad \epsilon \to 0, \tag{1.1}$$

*Part of this work was done while the author was a post–doctoral research associate at the Centre for Mathematical Sciences, Mathematical Statistics, Lund University, Sweden. The author is grateful to the Centre for its hospitality and stimulating research environment.

AMS 1991 subject classifications. Primary 65C05; secondary 65U05, 60F10, 60J15, 60K10.

Key words and phrases. Monte Carlo methods, rare event simulation, hitting probabilities, large deviations, Markov additive processes, Harris recurrent Markov chains, convex analysis.

where

$$T^{\epsilon}(A) = \inf\Big\{n : S_n \in \frac{A}{\epsilon}\Big\}. \tag{1.2}$$

It will be assumed that the mean drift of $\{S_n\}$ is directed away from A, so that the probabilities in (1.1) will tend to zero as $\epsilon \to 0$. Our objective will be to develop a numerical regime based on *importance sampling* which yields an efficient estimate for (1.1) for any fixed ϵ, and which has certain optimality properties as $\epsilon \to 0$.

The first analytical work on problems of this type seems to have appeared in Lundberg (1909), where a stochastic model was introduced for the capital fluctuations of an insurance company, and the risk faced by a company under this model was studied. In Lundberg's model, an insurance company gains capital from a constant stream of premiums inflow, and loses capital as a result of i.i.d. claims which arise at a Poisson rate. These assumptions imply that the total capital gain by time t, denoted S_t, is a Lévy process, assumed to have positive drift. The *ruin problem* then considers $\mathbf{P}\left\{S_t < -1/\epsilon, \text{ some } t \geq 0\right\}$, i.e., the probability that a company with an initial capital of $1/\epsilon$ will *ever* have negative total capital, or incur ruin. A classical result due to Cramér (1930) states

$$\mathbf{P}\Big\{S_t < -\frac{1}{\epsilon}, \text{ some } t \geq 0\Big\} \sim Ce^{-R/\epsilon} \text{ as } \epsilon \to 0 \tag{1.3}$$

for certain constants C and R.

Cramér's result and techniques were later extended to more general processes, and applied in queueing theory and, with some modification, in sequential analysis. An extension to higher dimensions was given in Collamore (1996a, b). There it was shown that if A is an arbitrary open subset of $\mathbb{R}^d$ and $S_1, S_2, \ldots$ are the sums of an i.i.d., Markov, or more general sequence of r.v.'s, then

$$\lim_{\epsilon \to 0} \epsilon \log \mathbf{P}\left\{T^{\epsilon}(A) < \infty\right\} = -\inf_{v \in A} I_{\mathcal{P}}(v), \tag{1.4}$$

where $I_{\mathcal{P}}$ is the support function of the d–dimensional surface $\{\alpha : \Lambda_{\mathcal{P}}(\alpha) = 0\}$ and $\Lambda_{\mathcal{P}}$ is the cumulant generating function of $\{S_n/n\}$. Further distributional properties of $T^{\epsilon}(A)$ were explored in Collamore (1998). This multidimensional problem is of current applied interest e.g. in risk theory, where it is of some concern to model the dependence of claims along different lines of an insurance company. Also, (1.4) serves as a preliminary study for queueing network problems which can be modelled as reflected random walk in $\mathbb{R}^d$, as in Borovkov and Mogulskii (1996).

While (1.3), (1.4) provide useful asymptotic results, which (for (1.3)) may be quite accurate when the limit is removed from the left–hand side, these estimates give no indication about the rate at which the convergence to the limit actually takes place. To circumvent this problem, numerical techniques were introduced for a closely related problem in sequential analysis by Siegmund (1976). Siegmund's approach utilized the numerical technique of importance sampling; namely, to estimate $\mathbf{P}(C)$, write

$$\mathbf{P}(C) = \mathbf{E}\left(\frac{d\mathbf{P}}{d\mathbf{Q}}(Z)\mathbf{1}_C(Z)\right), \text{ where } \mathcal{L}(Z) = \mathbf{Q}, \ \mathbf{1}_C = \text{indicator function on } C. \tag{1.5}$$

Then $\mathbf{P}(C)$ is numerically computed by simulating $\mathcal{E} \stackrel{\text{def}}{=} \frac{d\mathbf{P}}{d\mathbf{Q}}(Z)\mathbf{1}_C(Z)$ under the distribution $\mathbf{Q}$, and averaging the empirical samples of $\mathcal{E}$. In the context of the standard two–sided boundary crossing problem in sequential analysis, Siegmund showed that a judicious choice of $\mathbf{Q}$ leads to a much–reduced variance of $\mathcal{E}$ as compared with direct simulation. Moreover, he showed that there is a *unique* choice of $\mathbf{Q}$ which, in an appropriate asymptotic sense, is optimal. Extensions of Siegmund's algorithm to other large deviations problems in $\mathbb{R}^1$ were later given e.g. in Lehtonen and Nyrhinen (1992a, b), Bucklew, Ney and Sadowsky (1990).

The difficulty of extending Siegmund's algorithm beyond the one–dimensional setting has been documented in Glasserman and Wang (1997). They have shown by means of certain counterexamples that there is no hope of obtaining results like Siegmund's for the multidimensional problem in (1.4), for certain sets A. Further counterexamples in a queueing context are in Glasserman and Kou (1995). These counterexamples all show that the much used technique of minimizing the variational formula in Mogulskii's theorem [Dembo and Zeitouni (1998), Theorem 5.1] does *not* lead to any sort of efficient simulation regime, in general.

In this article, we establish an analogue of Siegmund's result and some related estimates for the multidimensional problem in (1.4), under the assumption that the set A in (1.4) is convex and the process $S_1, S_2, \dots$ consists of the sums of a Markov additive sequence of r.v.'s.

To state these results more precisely, let $X_1, X_2, \dots$ be a Markov chain on a general state space $(\mathbb{S}, \mathcal{S})$, let $\{F_n\}_{n\in\mathbb{Z}_+}$ be an i.i.d. sequence of random functions mapping $\mathbb{S}\times\mathbb{S} \to \mathbb{R}^d$, and let $\xi_n = F_n(X_{n-1}, X_n)$. (In the simplest setting, $\{\xi_n\}$ is itself i.i.d.) Let $S_n = X_1 + \cdots + X_n$, $n \geq 1$, and $S_0 = 0$. For $A \subset \mathbb{R}^d$ we consider

$$\mathbf{P}\big\{T^\epsilon(A) < \infty\big\} = \mathbf{P}\big\{S_n \in \frac{A}{\epsilon},\ \text{some } n \in \mathbb{Z}_+\big\} \text{ as } \epsilon \to 0$$

for the Markov additive process $\{(X_n, S_n) : n = 0, 1, \dots\}$. This process has a transition kernel $\mathcal{P}(x, E \times \Gamma) \stackrel{\text{def}}{=} \mathbf{P}\big\{(X_{n+1}, \xi_{n+1}) \in E \times \Gamma | X_n = x\big\}$.

The importance sampling technique suggests that we simulate $\mathbf{P}\{T^\epsilon(A) < \infty\}$ with *another* Markov additive sequence $\{(\tilde{X}_n, \tilde{S}_n) : n = 0, 1, \dots\}$ having transition kernel $\mathcal{Q} = \mathbf{P}\big\{(\tilde{X}_{n+1}, \tilde{\xi}_{n+1}) \in E \times \Gamma | \tilde{X}_n = x\big\}$. An adaptation of (1.5) then becomes

$$\mathbf{P}\left\{T^\epsilon(A) < \infty\right\} = \mathbf{E}_{\mathcal{Q}}\left(\mathcal{E}_{\mathcal{Q},\epsilon}\right)$$

for some "estimator" $\mathcal{E}_{\mathcal{Q},\epsilon}$ that is computed from the $\mathcal{Q}$–distributed sequence of simulated r.v.'s $\{X_0, \dots, X_{T^\epsilon(A)}; S_0, \dots, S_{T^\epsilon(A)}\}$. The main objective is to choose $\mathcal{Q}$ so that it minimizes $\mathrm{Var}_{\mathcal{Q}}(\mathcal{E}_{\mathcal{Q},\epsilon})$ as $\epsilon \to 0$, or equivalently $\mathbf{E}_{\mathcal{Q}}\big(\mathcal{E}^2_{\mathcal{Q},\epsilon}\big)$ as $\epsilon \to 0$.

Under the assumption that A is convex, our first result provides a large deviations estimate of the form

$$\lim_{\epsilon\to 0} \epsilon \log \mathbf{E}_{\mathcal{Q}}\big(\mathcal{E}^2_{\mathcal{Q},\epsilon}\big) = -\inf_{v\in\mathfrak{A}} I_{\mathcal{K}_{\mathcal{Q}}}(v), \tag{1.6}$$

for some subset $\mathfrak{A}$ of ∂A and some "rate function" $I_{\mathcal{K}_{\mathcal{Q}}}$. This establishes the *precise* correspondence between $\mathcal{Q}$ and the decay (or growth) rate of $\mathbf{E}\big(\mathcal{E}^2_{\mathcal{Q},\epsilon}\big)$ as $\epsilon \to 0$. The implication of this estimate is made clear in Example 3.1.4, where the level sets of $\mathbf{E}\big(\mathcal{E}^2_{\mathcal{Q},\epsilon}\big)$ as a function of $\mathcal{Q}$ are described by an explicit asymptotic formula.

Eq. (1.6) suggests that an efficient choice of $\mathcal{Q}$ should be one that maximizes $J(\mathcal{Q}) \stackrel{\text{def}}{=} \inf_{v\in\mathfrak{A}} I_{\mathcal{K}_\mathcal{Q}}(v)$. In the case that A is convex, we show there exists a *unique* choice of $\mathcal{Q}$ which maximizes $J(\mathcal{Q})$; moreover, under this optimal distribution, simulation is efficient and has "logarithmic efficiency" and very often "bounded relative error." This optimality is shown to be quite general and to extend to the case where the simulation distribution is allowed to be time–dependent. In conjuction with (1.6) we then obtain, in addition, a rather complete description of the robustness of the optimal distribution. This seems to be of some practical relevance. For example, in complex problems results such as ours serve only as guides; it is therefore necessary to ask how sensitive our simulation regime may be to slight perturbations in the problem, including slight changes in $\mathcal{Q}$. Also, in practical problems simulation is very often blind (see Bucklew (1998) for some discussion on "blind" simulation proceedures).

We conclude by observing that if A is a general set, then it is possible to partition A into a finite subcollection, $A_1, \ldots, A_l$, and simulate independently along the elements of this subpartition. We show that a useful partition can *always* be obtained. The basic idea is to partition A along the *level sets* of the function $I_\mathcal{P}$ in (1.4). The resulting estimator will generally be efficient and in some main cases will have "bounded relative error."

The above results can be easily generalized to finite time–horizon problems of the form $\mathbf{P}\left\{T^\epsilon(A) < K/\epsilon\right\}$, $K < \infty$, although the optimal simulation distribution may be different; the required modifications follow along the lines of Collamore (1998).

We will establish our results in some generality, at the level of Markov additive processes in general state space, as studied in a large deviations context by Ney and Nummelin (1987a, b), de Acosta (1988), de Acosta and Ney (1998), and references therein, and the seminal papers of Donsker and Varadhan (1975, 1976, 1983). Thus, our results differ from known importance sampling results given e.g. in Siegmund (1976) or Lehtonen and Nyrhinen (1992a, b), which focus on i.i.d. sums or the sums of a finite state space Markov chain, and Bucklew, Ney and Sadowsky (1990), where sums of a general state space Markov chain are considered, but under a strong uniform recurrence condition. The usefulness of this general approach is illustrated in Example 3.1.5, where we apply our results to the stationary ARMA(p, q) time series models. Further applications to other stationary, Markov, and semi–Markov processes are likewise possible.

To prove our results we will rely on the theory of convex analysis, as summarized in the classic book of Rockafellar (1970). This theory leads to separation properties for the rate functions in (1.4), (1.6), from which the optimal simulation distribution is obtained. Also, we will rely on the theory of non–negative operators, as described in Nummelin (1984). These latter results have been previously applied in Ney and Nummelin (1987a, b). In our setting, though, they will be used in a somewhat different way: We will make use of abstract renewal properties, and—in contrast with Ney and Nummelin's work—our renewal structure will *not* generally coincide with the inherent renewal structure of the Markov additive process, or the simulated process.

In the next section, we introduce Markov additive processes in general state space and provide some necessary background on these processes and non–negative kernels. The main results are stated formally in Section 3 and proved in Section 4.

2 Background

2.1 MA–processes: definition and regenerative property

We now give a precise description of the processes we consider. Let $\{X_n : n = 0, 1, \dots\}$ be a Markov chain on a countably generated general measurable space $(\mathbb{S}, \mathcal{S})$. Assume $\{X_n\}$ is aperiodic and irreducible with respect to a maximal irreducibility measure φ.

To this Markov chain adjoin an additive sequence $\{\xi_n\}$ such that $\{(X_n, \xi_n) : n = 1, 2, \dots\}$ is a Markov chain on $(\mathbb{S} \times \mathbb{R}^d, \mathcal{S} \times \mathcal{R}^d)$, where $\mathcal{R}^d$ denotes the Borel σ–algebra on $\mathbb{R}^d$. Let $S_n = \xi_1 + \dots + \xi_n$, $n = 1, 2, \dots$, and $S_0 = 0$. The sequence $\{(X_n, S_n) : n = 0, 1, \dots\}$ is a *Markov additive process* (abbr. "MA–process"). The transition kernel of this process is

$$\mathcal{P}(x, E \times \Gamma) \stackrel{\text{def}}{=} \mathbf{P}\left\{(X_{n+1}, \xi_{n+1}) \in E \times \Gamma | X_n = x\right\}, \tag{2.1}$$

for all $x \in \mathbb{S}, E \in \mathcal{S}, \Gamma \in \mathcal{R}^d$. Let $\mathfrak{F}_n$ denote the σ–algebra generated by $\{X_0, \dots, X_n, S_0, \dots, S_n\}$.

A φ–irreducible Markov chain *always* has a minorization [Nummelin (1984), Theorem 2.1]. Following Ney and Nummelin (1987a, b), we will work with a hypothesis which extends this minorization to MA–processes.

Minorization:

($\mathfrak{M}$) For some family of measures $\left\{h(x, \Gamma) : \Gamma \in \mathcal{R}^d\right\}$ on $\mathbb{R}^d$ and some probability measure $\left\{\nu(E \times \Gamma) : E \in \mathcal{S},\ \Gamma \in \mathcal{R}^d\right\}$ on $\mathbb{S} \times \mathbb{R}^d$,

$$h(x, \cdot) * \nu(E \times \cdot) \le \mathcal{P}(x, E \times \Gamma), \quad \text{for all } x \in \mathcal{S},\ E \in \mathcal{S},\ \Gamma \in \mathcal{R}^d.$$

[$*$ denotes convolution. We will often abbreviate the left–hand side by $h \circledast \nu$.] As in Ney and Nummelin (1987a, b), we will generally assume that either h or ν is independent of s [more precisely, $h(x, ds) = g(x)\eta_{\mathcal{O}}(ds)$, where $\eta_{\mathcal{O}}$ denotes point mass at the origin, or analogously for ν]. When this is the case, we will say that ($\mathfrak{M}'$) holds.

At certain times, we will strengthen this minorization to the following:

($\mathfrak{R}$) $a\nu(E \times \Gamma) \le \mathcal{P}(x, E \times \Gamma) \le b\nu(E \times \Gamma), \quad$ for all $x \in \mathcal{S},\ E \in \mathcal{S},\ \Gamma \in \mathcal{R}^d$, where ν is as in ($\mathfrak{M}$), and a, b are positive constants.

When ($\mathfrak{R}$) holds, the MA–process is said to be "uniformly recurrent."

Under ($\mathfrak{M}$), a regenerative structure can be deduced for the MA–process:

Lemma 2.1 *Let $\{(X_n, S_n)\}_{n \ge 0}$ be a MA–process satisfying ($\mathfrak{M}$). Then there exist r.v.'s $0 < T_0 < T_1 < \cdots$ and a decomposition $\xi_{T_i} = \xi'_{T_i} + \xi''_{T_i}$, $i = 0, 1, \dots$, with the following properties:*

(i) $\{T_{i+1} - T_i : i = 0, 1, \ldots\}$ *are i.i.d. and finite a.e.;*

(ii) *the random blocks* $\left\{X_{T_i}, \ldots, X_{T_{i+1}-1}, \xi''_{T_i}, \xi_{T_i+1}, \ldots, \xi_{T_{i+1}-1}, \xi'_{T_{i+1}}\right\}$ *are independent;*

(iii) $\mathbf{P}_x \left\{(X_{T_i}, \xi''_{T_i}) \in E \times \Gamma'' \mid \mathfrak{F}_{T_i-1}, \xi'_{T_i}\right\} = \nu(A \times \Gamma'')$, *for all* $E \in \mathcal{S}$ *and* $\Gamma'' \in \mathcal{R}^d$.

For Harris recurrent Markov chains, this lemma was established by Athreya and Ney (1978) and Nummelin (1978). The extension to MA–processes is in Ney and Nummelin (1984).

Remark 2.1.1 (i) If the function h in $(\mathfrak{M})$ is independent of x, i.e. if the lower bound of $(\mathfrak{R})$ holds, then $\mathbf{P}\{T_i = n, \text{ some } i \mid \mathfrak{F}_{n-1}\} \geq a$, where a is the positive constant in $(\mathfrak{R})$. Thus, in particular, $\mathbf{E}\,(T_{i+1} - T_i) < \infty$, $i \geq 0$, and $\mathbf{E}\,(T_0) < \infty$.

(ii) If h is independent of s, then $\xi'_{T_i} = 0$, $i \geq 0$; and if ν is independent of s, then $\xi''_{T_i} = 0$, $i \geq 0$. See Ney and Nummelin (1984).

Futher properties of Markov chains in general state space can be found in Nummelin (1984), Revuz (1975), and Meyn and Tweedie (1993). Further properties of MA–processes can be found in the large deviations papers of de Acosta (1988), de Acosta and Ney (1998), and especially Ney and Nummelin (1987a, b).

2.2 Nonnegative kernels, eigenvalues and eigenvectors

We will also need certain facts about nonnegative kernels, which we now summarize and apply in the context of MA–processes. For more complete explanations, see Nummelin (1984).

Let $\{K(x, E) : x \in \mathbb{S},\ E \in \mathcal{S}\}$ be a σ–finite nonnegative φ–irreducible kernel on a countably generated measurable space $(\mathbb{S}, \mathcal{S})$. For any function $h : \mathbb{S} \to \mathbb{R}$ and any measure ν on $(\mathbb{S}, \mathcal{S})$, let

$$Kh(x) = \int K(x, dy)h(y), \qquad \nu K(E) = \int \nu(dx)K(x, E),$$

$$(h \otimes \nu)(x, E) = h(x)\nu(E), \qquad \nu h(E) = \int_E \nu(dx)h(x), \qquad \nu h = \nu h(\mathbb{S}).$$

Assume

$$h \otimes \nu \leq K. \tag{2.2}$$

Define

$$G^{(\rho)} = \sum_{n=0}^{\infty} \rho^n K^n, \qquad G^{(\rho)}_{h,\nu} = \sum_{n=0}^{\infty} \rho^n (K - h \otimes \nu)^n,$$

$$b_n = \nu(K - h \otimes \nu)^{n-1} h, \qquad \hat{b}(\rho) = \sum_{n=1}^{\infty} \rho^n b_n.$$

We say that R is the *convergence parameter* of K if $G^{(\rho)}$ is "finite" for $\rho < R$ and "infinite" for $\rho > R$. [A precise definition can be found on pp. 27–8 of Nummelin (1984).] The kernel K with convergence parameter R is said to be *R–recurrent* if $G^{(R)}(x,E) = \infty$ for $x \in \mathbb{S}$, $\varphi(E) > 0$, and *R–transient* if this is not true. It can be shown that K is R–recurrent $\Longleftrightarrow \hat{b}(R) = 1$.

