BUSY PERIOD ANALYSIS, RARE EVENTS
AND TRANSIENT BEHAVIOR IN FLUID FLOW MODELS

SØREN ASMUSSEN
Aalborg University
Institute of Electronic Systems
Fr. Bajersv. 7, DK-9220 Aalborg, DENMARK

ABSTRACT

We consider a process $\{(J_t, V_t)\}_{t \geq 0}$ on $E \times [0, \infty)$, such that $\{J_t\}$ is a Markov process with finite state space E, and $\{V_t\}$ has a linear drift r_i on intervals where $J_t = i$ and reflection at 0. Such a process arises as a fluid flow model of current interest in telecommunications engineering for the purpose of modeling ATM technology. We compute the mean of the busy period and related first passage times, show that the probability of buffer overflow within a busy cycle is approximately exponential, and give conditioned limit theorems for the busy cycle with implications for quick simulation. Further, various inequalities and approximations for transient behavior are given. Also explicit expressions for the Laplace transform of the busy period are found. Mathematically, the key tool is first passage probabilities and exponential change of measure for Markov additive processes.

Key words: Markov Additive Process, Skip-free, Buffer Over-flow, Conditional Limit Theorems, First Passage Time, Change of Measure, Exponential Family, Likelihood Ratio Identity, Extreme Values, Quick Simulation, Asynchronous Transfer Mode.

AMS (MOS) subject classifications: 60K25, 94A05.

1. Introduction

Fluid flow processes can be seen as a class of applied probability models which in many ways is parallel to queues. From an application point of view, the historical origin is in both cases performance evaluation in telecommunication, with the difference being motivated in the change of technology: from switchboards in the days of Erlang to modern ATM (asynchronous transfer mode) devices. Mathematically, both class of models have fundamental relations to random walks and more general additive processes. For queues, the classical example is the reflected random walk representation of the $GI/G/1$ waiting time via the Lindley recursion ([5], Ch. III.7-8). More recently, the use of Markov-modulation for modeling bursty traffic has led into more general Markov additive processes (see e.g. [6], [7]) which are also the key tool we use for studying fluid flow models, by representing them as reflected versions of finite Markov additive processes with the additive component having the simplest possible structure of a pure linear drift.

Most of the applied literature deals with the computation of the steady-state distribution.

However, as in queueing theory, steady-state theory only tells part of the story and one may want to have some information on transient behavior as well. The purpose of this paper is to present a study of this aspect; in particular, we study the behavior within a busy cycle and bounds and approximations for the time-dependent state probabilities. Since Professor Takács is one of the main early contributors to busy period analysis and the study of transient behavior for queues (see e.g., his book [39]), and his work [40] on cycle maxima forms the foundation of an earlier study by the author ([9], with Perry), it is a great pleasure to present this contribution in the present volume.

The process $\{(J_t, V_t)\}_{t \geq 0}$ under study is defined by $\{J_t\}$ being an irreducible Markov process with finite state space E, and $\{V_t\}$ having piecewise linear paths with slope r_i on intervals where $J_t = i, V_t > 0$, and reflection at 0. The stability condition ensuring the existence of a limiting steady-state is

$$\sum_{i \in E} \tau_i r_i < 0, \tag{1.1}$$

where $\pi = (\pi_i)_{i \in E}$ is the stationary distribution of $\{J_t\}$, and we let (J, V) denote a pair of random variables having the limiting stationary distribution of (J_t, V_t). For the evaluation of the distribution of (J, V), see Anick et al. [2], Gaver & Lehoczky [19] for some early studies and Asmussen [8], Rogers [35] for more recent treatments and a more complete set of references (note, however, that the mathematically attractive features of the models have motivated purely theoretical papers like Barlow, Rogers & Williams [11]). We write $E_+ = \{i \in E : r_i > 0\}$, $E_- = \{i \in E : r_i < 0\}$ (for simplicity, it is assumed that $r_i \neq 0$ for all i though this assumption is not crucial, cf. [8]).

A summary of the results and organization of the paper is as follows. One of the main topics is various aspects of busy period behavior. A busy period of length $P_i = \inf\{t > 0 : V_t = 0\}$ starts from $V_0 = 0$ and $J_0 = i \in E_+$, and ends at the time P_i the process returns to 0. Our first result (Section 3) is an expression for the mean busy period $\mathbb{E}_i P_i$ given in terms of a set of linear equations; the equations involve quantities related to the steady-state solution. Besides its intrinsic interest, the mean busy cycle also enters in an essential way in the rare events analysis which is carried out in Section 5. Letting $M_V(T) = \max_{0 \leq t \leq T} V(t)$, the first result given there states that the cycle maximum $M_V(P_i)$ has an asymptotically exponential tail. The implications are that after suitable normalizations, the first time $\{V(t)\}$ exceeds a given large level u has an asymptotical exponential distribution, and $M_V(T)$ itself one of the classical extreme value distributions.

Section 7 deals with transient behavior, more precisely the study of $\mathbb{P}(V_T > u)$. We show that for large u and T, a certain time epoch of the form $T = u/\kappa'(\gamma)$ (with κ and γ defined in the body of the paper) plays a crucial role as the time at which $\mathbb{P}(V_T > u)$ approximately attains its stationary value $\mathbb{P}(V > u)$ (which in turn is approximately proportional to $e^{-\gamma u}$). For $T \ll u/\kappa'(\gamma)$, we determine the approximate form of $\mathbb{P}(V_T > u)$, and for $T \gg u/\kappa'(\gamma)$, we evaluate the difference $\mathbb{P}(V > u) - \mathbb{P}(V_T > u)$. Further results give a central limit estimate of $\mathbb{P}(V_T > u)$ when T is only moderately different from $u/\kappa'(\gamma)$, and an estimate of the rate of convergence $\mathbb{P}(V_T > u) \to \mathbb{P}(V > u)$ when u is fixed and only $T \to \infty$.

Whereas most results of the paper are inequalities or approximations, Section 8 contains a variety of exact results. In particular, we find the Laplace transform $\mathbb{E}_i e^{-\theta P_i}$ of the busy period and the related time $\tau_-(u)$ the system needs to empty from a large level u. However, the expressions involve a functional inversion and may appear too complicated to be useful for computational purposes (in fact, it does not seem not straightforward just to differentiate to derive the mean of P_i or $\tau_-(u)$). Nevertheless, $\mathbb{E}\tau_-(u)$ can be evaluated exactly.

Section 2 gives the preliminaries and a summary of the most relevant result from the

literature. In particular, some basic matrices occurring in the steady-state solution are introduced; they are of basic importance in the present paper as well, since the computational evaluation of the busy period/transient behavior results turns out to require either just these matrices, or matrices of just the same form but defined via duality in terms of time reversion, sign reversion or change of parameters. In Section 4, we introduce the basic technique used in most of the paper, change of measure via exponential families. In fact, some of the results show that this is not only a convenient mathematical tool but that the process in certain situations will behave precisely as if the parameters were changed in this way. In particular, Section 6 gives a precise description of this type of process behavior prior to exceedance of a large level in a busy cycle, a result which also determines the optimal change of measure in rare events simulation.

The results of the paper are exemplified via a simple two-state model in Section 9; this example may be read before the body of the paper to get a first impression of the flavor of the results. The Appendix contains two proofs deferred to there.

We finally mention that, though not developed in detail, most of the analysis of the present paper carries over to fluid models with Brownian noise which have received some recent attention, see in particular Gaver & Lehoczky [20], Kennedy & Williams [25], Asmussen [8], Rogers [35] and Karandikar & Kulkarni [28]. This means that on intervals where $J_t = i$, $\{S_t\}$ evolves as a Brownian motion with drift r_i and variance constant σ_i^2 depending on i. In some cases, the formulations have, however, to be slightly changed. In particular, the above definition of a busy period becomes trivial ($P_i \equiv 0$), so that instead one has to start the busy period at $x > 0$.

2. Preliminaries

As in [8], we represent $\{V_t\}$ as the reflected version

$$V_t = S_t - \min_{0 \le v \le t} S_v \tag{2.1}$$

of the net input process

$$S_t = \int_0^t r_{J_v} dv.$$

In particular, this means that $\{S_t\}$ is a continuous Markov additive process defined on an irreducible Markov jump process $\{J_t\}$ with a finite state space E (see e.g. Çinlar [16]).

An illustration of the connection between $\{V_t\}$ and $\{S_t\}$ is given in Figure 2.1. This figure shows also another fundamental tool of the paper (as well as of [8] and papers like [11], [35], [25]), two Markov processes

$$\left\{ J_{\tau_-(x)} \right\}_{x \ge 0}, \left\{ J_{\tau_+(x)} \right\}_{x \ge 0} \tag{2.2}$$

which are obtained by observing $\{J_t\}$ when $\{S_t\}$ is at a minimum or maximum. Here

$$\tau_+(x) = \inf\{t > 0 : S_t = x\}, \ \tau_-(x) = \inf\{t > 0 : S_t = -x\}$$

are the first passage times to levels $x > 0$, resp. $-x < 0$.