A function $r : \mathbb{S} \to [0,\infty]$ (not $\equiv \infty$) is *ρ–subinvariant* if $\rho K r \le r$, and *invariant* (with unique eigenvalue $\lambda = \rho^{-1}$) if $\rho K r = r$. If R is the convergence parameter of K, then the existence of invariant and subinvariant functions for K can be obtained under (2.2), as follows. If $\rho < R$ or if $\rho = R$ and K is R–transient, then a ρ–subinvariant function exists (given by $r(x) = (G^{(\rho)}h)(x)$). If K is R–recurrent, then an R–invariant function exists (given by $r(x) = (RG^{(R)}_{h,\nu}h)(x)$). [See Nummelin (1984), Proposition 5.2 and Theorem 5.1.]

Now specialize to the transformed Markov additive kernel $\hat{\mathcal{P}}(\alpha)$, where (for any kernel K)

$$\hat{K}(\alpha) = \hat{K}(x,E;\alpha) \stackrel{\text{def}}{=} \int_{\mathbb{R}^d} e^{\langle \alpha, s\rangle} K(x, E \times ds), \quad \alpha \in \mathbb{R}^d,\ x \in \mathbb{S},\ E \in \mathcal{S},$$
$$(\lambda_K(\alpha))^{-1} = \text{the convergence parameter of } \hat{K}(\alpha), \text{ and } \Lambda_K(\alpha) = \log \lambda_K(\alpha).$$

Let $\{T_i\}_{i \ge 0}$ and $\{(\xi_i', \xi_i'')\}_{i\ge 0}$ be given as in Lemma 2.1, and let

$$\tau \stackrel{d}{=} T_{i+1} - T_i, \qquad S_\tau \stackrel{d}{=} \left(\xi_{T_i+1} + \cdots + \xi_{T_{i+1}-1}\right) + \xi''_{T_i} + \xi'_{T_{i+1}},$$
$$\psi(\alpha,\zeta) = \mathbf{E}_\nu \left[e^{\langle \alpha, S_\tau\rangle - \zeta\tau} \right], \quad \text{all } \alpha \in \mathbb{R}^d,\ \zeta \in \mathbb{R},$$
$$\mathcal{U}_r = \left\{ \alpha : \psi(\alpha,\zeta) = 1, \text{ some } \zeta < \infty \right\}.$$

Observe that $(\mathfrak{M}) \Longrightarrow \hat{h}(\alpha)\otimes\hat{\nu}(\alpha) \le \hat{\mathcal{P}}(\alpha)$, where (for any function h and any measure ν)

$$\hat{h}(x;\alpha) = \int_{\mathbb{R}^d} e^{\langle \alpha, s\rangle} h(x, ds), \qquad \hat{\nu}(E;\alpha) = \int_{\mathbb{R}^d} e^{\langle \alpha, s\rangle} \nu(E \times ds).$$

Thus, under $(\mathfrak{M})$ the above theory for nonnegative kernels may be applied to $\hat{\mathcal{P}}(\alpha)$. This leads to certain representation formulas and other regularity properties for the relevant eigenvectors and eigenvalues, which we now describe.

Lemma 2.2 *Let $\{(X_n, S_n) : n = 0,1,\dots\}$ be a MA–process satisfying* $(\mathfrak{M})$.

(i) *If $\alpha \in \mathcal{U}_r$, then $\hat{\mathcal{P}}(\alpha)$ is $(\lambda_{\mathcal{P}}(\alpha))^{-1}$–recurrent. Moreover, the eigenvalue $\lambda_{\mathcal{P}}(\alpha)$ and invariant function $r_{\mathcal{P}}(\alpha)$ satisfy the following representation formulas:*

$$\psi(\alpha, \Lambda_{\mathcal{P}}(\alpha)) = 1, \quad r_{\mathcal{P}}(x;\alpha) = \mathbf{E}_x \left[e^{\langle \alpha, S_\tau\rangle - \Lambda_{\mathcal{P}}(\alpha)\tau} \right]. \tag{2.3}$$

(ii) *If* dom ψ *is open, then on* dom $\Lambda_{\mathcal{P}}$ *we have that $\alpha \in \mathcal{U}_r$ and $\Lambda_{\mathcal{P}}(\cdot)$ is analytic, and also $r_{\mathcal{P}}(x;\cdot)$ is finite and analytic on a set $\mathbb{F} \subset \mathbb{S}$ where $\varphi(\mathbb{F}^c) = 0$.*

(iii) *If $(\mathfrak{R})$ holds and $\alpha \in$ dom$\Lambda_{\mathcal{P}}$, then $\lambda_{\mathcal{P}}(\alpha)$ is an eigenvalue of $\hat{\mathcal{P}}(\alpha)$ and the associated invariant function $r_{\mathcal{P}}(\alpha)$ is analytic and uniformly positive and bounded on* dom $\Lambda_{\mathcal{P}}$ *(in particular, if $\hat{\nu}(\alpha) r_{\mathcal{P}}(\alpha) = 1$ then we have $a \le \lambda_{\mathcal{P}}(\alpha) r_{\mathcal{P}}(x;\alpha) \le b$).*

For the proofs, see Ney and Nummelin (1987a), Sections 3 and 4, and Iscoe, Ney and Nummelin (1985), Lemmas 3.1 and 3.4.

Remark 2.2.1 Using the split–chain construction described on p. 7 of Ney and Nummelin (1984), the quantities $\Lambda_{\mathcal{P}}(\alpha)$ and $r_{\mathcal{P}}(\cdot;\alpha)$ can be evaluated from (2.3) using direct simulation.

Remark 2.2.2 If the lower bound of $(\mathfrak{R})$ holds and $r_{\mathcal{P}}(\alpha)$ is a ρ–(sub)invariant function for $\hat{\mathcal{P}}(\alpha)$, then $r_{\mathcal{P}}(\alpha) \geq \rho a\,(\hat{\nu}(\alpha) r_{\mathcal{P}}(\alpha))$, which implies $r_{\mathcal{P}}(\alpha)$ is uniformly positive.

Finally, let $\mathcal{P}(x,\cdot) \ll \mathcal{Q}(x,\cdot)$ for all x, and define

$$\mathcal{K}_{\mathcal{Q}}(x, dy \times ds) = \left(\frac{d\mathcal{P}}{d\mathcal{Q}}(x, y \times s)\right)^2 \mathcal{Q}(x, dy \times ds). \tag{2.4}$$

Lemma 2.3 *Assume* $(\mathfrak{M})$. *Then:*

(i) $\left(\mathcal{P}(x, E \times \Gamma)\right)^2 \leq \mathcal{K}_{\mathcal{Q}}(x, E \times \Gamma)$, *for all* $x \in \mathbb{S}$, $E \in \mathcal{S}$, *and* $\Gamma \in \mathcal{R}^d$.

(ii) $\lambda^2_{\mathcal{P}}(\alpha) \leq \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)$, *for all* $\alpha \in \mathbb{R}^d$. *Moreover, if* $\alpha \in \mathcal{U}_r$, *then there is equality in this inequality if and only if*

$$\mathcal{Q}(x, dy \times ds) = e^{\langle \alpha, s\rangle - \Lambda_{\mathcal{P}}(\alpha)} \frac{r_{\mathcal{P}}(y;\alpha)}{r_{\mathcal{P}}(x;\alpha)} \mathcal{P}(x, dy \times ds) \tag{2.5}$$

φ*–a.e.* x, $\mathcal{P}$*–a.e.* (y,s), *where* $r_{\mathcal{P}}(\alpha)$ *is the* $(\lambda_{\mathcal{P}}(\alpha))^{-1}$*–invariant function for* $\hat{\mathcal{P}}(\alpha)$.

(iii) *If* $\mathcal{Q}$ *is defined as in* (2.5) *and* $\lambda_{\mathcal{P}}(\alpha)$, $\lambda_{\mathcal{P}}(\beta) < \infty$ *are eigenvalues of* $\hat{\mathcal{P}}(\alpha)$, $\hat{\mathcal{P}}(\beta)$, *resp., then* $\lambda_{\mathcal{K}_{\mathcal{Q}}}(\alpha + \beta) = \lambda_{\mathcal{P}}(\alpha)\lambda_{\mathcal{P}}(\beta)$, *and the associated invariant functions satisfy the equation* $r_{\mathcal{K}_{\mathcal{Q}}}(\alpha+\beta) = r_{\mathcal{P}}(\alpha) r_{\mathcal{P}}(\beta)$.

Proof (i) is established using Hölder's inequality.

For (ii), assume $\lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha) < \infty$, and let $r_{\mathcal{K}_{\mathcal{Q}}}$ be a $\left(\lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)\right)^{-1}$–(sub)invariant function for $\hat{\mathcal{K}}_{\mathcal{Q}}(2\alpha)$. Apply Hölder's inequality to the integral

$$\int_{\mathbb{S}\times\mathbb{R}^d} e^{\langle \alpha, s\rangle} r_{\mathcal{K}_{\mathcal{Q}}}(y; 2\alpha)^{\frac{1}{2}} \frac{d\mathcal{P}}{d\mathcal{Q}}(x, y\times s)\mathcal{Q}(x, dy \times ds)$$

to obtain

$$\hat{\mathcal{P}}(\alpha) r_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)^{\frac{1}{2}} \leq \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)^{\frac{1}{2}} r_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)^{\frac{1}{2}}. \tag{2.6}$$

Thus $r_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)^{\frac{1}{2}}$ is a $\left(\lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)\right)^{-\frac{1}{2}}$–subinvariant function for $\hat{\mathcal{P}}(\alpha)$. Hence $(\lambda_{\mathcal{P}}(\alpha))^2 \leq \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)$ [Nummelin (1984), Proposition 5.2].

Now suppose $\alpha \in \mathcal{U}_r$ and $(\lambda_{\mathcal{P}}(\alpha))^2 = \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)$. Then by (2.6), $r_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)^{\frac{1}{2}}$ is a $(\lambda_{\mathcal{P}}(\alpha))^{-1}$–subinvariant function for $\hat{\mathcal{P}}(\alpha)$. It follows that $r_{\mathcal{P}}(\alpha) = C\left(r_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)\right)^{\frac{1}{2}}$ φ–a.e., for some $C > 0$ [Nummelin (1984), Theorem 5.1]. Hence there is equality in (2.6), namely equality in Hölder's inequality, and—after normalizing so that $\mathcal{Q}$ is a probability measure—this implies (2.5).

Conversely, note that (2.5) implies

$$\mathcal{K}_{\mathcal{Q}}(x, E\times\Gamma) = \int_{E\times\Gamma} \lambda_{\mathcal{P}}(\alpha) e^{-\langle \alpha, s\rangle} \frac{r_{\mathcal{P}}(x;\alpha)}{r_{\mathcal{P}}(y;\alpha)} \mathcal{P}(x, dy\times ds), \quad \text{for all } E \in \mathcal{S},\ \Gamma \in \mathcal{R}^d, \tag{2.7}$$

and hence

$$\hat{\mathcal{K}}_{\mathcal{Q}}(2\alpha)\,(r_{\mathcal{P}}(\alpha))^2 = (\lambda_{\mathcal{P}}(\alpha))^2\,(r_{\mathcal{P}}(\alpha))^2\,. \tag{2.8}$$

It follows that $(\lambda_{\mathcal{P}}(\alpha))^2 = \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)$.

To establish (iii), repeat (2.7), (2.8) with "$r_{\mathcal{P}}(\alpha)r_{\mathcal{P}}(\beta)$" in place of "$r^2_{\mathcal{P}}(\alpha)$." □

3 Main results

3.1 Notation, hypotheses, and estimation

Given a MA–process $\{(X_n, S_n) : n = 0, 1, \dots\}$, we would like to evaluate $\mathbf{P}\{T^{\epsilon}(A) < \infty\}$, where $T^{\epsilon}(A)$ is defined as in (1.2). Suppose that we simulate for this quantity using simulated r.v.'s $\{(\tilde{X}^{\epsilon}_n, \tilde{S}^{\epsilon}_n) : n = 0, 1, \dots\}$ with transition kernel

$$\mathcal{Q}^{n,\epsilon}(x, E \times \Gamma) = \mathbf{P}\left\{(\tilde{X}^{\epsilon}_{n+1}, \tilde{\xi}^{\epsilon}_{n+1}) \in E \times \Gamma | \tilde{X}^{\epsilon}_n = x\right\}.$$

[Often it will be assumed that $\mathcal{Q}$ is independent of n, ϵ.] If $\mathcal{P}(x, \cdot) \ll \mathcal{Q}^{n,\epsilon}(x, \cdot)$ for all $n \in \mathbb{Z}_+$, $\epsilon > 0$, and $x \in \mathbb{S}$, then it follows from these definitions that

$$\begin{aligned}\mathbf{P}\{T^{\epsilon}(A) < \infty\} \;&=\; \sum_k \int_{\mathfrak{P}^{\epsilon}_k} \left(\prod_{i=1}^{k} \frac{d\mathcal{P}}{d\mathcal{Q}^{n,\epsilon}}(x_{i-1}, x_i \times s_i)\right) \\ &\qquad \cdot \mathcal{Q}^{n,\epsilon}(x_0, dx_1 \times ds_1) \cdots \mathcal{Q}^{n,\epsilon}(x_{k-1}, dx_k \times ds_k),\end{aligned} \tag{3.1}$$

where $\mathfrak{P}^{\epsilon}_k$ denotes all paths that *first* hit A/ϵ at time k, that is,

$$\mathfrak{P}^{\epsilon}_k = \left\{(x_0, \dots, x_k; s_0, \dots, s_k) : \sum_{j=1}^{l} s_i \in \frac{A}{\epsilon} \text{ for } l = k \text{ but not for } l < k\right\}. \tag{3.2}$$

It follows from (3.1) that

$$\mathcal{E}_{\mathcal{Q},\epsilon} \stackrel{\text{def}}{=} \left(\prod_{n=1}^{T^{\epsilon}(A)} \frac{d\mathcal{P}}{d\mathcal{Q}^{n,\epsilon}}(\tilde{X}^{\epsilon}_{n-1}, \tilde{X}^{\epsilon}_n \times \tilde{\xi}^{\epsilon}_n)\right) \mathbf{1}_{\{T^{\epsilon}(A) < \infty\}} \tag{3.3}$$

is an unbiased estimator for $\mathbf{P}\{T^{\epsilon}(A) < \infty\}$.

The efficiency of this estimator is measured by its variance, which we will study in an asymptotic sense as $\epsilon \to 0$. Since $\mathrm{Var}(\mathcal{E}_{\mathcal{Q},\epsilon}) = \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) - \left(\mathbf{E}(\mathcal{E}_{\mathcal{Q},\epsilon})\right)^2$, and

$$\mathbf{E}(\mathcal{E}_{\mathcal{Q},\epsilon}) = \mathbf{P}\{T^{\epsilon}(A) < \infty\}$$

has the asymptotic characterization given in (1.4) [Collamore (1996a), Theorems 2.1 and 2.2], it is sufficient to study the asymptotic behavior of

$$\mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) = \sum_k \int_{\mathfrak{P}^{\epsilon}_k} \mathcal{K}^{n,\epsilon}_{\mathcal{Q}}(x_0, dx_1 \times ds_1) \cdots \mathcal{K}^{n,\epsilon}_{\mathcal{Q}}(x_{k-1}, dx_k \times ds_k), \tag{3.4}$$

where for all n, ϵ,

$$\mathcal{K}_{\mathcal{Q}}^{n,\epsilon}(x_{n-1}, dx_n \times ds_n) \stackrel{\text{def}}{=} \left(\frac{d\mathcal{P}}{d\mathcal{Q}^{n,\epsilon}}(x_{n-1}, x_n \times s_n) \right)^2 \mathcal{Q}^{n,\epsilon}(x_{n-1}, dx_n \times ds_n). \tag{3.5}$$

Our objective will be to give estimation results for $\mathbf{E}(\mathcal{E}_{\mathcal{Q},\epsilon}^2)$ as $\epsilon \to 0$, and optimality results describing *which* transition kernels $\mathcal{Q}$ for the simulated r.v.'s minimize $\mathbf{E}(\mathcal{E}_{\mathcal{Q},\epsilon}^2)$ as $\epsilon \to 0$.

To state these results in a more formal way, we will first need to introduce some additional notation and hypotheses, as follows. Let ν and τ be given as in Section 2, and let

$$\begin{aligned}
\text{cone}\,(C) &= \{\zeta v : \zeta \geq 0,\, v \in C\}, \text{ for any } C \subset \mathbb{R}^d; \\
\text{cone}_{\,\delta}(C) &= \{\zeta v : \zeta \geq 0,\, \|v - w\| < \delta \|w\|, \text{ some } w \in C\}, \text{ for any } \delta > 0; \\
\mathfrak{S} &= \text{cone}\left(\text{Supp}_\nu \left(\frac{S_\tau}{\tau}\right) \right); \\
C^\perp &= \{v : \langle \alpha, v \rangle \leq 0, \text{ all } \alpha \in C\}, \text{ for any } C \subset \mathbb{R}^d; \\
\mathcal{H}(\alpha, a) &= \{v : \langle \alpha, v \rangle > a\}, \text{ for any } \alpha \in \mathbb{R}^d \text{ and } a \in \mathbb{R}; \\
\mathcal{L}_a f &= \{v : f(v) \leq a\}, \text{ for any } f : \mathbb{R}^d \to \mathbb{R} \text{ and } a \in \mathbb{R}.
\end{aligned}$$

For any nonnegative φ–irreducible kernel K, let

$$\begin{aligned}
\bar{\Lambda}(\alpha) &= \limsup_{n \to \infty} n^{-1} \log \hat{K}^n(X_0, \mathbb{S}; \alpha); \\
\Lambda^{(N)}(\alpha) &= \sup_{n \geq N} n^{-1} \log \hat{K}^n(X_0, \mathbb{S}; \alpha); \\
I_K(v) &= \sup\{\langle \alpha, v \rangle : \alpha \in \mathcal{L}_0 \Lambda_K\}; \quad \mathfrak{D}_K = \text{domain of } I_K; \\
I_K^{(c)}(v) &= \sup\{\langle \alpha, v \rangle : \alpha \in \mathcal{L}_c \Lambda_K\}; \quad \mathfrak{D}_K^{(c)} = \text{domain of } I_K^{(c)}; \\
\bar{I}_K(v) &= \sup\left\{\langle \alpha, v \rangle : \alpha \in \mathcal{L}_0 \bar{\Lambda}_K\right\}; \quad \text{resp. } \bar{I}_K^{(c)}(\cdot);
\end{aligned}$$

where, as before, $(\lambda_K(\alpha))^{-1}$ is the convergence parameter of $\hat{K}(\alpha)$ and $\Lambda_K(\alpha) = \log \lambda_K(\alpha)$. In the definitions of I_K, $I_K^{(c)}$, and $\bar{I}_K$, we follow the convention that the supremum over an empty set $= -\infty$.

For any set C, let $\mathbf{1}_C(\cdot)$, ri$C$, ∂C denote the indicator function on C, the relative interior of C, and the relative boundary of C, respectively. For any function f, let f^* denote the convex conjugate of f. [For definitions, see Rockafellar (1970).]

Hypotheses:

(H1) $\Lambda_{\mathcal{K}_{\mathcal{Q}}}^{(1)}(\alpha) < \infty$, for all $\alpha \in \mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}}}$.

(H2) $\text{cl}\, A \cap \text{cone}_{\,\delta}\left(\mathcal{L}_0 \Lambda_{\mathcal{P}}^*\right) = \emptyset$, for some $\delta > 0$.

For example, if $S_1, S_2, \ldots$ are the sums of an i.i.d. sequence of random variables, or the additive sums of a MA–process and $(\mathfrak{R})$ is satisfied under both $\mathcal{P}$ and $\mathcal{Q} \in \mathcal{C}_0$, then $\Lambda^{(N)}_{\mathcal{K}_\mathcal{Q}}(\alpha) < \infty$ for all $\alpha \in \operatorname{dom} \bar{\Lambda}_{\mathcal{K}_\mathcal{Q}}$ and $N \in \mathbb{Z}_+$; and so (H1) is then *always* satisfied. For (H2), observe that if $(\mathfrak{R})$ holds and $0 \in \operatorname{dom} \Lambda_\mathcal{P}$, or if $(\mathfrak{M}')$ holds and $\operatorname{dom} \psi$ is an open set, then

$$\Lambda^*_\mathcal{P}(v) = 0 \Longleftrightarrow \mathbf{E}_\pi(S_1) = v,$$

where π is the stationary measure of the Markov chain $\{X_n\}$. [Cf. Ney and Nummelin (1987a), Lemma 3.3 and Lemma 5.2.] Thus, hypothesis (H2) holds as long as the set A avoids an arbitrarily thin δ–cone about the mean ray $\{\zeta \mathbf{E}_\pi(S_1) : \zeta \geq 0\}$.

Definition We say that a simulation kernel $\mathcal{Q}^{n,\epsilon}$ belongs to the class $\mathcal{C}_0$ if $\mathcal{Q}^{n,\epsilon} = \mathcal{Q}$, independent of n and ϵ, and $\mathcal{P}(x,\cdot) \ll \mathcal{Q}(x,\cdot)$, for all $x \in \mathbb{S}$.

Definition If $A \subset \mathbb{R}^d$, then we say that $v \in \partial A$ is an *exposed point* of A if the ray $\{\zeta v : v > 1\} \subset \operatorname{int} A$, that is, the ray generated by v is an *interior* ray of A.

Theorem 3.1 *Let A be a convex open set intersecting* $\operatorname{ri} \mathfrak{S}$. *Let $\mathfrak{A}$ denote the exposed points of A. Assume that the MA–process $\{(X_n, S_n) : n = 0, 1, \ldots\}$ satisfies $(\mathfrak{M}')$ and has initial state $X_0 = x_0$. Suppose that simulation is performed with a kernel $\mathcal{Q} \in \mathcal{C}_0$. Then:*

(i) Lower bound.

$$\liminf_{\epsilon \to 0} \epsilon \log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \geq - \inf_{v \in \mathfrak{A}} I_{\mathcal{K}_\mathcal{Q}}(v). \tag{3.6}$$

(ii) Upper bound. *Further assume that hypothesis (H1) is satisfied, and:* (a) $\operatorname{cl} A \cap \operatorname{cone}_\delta \left(\mathcal{L}_0 \Lambda_{\mathcal{K}_\mathcal{Q}}\right)^\perp = \emptyset$, *for some* $\delta > 0$; (b) $\inf_\alpha \bar{\Lambda}_{\mathcal{K}_\mathcal{Q}}(\alpha) < 0$. *Then for φ a.e. x_0,*

$$\limsup_{\epsilon \to 0} \epsilon \log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \leq - \inf_{v \in \mathfrak{A}} \bar{I}_{\mathcal{K}_\mathcal{Q}}(v). \tag{3.7}$$

Remark 3.1.1 (i) If the lower bound of $(\mathfrak{R})$ is satisfied, then it follows from the definitions of $\Lambda_{\mathcal{K}_\mathcal{Q}}$, $\bar{\Lambda}_{\mathcal{K}_\mathcal{Q}}$, and of the convergence parameter that $I_{\mathcal{K}_\mathcal{Q}}(v) = \bar{I}_{\mathcal{K}_\mathcal{Q}}$ for all v. Thus the upper and lower bounds are the same in this case. There are also other examples where the upper and lower bounds are the same, as shown below in Example 3.1.5. More generally, it is well known in the context of large deviations for MA–processes that these bounds need *not* be the same; see Section 4 of de Acosta and Ney (1998).

(ii) The condition (a) in the statement of the upper bound can be viewed as a strengthening of (H2): It follows from the definitions that $\left(\mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_\mathcal{Q}}\right)^\perp = \mathcal{L}_0 \bar{I}_{\mathcal{K}_\mathcal{Q}}$. Since by Hölder's inequality $2\,(\mathcal{L}_0 \Lambda_\mathcal{P}) \supset \mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_\mathcal{Q}}$ [as in Lemma 2.3 (ii)], it can be shown that $\mathcal{L}_0 \bar{I}_{\mathcal{K}_\mathcal{Q}} \supset \mathcal{L}_0 I_\mathcal{P} = \{\zeta v : \zeta \geq 0,\, v \in \mathcal{L}_0 \Lambda^*_\mathcal{P}\}$.

If the lower bound of $(\mathfrak{R})$ holds, then it is sufficient to assume the weaker condition $\inf\{I_{\mathcal{K}_\mathcal{Q}} : v \in \mathfrak{A}\} > 0$; see the comments following the proof of the upper bound. Further weakening of this assumption (to the case where $\mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon})$ exhibits exponential *growth* as $\epsilon \to 0$) is not possible in general.

Remark 3.1.2 If the lower bound of $(\mathfrak{R})$ holds, then as an upper bound we actually obtain

$$\mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \le \text{const.} \cdot \exp\Big\{ -\inf_{v\in\mathfrak{A}} I_{\mathcal{K}_\mathcal{Q}}(v)/\epsilon \Big\}. \tag{3.8}$$

Remark 3.1.3 If $S_n = \xi_1 + \cdots + \xi_n$, where $\{\xi_n\}_{n\in\mathbb{Z}_+}$ is an i.i.d. sequence of random variables, then the quantities $\Lambda_\mathcal{P}$ and $\Lambda_{\mathcal{K}_\mathcal{Q}}$ which determine the rate functions on the right of (3.6), (3.7) can be simplified. In this setting, $\Lambda_\mathcal{P} = \bar{\Lambda}_\mathcal{P}$ may be identified as the cumulant generating function of ξ_n, namely,

$$\Lambda_\mathcal{P}(\alpha) = \int_{\mathbb{R}^d} e^{\langle\alpha,s\rangle}\mathcal{P}(ds),$$

where $\mathcal{P}$ is the probability law of ξ_n, and similarly for $\Lambda_{\mathcal{K}_\mathcal{Q}}$. Furthermore, any discussion of eigenvectors or subinvariant functions may be dropped, i.e., we may always take $r_\mathcal{P}(\cdot;\alpha) = 1$ and $r_{\mathcal{K}_\mathcal{Q}}(\cdot;\alpha) = 1$.

Example 3.1.4 Let $S_n = \xi_1 + \cdots + \xi_n$, where $\{\xi_n\}_{n\in\mathbb{Z}_+} \subset \mathbb{R}^2$ is an i.i.d. sequence of Normal r.v.'s with mean $m = (\mu,\mu)$ and covariance $S = \binom{1\ \sigma}{\sigma\ 1}$, where $\mu > 0$ and $0 \le \sigma < 1$. Let $A = \{(v_1,v_2) : v_1 < -1 \text{ and } v_2 < -1\}$. We consider the simulation of $\mathbf{P}\{T^\epsilon(A) < \infty\}$ using an exponentially tilted distribution of the form

$$\mathcal{Q}_\beta(ds) = e^{\langle\beta,s\rangle - \Lambda_\mathcal{P}(\beta)}\mathcal{P}(ds). \tag{3.9}$$

By Lemma 2.3 (iii), $\alpha \in \mathcal{L}_0\Lambda_{\mathcal{K}_\mathcal{Q}} \iff \Lambda_\mathcal{P}(\beta) + \Lambda_\mathcal{P}(\alpha-\beta) \le 0$. Since the cumulant generating function for a Normal(m,S) r.v. is $\Lambda_\mathcal{P}(\alpha) = \langle\alpha,m\rangle + \frac{1}{2}\langle\alpha,S\alpha\rangle$, it follows from a straightforward computation that

$$\mathcal{L}_0\Lambda_{\mathcal{K}_{\mathcal{Q}_{\bar\beta}}} = \left\{\bar\alpha :\ (1+\sigma)\Big(\bar\alpha_1 + \frac{\sqrt{2}\mu}{1+\sigma} - \bar\beta_1\Big)^2 + (1-\sigma)\Big(\bar\alpha_2 - \bar\beta_2\Big)^2 \le b\right\}, \tag{3.10}$$

where $b = -(1+\sigma)\bar\beta_1^2 - 2\sqrt{2}\mu\bar\beta_1 - (1-\sigma)\bar\beta_2^2 + 2\mu^2/(1+\sigma)$, and $\bar\alpha$, $\bar\beta$ denote the values of α, β in a coordinate system which has been rotated by angle $\pi/4$.

Our objective is to apply Theorem 3.1 to analyze the dependence of $\mathbf{E}(\mathcal{E}^2_{\mathcal{K}_{\mathcal{Q}_{\bar\beta}}})$ on $\bar\beta$. Thus we would like to study

$$J(\bar\beta) = -\inf_{v\in\partial A} I_{\mathcal{K}_{\mathcal{Q}_{\bar\beta}}}(v), \tag{3.11}$$

i.e. the rate function on the right of (3.6) and (3.7), as a function of $\bar\beta$. [If $J \ge 0$, the right–hand side of (3.7) must be taken to be infinity rather than as in (3.11).]

Suppose for simplicity that $\mu = 1/\sqrt{2}$ and $\sigma = 1/2$. Let $F_1 = \{\bar\beta : 3\big(\bar\beta_1 + 7/6\big)^2 + \big(\bar\beta_2 - 3/2\big)^2 \le 29/3\}$ and $F_2 = \{\bar\beta : 3\big(\bar\beta_1 + 7/6\big)^2 + \big(\bar\beta_2 + 3/2\big)^2 \le 29/3\}$.

By (3.10) and a straightforward computation, the level sets where $J(\bar\beta) = -r < 0$ are given by

$$\left(\bar\beta_1 + \frac{r}{2\sqrt{2}}\right)^2 + \frac{1}{6}\bar\beta_2^2 = r\left(\frac{\sqrt{2}}{3} - \frac{r}{8}\right), \tag{3.12}$$

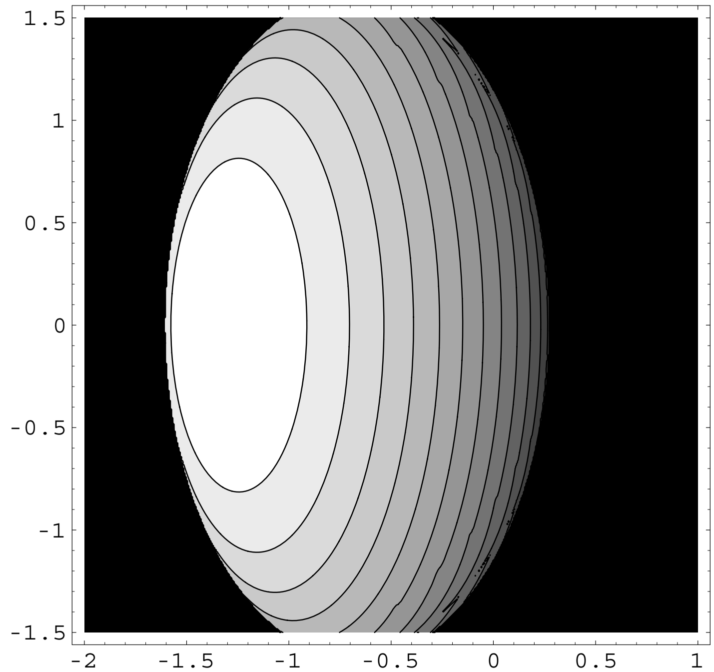

Figure 1: Let $a = \inf\{J(\bar\beta) : \bar\beta \in \mathbb{R}^2\}$, where J is defined as in (3.11). The figure illustrates the level lines $r = a + 0.25$, $a + 0.5$, $a + 0.75, \ldots$, with $\bar\beta_1$ on the horizontal axis, and $\bar\beta_2$ on the vertical axis. J is seen to increase rapidly to the left of its minimum at $(-4/3, 0)$. The black area indicates the region where $J = \infty$.

provided $\bar\beta \in F_1 \cap F_2$. On $(F_1 \cup F_2)^c$, we have $J(\bar\beta) = \infty$. The behavior of J on $(F_1 \cup F_2)^c - (F_1 \cap F_2)$ is somewhat more complicated but analytically tractable. In particular, $J \leq 0$ or $= \infty$; J is smaller than what would be predicted by (3.12), but larger than $\sup\{J(\bar\beta) : \bar\beta \in F_1 \cap F_2\}$. A graph of the the level sets of J is given in Figure 1.

The minimum value of J occurs at the maximum r for which the right–hand side of (3.12) ≥ 0, i.e. $r = 8\sqrt{2}/3$, and for this r we obtain by (3.12) that $\bar\beta = (-4/3, 0)$. The points where $J = \infty$ are all contained in the complement of the zero–set $\mathcal{L}_0\Lambda_{\mathcal{P}} = \{\bar\alpha : 3(\bar\alpha_1 + \frac{2}{3})^2 + \bar\alpha_2^2 \leq \frac{4}{3}\}$. This illustrates the fact that $J(\bar\beta)$ tends to be smaller on $\mathcal{L}_0\Lambda_{\mathcal{P}}$ as compared with $(\mathcal{L}_0\Lambda_{\mathcal{P}})^c$.

Example 3.1.5 Let $\{Y_n\}_{n\in\mathbb{Z}_+}$ be an ARMA(p,q) process in $\mathbb{R}^d$, namely,

$$Y_n = -(\phi_1 Y_{n-1} + \cdots + \phi_p Y_{n-p}) + W_n + \theta_1 W_{n-1} + \cdots + \theta_q W_{n-q}$$

for constants $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ satisfying appropriate regularity conditions [see Brockwell and Davis (1991), Chapter 3]. For simplicity, take $\{W_n\}_{n\in\mathbb{Z}_+}$ to be i.i.d. Normal$(0, S)$. As on p. 28 of Meyn and Tweedie (1993), we may then write $Y_n = F(X_n)$, where $\{X_n\}$ is a Markov chain taking values in $\mathbb{R}^{ld}$, where $l = \max\{p, q\}$, and this Markov chain can be shown to satisfy $(\mathfrak{M}')$ (by taking $h(x) = \text{const.} \cdot \mathbf{1}_{\mathcal{O}_\epsilon}(x)$, where $\mathcal{O}_\epsilon$ is an ϵ–neighborhood of the origin).

Assume that the past history of the process is known, or equivalently, the initial state of the Markov chain is $X_0 = x_0$ for some $x_0 \in \mathbb{R}^{ld}$. Let $m \in \mathbb{R}^d - \{0\}$ and $\xi_n = Y_n + m$;

and let $S_n = \xi_1 + \cdots + \xi_n$, $n \geq 1$, and $S_0 = 0$. Then $\{(X_n, S_n) : n = 0, 1, \ldots\}$ is a Markov additive process.

A simple computation gives

$$\bar{\Lambda}_{\mathcal{P}}(\alpha) = \langle \alpha, m \rangle + \sum_{j=0}^{\infty} \Psi_j \langle \alpha, S\alpha \rangle \tag{3.13}$$

for certain constants $\{\Psi_j\}$ [as given in Brockwell and Davis (1991), Theorem 3.1.1], and the upper bound of Theorem 3.1 is determined by (3.13). To see that the lower bound is likewise determined by (3.13), note

$$\begin{aligned} \Lambda_{\mathcal{P}}(\alpha) &= \limsup_{n\to\infty} \frac{1}{n} \log \mathbf{E}\left[e^{\langle \alpha, S_n \rangle} \mathbf{1}_{\mathcal{O}_\epsilon}(X_n)\right] \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \mathbf{E}\left[e^{\langle \alpha, S_n \rangle} \mathbf{1}_{A_{bn}}(X_{\lfloor bn \rfloor}) \mathbf{1}_{\mathcal{O}_\epsilon}(X_n)\right], \end{aligned} \tag{3.14}$$

where $b \in (0,1)$, $A_k = [-\sqrt{k}a, \sqrt{k}a]^p$, and a is a suitably large positive constant; the last step was obtained by an application of Hölder's inequality which separates the expectation over "$e^{\langle \alpha, S_n \rangle} \mathbf{1}_{\mathcal{O}_\epsilon}(\mathbf{X}_n)$" and "$\mathbf{1}^c_{A_{bn}}(\mathbf{X}_{\lfloor bn \rfloor})$." Finally, it can be shown that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{E}\left[e^{\langle \alpha, S_n - S_{\lfloor bn \rfloor} \rangle} \mathbf{1}_{\mathcal{O}_\epsilon}(X_n) \mid X_{\lfloor bn \rfloor} \in A_{cn}\right] = (1-b)\Lambda_{\mathcal{P}}(\alpha), \tag{3.15}$$

with convergence uniform in $c \leq b$. Substituting (3.15) into (3.14) and then letting $b \to 1$ yields $\Lambda_{\mathcal{P}}(\alpha) = \bar{\Lambda}_{\mathcal{P}}(\alpha)$, for all α.

Using (3.13) in place of (3.9), we may now proceed as in the previous example to determine the interdependence of $\mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon})$ and $\mathcal{Q}$.

For further applications to Markov and semi–Markov processes, see e.g. Iscoe, Ney and Nummelin (1985), Ney and Nummelin (1987a), and Meyn and Tweedie (1993).

3.2 Optimality

Our next objective is to find an optimal $\mathcal{Q} \in \mathcal{C}_0$ which maximizes the decay rate on the right of (3.6), (3.7).

Recall that $I_{\mathcal{P}}$ is the rate function describing the decay of $\mathbf{P}\{T^\epsilon(A) < \infty\}$ as $\epsilon \to 0$, as in (1.4). Our results will depend in an essential way on the following.

Lemma 3.2 *Let $A \subset \mathbb{R}^d$ be a convex set intersecting* ri $\mathfrak{S}$. *Assume* dom $\Lambda_{\mathcal{P}}$ *is open and (H2) and the lower bound of* $(\mathfrak{R})$ *are satisfied. Let $a = \inf_{v\in A} I_{\mathcal{P}}(v)$. Then:*

(i) *There exists an element $\alpha \in \partial(\mathcal{L}_0\Lambda_{\mathcal{P}})$ that determines a hyperplane separating A and $\mathcal{L}_a I_{\mathcal{P}}$, with $A \subset \{v : \langle \alpha_0, v \rangle \geq a\}$ and $\mathcal{L}_a I_{\mathcal{P}} \subset \{v : \langle \alpha_0, v \rangle \leq a\}$.*

(ii) *There exists a* unique *element $v_0 \in$* cl A *such that $I_{\mathcal{P}}(v_0) = a$.*

(iii) $\inf_{v\in A} I_{\mathcal{P}}(v) = \inf_{v\in\partial A} I_{\mathcal{P}}(v) = \langle \alpha_0, v_0 \rangle$.