On Figure 2.1, $E_+ = \{\text{spade, heart}\}$, $E_- = \{\text{diamond, club}\}$. The slopes r_i are ordered the natural way such that

$$r_\spadesuit > r_\heartsuit > 0 > r_\diamond > r_\clubsuit.$$
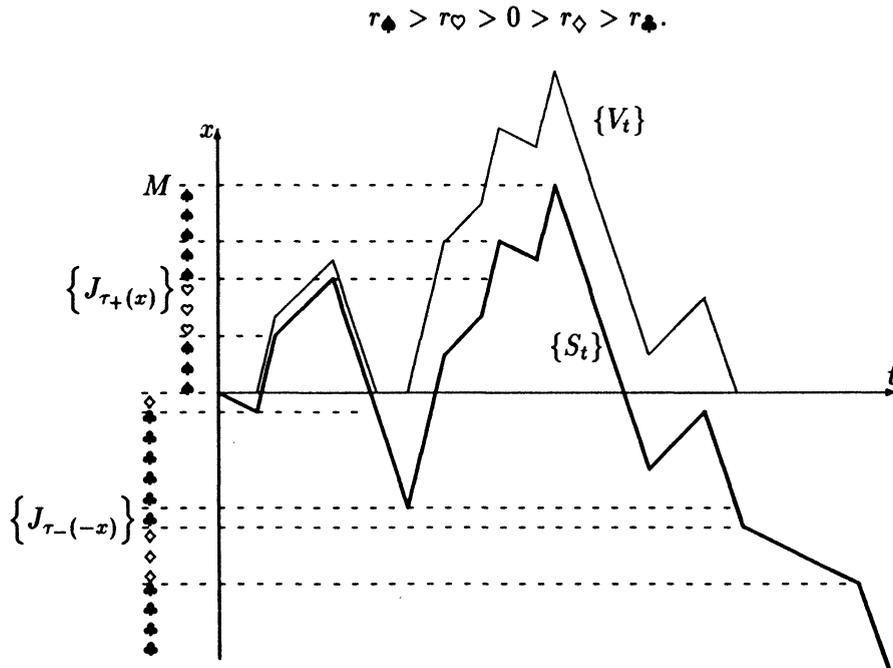
Figure 2.1

Let $\Lambda = (\lambda_{ij})_{i,j \in E}$ denote the intensity matrix for $\{J_t\}$ and write for brevity $\lambda_i = -\lambda_{ii}$ for the rate parameter of the exponential holding time in state i. We let T denote the matrix with ijth element $\lambda_{ij}/|r_i|$. Using the convention that for a given $E-$, E_+- or E_--vector $s = (s_i)$, Δ_s denotes the diagonal matrix with the s_i on the diagonal, we can write $T = \Delta_{|r|}^{-1} \Lambda$. We shall also use block-partitioned notation like

$$\Lambda = \begin{pmatrix} \Lambda^{(++)} & \Lambda^{(+-)} \\ \Lambda^{(-+)} & \Lambda^{(--)} \end{pmatrix}, \quad T = \begin{pmatrix} T^{(++)} & T^{(+-)} \\ T^{(-+)} & T^{(--)} \end{pmatrix}.$$

When we write say $\Delta_r^{-1}\Lambda^{(+-)}$, the convention is that dimensions should match. I.e., Δ_r is $E_+ \times E_+$ with the r_i, $i \in E_+$, on the diagonal. Similarly, the identity matrix I, the ith unit column vector e_i and the column vector e with all entries equal to 1 may have indices in E, E_+ or E_- depending on the context.

The intensity matrices for the Markov processes (2.2) are denoted $U^{(-)}: E_- \times E_-$, $U^{(+)}: E_+ \times E_+$, and we define matrices $\alpha^{(-+)}: E_- \times E_+$, $\alpha^{(+-)}: E_+ \times E_-$ by $\alpha_{ij}^{(-)} = \mathbb{P}_i(J_{\tau_+(0)} = j)$ etc. (it is trivial that $\mathbb{P}_i(J_{\tau_-(0)} = j) = 0$ if $j \in E_+$, that $\mathbb{P}_i(J_{\tau_-(0)} = j) = \delta_{ij}$ if $i,j \in E_-$ and similarly for $\mathbb{P}_i(J_{\tau_+(0)} = j)$). It is easy to see via an operational time argument ([8]) that, e.g.,

$$U^{(-)} = T^{(--)} + T^{(-+)}\alpha^{(+-)}, \tag{2.3}$$

$$\alpha^{(+-)} = \int_0^\infty e^{T^{(++)}y} T^{(+-)} e^{U^{(-)}y} dy \tag{2.4}$$

and similarly for $\boldsymbol{\alpha}^{(-+)}, \boldsymbol{U}^{(+)}$.

Algorithms for computing matrices like $\boldsymbol{\alpha}^{(+-)}, \boldsymbol{\alpha}^{(-+)}$ and thereby $\boldsymbol{U}^{(-)}, \boldsymbol{U}^{(+)}$ are discussed in [11], [8], [35] (this yields also the steady-state distribution in view of (2.10) below). Some are iterative, based upon functional equations provided by expressions like (2.3), (2.4), and others are based upon diagonalization ideas, delivering automatically matrices like $\boldsymbol{U}^{(-)}$ on diagonal form,

$$\boldsymbol{U}^{(-)} = \sum_{i \in E_-} s_i \boldsymbol{r}_i \boldsymbol{\ell}_i, \qquad (2.5)$$

where $\boldsymbol{\ell}_i \boldsymbol{U}^{(-)} = s_i \boldsymbol{\ell}_i, \boldsymbol{U}^{(-)} \boldsymbol{r}_i = s_i \boldsymbol{r}_i$. The numerical computation may be demanding, not least when the number of states in E is large (700-900 occurs in references like Anick et al. [2] or Stern & Elwalid [38]), but *from the point of view of the present paper, we shall consider this problem as settled and matrices of type* $\boldsymbol{\alpha}^{(+-)}, \boldsymbol{\alpha}^{(-+)}, \boldsymbol{U}^{(+)}, \boldsymbol{U}^{(-)}$ *as computable.*

For later reference, we quote also the Wiener-Hopf factorization identity ([11] or [35])

$$\Delta_r^{-1} \Lambda \begin{pmatrix} \boldsymbol{I} & \boldsymbol{\alpha}^{(+-)} \\ \boldsymbol{\alpha}^{(-+)} & \boldsymbol{I} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{\alpha}^{(+-)} \\ \boldsymbol{\alpha}^{(-+)} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{U}^{(+)} & 0 \\ 0 & \boldsymbol{U}^{(-)} \end{pmatrix}. \qquad (2.6)$$

We now introduce the time-reversed version $\{\tilde{J}_t, \tilde{S}_t\}$ of the Markov additive process $\{J_t, S_t\}$. We can write $\tilde{\Lambda}$ (i.e. the matrix with elements $\tilde{\lambda}_{ij} = \pi_j \lambda_{ji}/\pi_i$) as $\tilde{\Lambda} = \Delta_{1/\pi} \Lambda' \Delta_\pi$ (note that $\lambda_i = \tilde{\lambda}_i$ and $\pi_i = \tilde{\pi}_i$). Thus $\{\tilde{J}_t\}$ is defined in terms of $\tilde{\Lambda}$ rather than Λ, and $\{\tilde{S}_t\}$ is defined as $\{S_t\}$, with the same rates r_i but $\{J_t\}$ replaced by $\{\tilde{J}_t\}$. Further let $M_{\tilde{S}}(t) = sup_{0 \leq u \leq t} \tilde{S}_u$, $M_{\tilde{S}} = sup_{t \geq 0} \tilde{S}_t$.

Proposition 2.1 ([8]):

$$\mathbb{P}_j(J_t = i, V_t \in A) = \frac{\pi_i}{\pi_j} \mathbb{P}_i(\tilde{J}_t = j, M_{\tilde{S}}(t) \in A),$$

$$\mathbb{P}(V \in A, J = i) = \pi_i \mathbb{P}_i(M_{\tilde{S}} \in A).$$

Notation like

$$\tilde{\tau}_+(x), \tilde{\tau}_+(x), \left\{ \tilde{J}_{\tilde{\tau}_-(x)} \right\}, \left\{ \tilde{J}_{\tilde{\tau}_+(x)} \right\}, \tilde{U}^{(-)}, \tilde{U}^{(+)}, \tilde{\alpha}^{(-+)}, \tilde{\alpha}^{(+-)}$$

refers in an obvious way to $\{\tilde{S}_t\}$. In particular, since clearly $\{M_{\tilde{S}}(t) > x\} = \{\tilde{\tau}_+(x) \leq t\}$, Proposition 2.1 yields

$$\mathbb{P}_j(J_t = i, V_t > x) = \frac{\pi_i}{\pi_j} \mathbb{P}_i(\tilde{J}_t = j, \tilde{\tau}_+(x) \leq t), \qquad (2.7)$$

$$\mathbb{P}(V > x, J = i) = \pi_i \mathbb{P}_i(\tilde{\tau}_+(x) < \infty). \qquad (2.8)$$

Recall that a distribution F on $[0, \infty)$ is phase-type with phase generator U and initial vector $\boldsymbol{\alpha}$ if F is the distribution of the lifetime of a Markov process which has initial distribution $\boldsymbol{\alpha}$ and intensity matrix U, cf. [33]; if the mass $\boldsymbol{\alpha} e$ of $\boldsymbol{\alpha}$ is less than one, we adapt the convention that this corresponds to an atom of size $1 - \boldsymbol{\alpha} e$ in 0. From the above discussion, it follows immediately that (as shown in [8]) the distribution of the steady-state variable (J, V) is phase-type given $J = i$, with phase generator $\tilde{U}^{(+)}$ and initial vector $\tilde{\alpha}_i^{(-+)}$ for $i \in E_-$ and e_i' for $i \in E_+$. More precisely, for $i \in E_-$

$$\mathbb{P}(V > x, J = i) = \pi_i \cdot \widetilde{\alpha}_{i.}^{(+ -)} e^{\widetilde{U}^{(+)}x} e, \ x > 0, \tag{2.9}$$

$$\mathbb{P}(V = 0, J = i) = \pi_i (1 - \widetilde{\alpha}_{i.}^{(+ -)} e); \tag{2.10}$$

for $i \in E_+$, just replace $\widetilde{\alpha}_i^{(+ -)}$ by e'_i.

3. The Mean Busy Period

Let $\bar{P}$ be the matrix with ijth entry $\bar{p}_i = \mathbb{E}_i[P_i; J_{P_i} = j]$; then $\bar{P}e$ is the vector with ith entry $\mathbb{E}_i P_i$. We shall show that once $\alpha^{(+ -)}$ has been evaluated, the entries of $\bar{P}$ and $\bar{P}e$ can be computed as the solution of linear equations. We start with the case of $\bar{P}e$, which may be worthwhile treating separately because we get matrices of lower dimension than as for $\bar{P}$.

Theorem 3.1:

$$\bar{P}e = -(I + T^{(+ +)^{-1}}\alpha^{(+ -)}\Delta_{|r|}^{-1}\Lambda^{(- +)})^{-1}(\Lambda^{(+ +)^{-1}}e + T^{(+ +)^{-1}}\alpha^{(+ -)}\Delta_{|r|}^{-1}e).$$

Proof: We use a decomposition of the path $\{S_t\}_{0 \geq t \leq P_i}$ as indicated in Figure 3.1.
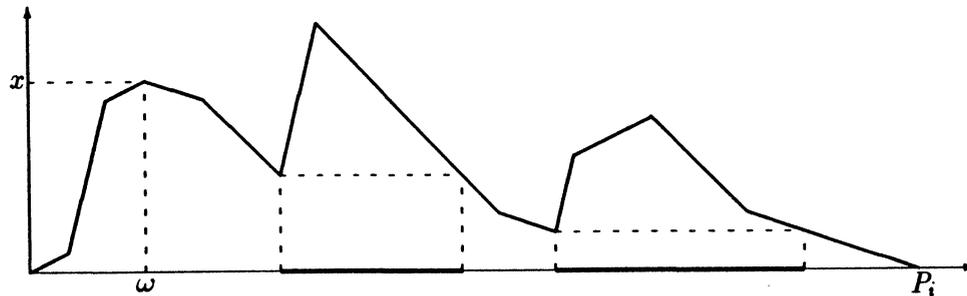
Figure 3.1

Here $\omega = inf\{t > 0: J_t \in E_-\}$, so that ω is phase-type with representation $(\Lambda^{(+ +)}, e'_i)$ w.r.t. $\mathbb{P}_i$ and

$$\mathbb{E}_i \omega = -e'_i \Lambda^{(+ +)^{-1}} e.$$

Similarly, an operational time argument ([8]) shows that S_ω is phase-type with representation $(T^{(+ +)}, e'_i)$, and that

$$\mathbb{P}_i(S_\omega \in dx, J_\omega = j) = e'_i e^{T^{(+ +)}x} T^{(+ -)} e_j dx \tag{3.1}$$

(note incidentally that this easily leads to (2.4)).

The post-ω path can be split up into two types of intervals, the first being intervals where $\{S_t\}_{t > 0}$ is a relative minimum and the second being sub-busy cycles (two on Figure 3.1; marked by bold lines on the time axis). Let the total lengths of intervals of the two types be ω_1, ω_2. If $J_\omega = j, S_\omega = x$, the values of $\{J_t\}$ observed on the ω_1-segment are distributed as

$$\left\{ J_{\tau_-}(x-y) \right\}_{0 \le y \le x}$$

starting from $J_{\tau_-}(0) = j$. Hence, the expected time in state $k \in E_-$ on the ω_1-segment is

$$\mathbb{E}\omega_{1;\,j,\,k}(x) = \int_0^x e_j' e^{U^{(-)}y} e_k \cdot \frac{1}{|r_k|} dy.$$

In particular, using (3.1) and (2.4), we get

$$
\begin{aligned}
\mathbb{E}_i \omega_{1;\,J_\omega,\,k}(S_\omega) &= \int_0^\infty e_i' e^{T^{(++)}x} T^{(+-)} dx \int_0^x e^{U^{(-)}y} e_k \cdot \frac{1}{|r_k|} dy \\[2mm]
&= -\int_0^\infty e_i' T^{(++)^{-1}} e^{T^{(++)}y} T^{(+-)} e^{U^{(-)}y} e_k \cdot \frac{1}{|r_k|} dy \\[2mm]
&= -e_i' T^{(++)^{-1}} \alpha^{(+-)} e_k \cdot \frac{1}{|r_k|}.
\end{aligned}
$$

Summing over k, we get

$$\mathbb{E}_i \omega_1 = -e_i' T^{(++)^{-1}} \alpha^{(+-)} \Delta_{|r|}^{-1} e.$$

Now when $J_t = k$ on the ω_1-segment, a sub-busy period of type $\ell \in E_+$ occurs at rate $\lambda_{k\ell}$. Hence

$$
\begin{aligned}
\mathbb{E}_i \omega_2 &= \sum_{k \in E_-} \sum_{\ell \in E_+} \mathbb{E}_i \omega_{1;\,J_\omega,\,k}(S_\omega) \lambda_{k\ell} \mathbb{E}_\ell P_\ell \\[2mm]
&= -e_i' T^{(++)^{-1}} \alpha^{(+-)} \Delta_{|r|}^{-1} \Lambda^{(-+)} \overline{P} e.
\end{aligned}
$$

Noting that $\mathbb{E}_i P_i = \mathbb{E}_i(\omega + \omega_1 + \omega_2)$, collecting terms and rewriting in matrix notation, the result follows by easy algebra. □

Remark 3.2: In [9], a somewhat similar argument is carried out in branching process language. As was kindly pointed out by Dr. S. Grishechkin, the process in question is not a branching process in the strict sense (some of the required independencies fail). However, the argument for expected values is correct. □

Now consider the more general case of $\overline{P}$.

Theorem 3.3:

$$
\begin{aligned}
0 &= \Delta_r^{-1} \Lambda^{(++)} \overline{P} + \Delta_r^{-1} \alpha^{(+-)} + \overline{P} \Delta_{|r|}^{-1} \Lambda^{(--)} + \overline{P} \Delta_{|r|}^{-1} \Lambda^{(-+)} \alpha^{(+-)} \\[2mm]
&\quad + \alpha^{(+-)} \Delta_{|r|}^{-1} \Lambda^{(-+)} \overline{P} + \alpha^{(+-)} \Delta_{|r|}^{-1}.
\end{aligned}
$$

Proof: We distinguish between the possibilities that $\{J_t\}$ has a state transition in $[0, dt/r_i)$, to k (say), or not. If $k \in E_-$ in the first case, the busy period will terminate within time $O(dt)$, so that the contribution from this is $O((dt)^2) = 0$. If $k \in E_+$, a sub-busy period starts from k

and coincides with P_i up to $O(dt)$ terms. Thus, the total contribution to $\bar{p}_{ij}$ from state transitions in $[0, dt/r_i)$ is

$$A_1 = \sum_{k \in E_+, k \neq i} \frac{\lambda_{ik}}{r_i} dt \bar{p}_{kj}.$$

In the second case, there are three contributions: the one A_2 from the initial segment; the one A_3 from the sub-busy period starting from i at time dt/r_i in level $V_{dt/r_i} = dt$ and ending at the next downcrossing of level dt; and the one A_r from the final segment after this downcrossing. The length of the initial segment is dt/r_i, and up to $O((dt)^2)$ terms, it provides a contribution if the sub-busy period ends in state j; thus $A_2 = dt/r_i \cdot \alpha_{ij}^{(+ -)}$. Let $k \in E_-$ denote the state in which level dt is downcrossed by the sub-busy period. If $k \neq j$, a contribution to A_3 can occur in two ways, either by a transition to j before time $dt/|r_k|$ or by a jump to some $\ell \in E_+$, in which case the following (second) sub-busy period must terminate in state j which occurs w.p. $\alpha_{\ell j}^{(+ -)}$. This second possibility also occurs if $k = j$, but then there is in addition a contribution from the event that no transition out of j occurs before time $dt/|r_j|$ after the downcrossing which occurs w.p. $1 - \lambda_j dt/|r_j|$. Thus

$$A_3 = \sum_{k \in E_-, k \neq j} \bar{p}_{ik} \frac{\lambda_{kj}}{|r_k|} dt + \sum_{k \in E_-, \ell \in E_+} \bar{p}_{ik} \frac{\lambda_{k\ell}}{|r_k|} dt \alpha_{\ell j}^{(+ -)} + \bar{p}_{ij}\left(1 - \frac{\lambda_j}{|r_j|} dt\right)$$

$$= \sum_{k \in E_-} \bar{p}_{ik} \frac{\lambda_{kj}}{|r_k|} dt + \sum_{k \in E_-, \ell \in E_+} \bar{p}_{ik} \frac{\lambda_{k\ell}}{|r_k|} dt \alpha_{\ell j}^{(+ -)} + \bar{p}_{ij}.$$