(iv) *The gradient of* $\Lambda_{\mathcal{P}}$ *at* α_0 *points in the same direction as* v_0*, that is,* $v_0 = \varrho \nabla \Lambda_{\mathcal{P}}(\alpha_0)$ *for some constant* $\varrho > 0$*. If* $\operatorname{int} \mathfrak{S} \neq \emptyset$*, then* α_0 *is the unique point of* $\partial(\mathcal{L}_0 \Lambda_{\mathcal{P}})$ *with these properties.*

For the proofs, see Collamore (1996a), Lemma 3.2, and Collamore (1998), Lemma 2.2. The uniqueness in (iv) is obtained from the strict convexity of $\Lambda_{\mathcal{P}}$ [Collamore (1996b), p. 38, and Ney and Nummelin (1987a), Corollary 3.3].

We note that if $\operatorname{dom} \psi$ is open, where ψ is defined as in Section 2, then $\operatorname{dom} \Lambda_{\mathcal{P}}$ is also open; see Lemma 3.2 of Ney and Nummelin (1987a).

To see how Theorem 3.1 and Lemma 3.2 may be applied to obtain an optimal simulation distribution which maximizes the decay rate on the right of (3.6), (3.7), we may reason as follows. First note by Lemma 2.3 (ii) that $\mathcal{L}_0 \Lambda_{\mathcal{K}_{\mathcal{Q}}} \subset 2(\mathcal{L}_0 \Lambda_{\mathcal{P}})$. Also, by Lemma 3.2 (iii), (iv),

$$\mathcal{L}_0 \Lambda_{\mathcal{P}} \cap \{\alpha : \langle \alpha - \alpha_0, v_0 \rangle \geq 0\} = \{\alpha_0\}.$$

Therefore $I_{\mathcal{K}_{\mathcal{Q}}}(v_0) \leq 2 I_{\mathcal{P}}(v_0)$, with equality $\Longleftrightarrow 2\alpha_0 \in \mathcal{L}_0 \Lambda_{\mathcal{K}_{\mathcal{Q}}}$. It follows by Lemma 2.3 (ii) that

$$I_{\mathcal{K}_{\mathcal{Q}}}(v_0) \leq 2 I_{\mathcal{P}}(v_0), \text{ with equality} \Longleftrightarrow \mathcal{Q} = \mathcal{Q}^* \quad \text{for } \varphi \times \mathcal{P} \textit{ a.e. } (x, y, s), \tag{3.16}$$

where

$$\mathcal{Q}^*(x, dy \times ds) = e^{\langle \alpha_0, s \rangle} \frac{r_{\mathcal{P}}(dy; \alpha_0)}{r_{\mathcal{P}}(x; \alpha_0)} \mathcal{P}(x, dy \times ds). \tag{3.17}$$

From (3.16) and Lemma 3.2 (ii) we then obtain:

$$\mathcal{Q} \neq \mathcal{Q}^* \Longrightarrow \inf_{v \in A} I_{\mathcal{K}_{\mathcal{Q}}}(v) < 2 I_{\mathcal{P}}(v_0) = 2 \inf_{v \in A} I_{\mathcal{P}}(v).$$

Conversely, if $\mathcal{Q} = \mathcal{Q}^*$ then by Lemma 2.3 (iii): $\mathcal{L}_0 \Lambda_{\mathcal{K}_{\mathcal{Q}^*}} = \{\alpha + \alpha_0 : \alpha \in \mathcal{L}_0 \Lambda_{\mathcal{P}}\}$. Hence $I_{\mathcal{K}_{\mathcal{Q}^*}}(v) = \langle \alpha_0, v \rangle + I_{\mathcal{P}}(v)$. It follows by Lemma 3.2 (i), (iii) that $\inf_{v \in A} I_{\mathcal{K}_{\mathcal{Q}^*}}(v) = 2 \inf_{v \in A} I_{\mathcal{P}}(v)$. In summary, from Theorem 3.1, Lemmas 2.3 and 3.2, and Remark 3.1.1 (i), may conclude the following.

Proposition 3.3 *Let* A *be a convex open set intersecting* $\operatorname{int} \mathfrak{S}$*. Suppose* $\operatorname{dom} \psi$ *is open and (H1), (H2), and the lower bound of* $(\mathfrak{R})$ *are satisfied. Let* x_0 *denote the initial state of* $\{X_n\}_{n \geq 0}$*, and let* $\mathcal{Q}^*$ *be the kernel defined in* (3.17)*. Suppose simulation is performed using a kernel* $\mathcal{Q} \in \mathcal{C}_0$*. Then:*

(i) *For* φ *a.e.* x_0*,*

$$\liminf_{\epsilon \to 0} \epsilon \log \mathbf{E}\left(\mathcal{E}^2_{\mathcal{Q},\epsilon}\right) \geq \lim_{\epsilon \to 0} \epsilon \log \mathbf{E}\left(\mathcal{E}^2_{\mathcal{Q}^*,\epsilon}\right), \tag{3.18}$$

with equality if and only if $\mathcal{Q} = \mathcal{Q}^*$ *for* $\varphi \times \mathcal{P}$ *a.e.* (x, y, s)*. Thus,* $\mathcal{Q}^*$ *is the* unique *kernel in* $\mathcal{C}_0$ *which minimizes* $\mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon})$ *as* $\epsilon \to 0$*.*

(ii) *For φ a.e. x_0,*

$$\lim_{\epsilon \to 0} \epsilon \log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q}^*,\epsilon}) = -2 \inf_{v \in A} I_{\mathcal{P}}(v). \tag{3.19}$$

Eqs. (3.19) and (1.4) imply that simulation performed under the distribution $\mathcal{Q}^*$ has "logarithmic efficiency" [Asmussen (1999), p. 46]. Moreover, by Remark 3.1.2 and a sharp form of (1.4)—available at least for the case that $\{\xi_n\}$ is i.i.d. and A is convex with smooth boundary [Borovkov (1997)]—one obtains the stronger property of "bounded relative error," i.e.,

$$\limsup_{\epsilon \to 0} \frac{\text{Var}(\mathcal{E}_{\mathcal{Q},\epsilon})}{\mathbf{E}(\mathcal{E}_{\mathcal{Q},\epsilon})^2} < \infty.$$

It is natural to expect that this stronger property also holds under ($\mathfrak{R}$). This property of "bounded relative error" is the strongest known property for nontrivial rare event simulation problems. [See Asmussen (1999).]

In the next theorem, we show that the optimality of $\mathcal{Q}^* \in \mathcal{C}_0$ is in fact more general, and extends to time–dependent simulation regimes of a larger class $\mathcal{C}_\varrho$, defined as follows.

Definitions (i) We say that a family of probability measures $\{\mathcal{Q}^{n,\epsilon}(x, E \times \Gamma) : E \in \mathcal{S}, \Gamma \in \mathcal{R}^d\}$ belongs to a class $\mathcal{C}$ if $\mathcal{Q}^{n,\epsilon} = \mathcal{Q}^{(n\epsilon)}$ for all n and ϵ, for some family $\{\mathcal{Q}^{(t)}(x, E \times \Gamma) : E \in \mathcal{S}, \Gamma \in \mathcal{R}^d\}$, and

$$\mathcal{P}(x, \cdot) \ll \mathcal{Q}^{(t)}(x, \cdot), \text{ all } x \in \mathcal{S} \text{ and } t \geq 0.$$

(ii) We say that a family $\{\mathcal{Q}^{n,\epsilon}(x, E \times \Gamma) : E \in \mathcal{S}, \Gamma \in \mathcal{R}^d\}$ belongs to a class $\mathcal{C}_\varrho$ if it belongs to $\mathcal{C}$, and

$$\mathcal{Q}^{(t)} = \mathcal{Q}^{(\varrho - \Delta)}, \text{ all } t \geq \varrho - \Delta,$$

where ϱ is the constant given in Lemma 3.2 (iv) and Δ is some positive constant.

The significance of the constant ϱ is made clear in Theorem 2 of Collamore (1998), where it is shown that asymptotically $\epsilon T^\epsilon(A) \to \varrho$ in probability, conditioned on $\{T^\epsilon(A) < \infty\}$, i.e. ϱ/ϵ is the "most likely" first passage time of the process $\{(X_n, S_n) : n = 0, 1, \dots\}$ into the set A/ϵ.

We note that the scaling of the form $\mathcal{Q}^{n,\epsilon} = \mathcal{Q}^{(n\epsilon)}$ coincides with the standard large deviations scaling appearing in Donsker and Varadhan (1975, 1976, 1983), Freidlin and Wentzell (1984), and essentially all subsequent work; it appeared in the context of the present problem in Collamore (1998).

For notational convenience, we will from now on write "$\mathcal{Q} \in \mathcal{C}$" to mean that the family $\{\mathcal{Q}^{n,\epsilon}(x, E \times \Gamma) : E \in \mathcal{S}, \Gamma \in \mathcal{R}^d\}$ belongs to the class $\mathcal{C}$, and likewise for members of $\mathcal{C}_\varrho$.

Definition Let $\mathcal{Q} \in \mathcal{C}$. Then we say that t_0 is a *continuity point* of $\mathcal{Q}$ if for any $\Delta > 0$ there exists a positive constant γ such that for all $|t - t_0| \leq \gamma$,

$$\mathcal{Q}^{(t)}(x, \cdot) \ll \mathcal{Q}^{(t_0)}(x, \cdot), \text{ all } x \in \mathbb{S},$$

and for any $n \in \mathbb{Z}_+$, $\epsilon > 0$ with $n\epsilon = t$:

$$\text{(i)} \qquad \mathbf{E}_{\pi^*}\left[\left(\frac{d\mathcal{Q}^{(n\epsilon)}}{d\mathcal{Q}^{(t_0)}}(X_n^*, X_{n+1}^* \times \xi_{n+1}^*)\right)\right] \le \Delta; \tag{3.20}$$

$$\text{(ii)} \qquad \mathbf{E}_x\left[\left(\frac{d\mathcal{Q}^{(n\epsilon)}}{d\mathcal{Q}^{(t_0)}}(X_n^*, X_{n+1}^* \times \xi_{n+1}^*)\right)\right] \le K < \infty, \quad \text{for all } x \in \mathbb{S}; \tag{3.21}$$

where $\{(X_n^*, S_n^*) : n = 0, 1, \dots\}$ is a MA–process having the transition kernel $\mathcal{Q}^*$ in (3.17) and π^* is the stationary measure of $\{X_n^*\}$.

For example, if $(\mathfrak{R})$ holds and $\mathcal{Q} \in \mathcal{C}$ has the form

$$\mathcal{Q}^{(t)}(x, dy \times ds) = e^{\langle \alpha_t, s\rangle - \Lambda_{\mathcal{P}}(\alpha_t)} \frac{r_{\mathcal{P}}(y; \alpha_t)}{r_{\mathcal{P}}(x; \alpha_t)} \mathcal{P}(x, dy \times ds), \tag{3.22}$$

where $\alpha_t = f(t)$ for some continuous function $f : [0, \infty) \to \text{int}\,(\text{dom}\,\Lambda_{\mathcal{P}})$, then *all* points $t \in [0, \infty)$ are continuity points of $\mathcal{Q}$.

Definition We say that $A \subset \mathbb{R}^d$ is a *semi–cone* if $v \in \partial A \Longrightarrow \{\zeta v : \zeta > 1\} \subset \text{int}\, A$, that is, the ray generated by any point on its relative boundary is an interior ray of A.

Theorem 3.4 *Let A be a convex open semi–cone intersecting* int $\mathfrak{S}$. *Assume that* dom $\Lambda_{\mathcal{P}}$ *is open and (H2) and* $(\mathfrak{R})$ *are satisfied. Let x_0 denote the initial state of $\{X_n\}_{n\ge 0}$, and let $\mathcal{Q}^* \in \mathcal{C}_0$ be the kernel defined in* (3.17). *If simulation is performed using a family of measures $\mathcal{Q} \in \mathcal{C}$, then for φ–a.e. x_0,*

$$\liminf_{\epsilon \to 0} \epsilon \log \mathbf{E}\big(\mathcal{E}^2_{\mathcal{Q},\epsilon}\big) \ge \lim_{\epsilon \to 0} \epsilon \log \mathbf{E}\big(\mathcal{E}^2_{\mathcal{Q}^*,\epsilon}\big). \tag{3.23}$$

Moreover, if we do not have $\mathcal{Q}^{(t_0)} = \mathcal{Q}^*$, for $\varphi \times \mathcal{P}$ a.e. (x, y, s), at all continuity points of $\mathcal{Q}$ in $[0, \varrho]$, then there is strict *inequality in* (3.23). *Thus, $\mathcal{Q}^*$ is the essentially unique element of $\mathcal{C}_\varrho$ which minimizes $\mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon})$ as $\epsilon \to 0$.*

If $\mathcal{Q} \ne \mathcal{Q}^*$ at a continuity point t_0 which is outside of $[0, \varrho]$, then we do not necessarily obtain strict inequality in (3.23); thus, the logarithmic–level optimality of $\mathcal{Q}^*$ in Theorem 3.4 cannot be extended from $\mathcal{C}_\varrho$ to $\mathcal{C}$.

Remark 3.4.1 In fact what needs to be minimized in the above discussion is the number of *random variables* that need to be generated, that is,

$$\epsilon \log \left[\text{Var}\big(\mathcal{E}^2_{\mathcal{Q},\epsilon}\big)\mathbf{E}_{\mathcal{Q}}\big(T^\epsilon(A)\big)\right] \text{ as } \epsilon \to 0 \tag{3.24}$$

[cf. Siegmund (1976), p. 676, or Collamore (1996b), Lemma 5.2]. But if A is a semi–cone, then

$$\lim_{\epsilon \to 0} \epsilon \log \mathbf{E}_{\mathcal{Q}}\big(T^\epsilon(A)\big) = 0 \tag{3.25}$$

[Collamore (1996b), Lemma 5.3]. Thus simulation under $\mathcal{Q}^*$ is efficient and optimal.

If A is not a semi–cone, then the situation is more complicated; in particular, we need not have (3.25) in this case. It may then be preferable to simulate with $\mathcal{Q} \in \mathcal{C}$, where $\mathcal{Q}^{(t)} = \mathcal{Q}^*$ for $t \in [0, \varrho]$, but $\mathcal{Q}^{(t)} \neq \mathcal{Q}^*$ for $t \geq t_0$, some $t_0 > \varrho$. By a judicious choice of $\mathcal{Q}$, we may often obtain both (3.25) and (3.19).

3.3 General sets

Finally, suppose that A is an arbitrary open subset of $\mathbb{R}^d$. In this case, we will show that we can partition A into subsets $A_1, \ldots, A_l$ and apply the techniques of Theorem 3.1 to efficiently simulate for $\mathbf{P}\left\{T^\epsilon(A),\ S_{T^\epsilon(A)} \in A_i\right\}$, for $i = 1, \ldots, l$.

For any $\alpha \in \operatorname{dom} \Lambda_{\mathcal{P}}$, let

$$\mathcal{Q}_\alpha(x, dy \times ds) = e^{\langle \alpha, s\rangle - \Lambda_{\mathcal{P}}(\alpha)} \frac{r_{\mathcal{P}}(y;\alpha)}{r_{\mathcal{P}}(x;\alpha)} \mathcal{P}(x, dy \times ds).$$

Let $B \subset A$, and let $\mathfrak{P}_k^\epsilon$ denote the paths which first hit A/ϵ at time k, as defined formally in (3.2). If simulation is performed using a kernel $\mathcal{Q} \in \mathcal{C}_0$, then

$$\begin{aligned} \mathbf{P}\left\{T^\epsilon(A) < \infty,\ \epsilon S_{T^\epsilon(A)} \in B\right\} &= \sum_k \int_{\mathfrak{P}_k^\epsilon} \left(\prod_{i=1}^k \frac{d\mathcal{P}}{d\mathcal{Q}}(x_{i-1}, x_i \times s_i) \right) \mathbf{1}_B\left(\epsilon S_k\right) \\ &\qquad \cdot \mathcal{Q}(x_0, dx_1 \times ds_1) \cdots \mathcal{Q}(x_{k-1}, dx_k \times ds_k). \quad (3.26) \end{aligned}$$

Hence

$$\mathcal{E}_{\mathcal{Q},\epsilon}(B) \stackrel{\text{def}}{=} \prod_{n=1}^{T^\epsilon(A)} \left(\frac{d\mathcal{P}}{d\mathcal{Q}}(\tilde{X}_{n-1}, \tilde{X}_n \times \tilde{\xi}_n) \right) \mathbf{1}_B\left(\epsilon S_{T^\epsilon(A)}\right) \qquad (3.27)$$

is an unbiased estimator for $\mathbf{P}\left\{T^\epsilon(A) < \infty,\ \epsilon \tilde{S}_{T^\epsilon(A)} \in B\right\}$, where $\{\tilde{X}_n, \tilde{S}_n : n = 0, 1, \ldots\}$ is a MA–process having transition kernel $\mathcal{Q}$, and we would like to minimize $\mathbf{E}\big(\mathcal{E}^2_{\mathcal{Q},\epsilon}(B)\big)$ as $\epsilon \to 0$.

Proposition 3.5 *Let $\Delta > 0$ and $A \subset \mathbb{R}^d$, and suppose that* $\operatorname{dom} \Lambda_{\mathcal{P}}$ *is open and (H1), (H2), and the lower bound of* $(\mathfrak{R})$ *are satisfied. Then:*

(i) *For some finite subset $\{\alpha_1, \ldots, \alpha_l\}$ of $\partial(\mathcal{L}_0 \Lambda_{\mathcal{P}})$, the collection $\left\{\mathcal{H}(\alpha_i, a - \Delta) : i = 1, \ldots, l\right\}$ is an open cover for A.*

(ii) *Let $A_1 = A \cap \mathcal{H}(\alpha_1, a - \Delta)$, $A_2 = \big(A \cap \mathcal{H}(\alpha_2, a - \Delta)\big) - A_1$, and so on for $A_3, \ldots, A_l$. Then $\{A_1, \ldots, A_l\}$ is a partition of A, and for each i,*

$$\mathbf{E}\big(\mathcal{E}^2_{\mathcal{Q}_{\alpha_i},\epsilon}(A_i)\big) \leq C \exp\left\{-2\big(\inf_{y \in A} \tilde{I}(y) - \Delta\big)/\epsilon\right\} \qquad (3.28)$$

for a certain positive constant C. In the event that A is a finite union of disjoint convex sets $\{A'_1, \ldots, A'_k\}$, then we may instead take $A_i = A'_i$, and then (3.28) holds with $\Delta = 0$.

4 Proofs

Some further notation from convex analysis:

For any convex function f, let f^*, $\mathrm{cl}\, f$, $f0^+(\cdot)$, 0^+f, $\mathrm{dom}\, f$, and $\partial f(\cdot)$ denote the convex conjugate of f, the closure of f, the recession function of f, the recession cone of f, the domain of f, and the subgradient set of f, respectively.

For any convex set C, let C°, 0^+C, $\mathrm{aff}C$, $\mathrm{ri}C$, and ∂C denote the polar of C, the recession cone of C, the affine hull of C, the relative interior of C, and the relative boundary of C, respectively. [For definitions, see Rockafellar (1970).]