Finally, decomposing A_4 as a contribution from a second sub-busy period and a passage to level 0 without state transitions yields

$$A_4 = \sum_{k \in E_-, \ell \in E_+} \alpha_{ik}^{(+ -)} \frac{\lambda_{k\ell}}{|r_k|} dt \bar{p}_{\ell j} + \alpha_{ij}^{(+ -)} \frac{1}{|r_j|} dt.$$

Writing

$$\bar{p}_{ij} = A_1 + \left(1 - \frac{\lambda_i}{r_i} dt\right)(A_2 + A_3 + A_4),$$

subtracting $\bar{p}_{ij}$ from both sides and dividing by dt, we get

$$0 = \sum_{k \in E_+, k \neq i} \frac{\lambda_{ik}}{r_i} \bar{p}_{kj} + \frac{\alpha_{ij}^{(+ -)}}{r_i} + \sum_{k \in E_-} \bar{p}_{ik} \frac{\lambda_{kj}}{|r_k|}$$

$$+ \sum_{k \in E_-, \ell \in E_+} \bar{p}_{ik} \frac{\lambda_{k\ell}}{|r_k|} \alpha_{\ell j}^{(+ -)} - \bar{p}_{ij} \frac{\lambda_i}{r_i}$$

$$+ \sum_{k \in E_-, \ell \in E_+} \alpha_{ik}^{(+ -)} \frac{\lambda_{k\ell}}{|r_k|} \bar{p}_{\ell j} + \alpha_{ij}^{(+ -)} \frac{1}{|r_j|}.$$

Absorbing the fifth term into the first sum as the $k = i$ term and rewriting in matrix notation, the result follows.   □

Note that the matrix identity in Theorem 3.3 is of dimension $E_+ \times E_-$ and depends linearly on the elements $\bar{p}_{ij}$ of $\bar{P}$, so that indeed we have $E_+ \times E_-$ unknowns and as many linear equations.

Define the busy cycle C_i starting from $J_0 = i$, $V_0 = S_0 = 0$ as

$$C_i = inf\{t > P_i : S_t > 0\} = inf\{t > P_i : J_t \in E_+\}.$$

Proposition 3.4:

$$\mathbb{E}_i C_i = \mathbb{E}_i P_i - e_i' \alpha_{i.}^{(+)} \Lambda^{(--)^{-1}} e.$$

Proof: Obviously, the idle period $C_i - P_i$ is phase-type with phase generator $\Lambda^{(--)}$. The initial vector is the distribution of J_{P_i}, which is just $\alpha_{i.}^{(+)}$. Thus, the results follow from general formulas for the mean of phase-type distributions. □

The busy cycles C_i are not regenerative for $\{(V_t, J_t)\}$ but semi-regenerative. A proper regenerative cycle C_i^* is defined by fixing $i \in E_+$ and adding up cycles until a second cycle of type i occurs. That is,

$$C_i^* = inf\{t > 0 : V_t = 0, J_t = i\}.$$

Proposition 3.5: *For* $i \in E_+$,

$$\mathbb{E}_i C_i^* = \frac{1}{\eta_i} \sum_{j \in E_+} \eta_j \mathbb{E}_j C_j, \tag{3.2}$$

where

$$\eta_j = \frac{\sum_{i \in E_-} \pi_i (1 - \tilde{\alpha}_{i.}^{(+-)} e) \lambda_{ij}}{\sum_{i \in E_-, k \in E_+} \pi_i (1 - \tilde{\alpha}_{i.}^{(+-)} e) \lambda_{ik}}. \tag{3.3}$$

Proof: Consider the E_+-valued discrete-time Markov chain obtained by observing $\{J_t\}$ just after the beginning of busy cycles, and let $\boldsymbol{\eta} = (\eta_j)_{j \in E_+}$ denote its stationary distribution. Then (3.2) holds by general results on semi-regenerative process ([5], p. 228), so we only have to verify the asserted expression for η_j. But consider a large time interval $[0, T]$. Conditioning upon the state $i \in E_-$ of $\{J_t\}$ just before a busy cycle starts from $j \in E_+$ shows that the expected number of such cycles is

$$\sum_{i \in E_-} \int_0^T \mathbb{P}(V_t = 0, J_t = i) \lambda_{ij} dt \approx T \sum_{i \in E_-} \pi_i (1 - \tilde{\alpha}_{i.}^{(+-)} e) \lambda_{ij},$$

cf. (2.1). Hence the proportion of busy cycles starting from j among all busy cycles is approximately given by (3.3), and from this the result follows by letting $T \to \infty$ (the argument is essentially "conditional PASTA", cf. [18]). □

4. Change of Measure via Exponential Families

4.1 Moments and Cumulants

We first introduce a suitable matrix generalization of the m.g.f. Define F_t as the measure-valued matrix with ijth entry $F_t[i, j; x] = \mathbb{P}_i[S_t \le x; J_t = j]$, and $\widehat{F}_t[s]$ as the matrix with ijth entry $\widehat{F}_t[i, j; s] = \mathbb{E}_i[e^{sS_t}; J_t = j]$ (thus, $\widehat{F}[s]$ may be viewed as the matrix m.g.f. of F_t defined by entrywise integration). Let further $K[s] = \Lambda + s\Delta_r$.

Proposition 4.1 ([8]): $\widehat{F}_t[s] = e^{tK[s]}$.

Since obviously $\widehat{F}_t[s]$ is strictly positive (and defined for all real s), it follows that $K[s]$ has a simple and unique eigenvalue $\kappa(s)$ with maximal real part, such that the corresponding left and right eigenvectors $\boldsymbol{\nu}^{(s)}$, $\boldsymbol{h}^{(s)}$ may be taken with strictly positive components. We shall use the normalization $\boldsymbol{\nu}^{(s)}\boldsymbol{e} = \boldsymbol{\nu}^{(s)}\boldsymbol{h}^{(s)} = 1$. Note that since $K[0] = \Lambda$, we have $\boldsymbol{\nu}^{(0)} = \boldsymbol{\pi}$, $\boldsymbol{h}^{(0)} = \boldsymbol{e}$.

The following result, which is proved in the Appendix, shows that the function $\kappa(s)$ plays the role of an appropriate generalization of the cumulant g.f. as well as it shows how to compute the asymptotic mean and variance directly from the model parameters. Its origin is results for discrete-time Markov additive processes obtained by Keilson & Wishart [29], [30]; similar results for Markov-modulated $M/G/1$ queues are in Asmussen [6].

Theorem 4.2: *The function $\kappa(s)$ is strictly convex with*

$$\lim_{s \to \infty} \kappa'(s) = \max_{i \in E_+} r_i, \quad \lim_{s \to -\infty} \kappa'(s) = \min_{i \in E_-} r_i.$$

Furthermore

$$\kappa'(0) = \lim_{t \to \infty} \frac{\mathbb{E}_i S_t}{t} = \boldsymbol{\pi}\Delta_r \boldsymbol{e} = \sum_{i \in E} \pi_i r_i, \tag{4.1}$$

$$\kappa''(0) = \lim_{t \to \infty} \frac{Var_i S_t}{t} = -2\kappa'(0)^2 - 2\boldsymbol{\pi}\Delta_r D \Delta_r \boldsymbol{e}, \tag{4.2}$$

where D is the matrix $(\lambda - \boldsymbol{e}\boldsymbol{\pi})^{-1}$.

The function $\kappa(s)$ is finite for all θ, has $\kappa'(0) < 0$ (cf. (1.1), (4.1)) and converges to ∞ as $s{\uparrow}\infty$. In particular, a $\gamma > 0$ with $\kappa(\gamma) = 0$, a $\gamma_0 > 0$ with $\kappa'(\gamma_0) = 0$ and (for $y > 1/ \max_{i \in E_+} r_i$) an $\alpha_y > \gamma_0$ with $\kappa'(\alpha_y) = \frac{1}{y}$ exist, see Figure 4.1. Since γ plays a special role, we write $\boldsymbol{h} = \boldsymbol{h}^{(\gamma)}$.
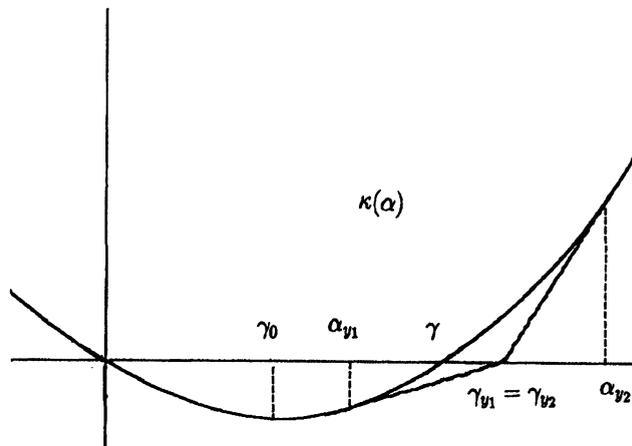
Figure 4.1

Proposition 4.3: *Let θ be fixed. Then for any $i,t, \mathbb{E}_i e^{\theta S_t} h_{J_t}^{(\theta)} = e^{t\kappa(\theta)} h_i^{(\theta)}$. In particular, $\{e^{\theta S_t - t\kappa(\theta)} h_{J_t}^{(\theta)}\}_{t \geq 0}$ is a martingale.*

Proof: For the first assertion, just note that

$$\mathbb{E}_i e^{\theta S_t} h^{(\theta)}_{J_t} = e'_i \widehat{F}_t[\theta] \; h^{(\theta)} = e'_i e^{tK[\theta]} h^{(\theta)} = e'_i e^{t\kappa(\theta)} h^{(\theta)} = e^{t\kappa(\theta)} h^{(\theta)}_i.$$

Letting $\mathfrak{F}_t = \sigma(J_v, S_t : 0 \le v \le t)$, it then follows that

$$\mathbb{E}[e^{\theta S_{t+v} - (t+v)\kappa(\theta)} h^{(\theta)}_{J_{t+v}} \mid \mathfrak{F}_t] = e^{\theta S_t - t\kappa(\theta)} \mathbb{E}[e^{\theta(S_{t+v} - S_t) - v\kappa(\theta)} h^{(\theta)}_{J_{t+v}} \mid \mathfrak{F}_t]$$

$$= e^{\theta S_t - t\kappa(\theta)} \mathbb{E}_{J_t}[e^{\theta S_v - v\kappa(\theta)} h^{(\theta)}_{J_v}] = e^{\theta S_t - t\kappa(\theta)} h^{(\theta)}_{J_t}.$$

$\square$

It follows from the below results (e.g. Theorem 5.1) that a large value of h_i can be interpreted as i being a state such that starting from $J_0 = i$, $\{V_t\}$ grows rapidly in its initial phase (before $\{J_t\}$ reaches equilibrium).

For the time-reversed process $\{(\widetilde{S}_t, \widetilde{J}_t)\}, \widetilde{K} = \Delta_{\boldsymbol{\pi}}^{-1} K' \Delta_{\boldsymbol{\pi}}$. From this it follows easily that $\widetilde{\kappa} = \kappa$ (in particular, γ, γ_0, etc. remain the same), whereas $\widetilde{\boldsymbol{\nu}} = h' \Delta_{\boldsymbol{\pi}}, \widetilde{h} = \Delta_{\boldsymbol{\pi}}^{-1} \boldsymbol{\nu}'$. A large value of $\widetilde{h}_i$ can be interpreted as i being a state for which $J_t = i$ is likely to occur for large values of V_t, cf. e.g. Corollary 4.7 and Theorem 7.1 below.

4.2 Likelihood Ratio Identities

We now turn to the construction of an exponential family of fluid models, such that the θ member has a changed intensity matrix $\boldsymbol{\Lambda}(\theta) = (\lambda^{(\theta)}_{ij})_{i,j \in E}$, but is otherwise unchanged (in particular, the r_i are the same), and that the case $\theta = 0$ corresponds to the given process, i.e. $\boldsymbol{\Lambda}(0) = \boldsymbol{\Lambda}$.

The relevant choice turns out to be

$$\boldsymbol{\Lambda}(\theta) = \Delta_{h(\theta)}^{-1} K[\theta] \Delta_{h(\theta)} - \kappa(\theta) I = \Delta_{h(\theta)}^{-1} \boldsymbol{\Lambda} \Delta_{h(\theta)} + \theta \Delta_r - \kappa(\theta) I.$$

That is,

$$\lambda^{(\theta)}_{ij} = \begin{cases} \lambda_{ij} \dfrac{h^{(\theta)}_j}{h^{(\theta)}_i} & i \ne j \\[2ex] \lambda_{ii} + \theta r_i - \kappa(\theta) & i = j \end{cases} .$$

Proposition 4.4: $\boldsymbol{\Lambda}(\theta)$ *is an intensity matrix. The stationary distribution* $\boldsymbol{\pi}^{(\theta)}$ *is given by* $\pi^{(\theta)}_i = \nu^{(\theta)}_i h^{(\theta)}_i$, *i.e.* $\boldsymbol{\pi}^{(\theta)} = \boldsymbol{\nu}^{(\theta)} \Delta_{h(\theta)}$.

Proof: Since the off-diagonal elements are non-negative, it suffices to verify $\boldsymbol{\Lambda}(\theta) e = 0$. But

$$\boldsymbol{\Lambda}(\theta) e = \Delta_{h(\theta)}^{-1}(K[\theta] - \kappa(\theta)I) \Delta_{h(\theta)} e$$

$$= \Delta_{h(\theta)}^{-1}(K[\theta] - \kappa(\theta)I) h^{(\theta)}$$

$$= \Delta_{h(\theta)}^{-1}(\kappa(\theta)I - \kappa(\theta)I) h^{(\theta)} = 0.$$

Similarly, the components of $\boldsymbol{\nu}^{(\theta)} \Delta_{h(\theta)}$ are non-negative, sum to one in view of $\boldsymbol{\nu}^{(\theta)} h^{(\theta)} = 1$, and we have

$$\nu^{(\theta)}\Delta_{\pmb{h}^{(\theta)}}\Lambda(\theta) = \nu^{(\theta)}\Delta_{\pmb{h}^{(\theta)}}\Delta^{-1}_{\pmb{h}^{(\theta)}}(K[\theta] - \kappa(\theta)I)\Delta_{\pmb{h}^{(\theta)}}$$

$$= \nu^{(\theta)}(K[\theta] - \kappa(\theta)I)$$

$$= (\kappa(\theta)I - \kappa(\theta)I)\nu^{(\theta)} = 0. \qquad\qquad \square$$

The idea behind the likelihood ratio method is basically to change the mean drift

$$\lim_{t\to\infty}\frac{S_t}{t} = \sum_{i\,\in\,E}\pi_i r_i = \kappa'(0)$$

of $\{S_t\}$ from negative to positive values, thereby giving rare events like $\{\tau_+(u)\}$ $\mathbb{P}_{\theta,i}$-probability one. The following result shows that this is attained for $\theta \geq \gamma_0$. Let $\kappa_\theta(\alpha)$ denote the cumulant g.f. for the $\mathbb{P}_{\theta,i}$-process.

Proposition 4.5:   (a) $\kappa_\theta(\alpha) = \kappa(\alpha + \theta) - \kappa(\theta)$;

                   (b) $S_t \to \kappa'(\theta)\mathbb{P}_{\theta,i}$-a.s.

Proof: Obviously,

$$(\mathbb{E}_{\theta,i}[e^{\alpha S_T}; J_T = j])_{i,j\,\in\,E} = exp\{T(\Lambda(\theta) + \alpha\Delta_r)\}$$

$$= \Delta^{-1}_{\pmb{h}^{(\theta)}}exp\{T(\Lambda + (\alpha + \theta)\Delta_r - \kappa(\theta)I)\}\Delta_{\pmb{h}^{(\theta)}}$$

$$= e^{-T\kappa(\theta)}\Delta^{-1}_{\pmb{h}^{(\theta)}}\widehat{F}[\alpha + \theta]\Delta_{\pmb{h}^{(\theta)}},$$

from which (a) follows. Differentiating w.r.t. α and letting $\alpha = 0$ yields (b). $\square$

Now let $\mathbb{P}_{\theta,i}$ be the governing probability measure for the fluid model which is governed by $\Lambda(\theta)$ and the r_i and has initial environment $J_0 = i$. In the Appendix, we show the following likelihood ratio identity, which is our fundamental tool in the following. Its origin is results for discrete-time Markov additive processes obtained by Bellman [12], Tweedie [41] and Miller [32]; a similar identity for Markov-modulated $M/G/1$ queues is exploited in Asmussen [6] (cf. also Asmussen & Rolski [10]). For a survey of the likelihood ratio method for simple queues and random walks, see [5] Ch. XII.