Also, we adopt the same terminology that was already introduced at the beginning of Sections 2 and 3.

4.1 Proof of Theorem 3.1: Upper Bound

The proof of the upper bound is based on the following convexity lemma, which shows that the separation property described in Lemma 3.2 (i) is in fact quite general.

Lemma 4.1 *Let $A \subset \mathbb{R}^d$ be a convex open set, and let Λ be a closed convex function. Let $I(v) = \sup\{\langle \alpha, v\rangle : \alpha \in \mathcal{L}_0\Lambda\}$, and let $a = \inf_{v\in A} I(v)$. Assume $0 < a < \infty$. Then there exists $\theta \in \mathcal{L}_0\Lambda$ such that*

$$A \subset \{v : \langle \theta, v\rangle \geq a\} \quad \textit{and} \quad \mathcal{L}_a I \subset \{v : \langle \theta, v\rangle \leq a\}. \tag{4.1}$$

Proof Since I is a positively homogeneous convex function and $a > 0$, the sets A and $\mathcal{L}_a I$ are convex and have no common points in their relative interiors. Hence there exists a separating hyperplane [Rockafellar (1970), Theorem 11.3], i.e., for some $\beta \in \mathbb{R}^d - \{0\}$,

$$A \subset \{v : \langle \beta, v\rangle \geq b\} \quad \text{and} \quad \mathcal{L}_a I \subset \{v : \langle \beta, v\rangle \leq b\}, \tag{4.2}$$

where $b \in \mathbb{R}$; in fact, $b \geq 0$ because the definition of I implies $I(0) = 0$, so $0 \in \mathcal{L}_a I$.

Let $c > 0$, and define $f = I - c$. Then $\mathcal{L}_0 f = \mathcal{L}_c I$, and $f^* = \mathbf{1}_{\mathcal{L}_0\Lambda} + c$ [Rockafellar (1970), Theorem 12.2]. An application of Theorems 13.5 and 9.7 of Rockafellar (1970) then gives

$$\mathbf{1}^*_{\mathcal{L}_c I}(\beta) = \inf\{(f^*\gamma)(\beta) \mid \gamma > 0 \text{ or } \gamma = 0^+\}, \tag{4.3}$$

where $(f^*\gamma)(\cdot) = \gamma f^*(\cdot/\gamma)$, for all $\gamma > 0$, and f^*0^+ is the recession function of f^*.

Note that $f^*(\cdot) \in \{c, \infty\} \implies f^*0^+(\cdot) \in \{0, \infty\}$ [Rockafellar (1970), Theorem 8.5]. But we cannot have $f^*0^+(\beta) = 0$: Otherwise, (4.3) would imply $\mathbf{1}^*_{\mathcal{L}_c I}(\beta) = 0$ for *all* $c > 0$. Then $\mathcal{L}_c I \subset \{v : \langle \beta, v\rangle \leq 0\}$, all $c > 0$. Since A is open, we would then obtain $A \cap \mathrm{dom}\, I = \emptyset$, by (4.2), which is contrary to hypothesis. Therefore $f^*0^+(\beta) = \infty$; thus the point $\gamma = 0^+$ can be removed from the infimum in (4.3).

Since $f^*(\cdot) \in \{c, \infty\}$, we now conclude from (4.3) that

$$\mathbf{1}^*_{\mathcal{L}_c I}(\beta) = \gamma c, \text{ where } \gamma = \inf\{\tilde{\gamma} : \frac{\beta}{\tilde{\gamma}} \in \mathcal{L}_0\Lambda\}. \tag{4.4}$$

Setting $c = a$ yields $\mathcal{L}_a I \subset \{v : \langle \beta, v \rangle \le \gamma a\}$. Hence the constant b in (4.2) is $\ge \gamma a$. Next suppose $b \ge \gamma a'$, where $a' > a$. Then $\mathcal{L}_{a'} I \subset \{v : \langle \beta, v \rangle \le b\}$. By (4.2) it follows that $\inf_{v \in A} I(v) \ge a'$, since A is open, and this contradicts the definition of a. We conclude $b = \gamma a$. Observe also that $\gamma > 0$, because otherwise (4.2) and (4.4) would imply $A \cap \mathrm{dom} I = \emptyset$, contrary to hypothesis. The required result now follows from (4.2) and the previous paragraph by setting $\theta = \beta / \gamma$. □

Proof of Theorem 3.1. Upper Bound. If $\mathcal{Q} = \mathcal{P}$, then the result follows trivially from (1.4); we assume from now on that this is not true.

For any $\Delta > 0$, let

$$\mathcal{K}_{\mathcal{Q}}^{\Delta}(x, dy \times ds) = \mathcal{K}_{\mathcal{Q}}(x, dy \times ds) + \Delta \eta_{x_0}(dy) \eta_{\mathcal{O}}(ds), \tag{4.5}$$

where η_{x_0} denotes a measure on $\mathbb{S}$ having point mass at x_0, and $\eta_{\mathcal{O}}$ denotes a measure on $\mathbb{R}^d$ having point mass at the origin. For shorthand notation, let $(\lambda_\Delta(\alpha))^{-1}$ denote the convergence parameter of $\mathcal{K}_{\mathcal{Q}}^{\Delta}(\alpha)$ and $\Lambda_\Delta(\alpha) = \log \lambda_\Delta(\alpha)$.

First we show

$$\limsup_{\Delta \to 0} \Lambda_\Delta(\alpha) \le \bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}}}(\alpha), \text{ for all } \alpha \in \mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}}}. \tag{4.6}$$

To establish this fact, note

$$\begin{aligned} \left(\hat{\mathcal{K}}_{\mathcal{Q}}^{\Delta}\right)^k (x_0, \mathbb{S}; \alpha) \quad = \quad & \sum_{(i_1, \ldots, i_k) \in \mathfrak{I}} \int_{(\mathbb{S} \times \mathbb{R}^d)^k} e^{<\alpha, s>} \mathcal{J}^{(i_1)}(x_0, dx_1 \times \cdot) * \\ & \cdots * \mathcal{J}^{(i_k)}(dx_{k-1}, dx_k \times \cdot), \end{aligned} \tag{4.7}$$

where $\mathcal{J}^{(0)} = \mathcal{K}_{\mathcal{Q}}$ and $\mathcal{J}^{(1)} = \Delta \eta_{x_0} \eta_{\mathcal{O}}$, and $\mathfrak{I}$ consists of all elements of the form $(i_1, \ldots, i_k)$ where either $i_j = 0$ or $i_j = 1$. Fix $\alpha \in \mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}}}$ and $N \in \mathbb{Z}_+$, and let

$$\log b_N = \left[\Lambda_{\mathcal{K}_{\mathcal{Q}}}^{(1)}(\alpha) - \Lambda_{\mathcal{K}_{\mathcal{Q}}}^{(N)}(\alpha)\right] < \infty \quad \text{(by (H1))}.$$

Observe that any product $\mathcal{J}^{(1)} * \cdots * \mathcal{J}^{(k)}$ which has n "$\Delta \eta_{x_0} \eta_{\mathcal{O}}$" terms contains at most $n+1$ products of the form $\Delta \eta_{x_0} \eta_{\mathcal{O}} * \mathcal{K}_{\mathcal{Q}}^{(i)} * \cdots * \mathcal{K}_{\mathcal{Q}}^{(i+j)} * \Delta \eta_{x_0} \eta_{\mathcal{O}}$ where $j < N$. It follows that for a product of this form,

$$\begin{aligned} \int_{(\mathbb{S} \times \mathbb{R}^d)^k} & e^{<\alpha, s>} \mathcal{J}^{(i_1)}(x_0, dx_1 \times \cdot) * \cdots * \mathcal{J}^{(i_k)}(dx_{k-1}, dx_k \times \cdot) \\ & \le b_N^{n+1} \Delta^n \left(\lambda_{\mathcal{K}_{\mathcal{Q}}}^{(N)}(\alpha)\right)^{k-n}. \end{aligned} \tag{4.8}$$

Summing all terms in (4.7) gives $\lambda_\Delta(\alpha) \le \lambda_{\mathcal{K}_{\mathcal{Q}}}^{(N)}(\alpha) + b_N \Delta$, by the definition of the convergence parameter. Letting $\Delta \to 0$ and then $N \to \infty$ establishes (4.6).

We now apply Lemma 4.1 with $\Lambda(\alpha) = \Lambda_{\mathcal{K}_{\mathcal{Q}}}(\alpha) - c$ where $c < 0$. Note that the hypotheses of the lemma are satisfied. First, $\Lambda_{\mathcal{K}_{\mathcal{Q}}}$ is convex [Lemma 4.2 (ii) below]. Secondly, under our assumptions, $\mathrm{cl}\, A \cap \mathrm{cone}_\delta \left(\mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}}}\right)^\perp = \emptyset$. Now $\mathcal{Q} \ne \mathcal{P} \implies \bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}}}(0) > 0$ [as in Lemma 2.3], and then

$$\mathrm{cone}_{\,\delta} \left(\mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}}}\right)^\perp \supset \left(\mathcal{L}_c \bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}}}\right)^\perp = \mathcal{L}_0 \bar{I}_{\mathcal{K}_{\mathcal{Q}}}^{(c)}$$

for sufficiently large $c < 0$ [the last step was obtained from the definition of $\bar{I}^{(c)}_{\mathcal{K}_\mathcal{Q}}$]. Consequently $\inf\{\bar{I}^{(c)}_{\mathcal{K}_\mathcal{Q}}(v) : v \in A\} > 0$. Finally, if $c < 0$ is sufficiently large, then by Lemma 4.3 (i) below,

$$\operatorname{ri} \mathfrak{D}^{(c)}_{\mathcal{K}_\mathcal{Q}} \supset \operatorname{ri} \mathfrak{D}_\mathcal{P} = \operatorname{ri} \mathfrak{S}.$$

[The last step follows from Theorem 13.5 of Rockafellar (1970), as noted in the "Remark" following Theorem 2.1 of Collamore (1996a).] We conclude $0 < \inf\{I^{(c)}_{\mathcal{K}_\mathcal{Q}}(v) : v \in A\} < \infty$, for all negative $c \geq$ some c_0.

Let θ be the element obtained in Lemma 4.1 when $\Lambda(\alpha) = \Lambda_{\mathcal{K}_\mathcal{Q}}(\alpha) - c$ and $c \in [c_0, 0)$. By (4.6), there exists a $\Delta > 0$ such that $\Lambda^{\Delta}_{\mathcal{K}_\mathcal{Q}}(\theta) \leq 0$. Moreover, there exists a $\big(\lambda_\Delta(\theta)\big)^{-1}$–(sub)invariant function, $r_\Delta(\theta)$, for the resulting kernel $\hat{\mathcal{K}}^{\Delta}_{\mathcal{Q}}(\theta)$ [Nummelin (1984), Proposition 5.2 and Theorem 5.1, or Section 2.2 above]. Define

$$\mathcal{R}_\theta(x, dy \times ds) = \frac{e^{\langle \theta, s\rangle} r_\Delta(dy;\theta)}{r_\Delta(x;\theta)} \mathcal{K}^{\Delta}_{\mathcal{Q}}(x, dy \times ds). \tag{4.9}$$

Since $r_\Delta(\theta)$ is $\big(\lambda_\Delta(\theta)\big)^{-1}$–(sub)invariant and $\Lambda_\Delta(\theta) \leq 0$, $\mathcal{R}_\theta$ is itself a Markov additive (sub)probability kernel.

Let $\mathfrak{P}^\epsilon_k$ denote the paths which first hit A/ϵ at time $T^\epsilon(A) = k$. Then by (3.4) and the definition of $\mathcal{K}^{\Delta}_{\mathcal{Q}}$,

$$\begin{aligned}
\mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \quad &\leq \quad \sum_{k=1}^{\infty} \int_{\mathfrak{P}^\epsilon_k} \mathcal{K}^{\Delta}_{\mathcal{Q}}(x_0, dx_1 \times ds_1) \cdots \mathcal{K}^{\Delta}_{\mathcal{Q}}(x_{k-1}, dx_k \times ds_k) \\
&\leq \quad \sum_{k=1}^{\infty} \int_{\mathfrak{P}^\epsilon_k} \frac{r_\Delta(x_0;\theta) e^{-\langle \theta, s_1 + \cdots + s_k\rangle}}{r_\Delta(x_{T^\epsilon(A)};\theta)} \\
&\qquad \cdot \mathcal{R}_\theta(x_0, dx_1 \times ds_1) \cdots \mathcal{R}_\theta(x_{k-1}, dx_k \times ds_k).
\end{aligned} \tag{4.10}$$

Note that $r_\Delta(\cdot;\theta)$ is uniformly positive [Remark 2.2.2, since (4.5) yields a minorization $\Delta\nu_{x_0}\nu_0$]. Also, $r_\Delta(\cdot;\theta) < \infty$ φ a.e. [Nummelin (1984), Proposition 5.1]. Thus the ratio $\big(r_\Delta(x_0;\theta)/r_\Delta(X_{T^\epsilon(A)};\theta)\big)$ in (4.10) is deterministically bounded.

Next, observe by Lemma 4.1 that $\langle \theta, \tilde{S}_{T^\epsilon(A)}\rangle \geq \inf\{\bar{I}^{(c)}_{\mathcal{K}_\mathcal{Q}}(v) : v \in A\}/\epsilon$; and since $\bar{I}^{(c)}_{\mathcal{K}_\mathcal{Q}}$ is positively homogeneous,

$$\inf\{\bar{I}^{(c)}_{\mathcal{K}_\mathcal{Q}}(v) : v \in A\} = \inf\{\bar{I}^{(c)}_{\mathcal{K}_\mathcal{Q}}(v) : v \in \mathfrak{A}\},$$

provided that the infimum on the left is ≥ 0. Therefore, the integrand in (4.10) is $\leq$ const. $\cdot \exp\{-\inf_{v\in\mathfrak{A}} \bar{I}^{(c)}_{\mathcal{K}_\mathcal{Q}}(v)/\epsilon\}$. Since $\mathcal{R}_\theta$ is a subprobability kernel, we obtain by letting $\epsilon \searrow 0$ in (4.10) that

$$\limsup_{\epsilon \to 0} \epsilon \log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \leq -\inf_{v \in \mathfrak{A}} \bar{I}^{(c)}_{\mathcal{K}_\mathcal{Q}}(v). \tag{4.11}$$

It remains to show

$$\inf\{\bar{I}^{(c)}_{\mathcal{K}_\mathcal{Q}}(v) : v \in \mathfrak{A}\} \nearrow \inf\{\bar{I}_{\mathcal{K}_\mathcal{Q}}(v) : v \in \mathfrak{A}\} \quad \text{as } c \to 0. \tag{4.12}$$

To this end, note that the assumption $\mathrm{cl}\, A \cap \mathrm{cone}_\delta\big(\mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_\mathcal{Q}}\big)^\perp = \emptyset \Longrightarrow$ the level sets of $\bar{I}^{(c)}_{\mathcal{K}_\mathcal{Q}}$ are compact on the restricted set $\big(\mathrm{cone}_\delta(\mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_\mathcal{Q}})^\perp\big)^c$ when $c < 0$ is sufficiently large [cf. Collamore (1996b), Lemma 3.1]. Let a denote the limit on the left–hand side of (4.12). The sets $\mathrm{cl}\, A \cap \mathcal{L}_a \bar{I}^{(c)}_{\mathcal{K}_\mathcal{Q}}$ are then compact and monotonically decreasing to $\mathrm{cl}\, A \cap \mathcal{L}_a \bar{I}_{\mathcal{K}_\mathcal{Q}}$ as $c \nearrow 0$, implying (4.12). □

We remark that if the lower bound of $(\mathfrak{R})$ holds, then the above proof can be simplified. In that case, we may apply the measure transformation (4.9) directly to $\mathcal{K}_\mathcal{Q}$, and Lemma 4.1 directly with $c = 0$. This approach leads to some improvement in hypothesis, namely the assumption "$A \cap \mathrm{cone}_\delta\big(\mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_\mathcal{Q}}\big)^\perp = \emptyset$" may then be weakened to "$\inf\{I_{\mathcal{K}_\mathcal{Q}}(v) : v \in \mathfrak{A}\} > 0$."

4.2 Proof of Theorem 3.1: Lower Bound

We begin by introducing a splitting and truncation of $\mathcal{K}_\mathcal{Q}$, as follows.

Let $h \circledast \nu$ be the minorization in $(\mathfrak{M}')$, and note under $(\mathfrak{M}')$ that either $h(x, ds)$ or $\nu(E, ds)$ is independent of s. Thus $(h \circledast \nu)^2 = g \circledast \mu$, where $g = h^2$ and $\mu = \nu^2$. Hence by Lemma 2.3 (i), $(g \circledast \nu) \le \mathcal{K}_\mathcal{Q}$. This implies the minorization

$$\hat{g}(\alpha) \otimes \hat{\mu}(\alpha) \le \hat{\mathcal{K}}_\mathcal{Q}(\alpha), \quad \text{for all } \alpha. \tag{4.13}$$

Define

$$\bar{\mathcal{K}}_\mathcal{Q}(x, dy \times ds) = \frac{\mathcal{K}_\mathcal{Q}(x, dy \times ds) - (g \circledast \mu)(x, dy \times ds)}{1 - g(x, \mathbb{R}^d)} \quad (\ge 0),$$

and observe under this definition that

$$\mathcal{K}_\mathcal{Q}(x, dy \times ds) = (g \circledast \mu)(x, dy \times ds) + \big(1 - g(x, \mathbb{R}^d)\big) \bar{\mathcal{K}}_\mathcal{Q}(x, dy \times ds). \tag{4.14}$$

Enlarge $(\mathbb{S}, \mathcal{S})$ to $(\tilde{\mathbb{S}}, \tilde{\mathcal{S}})$, where $\tilde{\mathbb{S}} = \mathbb{S} \times \{0, 1, 2, \dots\}$ and $\tilde{\mathcal{S}}$ is the natural extension of $\mathcal{S}$ to $\tilde{\mathbb{S}}$; and for $M \in \mathbb{Z}_+$, define truncated versions g_M, h_M, $\bar{\mathcal{K}}^M_\mathcal{Q}$, $\mathcal{K}^M_\mathcal{Q}$ by:

$$\begin{aligned}
g_M\big((x,i), ds\big) &= \Big(\frac{M}{M+1}\Big)\, \mathbf{1}_{[0,M)}(i) \mathbf{1}_{[-M,M]^d}(s) g(x, ds), \\
\mu_M\big((dy,j), ds\big) &= \mathbf{1}_0(j) \delta_{[-M,M]^d}(s) \mu\big(dy \times ds\big), \\
\bar{\mathcal{K}}^M_\mathcal{Q}\big((x,i),(dy,j) \times ds\big) &= \Big[\mathbf{1}_{i+1}(j)\, \mathbf{1}_{(0,M]}(j)\, \mathbf{1}_{[-M,M]^d}(s)\, \bar{\mathcal{K}}_\mathcal{Q}(x, dy \times ds)\Big] \wedge M, \\
\mathcal{K}^M_\mathcal{Q}\big((x,i),(dy,j) \times ds\big) &= \frac{M}{M+1}\Big[g_M \circledast \mu_M\big((x,i),(dy,j) \times ds\big) + \\
&\qquad \big(1 - g(x, \mathbb{R}^d)\big) \bar{\mathcal{K}}^M_\mathcal{Q}\big((x,i),(dy,j) \times ds\big)\Big].
\end{aligned}$$

Note that $\mathcal{K}^M_\mathcal{Q}$ is strictly increasing,

$$\mathcal{K}^M_\mathcal{Q}\big((x \times \mathbb{N}), (E \times \mathbb{N}) \times \Gamma\big) \le \mathcal{K}_\mathcal{Q}(x, E \times \Gamma), \quad \text{all } E \in \mathcal{S},\ \Gamma \in \mathcal{R}^d \tag{4.15}$$

(where $\mathbb{N}$ denotes the set of natural numbers), and

$$\mathcal{K}^M_\mathcal{Q}\big((x \times \mathbb{N}), (E \times \mathbb{N}) \times \Gamma\big) \nearrow \mathcal{K}_\mathcal{Q}(x, E \times \Gamma) \quad \text{as } M \to \infty. \tag{4.16}$$

It can be easily shown that $\mathcal{K}_Q^M$ is irreducible with respect to a maximal irreducibility measure $\varphi_M \nearrow \varphi$ as $M \to \infty$.