Proposition 4.6: *Let τ be any stopping time and let $G \in \mathcal{F}_\tau, G \subseteq \{\tau < \infty\}$. Then*

$$\mathbb{P}_i G = \mathbb{P}_{0;i} G = h_i^{(\theta)}\mathbb{E}_{\theta,i}\left[\frac{1}{h^{(\theta)}_{J_\tau}}exp\{-\theta S_\tau + \tau\kappa(\theta)\}; G\right]. \qquad (4.3)$$

Here is a first quick application of the likelihood ratio method. For each θ, define

$$C_-(\theta) = \frac{1}{max_{j\,\in\,E_+}h_j^{(\theta)}}, \quad C_+(\theta) = \frac{1}{min_{j\,\in\,E_+}h_j^{(\theta)}}$$

and similarly for $\widetilde{C}_+(\theta), \widetilde{C}_+(\theta)$. Then:

Corollary 4.7: $\pi_i\widetilde{h}_i^{(\gamma)}\widetilde{C}_+(\gamma)e^{-\gamma u} \leq \mathbb{P}(V > u, J = i) \leq \pi_i\widetilde{h}_i^{(\gamma)}\widetilde{C}_-(\gamma)e^{-\gamma u}.$

Proof: Applying (4.3) (in its time-reversed version) with $\tau = \widetilde{\tau}_+(u)$, $G = \{\widetilde{\tau}_+(u) < \infty\}$, and noting that $\widetilde{S}_{\widetilde{\tau}_+(u)} = u$ by the skip-free property, (2.9) yields

$$\mathbb{P}(V > u, J = i) = \pi_i\mathbb{P}_i(\widetilde{\tau}_+(u) < \infty) = \pi_i\widetilde{h}_i^{(\theta)}e^{-\theta u}\mathbb{E}_{\theta,i}\frac{e^{\widetilde{\tau}_+(u)\kappa(\theta)}}{\widetilde{h}^{(\theta)}_{J_{\widetilde{\tau}_+(u)}}}. \qquad (4.4)$$

Letting $\theta = \gamma$ (so that $\kappa(\theta) = 0$) yields

$$\mathbb{P}(V > u, J = i) = \pi_i \widetilde{h}_i e^{-\gamma u} \mathbb{E}_{\gamma;i} \frac{1}{\widetilde{h}_{J_{\widetilde{\tau}_+(u)}}}. \tag{4.5}$$

From this the corollary follows by trivial estimates. □

Compared to the exact solutions in the literature, the advantage of Corollary 4.7 is of course that less computations are required. For example, we can compute $\kappa(\theta)$ by Elsner's algorithm ([33]) which automatically gives us also $h^{(\theta)}$. In queueing theory, corresponding inequalities for the $GI/G/1$ queue have been derived by Kingman and Ross (see, e.g., the survey in Stoyan [37]).

In fact, the argument in the proof of Corollary 4.7 can be strengthened to show that $\mathbb{P}(V > u, J = i)$ is asymptotically exponential. This fact follows of course from the phase-type form of $\mathbb{P}(V > u, J = i)$ (see [33]), but we shall give the result anyway since the proof is short, given some auxiliary results (that are needed below for other purposes) and since the form of the constants which come out in this way is more suitable for comparison with other results of the paper.

We first need to introduce the matrix $U^{(+)}(\theta)$, defined as $U^{(+)}$ but with the $\mathbb{P}_i$ replaced by the $\mathbb{P}_{\theta;i}$ (similar notation like $\alpha^{(+-)}(\theta)$ etc. is used in the following). When $\theta \geq \gamma_0, U^{(+)}(\theta)$ is a proper intensity matrix and thus has a unique stationary distribution $\zeta^{(+)}(\theta)$.

Lemma 4.8: (a) $U^{(+)}(\gamma) = \Delta_h^{-1} U^{(+)} \Delta_h + \gamma I$, $\alpha^{(+-)}(\gamma) = \Delta_h^{-1} \alpha^{(+-)} \Delta_h$;

(b)
$$\zeta^{(+)}(\gamma) = \pi^{(\gamma)} \Delta_r \begin{pmatrix} I \\ \alpha^{(-+)} \end{pmatrix} / \pi^{(\gamma)} \Delta_r e. \tag{4.6}$$

Proof:

$$e_i' e^{U^{(+)}(\gamma)u} e_j = \mathbb{P}_{\gamma;i}(J_{\tau_+(u)} = j)$$

$$= \frac{h_j}{h_i} \mathbb{E}_i [e^{\gamma S_{\tau_+(u)}}; J_{\tau_+(u)} = j]$$

$$= \frac{h_j}{h_i} e^{\gamma u} e_i' e^{(U^{(+)} + \gamma I)u} e_j,$$

which implies the asserted expression for $U^{(+)}(\gamma)$. Similarly

$$\alpha_{ij}^{(+-)}(\gamma) = \mathbb{P}_{\gamma;i}(J_{\tau_-(0)} = j) = \frac{h_j}{h_i} \mathbb{P}_i(J_{\tau_-(0)} = j) = \frac{h_j}{h_i} \alpha_{ij}^{(+-)}.$$

From the Wiener-Hopf identity (2.6) and $\pi^{(\gamma)} \Lambda(\gamma) = 0$, we obtain

$$0 = \pi^{(\gamma)} \Delta_r \Delta_r^{-1} \Lambda(\gamma) \begin{pmatrix} I & \alpha^{(+-)}(\gamma) \\ \alpha^{(-+)}(\gamma) & I \end{pmatrix}$$

$$= \pi^{(\gamma)} \Delta_r \begin{pmatrix} I & \alpha^{(+-)}(\gamma) \\ \alpha^{(-+)}(\gamma) & I \end{pmatrix} \begin{pmatrix} U^{(+)}(\gamma) & 0 \\ 0 & U^{(-)}(\gamma) \end{pmatrix}.$$

From this it follows that the r.h.s. of (4.6) is indeed a left eigenvector of $U^{(+)}(\gamma)$ corresponding

to the eigenvalue 0. Thus the result follows from

$$\pi^{(\gamma)}\Delta_r\begin{pmatrix} I \\ \alpha^{(-+)} \end{pmatrix}e = \pi^{(\gamma)}\Delta_r\begin{pmatrix} e \\ e \end{pmatrix} = \pi^{(\gamma)}\Delta_r e$$

(using that $\alpha^{(-+)}e = e$). □

Corollary 4.9: $\mathbb{P}(V > u, J = 1) \approx \pi_i\tilde{h}_i\tilde{D}e^{-\gamma u}$, $u\to\infty$, where $\tilde{D} = \tilde{\zeta}^{(+)}(\gamma)\Delta_{\tilde{h}}^{-1}e$.

Proof: By (4.5), the result holds with $\tilde{D} = lim_{u\to\infty}\mathbb{E}_{\gamma,i}1/\tilde{h}_{J_{\tilde{\tau}_+(u)}}^{(\theta)}$; the limit exists because

$\{J_{\tilde{\tau}_+(u)}\}$ is an irreducible Markov process. Since the limiting stationary distribution is $\tilde{\zeta}^{(+)}(\gamma)$, the result follows. □

5. Cycle Maxima and Rare Events

The distribution of the cycle maximum

$$M_V(P_i) = \sup_{0 \le t \le P_i} V_t = \sup_{0 \le t \le P_i} S_t$$

is of interest for a variety of reasons: if x is the buffer size, $\mathbb{P}(M_V(P_i) > x)$ can be interpreted as the probability of buffer overflow within a busy cycle; and the set of $\mathbb{P}_i$-distributions of $M_V(P_i)$ lead to the extreme value behavior of $\{V_t\}$ as explained below.

Note that the $\mathbb{P}_i$-distributions of $M_V(P_i)$ is only non-trivial for $i \in E_+$ (if $u \in E_-$, then $\mathbb{P}_i(M_V(P_i) = 0) = 1$). Our main result on the cycle maximum is the following:

Theorem 5.1: *For* $i \in E_+$,

$$\mathbb{P}_i(M_V(P_i) > u) \approx D_C h_i e^{-\gamma u}, u\to\infty, \tag{5.1}$$

where

$$D_C = (1 - \Delta_h^{-1}\alpha^{(+-)}h)\zeta^{(+)}(\gamma)\Delta_h^{-1}e.$$

Proof: In just the same way as in (4.4), we have

$$\mathbb{P}_i(M_V(P_i) > u) = \mathbb{P}_i(\tau_+(u) < P_i)$$

$$= h_i\mathbb{E}_{\gamma;i}\left[\frac{e^{-\gamma S_{\tau_+(u)}}}{h_{J_{\tau_+(u)}}}; \tau_+(u) < P_i\right]$$

$$= h_i e^{-\gamma u}\mathbb{E}_{\gamma;i}\left[\frac{1}{h_{J_{\tau_+(u)}}}; \tau_+(u) < P_i\right]$$

$$\approx h_i e^{-\gamma u}\mathbb{E}_{\gamma;i}\left[\frac{1}{h_{J_{\tau_+(u)}}}\right]\mathbb{P}_{\gamma;i}(P_i = \infty),$$

using $\mathbb{P}_{\gamma;i}(\tau_+(u){\to}\infty) = 1$ in the last step. Thus the result follows with the preliminary expression

$$D_C = \mathbb{P}_{\gamma;i}(P_i = \infty) \lim_{u\to\infty}\mathbb{E}_{\gamma;i}\left[\frac{1}{h_{J_{\tau_+(u)}}}\right] \tag{5.2}$$

for D_C. But by Lemma 4.8,

$$\mathbb{P}_{\gamma;i}(P_i < \infty) = \alpha^{(+\,-)}(\gamma)e = \Delta_h^{-1}\alpha^{(+\,-)}\Delta_h e = \Delta_h^{-1}\alpha^{(+\,-)}h,$$

and as in the proof of Corollary 4.9, the limit in (5.2) is $\zeta^{(+)}(\gamma)\Delta_h^{-1}e$. $\quad\square$

The study of cycle maxima in queueing theory was initiated by Takács [40], who found the exact distribution for the $M/G/1$ queue (for a simple proof of his result, see Asmussen & Perry [9]). For fluid models, one can as in [9] find a representation of the exact distribution of $M_V(P_i)$ in terms of the lifetime of a non-homogeneous Markov process, the time-dependent intensities of which can be expressed in terms of the matrices $\widetilde{\alpha}^{(-\,+)}, \widetilde{U}^{(+)}, U^{(-)}$, but we shall not give the details.

The $GI/G/1$ analog of Theorem 5.1 was obtained by Iglehart [24] and extended to more general queues in [9].

We shall not apply Theorem 5.1 to rare events analysis. To this end, we need first to translate Theorem 5.1 into a similar statement on the maximum $M_V(C_i^*)$ of $\{V_t\}$ within the regenerate cycle C_i^* defined in Section 3.

Lemma 5.2: $\mathbb{P}_i(M_V(C_i^*) > u) \approx D_i^* e^{-\gamma u}$, *where*

$$D_i^* = \frac{D_C}{\eta_i}\sum_{j\in E_+}\eta_j h_j.$$

Proof: Use Theorem 5.1 and [9], Proposition 10.1. $\quad\square$

Now let $\tau_V(u) = inf\{t > 0: V_t \geq u\}$ be the first occurrence of the rare event $\{V_t \geq u\}$.

Corollary 5.3: *As* $u{\to}\infty$, $e^{-\gamma u}\tau_V(u)$ *is asymptotically exponential with rate parameter* $\mathbb{E}_i C_i^*/D_i^*$. *That is, for all* $x \geq 0$ *and all* $j \in E$,

$$\mathbb{P}_{j,x}(e^{-\gamma u}\tau_V(u) > t){\to}e^{-t\mathbb{E}_i C_i^*/D_i^*}.$$

Proof: This follows immediately from Lemma 5.2 and standard results on rare events in regenerative processes (e.g. Gnedenko & Kovalenko [22], p. 135). $\quad\square$

Next consider the extreme value $M_V(T)$.

Corollary 5.4: *As* $T{\to}\infty$,

$$\gamma M_V(T) - logT - log\frac{\sum_{j\in E_+}\eta_j D_C h_j}{\sum_{j\in E_+}\eta_j\mathbb{E}_j C_j} \xrightarrow{\mathcal{D}} H,$$

where H has the extreme value distribution $\mathbb{P}(H \leq x) = e^{-e^{-x}}$, $-\infty < x < \infty$.

Proof: By Lemma 5.2 and [9], Corollary 10.1. $\quad\square$

Note that (by general regenerative processes theory) Corollaries 5.3, 5.4 hold for arbitrary starting values $J_0 = i$, $V_0 = x$.

The $GI/G/1$ version of Corollary 5.4 was proved in [24] and extended to more general queues in [9].

6. Conditioned Limit Theorems. Quick Simulation

For the process $\{V_t\}$ to reach level u within the busy cycle, it must have behaved atypical because of the negative mean drift condition (1.1). Thus one may ask what more precisely this atypical behavior looks like. The following theorem tells that the answer is "like if the intensities were switched from Λ to $\Lambda(\gamma)$". To make this precise, define

$$N_{kj}(u) = \sum_{t \le \tau_+(u)} I(J_{t-} = k, J_t = j), \quad N_k(u) = \sum_{j \ne k} N_{kj}(u), \quad N(u) = \sum_{k \in E} N_k(u),$$

$$T_k(u) = \int_0^{\tau_+(u)} I(J_t = k)dt.$$

Thus, e.g., $T_k(u)$ is the time spent in state k before $\tau_+(u)$, $N_{kj}(u)$ is the number of jumps from k to j, and $N_{kj}(u)/T_k(u)$ is the average rate of such jumps.

Theorem 6.1: *The following convergences all hold in* $\mathbb{P}_i(\cdot \mid \tau_+(u) < P_i)$*-probability:*

(a) $T_k(u)/\tau_+(u) \to \pi_k^{(\gamma)}$;

(b) $N_{kj}(u)/N_k(u) \to \lambda_{kj}^{(\gamma)}/\lambda_k^{(\gamma)}$;

(c) $N_{kj}(u)/T_k(u) \to \lambda_{kj}^{(\gamma)}$;

(d) $sup_{0 \le t < \infty} | F_u^{(k)}(t) - (1 - e^{-\lambda_k^{(\gamma)}t}) | \to 0$,   *where* $F_u^{(k)}(\cdot)$ *is the empirical distribution of the holding times of state* k *prior to* $\tau_+(u)$;

(e) $\dfrac{1}{\lfloor \tau_+(u)/T \rfloor} \sum_{n=1}^{\lfloor \tau_+(u)/T \rfloor} \varphi(\{J_t\}_{(n-1)T \le t \le nT}) \to \mathbb{E}_{\gamma;\pi}(\gamma)\varphi(\{J_t\}_{0 \le t \le T})$,

   where $\varphi(\cdot)$ *is a measurable functional* $D_E[0,T] \to \mathbb{R}$;

(f) $\tau_+(u)/u \to 1/\kappa'(\gamma)$.

For the proof, we need a lemma:

Lemma 6.2: *Let* $\Psi(u)$ *be a* $\mathcal{F}_{\tau_+(u)}$*-measurable r.v. such that* $\Psi(u) \xrightarrow{a.s.} a$ $\mathbb{P}_{\gamma;i}$*-a.s. as* $u \to \infty$ *for some constant* a. *Then* $\mathbb{E}[\Psi(u) \mid \tau_+(u) < P_i] \to a$, $u \to \infty$.

Proof:

$$\mathbb{E}_i[\Psi(u); \tau_+(u) < P_i] = h_i \mathbb{E}_{\gamma;i}\left[\Psi(u) \frac{e^{-\gamma S_{\tau_+(u)}}}{h_{J_{\tau_+(u)}}}; \tau_+(u) < P_i \right]$$

$$= h_i e^{-\gamma u} \mathbb{E}_{\gamma;i}\left[\frac{\Psi(u)}{h_{J_{\tau_+(u)}}}; \tau_+(u) < P_i \right]$$

$$\approx h_i e^{-\gamma u} \mathbb{E}_{\gamma;i}\left[\frac{a}{h_{J_{\tau_+(u)}}}\right] \mathbb{P}_{\gamma;i}(P_i = \infty)$$

$$\approx a \mathbb{P}_i(\tau_+(u) < P_i)$$

(cf. the proof of Theorem 5.1 for the last step). □

Proof of Theorem 6.1: Assertions (a)-(e) follow by easy combinations of Lemma 6.2 and the law of large numbers for Markov processes. E.g.,

$$\frac{\sum_{t \le T} I(J_{t-} = k, J_t = j)}{\int_0^T I(J_t = k)dt} \to \lambda_{kj}^{(\gamma)} \quad \mathbb{P}_{\gamma;i} - \text{a.s.}$$

Thus letting $T = \tau_+(u)$, $\Psi(u) = I(\,|\,N_{kj}(u)/T_k(u)) - \lambda_{kj}^{(\gamma)}\,| > \epsilon)$, the assumptions of Lemma 6.2 are satisfied with $a = 0$, and (c) follows.

For (f), let similarly $\Psi(u) = I(\,|\,\tau_+(u)/u - 1/\kappa'(\gamma)\,| > \epsilon)$, and appeal to Lemma 7.2 below. □

Theorem 6.1 may be seen as an analog of $GI/G/1$ results of Asmussen [4] (see in particular Theorem 5.1 of that paper). See also Anantharam [1].

The result has implications for quick simulation. Assume we want to estimate the probability $\mathbb{P}_i(\tau_+(u) < P_i)$ of buffer overflow within a busy cycle by simulation. The crude Monte Carlo method has the typical problem of rare events simulation (a low relative precision so that an excessive number of replications is needed), and thus we may want to speed up the simulation by a change of measure. Formally, the simulation can be seen as picking a point at random from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set of all sample paths $\{(J_t, V_t)\}_{0 \le t \le \tau_+(u) \wedge P_i}$, $\mathcal{F}$ the obvious σ-field and $\mathbb{P}$ the restriction of $\mathbb{P}_i$ to $(\Omega, \mathcal{F})$. The change of measure amounts to simulating from a different $\tilde{\mathbb{P}}$, i.e. to use importance sampling, and by general results from that area, the optimal $\tilde{\mathbb{P}}$ is given by $\mathbb{P}_i(\cdot\,|\,\tau_+(u) < P_i)$. This choice is not practicable, one among many reasons being that the likelihood ratio involves the unknown probability $\mathbb{P}_i(\tau_+(u) < P_i)$. However, by Theorem 6.1

$$\mathbb{P}_i(\cdot\,|\,\tau_+(u) < P_i) \approx \mathbb{P}_{\gamma;i}(\cdot\,|\,\tau_+(u) < P_i),$$

which suggests to simulate simply from $\mathbb{P}_{\gamma;i}$; the transition from $\mathbb{P}_{\gamma;i}(\cdot\,|\,\tau_+(u) < P_i)$ to $\mathbb{P}_{\gamma;i}$ involves no asymptotic loss of efficiency since the $\mathbb{P}_{\gamma;i}$-probability of the conditioning event $\{\tau_+(u) < P_i\}$ has a strictly positive limit (viz., $\mathbb{P}_{\gamma;i}(P_i = \infty)$), in contrast to what is the case for $\mathbb{P}_i$.