The kernel $\mathcal{K}_Q^M$ has a minorization, namely $g_M \circledast \mu_M \leq \mathcal{K}_Q^M$, which implies

$$\hat{g}_M(\alpha) \otimes \hat{\mu}_M(\alpha) \leq \hat{\mathcal{K}}_Q^M(\alpha), \quad \text{for all } \alpha. \tag{4.17}$$

For shorthand notation, let $(\lambda_M(\alpha))^{-1}$ denote the convergence parameter of $\hat{\mathcal{K}}_Q^M(\alpha)$, $\Lambda_M(\alpha) = \log \lambda_M(\alpha)$, and let $\mathfrak{D}_M = \mathfrak{D}_{\mathcal{K}_Q^M}$ and $I_M^{(c)} = I_{\mathcal{K}_Q^M}^{(c)}$.

Our main reason for introducing the above truncation is because the transformed kernel $\hat{\mathcal{K}}_Q^M(\alpha)$ has eigenvectors which are *bounded*, and the following regularity properties also hold.

Lemma 4.2 *Let the kernels $\mathcal{K}_Q^M$, $\mathcal{K}_Q$ and functions λ_M, Λ_M, and $\Lambda_{\mathcal{K}_Q}$ be defined as above. Then:*

(i) *Λ_M is convex and analytic, and $\Lambda_M(\alpha) \nearrow \Lambda_{\mathcal{K}_Q}(\alpha)$ as $M \to \infty$, for all α.*

(ii) *$\Lambda_{\mathcal{K}_Q}$ is convex and lower semicontinuous.*

(iii) *For any α, there exists a $(\lambda_M(\alpha))^{-1}$–invariant function, $r_M(\cdot;\alpha)$, for the kernel $\hat{\mathcal{K}}_Q^M(\alpha)$. Moreover, the function $r_M(\cdot;\alpha)$ is positive and bounded.*

Proof (i) Following Iscoe, Ney and Nummelin (1985), Lemma 3.4, introduce the generating function

$$\begin{aligned}\psi_M(\alpha,\zeta) &= \sum_{n=1}^{\infty} \int_{s\in\mathbb{R}^d,\, x\in\tilde{\mathbb{S}}} e^{\langle\alpha,s\rangle-\zeta n} \Big(\mu_M \circledast \big(\mathcal{K}_Q^M - g_M \circledast \mu_M\big)^{n-1} \circledast g_M\Big)(x, \tilde{\mathbb{S}} \times ds) \\ &= \sum_{n=1}^{\infty} e^{-\zeta n}\, \hat{\mu}_M(\alpha)\, \Big(\hat{\mathcal{K}}_Q^M(\alpha) - \hat{g}_M(\alpha) \otimes \hat{\mu}_M(\alpha)\Big)^{n-1}\, \hat{g}_M(\alpha).\end{aligned} \tag{4.18}$$

Then $\Lambda_M(\alpha) = \inf\{\zeta : \psi_M(\alpha,\zeta) < 1\}$ [Nummelin (1984), Proposition 4.7 (i)]. Note by the construction of $\mathcal{K}_Q^M$ that the sum on the right of (4.18) is actually finite; consequently,

$$\psi_M\big(\alpha, \Lambda_M(\alpha)\big) = 1. \tag{4.19}$$

The convexity of Λ_M follows from (4.19) and the convexity of ψ_M. Since ψ_M is analytic on $\mathbb{R}^{d+1}$, the analyticity of Λ_M follows from (4.19) and the implicit function theorem. The convergence $\Lambda_M \nearrow \Lambda_{\mathcal{K}_Q}$ is obtained as in Lemma 3.3 (i) of Ney and Nummelin (1987b).

(ii) $\Lambda_{\mathcal{K}_Q}$ is convex because (by (i)) it is a limit of convex functions, and lower semicontinous since the analytic functions $\Lambda_M \nearrow \Lambda$ as $M \to \infty$.

(iii) Since (4.19) holds, $\hat{\mathcal{K}}_Q^M$ is $(\lambda_M(\alpha))^{-1}$–recurrent [Nummelin (1984), Proposition 4.3]. Hence a $(\lambda_M(\alpha))^{-1}$–invariant function exists and is given by

$$r_M(\alpha) = \sum_{n=0}^{\infty} e^{-\zeta(n+1)}\, \Big(\hat{\mathcal{K}}_Q^M(\alpha) - \hat{g}_M(\alpha) \otimes \hat{\mu}_M(\alpha)\Big)^n \hat{g}_M(\alpha) \tag{4.20}$$

[Nummelin (1984), Theorem 5.1]. By the construction of $\mathcal{K}_Q^M$, the sum and individual terms on the right are finite, hence $r_M(\cdot;\alpha)$ is bounded. Finally, positivity of $r_M(\cdot;\alpha)$ is obtained from Nummelin (1984), Proposition 5.1. □

Lemma 4.3 *Let the kernels $\mathcal{P}$, $\mathcal{K}_{\mathcal{Q}}^{M}$ and $\mathcal{K}_{\mathcal{Q}}$ be defined as above. Then:*

(i) *For any $b \geq \inf_\alpha \Lambda_{\mathcal{P}}(\alpha)$ and $c \geq \inf_\alpha \Lambda_{\mathcal{K}_{\mathcal{Q}}}(\alpha)$, $\mathfrak{D}_{\mathcal{P}}^{(b)} \subset \mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}}^{(c)}$.*

(ii) *For any $c_M \geq \inf_\alpha \Lambda_M(\alpha)$ and $c \geq \inf_\alpha \Lambda_{\mathcal{K}_{\mathcal{Q}}}(\alpha)$, $\mathfrak{D}_M^{(c)} \nearrow \mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}}^{(c)}$ as $M \to \infty$.*

(iii) *For any $c \in \mathbb{R}$ and $v \in \operatorname{ri} \mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}}^{(c)}$, $I_M^{(c)}(v) \searrow I_{\mathcal{K}_{\mathcal{Q}}}^{(c)}(v)$ as $M \to \infty$.*

(iv) *For any $c \in \mathbb{R}$ and $v \in \operatorname{ri} \mathfrak{D}_M^{(c)}$, the supremum in the definition of $I_M^{(c)}$ is* achieved *at a point $\theta \in \mathcal{L}_c\Lambda_M \cap \operatorname{aff} \mathfrak{D}_M$. Moreover, if $c > \inf_\alpha \Lambda_M(\alpha)$, then for some positive constant $\rho = \rho(v, c)$, we have $\rho \nabla \Lambda_M(\theta) = v$.*

Proof (i) Note

$$\big(\mathfrak{D}_{\mathcal{P}}^{(b)}\big)^\circ = 0^+\big(\mathcal{L}_b\Lambda_{\mathcal{P}}\big) \text{ and } \big(\mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}}^{(c)}\big)^\circ = 0^+\big(\mathcal{L}_c\Lambda_{\mathcal{K}_{\mathcal{Q}}}\big). \tag{4.21}$$

[Rockafellar (1970), Theorem 14.2, applied to $\mathbf{1}^*_{\mathcal{L}_b\Lambda_{\mathcal{P}}}$ and $\mathbf{1}^*_{\mathcal{L}_c\Lambda_{\mathcal{K}_{\mathcal{Q}}}}$. By Rockafellar (1970), Theorem 8.7, $0^+(\mathbf{1}_{\mathcal{L}_b\Lambda_{\mathcal{P}}})$, $0^+(\mathbf{1}_{\mathcal{L}_c\Lambda_{\mathcal{K}_{\mathcal{Q}}}})$ may be identified with $0^+(\mathcal{L}_b\Lambda_{\mathcal{P}})$, $0^+(\mathcal{L}_c\Lambda_{\mathcal{K}_{\mathcal{Q}}})$, resp.] Now set $b = c/2$. Since $\mathcal{L}_c\Lambda_{\mathcal{K}_{\mathcal{Q}}} \subset 2(\mathcal{L}_{c/2}\Lambda_{\mathcal{P}})$ [Lemma 2.3 (ii)], it follows from (4.21) that $\big(\mathfrak{D}_{\mathcal{P}}^{(c/2)}\big)^\circ \supset \big(\mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}}^{(c)}\big)^\circ$, hence $\mathfrak{D}_{\mathcal{P}}^{(c/2)} \subset \mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}}^{(c)}$. For a general $b \geq \inf_\alpha \Lambda_{\mathcal{P}}(\alpha)$, observe that $0^+\big(\mathcal{L}_b\Lambda_{\mathcal{P}}\big) = 0^+\big(\mathcal{L}_{c/2}\Lambda_{\mathcal{P}}\big)$ [Rockafellar (1970), Theorem 8.7], hence $\mathfrak{D}_{\mathcal{P}}^{(b)} = \mathfrak{D}_{\mathcal{P}}^{(c/2)}$.

(ii) The proof is analogous to (i), once it is observed that

$$\big(0^+\Lambda_{\mathcal{K}_{\mathcal{Q}}}\big) = \bigcap_M \big(0^+\Lambda_M\big) \tag{4.22}$$

[Rockafellar (1970), Corollary 8.3.3 and Theorem 8.7].

(iii) First assume $v \in \operatorname{int} \mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}}$. Let

$$\mathcal{W}_M = \Big\{\alpha \in \mathcal{L}_c\Lambda_M : \langle \alpha, v\rangle \geq I_M^{(c)}(v)\Big\}, \quad \mathcal{W} = \Big\{\alpha \in \mathcal{L}_c\Lambda_{\mathcal{K}_{\mathcal{Q}}} : \langle \alpha, v\rangle \geq I_{\mathcal{K}_{\mathcal{Q}}}^{(c)}(v)\Big\}.$$

Since $\mathcal{L}_c\Lambda_M \searrow \mathcal{L}_c\Lambda_{\mathcal{K}_{\mathcal{Q}}}$ monotonically as $M \to \infty$, $\bigcap_M \mathcal{W}_M = \mathcal{W} = \partial I_{\mathcal{K}_{\mathcal{Q}}}^{(c)}(v)$ (where the last step follows from Theorem 23.5 of Rockafellar (1970) and the definition of $I_{\mathcal{K}_{\mathcal{Q}}}^{(c)}$). Now $v \in \operatorname{int} \mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}} \Longrightarrow \partial I_{\mathcal{K}_{\mathcal{Q}}}^{(c)}(v)$ is compact [Rockafellar (1970), Theorem 23.4]. Hence the convergence $\mathcal{W}_M \searrow \mathcal{W}$ implies

$$\mathcal{W}_M \subset \{z : \|z - w\| < \Delta,\ w \in \mathcal{W}\}, \quad M \geq \text{ some } M_0(\Delta), \text{ for any } \Delta > 0.$$

Thus $I_M^{(c)}(v) \leq I_{\mathcal{K}_{\mathcal{Q}}}^{(c)}(v) + \Delta\|v\|$, all $M \geq M_0(\Delta)$. Conversely, $\mathcal{L}_0\Lambda_M \supset \mathcal{L}_0\Lambda_{\mathcal{K}_{\mathcal{Q}}} \Longrightarrow I_M^{(c)}(v) \geq I_{\mathcal{K}_{\mathcal{Q}}}^{(c)}(v)$, all M. We conclude $I_M^{(c)}(v) \searrow I_{\mathcal{K}_{\mathcal{Q}}}^{(c)}(v)$.

Next assume $v \in \operatorname{ri} \mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}}$. Then $v \in \operatorname{aff} \mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}}$, hence (by (i)) $v \in \operatorname{aff} \mathfrak{D}_M$ for sufficiently large M. Thus $\langle \alpha, v\rangle = 0$, all $\alpha \in (\operatorname{aff} \mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}})^\perp$ *and* all $\alpha \in (\operatorname{aff} \mathfrak{D}_M)^\perp$, $M \geq$ some M_0. Using this fact, we may then proceed as in the previous paragraph, replacing $I_{\mathcal{K}_{\mathcal{Q}}}^{(c)}$, $I_M^{(c)}$ with their restrictions to $\operatorname{aff} \mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}}$.

(iv) Let $\breve{v}$, $\breve{\Lambda}_M$, $\breve{I}_M$, and $\breve{\mathfrak{D}}_M$ denote the restrictions of v, Λ_M, I_M, and $\mathfrak{D}_M$ to $\operatorname{aff} \mathfrak{D}_M$. Then $\breve{v} \in \operatorname{int} \breve{\mathfrak{D}}_M$. Hence $\partial \breve{I}_M(v) \neq \emptyset$ [Rockafellar (1970), Theorem 23.4]. This implies

$$\breve{I}_M(\breve{v}) \stackrel{\text{def}}{=} \sup_{\beta \in \mathcal{L}_c\breve{\Lambda}_M} \langle \beta, \breve{v}\rangle = \langle \breve{\theta}, \breve{v}\rangle, \text{ some } \breve{\theta} \in \partial(\mathcal{L}_c\breve{\Lambda}_M) \tag{4.23}$$

[Rockafellar (1970), Theorem 23.5]. Since $\langle \alpha, v \rangle = 0$ for all $\alpha \in (\operatorname{aff} \mathfrak{D}_M)^\perp$, (4.23) also holds with I_M in place of $\breve{I}_M$, etc., and θ in place of $\breve{\theta}$, where $\theta = \breve{\theta}$ on aff $\mathfrak{D}_M$ and $\theta_M = 0$ on $(\operatorname{aff} \mathfrak{D}_M)^\perp$.

Finally, observe by (4.23) that $\breve{\theta}$ is normal to $\mathcal{L}_c \breve{\Lambda}_M$, hence θ is normal to $\mathcal{L}_c \Lambda_M$. If $c > \inf_\alpha \Lambda_M(\alpha)$, then it follows from Corollary 23.7.1 of Rockafellar (1970) that $v = \rho \nabla \Lambda(\theta)$, for some positive constant ρ. □

Lemma 4.4 *Let $\{(X_n, S_n) : n = 0, 1, \dots\}$ be a MA–process on $\mathbb{S} \times \mathbb{R}^1$ satisfying $(\mathfrak{M}')$. Let $\mathcal{P}$ denote the transition kernel, and assume the additive components $\{\xi_n\}$ and regeneration times $\{\tau_i\}$ are bounded, and $\mathbf{E}_\pi(S_1) = 0$, where π is the stationary distribution of $\{X_n\}$. Then for any $\Delta > 0$ and $K > 0$,*

$$\lim_{\epsilon \to 0} \epsilon \log \mathbf{P} \left\{ \max_{0 \le n \le \lfloor \frac{K}{\epsilon} \rfloor} |S_n| \ge \frac{\Delta}{\epsilon} \right\} = - \inf_{t \in (0, \frac{K}{\epsilon}]} \left\{ \inf_{v = \pm \Delta} t \Lambda^*_{\mathcal{P}}(\frac{v}{t}) \right\} < 0, \tag{4.24}$$

where $(\lambda_{\mathcal{P}}(\alpha))^{-1}$ is the convergence parameter of $\hat{\mathcal{P}}(\alpha)$.

Proof See Collamore (1998), Theorem 1. Since $\Lambda^*_{\mathcal{P}}(v) = 0 \iff v = \mathbf{E}_\pi(S_1) = 0$, the right–hand side of (4.24) is < 0.

We remark that hypothesis (H2) of Collamore (1998) is not needed when the time interval ($= [0, K]$ in this case) is bounded. Also, hypothesis (H0) of that paper is satisfied, by the results of Ney and Nummelin (1987b). The "s–set" assumption in Theorem 2 of Ney and Nummelin (1987b) is *not* needed, because $\{\xi_n\}$ and $\{\tau_i\}$ are bounded; hence $r_{\mathcal{P}}(\cdot; \alpha)$ is bounded below for all α, by Lemma 2.2 (i), and inspection of the proof shows that the "s–set" condition is unnecessary in that case. □

Proof of Theorem 3.1. Lower Bound. Case 1: $\mathcal{L}_0 \Lambda_{\mathcal{K}_\mathcal{Q}} \neq \emptyset$.

Let $v \in \mathfrak{A} \cap \operatorname{ri} \mathfrak{D}_{\mathcal{K}_\mathcal{Q}}$. Then $v \in \operatorname{ri} \mathfrak{D}_M$ for sufficiently large M [Lemma 4.3 (i)]. Assume M has been chosen so that this is true. Then by Lemma 4.3 (iv), there exists $\theta \in \partial(\mathcal{L}_0 \Lambda_M)$ and a positive constant ρ such that $\rho \nabla \Lambda_M(\theta) = v$.

Define

$$\mathcal{R}_\theta(x, dy \times ds) = \frac{e^{\langle \theta, s \rangle} r_M(y; \theta)}{r_M(x; \theta)} \mathcal{K}^M_\mathcal{Q}(x, dy \times ds), \tag{4.25}$$

and observe that $\Lambda_M(\theta) = 0 \Longrightarrow \mathcal{R}_{\theta_M}$ is itself a Markov additive probability kernel.

Let $\mathfrak{P}^\epsilon_k$ denote the paths which first hit A/ϵ at time $T^\epsilon(A) = k$, and let $\tilde{x}_0 = (x_0, 0)$. Then (3.4) and (4.15) yield

$$\begin{aligned} \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) &\ge \sum_{k=1}^\infty \int_{\mathfrak{P}^\epsilon_k} \mathcal{K}^M_\mathcal{Q}(\tilde{x}_0, dx_1 \times ds_1) \cdots \mathcal{K}^M_\mathcal{Q}(x_{k-1}, dx_k \times ds_k) \\ &= \sum_{k=1}^\infty \int_{\mathfrak{P}^\epsilon_k} \frac{e^{-\langle \theta, s_1 + \cdots + s_k \rangle} r_M(\tilde{x}_0; \theta)}{r_M(x_k; \theta)} \\ &\qquad \cdot \mathcal{R}_\theta(\tilde{x}_0, dx_1 \times ds_1) \cdots \mathcal{R}_\theta(x_{k-1}, dx_k \times ds_k). \end{aligned} \tag{4.26}$$

To analyze the quantity on the right, note $\mathbf{E}_{\pi_\theta}(\xi_n) = \nabla\Lambda_M(\theta) = v/\rho$, where π_θ is the stationary distribution under the measure $\mathcal{R}_\theta$ [Ney and Nummelin (1987a), Lemma 3.3]. Thus, the expected time for the $\mathcal{R}_\theta$–process to reach the point $v/\epsilon \in (\partial A)/\epsilon$ is $\approx \rho/\epsilon$. Also, since $v \in \mathfrak{A}$, the straight–line path $[0, v]$ contains no points *other* than v in the convex set $\mathrm{cl}\, A$. Hence by Lemma 4.4,

$$\mathbf{P}_{\mathcal{R}_\theta}\{T^\epsilon(A) \le (\rho - \Delta)/\epsilon\} \to 0 \text{ as } \epsilon \to 0, \text{ for any } \Delta > 0;$$

in words, the process stays near its central tendency and therefore does *not* enter A/ϵ before the expected time of ρ/ϵ. By an analogous argument, we also obtain

$$\mathbf{P}_{\mathcal{R}_\theta}\left\{T^\epsilon(A) \le \rho/\epsilon,\ S_{T^\epsilon(A)} \in B(v,\Delta)^c/\epsilon\right\} \to 0 \text{ as } \epsilon \to 0.$$

Finally, by the central limit theorem for MA–processes,

$$\liminf_{\epsilon\to 0} \mathbf{P}_{\mathcal{R}_\theta}\{T^\epsilon(A) \le \rho/\epsilon\} \ge \text{const.} > 0.$$

Putting these together yields

$$\liminf_{\epsilon\to 0} \mathbf{P}\left\{\epsilon T^\epsilon(A) \in (\rho - \Delta, \rho],\ \epsilon S_{T^\epsilon(A)} \in B(v,\Delta)\right\} \ge \text{const.} > 0. \tag{4.27}$$

Since $r_M(\cdot;\theta)$ is positive and bounded, by Lemma 4.2 (iii), it follows from (4.26) and (4.27) that

$$\liminf_{\epsilon\to 0} \epsilon \log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \ge -\langle\theta, v\rangle - \Delta\|\theta\| = -I_M(v) - \Delta\|\theta\|. \tag{4.28}$$

Now let $\Delta \to 0$ and then $M \to \infty$. From Lemma 4.3 (iii) and (4.28), we then obtain

$$\liminf_{\epsilon\to 0} \epsilon \log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \ge -I_{\mathcal{K}_\mathcal{Q}}(v). \tag{4.29}$$

The required lower bound follows by taking the supremum in (4.29) of left and right–hand sides over $v \in \mathfrak{A} \cap \mathrm{ri}\,\mathfrak{D}_{\mathcal{K}_\mathcal{Q}}$, and observing by Lemma 4.3 (i) and the definition of $\mathfrak{A}$ that $A \cap \mathrm{ri}\,\mathfrak{D}_\mathcal{P} \neq \emptyset \Longrightarrow \mathfrak{A} \cap \mathrm{ri}\,\mathfrak{D}_{\mathcal{K}_\mathcal{Q}} \neq \emptyset$. Hence $\inf\left\{I_{\mathcal{K}_\mathcal{Q}}(v) : v \in \mathfrak{A} \cap \mathrm{ri}\,\mathfrak{D}_{\mathcal{K}_\mathcal{Q}}\right\} = \inf\left\{I_{\mathcal{K}_\mathcal{Q}}(v) : v \in \mathfrak{A}\right\}$ [cf. Collamore (1996a), the last paragraph in the proof of Theorem 2.1].

Case 2: $\mathcal{L}_0\Lambda_{\mathcal{K}_\mathcal{Q}} = \emptyset$.

Let $M_0 = \min\{M \in \mathbb{Z}_+ : \mathcal{L}_0\Lambda_M = \emptyset\} \le \infty$, and assume first that $M_0 < \infty$.

For each M, let $c_M = \inf_{\alpha\in\mathbb{R}^d} \Lambda_M(\alpha)$, and let $c = \inf_{\alpha\in\mathbb{R}^d} \Lambda_{\mathcal{K}_\mathcal{Q}}(\alpha)$. Then $\mathfrak{D}_M^{(c_M)} \searrow \mathfrak{D}_{\mathcal{K}_\mathcal{Q}}^{(c)}$ as $M \to \infty$; hence $A \cap \mathrm{ri}\,\mathfrak{D}_\mathcal{P} \neq \emptyset \Longrightarrow A \cap \mathrm{ri}\,\mathfrak{D}_M^{(c_M)} \neq \emptyset$ for sufficiently large M [Lemma 4.3 (ii)].

Let $M \ge M_0$ be chosen such that $A \cap \mathrm{ri}\,\mathfrak{D}_M^{(c_M)} \neq \emptyset$, and let $v \in \mathfrak{A} \cap \mathrm{ri}\,\mathfrak{D}_M^{(c_M)}$. Let $d_j = c_M + 1/j$, $j = 1, 2, \ldots$. Then by Lemma 4.3 (iv), there exist elements $\theta_j \in \partial(\mathcal{L}_{d_j}\Lambda_M) \cap \mathrm{aff}\,\mathfrak{D}_M$ and positive constants ρ_j such that $\rho_j \nabla\Lambda_M(\theta_j) = v$.

For each j, introduce the Markov additive probability kernel

$$\mathcal{R}_{\theta_j}(x, dy \times ds) = \frac{e^{\langle\theta_j, s\rangle - \Lambda_M(\theta_j)} r_M(y;\theta_j)}{r_M(x;\theta_j)} \mathcal{K}_\mathcal{Q}^M(x, dy \times ds), \tag{4.30}$$

and reason as in (4.26) to obtain

$$\mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \geq \sum_{k=1}^{\infty} \int_{\mathfrak{P}^\epsilon_k} \frac{e^{-\langle \theta_j, s_1+\cdots+s_k\rangle + k\Lambda_M(\theta_j)} r_M(x_0;\theta_j)}{r_M(x_k;\theta_j)} \cdot \mathcal{R}_{\theta_j}(x_0, dx_1 \times ds_1) \cdots \mathcal{R}_{\theta_j}(x_{k-1}, dx_k \times ds_k). \qquad (4.31)$$

It follows from (4.27) and (4.31) that

$$\liminf_{\epsilon\to 0} \epsilon \log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \geq -\langle \theta_j, v\rangle + \rho_j \Lambda_M(\theta_j). \qquad (4.32)$$

We now distinguish two possible cases. First, suppose that $\{\theta_j\}$ converges (possibly after passing to a subsequence) to some element $\hat{\theta} \in \mathbb{R}^d$. Then the infimum in the definition of c_M is *achieved* at $\hat{\theta}$; hence $\Lambda_M(\hat{\theta}) = c_M > 0$ and $\nabla\Lambda_M(\hat{\theta}) = 0$. But then $\lim_{j\to\infty} \rho_j = \lim_{j\to\infty} (v/\nabla\Lambda_M(\theta_j)) = \infty$. Letting $j \to \infty$ in (4.32), we conclude

$$\lim_{\epsilon\to 0} \epsilon \log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) = \infty. \qquad (4.33)$$

Next, suppose that $\{\theta_j\}$ does not converge along any subsequence. Let $\beta_j = \theta_j/\|\theta_j\|$, and observe that (possibly after passing to a subsequence) $\beta_j \to \hat{\beta} \in \mathbf{S}^{d-1}$ and $\|\theta_j\| \to \infty$ as $j \to \infty$. Then $\hat{\beta} \in 0^+\Lambda_M$ [Rockafellar (1970), Theorems 8.2 and 8.7]. Hence $\hat{\beta} \in \left(\mathfrak{D}_M^{(c_M)}\right)^\circ$ (as in the proof of Lemma 4.3 (i)). Since the θ_j's were chosen in aff $\mathfrak{D}_M$, it follows that $\hat{\beta} \in \mathfrak{D}^\circ_M \cap \text{aff}\, \mathfrak{D}_M$. Then $v \in \text{ri}\, \mathfrak{D}_M \implies \langle \hat{\beta}, v\rangle < 0$. Hence $\langle \beta_j, v\rangle \leq b < 0$, for all $j \geq$ some j_0. But then

$$-\langle \theta_j, v\rangle = -\|\theta_j\| \langle \beta_j, v\rangle \to \infty \text{ as } j \to \infty. \qquad (4.34)$$

Thus, letting $j \to \infty$ in (4.32), we again obtain (4.33).

Finally suppose $M_0 = \infty$. In this case, the elements of $\{\mathcal{L}_0\Lambda_M : M = 1, 2, \ldots\}$ are nonempty and monotonically decreasing to $\bigcap_M \mathcal{L}_0\Lambda_M = \mathcal{L}_0\Lambda_{\mathcal{K}_\mathcal{Q}} = \emptyset$. Then

$$\inf\{\|v\| : v \in \mathcal{L}_0\Lambda_M\} \to \infty \text{ as } M \to \infty. \qquad (4.35)$$

Since Λ_M is strictly increasing, which (with $M_0 = \infty$) implies $\inf_\alpha \Lambda_M(\alpha) < 0$, $\forall M$, we may apply Lemma 4.3 (iv) to obtain elements $\theta_M \in \left(\partial(\mathcal{L}_0\Lambda_M) \cap \text{aff}\, \mathfrak{D}_M\right)$ and positive constants ρ_M such that $\rho_M \nabla\Lambda_M(\theta_M) = v$. Then (4.35) implies $\|\theta_M\| \to \infty$, and (possibly after passing to a subsequence) $\beta_M \stackrel{\text{def}}{=} \theta_M/\|\theta_M\| \to \hat{\beta} \in \bigcap_M 0^+\Lambda_M = \mathfrak{D}^\circ_{\mathcal{K}_\mathcal{Q}}$ [Lemma 4.3 (ii) and its proof]. Then (4.32), (4.34) (with "M" in place of "j") give (4.33), as before. □

4.3 Proofs of Theorem 3.4 and Proposition 3.5

Next we turn to the proof of Theorem 3.4.

Let $\mathcal{Q}^*$ be the kernel described in (3.17), and let $\mathcal{Q} \in \mathcal{C}$. By the Radon-Nikodym Theorem, we may write $\mathcal{Q}^{(t)} = \mathcal{R}^{(t)} + \mathcal{V}^{(t)}$, where $\mathcal{R}^{(t)} \ll \mathcal{Q}^*$ and $\mathcal{V}^{(t)} \perp \mathcal{Q}^*$. Now define

$$Z^\epsilon_{n+1} = \log\left(\frac{d\mathcal{R}^{(\epsilon n)}}{d\mathcal{Q}^*}(X^*_n, X^*_{n+1} \times \xi^*_{n+1})\right), \; n = 1, 2, \ldots, \qquad (4.36)$$

where $\{(X_n^*, S_n^*) : n = 0, 1, \dots\}$ denotes a MA–process with transition kernel $\mathcal{Q}^*$. Let

$$W_n^\epsilon = Z_1^\epsilon + \cdots + Z_n^\epsilon, \quad n = 1, 2, \dots, \quad \text{and} \quad W_0^\epsilon = 0.$$

The proof of Theorem 3.4 will rely on the following.

Lemma 4.5 (i) *For any fixed* ϵ, $\{W_n^\epsilon\}_{n\geq 0}$ *is a submartingale.*

(ii) *If* $\bar{Z}_n^\epsilon = Z_n^\epsilon \vee 0 - 1$, $\bar{W}_n^\epsilon = \bar{Z}_1^\epsilon + \cdots + \bar{Z}_n^\epsilon$ *for* $n \geq 1$, *and* $\bar{W}_n^\epsilon = 0$, *then* $\{\bar{W}_n^\epsilon\}$ *is a submartingale.*

(iii) *Suppose* $\mathcal{Q} \in \mathcal{C}_0$, *so that* $\{Z_n^\epsilon\}$, $\{W_n^\epsilon\}$ *are actually independent of* ϵ, *and let* $Z_n^M = Z_n^\epsilon \vee (-M)$ *and* $W_n^M = Z_1^M + \cdots + Z_n^M$ *for* $n \geq 1$, *and* $W_0^M = 0$. *Assume that the lower bound of* $(\mathfrak{R})$ *is satisfied, and that we do not have* $\mathcal{Q} = \mathcal{Q}^*$ *for* $\varphi \times \mathcal{P}$ *a.e.* (x, y, s). *Then, for certain positive constants* M_0 *and* D,

$$\limsup_{n\to\infty} \frac{1}{n} \mathbf{E}_{\mathcal{Q}^*}\left(W_n^M\right) \leq -D, \quad \text{for all} \quad M \geq M_0. \tag{4.37}$$

Proof (i) Jensen's inequality implies that $\mathbf{E}\left(W_{n+1}^\epsilon | X_n^* = x\right) \leq 0$ for *all* x; hence $\{W_n^\epsilon\}$ is a submartingale.

(ii) This follows by a similar argument and the inequality $(\log s) \leq s$.

(iii) Let $\{\tilde{X}_n^*, \tilde{S}_n^*, \tilde{W}_n : n = 0, 1, \dots\}$ be an independent copy of $\{X_n^*, S_n^*, W_n : n = 0, 1, \dots\}$, but assume that the initial measure of $\tilde{X}_n^*$ is

$$\pi^* = \text{ the stationary measure of } \{X_n^*\} \text{ under the transition kernel } \mathcal{Q}^*.$$

Let $\{T_i\}_{i\in\mathbb{N}}$ and $\{\tilde{T}_i\}_{i\in\mathbb{N}}$ denote the respective regeneration times, as described in Lemma 2.1, and let

$$\mathfrak{T} = \inf\left\{n : T_i = n \text{ and } \tilde{T}_j = n, \text{ some } i, j \in \mathbb{N}\right\}$$

denote the coupling time.

First note that if we do not have $\mathcal{Q} = \mathcal{Q}^*$ for $\varphi \times \mathcal{P}$ a.e. (x, y, s), then by Jensen's inequality $\mathbf{E}_{\pi^*}(Z_n^\epsilon) < 0$. By the monotone convergence $Z_n^M \searrow Z_n^\epsilon$ as $M \to \infty$, it follows that

$$\mathbf{E}_{\pi^*}(Z_n^M) \leq -D' < 0, \quad \text{for all } M \geq \text{ some } M_0'. \tag{4.38}$$

Consequently

$$\mathbf{E}(\tilde{W}_n^M) \leq -nD', \quad \text{all } M \geq \text{ some } M_0'. \tag{4.39}$$

Let $\mathfrak{T}_n = \mathfrak{T} \wedge n$, and observe

$$\mathbf{E}\left(W_n^M\right) = \mathbf{E}\left(\tilde{W}_n^M\right) + \mathbf{E}\left(W_{\mathfrak{T}_n}^M - \tilde{W}_{\mathfrak{T}_n}^M\right). \tag{4.40}$$

Also, by a slight variant of (i), $\{W_n^M - n\epsilon\}$ is a submartingale for all $M \geq \log\{1/(e^\epsilon - 1)\}$, and analogously for (ii). Hence, letting $\epsilon = D'/2$, we obtain by the optional sampling theorem that

$$\mathbf{E}\big(W_{\mathfrak{T}_n}^M\big) = \frac{D'}{2} \cdot \mathbf{E}(\mathfrak{T}) = C < \infty, \quad \text{all } M \geq \text{ some } M_0''; \tag{4.41}$$

and since $\tilde{W}_n^M \geq -M$ we also have

$$-\mathbf{E}\big(\tilde{W}_{\mathfrak{T}_n}^M\big) \leq M\mathbf{E}\big(\mathfrak{T}\big) = C' < \infty, \tag{4.42}$$

where C, C' are $< \infty$ by Remark 2.1.1 (i). The required result is then obtained by substituting (4.41) and (4.42) into (4.40). □

Lemma 4.6 *Let $A \subset \mathbb{R}^d$ be a convex semi–cone intersecting* $\mathrm{ri}\,\mathfrak{S}$. *Assume* $\mathrm{dom}\,\Lambda_{\mathcal{P}}$ *is open, and that (H2) and* $(\mathfrak{R})$ *are satisfied. Let ϱ be given as in Lemma 3.2, and let $\tau \stackrel{d}{=} \tau_i$, where $\tau_i = T_{i+1} - T_i$ are the interregeneration times described in Lemma 2.1. Define* $\mathcal{I}^\epsilon(A) = \inf\{i : T_i \geq T^\epsilon(A)\}$. *Then for φ a.e. x_0,*

$$\lim_{\epsilon \to 0} \epsilon \mathbf{E}_{\mathcal{Q}^*}\big(\mathcal{I}^\epsilon(A)\big) = \frac{\varrho}{\mathbf{E}(\tau)}. \tag{4.43}$$

Proof *Lower bound.* First intoduce a truncation on the additive components; namely, let $M > 0$ and define

$$\xi_n^M = \xi_n^* \quad \text{if } \langle \alpha_0, \xi_n^* \rangle \geq -M, \qquad \text{and} \qquad \xi_n^M = \xi_n^* \frac{M}{\|\langle \alpha_0, \xi_n^* \rangle\|} \quad \text{otherwise}. \tag{4.44}$$

Let $S_n^M = \xi_1^M + \cdots + \xi_n^M$, $n = 0, 1, \ldots$, and let $S_0^M = 0$. Also let $B = \mathcal{H}(\alpha_0, a)$, where α_0 and a are given as in Lemma 3.2. Then by Lemma 3.2 (i), $\mathcal{I}^\epsilon(A) \geq \mathcal{I}^{M,\epsilon}(B)$ (where $\mathcal{I}^{M,\epsilon}(\cdot)$ denotes the stopping time with respect to the truncated process $\{S_n^M\}_{n \in \mathbb{N}}$).

Note dom $\Lambda_{\mathcal{P}}$ open $\Longrightarrow$ $0 \in$ int (dom $\Lambda_{\mathcal{Q}^*}$) [Lemma 2.3 (iii)]. Hence by the optional sampling theorem,

$$\begin{aligned} \mathbf{E}_{\mathcal{Q}^*}\big(\langle \alpha_0, S^M_{T_{\mathcal{I}^{M,\epsilon}(B)}} \rangle\big) &= \mathbf{E}_{\mathcal{Q}^*}\big(\langle \alpha_0, S^M_{T_0} \rangle\big) \\ &\quad + \mathbf{E}_{\mathcal{Q}^*}\big(\langle \alpha_0, S^M_{T_{i+1}} - S^M_{T_i} \rangle\big) \mathbf{E}_{\mathcal{Q}^*}\left(\big(\mathcal{I}^{M,\epsilon}(B)\big) - 1\right). \end{aligned} \tag{4.45}$$

Also, under the above truncation,

$$\mathbf{E}_{\mathcal{Q}^*}\big(\langle \alpha_0, S^M_{T_{\mathcal{I}^{M,\epsilon}(B)}} - S^M_{T^{M,\epsilon}(B)} \rangle\big) \geq -M\mathbf{E}_{\mathcal{Q}^*}\big(T_{\mathcal{I}^{M,\epsilon}(B)} - T^{M,\epsilon}(B)\big) \geq -MC. \tag{4.46}$$

where by Remark 2.1.1 (i), the constant $C < \infty$. Since $S^M_{T^{M,\epsilon}(B)} \geq a/\epsilon$, it follows from (4.45) and (4.46) that

$$\liminf_{\epsilon \to 0} \epsilon \mathbf{E}_{\mathcal{Q}^*}\big(\mathcal{I}^{M,\epsilon}(B)\big) \geq a/\mathbf{E}_{\mathcal{Q}^*}\big(\langle \alpha_0, S^M_{T_{i+1}} - S^M_{T_i} \rangle\big), \tag{4.47}$$

provided that

$$\mathbf{E}_{\mathcal{Q}^*}\big(\langle \alpha_0, S^M_{T_0} \rangle\big) \geq -M\mathbf{E}\big(T_0\big) < \infty \quad \text{for } \varphi \text{ a.e. } x_0. \tag{4.48}$$

Finally observe by Lemma 3.3 of Ney and Nummelin (1987a) and Lemma 3.2 (iv),

$$\mathbf{E}_{\mathcal{Q}^*}\left(S^*_{T_{i+1}} - S^*_{T_i}\right) = \nabla\Lambda_{\mathcal{P}}(\alpha_0)\cdot\mathbf{E}(\tau) = \frac{v_0}{\varrho}\mathbf{E}(\tau), \quad i = 0, 1, \ldots . \tag{4.49}$$

Then (4.48) holds, and the required result follows from (4.47), (4.49), Lemma 3.2 (iii), and the monotone convergence $\langle \alpha_0, S^M_{T_{i+1}} - S^M_{T_i}\rangle \searrow \langle \alpha_0, S^*_{T_{i+1}} - S^*_{T_i}\rangle$ as $M \to \infty$.

Upper bound. First assume $d > 1$. Let $t > 0$, and observe since A is a semi–cone that $(1+t)v_0$ is an interior point of A. Choose $w^{(1)}, \ldots, w^{(d)} \in A$ such that the convex hull of $\{v_0, w^{(1)}, \ldots, w^{(d)}\}$ contains a neighborhood of $(1+t)v_0$. Let $v^{(k)} = v_0 + w^{(k)}$; let $\mathcal{J}^{(k)} =$ the hyperplane containing $\{v_0, v^{(1)}, \ldots, v^{(d)}\} - \{v^{(k)}\}$; and let $\mathcal{H}^{(k)} =$ the open halfspace determined by $\mathcal{J}^{(k)}$ which contains the point $(1+t)v_0$. It follows from this constuction and the semi–cone property that $0 \notin \mathcal{H}^{(k)}$ and $\bigcap_{k=1}^{d} \mathcal{H}^{(k)} \subset A$.

Let $\mathcal{I}^\epsilon_j = \inf\{i : S^*_{T_i} \in \mathcal{H}^{(k)}, \text{ all } j \geq i\}$. By (4.49), the expected time for $\{S^*_{T_i}\}_{i\in\mathbb{N}}$ to reach $\mathcal{H}^{(k)}$ is $i = \varrho/\mathbf{E}(\tau)$. Hence by a simple one–dimensional change of measure argument,

$$\lim_{\epsilon\to 0}\mathbf{E}_{\mathcal{Q}^*}\left[\mathcal{I}^\epsilon_j;\ \mathcal{I}^\epsilon_j \geq \frac{1}{\epsilon}\left(\frac{\varrho+\Delta}{\mathbf{E}(\tau)}\right)\right] = 0, \quad \text{for all } \Delta > 0. \tag{4.50}$$

Since $\mathcal{I}^\epsilon(A) \leq \max\left\{\mathcal{I}^\epsilon_1, \ldots, \mathcal{I}^\epsilon_d\right\}$, the upper bound is obtained from (4.50).

Finally, if $d = 1$ then the upper bound can be obtained directly from (4.50). □

Proof of Theorem 3.4. Following Asmussen and Rubenstein (1995), Theorem 17.6, first observe

$$d\mathcal{K}_{\mathcal{Q}^{(\epsilon n)}} \stackrel{\text{def}}{=} \left(\frac{d\mathcal{P}}{d\mathcal{Q}^{(\epsilon n)}}\right)^2 d\mathcal{Q}^{(\epsilon n)} = \left(\frac{d\mathcal{P}}{d\mathcal{Q}^*}\right)^2\left(\frac{d\mathcal{Q}^*}{d\mathcal{Q}^{(\epsilon n)}}\right) d\mathcal{Q}^*. \tag{4.51}$$

Also by the Radon–Nikodym Theorem and the definition of $\mathcal{R}^{(\epsilon n)}$,

$$\frac{d\mathcal{Q}^*}{d\mathcal{Q}^{(\epsilon n)}} = \frac{d\mathcal{Q}^*}{d\mathcal{R}^{(\epsilon n)}} = \left(\frac{d\mathcal{R}^{(\epsilon n)}}{d\mathcal{Q}^*}\right)^{-1} \qquad \mathcal{Q}^*\ a.e. \tag{4.52}$$

From (4.51) and (4.52) it follows that

$$\begin{aligned}\mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) &= \mathbf{E}_{\mathcal{Q}^*}\left[\frac{r^2_{\mathcal{P}}(x_0;\alpha_0)}{r^2_{\mathcal{P}}(X_{T^\epsilon(A)};\alpha_0)}\, e^{-2\left\langle \alpha_0, S^*_{T^\epsilon(A)}\right\rangle - W^\epsilon_{T^\epsilon(A)}}\right] \\ &\geq \exp\left\{-2\mathbf{E}_{\mathcal{Q}^*}\left(\left\langle \alpha_0, S^*_{T^\epsilon(A)}\right\rangle\right) - \mathbf{E}_{\mathcal{Q}^*}(W^\epsilon_{T^\epsilon(A)}) + C'\right\},\end{aligned} \tag{4.53}$$

by Jensen's inequality, where C' is a finite constant obtained from the uniform positivity and boundedness of $r_{\mathcal{P}}(\alpha_0)$ as described in Lemma 2.2 (iii).

Let $T_0, T_1, \ldots$ denote the regeneration times in Lemma 2.1 generated by the MA–process $\{(X_n, S_n)\}$; let $\{\tau_i\}$ denote the interregeneration times; and let

$$\mathcal{I}^\epsilon(A) = \inf\{i : T_i \geq T^\epsilon(A)\}.$$

Introduce the truncation $\left\{(\xi^M_n, S^M_n) : n = 0, 1, \ldots\right\}$ of $\left\{(\xi^*_n, S^*_n) : n = 0, 1, \ldots\right\}$ that was described above in (4.44), and observe under this truncation that (4.46) holds with A

in place of B and $\mathcal{I}^\epsilon(\cdot)$ in place of $\mathcal{I}^{M,\epsilon}(\cdot)$. It follows from (4.52), (4.53), and the definition of $\{S_n^M\}$ that

$$\log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \geq -2\mathbf{E}_{\mathcal{Q}^*}\big(\langle \alpha_0, S^M_{T_{\mathcal{I}^\epsilon(A)}}\rangle\big) - \mathbf{E}_{\mathcal{Q}^*}(W^\epsilon_{T^\epsilon(A)}) + C, \tag{4.54}$$

where $C \in (-\infty, \infty)$.

By the optional sampling theorem and Lemma 4.5 (i), (ii),

$$\mathbf{E}_{\mathcal{Q}^*}(W^\epsilon_{T^\epsilon(A)}) \leq 0. \tag{4.55}$$

Also, by the optional sampling theorem

$$\mathbf{E}_{\mathcal{Q}^*}\big(\langle \alpha_0, S^M_{T_{\mathcal{I}^\epsilon(A)}}\rangle\big) = \mathbf{E}_{\mathcal{Q}^*}\big(\langle \alpha_0, S^M_{T_0}\rangle\big) + \mathbf{E}_{\mathcal{Q}^*}\big(\langle \alpha_0, S^M_{T_{i+1}} - S^M_{T_i}\rangle\big)\mathbf{E}_{\mathcal{Q}^*}\big((T_{\mathcal{I}^\epsilon(A)}) - 1\big). \tag{4.56}$$

Then by (4.48), (4.49), Lemma 4.6, and the monotone convergence $\langle \alpha_0, S^M_{T_{i+1}} - S^M_{T_i}\rangle \searrow \langle \alpha_0, S^*_{T_{i+1}} - S^*_{T_i}\rangle$,

$$\lim_{\epsilon\to 0} \epsilon \mathbf{E}_{\mathcal{Q}^*}\big(\langle \alpha_0, S^M_{T_{\mathcal{I}^\epsilon(A)}}\rangle\big) \searrow \langle \alpha_0, v_0\rangle \quad \text{as } M \to \infty. \tag{4.57}$$

From (4.54), (4.55) and (4.57) we conclude

$$\liminf_{\epsilon\to 0} \epsilon \log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \geq -2\langle \alpha_0, v_0\rangle. \tag{4.58}$$

In view of Lemma 3.2 (iii) and Proposition 3.3, this implies

$$\liminf_{\epsilon\to 0} \epsilon \log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q},\epsilon}) \geq \lim_{\epsilon\to 0} \epsilon \log \mathbf{E}(\mathcal{E}^2_{\mathcal{Q}^*,\epsilon}). \tag{4.59}$$

It remains to show that if $\mathcal{Q}^{(t_0)} \neq \mathcal{Q}^*$ at some continuity point $t_0 \in [0, \varrho]$, then there is strict inequality in (4.59). Suppose now that $t_0 \in [0, \varrho]$ is a continuity point and $\mathcal{Q}^{(t_0)} \neq \mathcal{Q}^*$. Then the continuity property (3.20) is satisfied in some interval $[\zeta_1, \zeta_2] \subset [0, \varrho)$. Let D and M_0 be the constants obtained in Lemma 4.5 (iii) when $\mathcal{Q}_0 = \mathcal{Q}^{(t_0)}$. Assume that the interval $[\zeta_1, \zeta_2]$ has been chosen sufficiently small so that (3.20) holds with $\Delta = D/2$.

Decompose the r.v. Z_n^ϵ into a sum of two terms, namely,

$$\begin{aligned} U_n &= \log\left(\frac{d\mathcal{R}^{(t_0)}}{d\mathcal{Q}^*}(X_n^*, X_{n+1}^* \times \xi_{n+1}^*)\right), \ n = 0, 1, \dots, \\ V_n^\epsilon &= \log\left(\frac{d\mathcal{R}^{(\epsilon n)}}{d\mathcal{R}^{(t_0)}}(X_n^*, X_{n+1}^* \times \xi_{n+1}^*)\right), \ n = 0, 1, \dots. \end{aligned}$$

For $M > 0$, let $U_n^M = U_n \vee (-M)$, $V_n^{M,\epsilon} = V_n^\epsilon \vee (-M)$,

$$R_n^M = U_0^M + \cdots + U_n^M, \qquad Z_n^{M,\epsilon} = U_n^M + V_n^\epsilon,$$

and $W_n^{M,\epsilon} = Z_0^{M,\epsilon} + \cdots + Z_n^{M,\epsilon}$.

Now by Lemma 4.5 (iii),

$$\limsup_{\epsilon\to 0} \epsilon \mathbf{E}_{\mathcal{Q}^*}\Big(R^M_{\lfloor \frac{\zeta_2}{\epsilon}\rfloor} - R^M_{\lfloor \frac{\zeta_1}{\epsilon}\rfloor}\Big) \leq -\Big(\frac{\zeta_2 - \zeta_1}{\zeta_2}\Big) D, \quad \text{all } M \geq M_0. \tag{4.60}$$

Hence from the continuity properties (3.20) and (3.21) and a straightforward variant of Lemma 4.5 (iii), we obtain

$$\limsup_{\epsilon\to 0} \epsilon \mathbf{E}_{\mathcal{Q}^*}\left(W^{M,\epsilon}_{\lfloor\frac{\zeta_2}{\epsilon}\rfloor} - W^{M,\epsilon}_{\lfloor\frac{\zeta_1}{\epsilon}\rfloor}\right) \le -D' < 0, \quad \text{all } M \ge M_0. \tag{4.61}$$

Moreover by the optional sampling theorem and Lemma 4.5 (i),

$$\mathbf{E}_{\mathcal{Q}^*}\left(W^{\epsilon}_{T^\epsilon(A)\wedge\lfloor\frac{\zeta_1}{\epsilon}\rfloor}\right) \le 0, \tag{4.62}$$

and

$$\mathbf{E}_{\mathcal{Q}^*}\left(W^{\epsilon}_{T^\epsilon(A)} - W^{\epsilon}_{\lfloor\frac{\zeta_2}{\epsilon}\rfloor};\, T^\epsilon(A) > \lfloor\frac{\zeta_2}{\epsilon}\rfloor\right) \le 0. \tag{4.63}$$

It follows from (4.61)–(4.63) and the definition of $\{W^{M,\epsilon}_n\}$ that

$$\limsup_{\epsilon\to 0} \epsilon \mathbf{E}\left(W^{\epsilon}_{T^\epsilon(A)}\right) \le -D' - \liminf_{\epsilon\to 0} \epsilon \mathbf{E}_{\mathcal{Q}^*}\left(W^{M,\epsilon}_{\lfloor\frac{\zeta_2}{\epsilon}\rfloor} - W^{M,\epsilon}_{T^\epsilon(A)\wedge\lfloor\frac{\zeta_1}{\epsilon}\rfloor};\, T^\epsilon(A) < \frac{\zeta_2}{\epsilon}\right). \tag{4.64}$$

Since $\zeta_2 < \varrho$, $\mathbf{P}\{T^\epsilon(A) < \lfloor\zeta_2/\epsilon\rfloor\} \overset{\exp}{\to} 0$ [Collamore (1998), Theorem 1; cf. (4.49) and the proof of Lemma 4.4]. Also, from the above definitions the last integrand in (4.64) is bounded below by $-2M\,[(\zeta_2-\zeta_1)/\epsilon+1]$. We conclude that the last term on the right of (4.64) can actually be dropped.

Using (4.64) in place of (4.55) now gives strict inequality in (4.59), as desired. □

Proof of Proposition 3.5 (i) By definition,

$$\left(\mathcal{L}_b I_{\mathcal{P}}\right)^c = \bigcup_{\alpha\in\mathcal{L}_0\Lambda_{\mathcal{P}}} \mathcal{H}(\alpha, b), \quad \text{for all } b \ge 0.$$

Hence $\{\mathcal{H}(\alpha, a-\Delta)\}_{\alpha\in\mathcal{L}_0\Lambda_{\mathcal{P}}}$ is an open cover for $\mathfrak{B} \overset{\text{def}}{=} \partial(\mathcal{L}_a I_{\mathcal{P}}) \cap \text{cone}_{\,\delta}\left(\mathcal{L}_0\Lambda^*_{\mathcal{P}}\right)^c$.

The set $\mathfrak{B}$ is compact, since $I_{\mathcal{P}}$ is positively homogeneous and strictly positive on the compact set $\mathbf{S}^{d-1} \cap \text{cone}_{\,\delta}\left(\mathcal{L}_0\Lambda^*_{\mathcal{P}}\right)^c$ [Collamore (1996b), Lemma 3.1]. Hence there exists a finite subcover for $\mathfrak{B}$. This subcover also covers $\left(\mathcal{L}_a I_{\mathcal{P}}\right)^c \cap \text{cone}_{\,\delta}\left(\mathcal{L}_0\Lambda^*_{\mathcal{P}}\right)^c$, and hence A.

(ii) This is established in the same way as the upper bound of Theorem 3.1, with α_i in place of θ. (See also the comments following the proof of this upper bound.) In the case where A is a finite union of convex sets, choose the α_i's to be the elements obtained in Lemma 3.2 when $A = A'_i$, $i = 1, \ldots, k$, and then proceed as in the proof of the upper bound of Theorem 3.1. □

References

Asmussen, S. (1999). *Stochastic Simulation with a View Towards Stochastic Processes.* Aarhus, Denmark: MaPhySto Lecture Notes.

Asmussen, S. and R. Y. Rubinstein (1995). Steady–state rare events simulation in queueing models and its complexity properties. In J. Dshalalow (Ed.), *Advances in Queueing: Models, Methods and Problems*, pp. 429–461. Boca Raton, FL: CRC Press.

Athreya, K. B. and P. Ney (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245**, 493–501.

Borovkov, A. A. (1997). Limit theorems for time and place of the first boundary passage by a multidimensional random walk. *Doklady* **55**, 254–256.

Borovkov, A. A. and A. A. Mogulskii (1996). Large deviations for stationary Markov chains in a quarter plane. In *Probability Theory and Mathematical Statistics (Tokyo, 1995)*. River Edge, NJ: World Sci. Publishing.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (2nd ed.). New York: Springer Verlag.

Bucklew, J. (1998). The blind simulation problem and regenerative processes. *IEEE Trans. Inform. Theory* **44**, 2877–2891.

Bucklew, J. A., P. Ney, and J. S. Sadowsky (1990). Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *J. Appl. Prob.* **27**, 44–59.

Collamore, J. F. (1996a). Hitting probabilities and large deviations. *Ann. Probab.* **24**, 2065–2078.

Collamore, J. F. (1996b). *Large Deviation Techniques for the Study of the Hitting Probabilities of Rare Sets*. Ph.D. dissertation, University of Wisconsin-Madison.

Collamore, J. F. (1998). First passage times of general sequences of random vectors: a large deviations approach. *Stochastic Process. Appl.* **78**, 97–130.

Cramér, H. (1930). On the mathematical theory of risk. In *Scand. Jubilee Volume*, Stockholm.

de Acosta, A. (1988). Large deviations for vector–valued functionals of a Markov chain: lower bounds. *Ann. Probab.* **16**, 925–960.

de Acosta, A. and P. Ney (1998). Large deviation lower bounds for arbitrary additive functionals of a Markov chain. *Ann. Probab.* **26**, 1660–1682.

Dembo, A. and O. Zeitouni (1998). *Large Deviations Techniques and Applications* (2nd ed.). Berlin: Springer–Verlag.

Donsker, M. D. and S. R. S. Varadhan (1975, 1976, 1983). Asymptotic evaluation of certain Markov process expectations for large time. I, II, III, IV. *Comm. Pure Appl. Math.* **28**, 1–47; **28**, 279–301; **29**, 389–461; **36**, 183–212.

Freidlin, M. I. and A. D. Wentzell (1984). *Random Perturbations of Dynamical Systems*. New York: Springer-Verlag.

Glasserman, P. and S. Kou (1995). Analysis of an importance sampling estimator for tandem queues. *ACM Trans. Modeling Comp. Simulation* **4**, 22–42.

Glasserman, P. and Y. Wang (1997). Counterexamples in importance sampling for large deviations probabilities. *Ann. Appl. Probab.* **7**, 731–746.

Iscoe, I., P. Ney, and E. Nummelin (1985). Large deviations of uniformly recurrent Markov additive processes. *Adv. in Appl. Math.* **6**, 373–412.

Lehtonen, T. and H. Nyrhinen (1992a). On asymptotically efficient simulation of ruin probabilities in a Markovian environment. *Scand. Actuarial J.*, 60–75.

Lehtonen, T. and H. Nyrhinen (1992b). Simulating level-crossing probabilities by importance sampling. *Adv. Appl. Prob.* **24**, 858–874.

Lundberg, F. (1909). Zur Theorie der Rückversicherung. In *Verhandl. Kongr. Versicherungsmath*, Wien.

Meyn, S. P. and R. L. Tweedie (1993). *Markov Chains and Stochastic Stability.* Berlin: Springer-Verlag.

Ney, P. and E. Nummelin (1984). Some limit theorems for Markov additive processes. In *Proc. of Symposium on semi–Markov processes*, Brussels.

Ney, P. and E. Nummelin (1987a). Markov additive processes I. Eigenvalue properties and limit theorems. *Ann. Probab.* **15**, 561–592.

Ney, P. and E. Nummelin (1987b). Markov additive processes II. Large deviations. *Ann. Probab.* **15**, 593–609.

Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete* **43**, 309–318.

Nummelin, E. (1984). *General Irreducible Markov Chains and Non–negative Operators.* Cambridge: Cambridge University Press.

Revuz, D. (1975). *Markov Chains.* Amsterdam: North–Holland Publishing Company.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton: Princeton Univ. Press.

Siegmund, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* **4**, 673–684.

EURANDOM
EINDHOVEN UNIVERSITY OF TECHNOLOGY
P. O. BOX 513
5600 MB EINDHOVEN
THE NETHERLANDS
E-MAIL: collamore@eurandom.tue.nl