The corresponding simulation estimator is

$$e^{-\gamma u}\frac{h_i}{h_{J_{\tau_+(u)}}}I(\tau_+(u) < P_i).$$

Obviously, its $\mathbb{P}_{\gamma;i}$-variance is $O(e^{-2\gamma u})$, i.e. of the same order of magnitude as $\mathbb{P}_i(\tau_+(u) < P_i)^2$ (this is roughly the optimality criterion used in Chang *et al.* [15]).

Estimation of the steady-state probability $\mathbb{P}(V > u, J = i)$ can be carried out in a similar way by simulating $\{(\tilde{J}_t, \tilde{S}_t)\}$ from $\mathbb{P}_{\gamma;i}$ and using the estimator

$$e^{-\gamma u} \frac{\pi_i \tilde{h}_i}{\tilde{h}_{\tilde{J}_{\tilde{\tau}_+(u)}}},$$

cf. (2.8). There is a straightforward analog of Theorem 6.1 for that setting too. When estimating $\mathbb{P}(V_T > u, J_T = i)$, the results of the next section suggest to simulate $\{(\tilde{J}_t, \tilde{S}_t)\}$ from $\mathbb{P}_{\alpha_y; i}$, where $y = T/u$.

Note that the approach to rare events simulation is most of the literature (e.g. Bucklew, Ney & Sadowski [14], Parekh & Walrand [34] and Cottrell, Fort & Malgouyres [17]) takes a somewhat different approach via the general theory of large deviations. For fluid models, see in particular Kesidis & Walrand [26].

7. Inequalities and Approximations for Transient Behavior

Define

$$\omega_\alpha^2 = \frac{\kappa''(\alpha)}{\kappa'(\alpha)^3}.$$

For computational purposes, note that by Theorem 4.2 and Propositions 4.4, 4.5

$$\kappa'(\alpha) = \kappa'_\alpha(0) = \pi^{(\alpha)} \Delta_r e = \nu^{(\alpha)} \Delta_{h^{(\alpha)}} \Delta_r e,$$

$$\kappa''(\alpha) = \kappa''_\alpha(0) = -2\kappa'(\alpha)^2 - 2\nu^{(\alpha)} \Delta_{h^{(\alpha)}} \Delta_r (\Lambda(\alpha) - e\nu^{(\alpha)} \Delta_{h^{(\alpha)}})^{-1} \Delta_r e,$$

Theorem 7.1: *As $u \to \infty$,*

$$\mathbb{P}_j(V_T > u, J_T = i) \approx \pi_i \tilde{h}_i \tilde{\zeta}^{(+)} \Delta_{\tilde{h}}^{-1} e \cdot e^{-\gamma u} \Phi\left(\frac{T - u/\kappa'(\gamma)}{\sqrt{u\omega_\gamma^2}}\right) \tag{7.1}$$

in the sense that if $T = T(u)$ varies with u in such a way that

$$y = \lim_{u \to \infty} \frac{T(u) - u/\kappa'(\gamma)}{\sqrt{u\omega_\gamma^2}}$$

exists, then (7.1) is $\pi_i \tilde{h}_i / \pi_j e^{-\gamma u} \Phi(y) + O(e^{-\gamma u})$.

For the proof, we shall need some lemmas.

Lemma 7.2: *For any $\theta \geq \gamma_0$, it holds w.r.t. $\mathbb{P}_{\theta; i}$ that $\tilde{\tau}_+(u)/u \xrightarrow{a.s.} 1/\kappa'(\theta)$ and that*

$$\frac{\tilde{\tau}_+(u) - u/\kappa'(\theta)}{\sqrt{u\omega_\theta^2}} \xrightarrow{\mathcal{D}} N(0,1). \tag{7.2}$$

Proof: First note that $\{\tilde{S}_t\}$ is a cumulative process with asymptotic mean and variance given by Theorem 4.2. Hence by general results on cumulative processes ([5] pp. 136-137),

$$\frac{\tilde{S}_t}{t} \xrightarrow{a.s.} \kappa'(\theta), \qquad \frac{\tilde{S}_t - t\kappa'(\theta)}{\sqrt{t}} \xrightarrow{\mathcal{D}} N(0, \kappa''(\theta)).$$

Letting $t = \tilde{\tau}_+(u)$ in the first limit and noting that $\tilde{S}_{\tilde{\tau}_+(u)} = u$ yields $\tilde{\tau}_+(u)/u \xrightarrow{a.s.} 1/\kappa'(\theta)$, and (7.2) then follows by applying Anscombe's theorem to the second limit.   □

Lemma 7.3: $\tilde{\tau}_+(u)$ and $\tilde{J}_{\tilde{\tau}_+(u)}$ *are asymptotically independent. That is,*

$$\mathbb{E}_{\alpha;i}\left[f(\tilde{h}^{(\alpha)}_{\tilde{\tau}_+(u)}); \tilde{\tau}_+(u) \le u/\kappa'(\alpha) + zu^{1/2}\omega_\alpha\right] \to \Phi(z)\mathbb{E}_{\alpha;i}f(\tilde{h}^{(\alpha)}_{\tilde{\tau}_+(\infty)}).$$

Proof: Easy along the lines of the proof of Stam's lemma in [5], pp. 271-272.   □

Lemma 7.4: *Let* $\kappa'(\alpha) > 0$ *and assume that* $T = u/\kappa'(\alpha) + z\sqrt{u}/\omega_\alpha$. *Then*

$$\mathbb{P}_j(V_T > u, J_T = i) \approx \pi_i \tilde{h}^{(\alpha)}_i e^{-\alpha u} \mathbb{E}_{\alpha;i}\left[\frac{e^{\tilde{\tau}_+(u)\kappa(\alpha)}}{\tilde{h}^{(\alpha)}_{\tilde{J}_{\tilde{\tau}_+(u)}}}; \tilde{\tau}_+(u) \le T\right].$$

Proof: By (2.7),

$\mathbb{P}_j(V_T > u, J_T = i)$

$$= \frac{\pi_i}{\pi_j}\mathbb{P}_i(\tilde{\tau}_+(u) \le T, \tilde{J}_T = j)$$

$$= \frac{\pi_i \tilde{h}^{(\alpha)}_i}{\pi_j \tilde{h}^{(\alpha)}_j}\mathbb{E}_{\alpha;i}\left[e^{-\alpha \tilde{S}_T + T\kappa(\alpha)}; \tilde{\tau}_+(u) \le T, \tilde{J}_T = j\right]$$

$$= \frac{\pi_i \tilde{h}^{(\alpha)}_i}{\pi_j \tilde{h}^{(\alpha)}_j}\mathbb{E}_{\alpha;i}\left[I(\tilde{\tau}_+(u) \le T)e^{-\alpha \tilde{S}_{\tilde{\tau}_+(u)} + \tilde{\tau}_+(u)\kappa(\alpha)}\right.$$

$$\left.\times \mathbb{E}_\alpha[I(\tilde{J}_T = j)e^{-\alpha(\tilde{S}_T - \tilde{S}_{\tilde{\tau}_+(u)}) + (T - \tilde{\tau}_+(u))\kappa(\alpha)} \mid \mathcal{F}_{\tilde{\tau}_+(u)}]\right]. \tag{7.3}$$

Now

$$\mathbb{E}_{\alpha;k}\left[e^{\theta \tilde{S}_v}; \tilde{J}_v = j\right] = \frac{\tilde{h}^{(\alpha)}_j}{\tilde{h}^{(\alpha)}_k}e^{-v\kappa(\alpha)}\mathbb{E}_k\left[e^{(\alpha + \theta)\tilde{S}_v}; \tilde{J}_v = j\right],$$

$$\mathbb{E}_{\alpha;k}\left[e^{-\alpha \tilde{S}_v + v\kappa(\alpha)}; \tilde{J}_v = j\right] = \frac{\tilde{h}^{(\alpha)}_j}{\tilde{h}^{(\alpha)}_k}\mathbb{P}_k(\tilde{J}_v = j)$$

$$\to \frac{\tilde{h}^{(\alpha)}_j}{\tilde{h}^{(\alpha)}_k}\pi_j, \quad v \to \infty.$$

Letting $v = T - \tilde{\tau}_+(u)$ and noting that $v \to \infty$ conditionally upon $\mathcal{F}_{\tilde{\tau}_+(u)}$ (and $\tilde{\tau}_+(u) \le T$) according to (7.2) (it follows that asymptotically (7.3) becomes

$$\pi_i \tilde{h}^{(\alpha)}_i \mathbb{E}_{\alpha;i}\left[\frac{I(\tilde{\tau}_+(u) \le T)e^{-\alpha \tilde{S}_{\tilde{\tau}_+(u)} + \tilde{\tau}_+(u)\kappa(\alpha)}}{\tilde{h}^{(\alpha)}_{\tilde{\tau}_+(u)}}\right]$$

$$= \pi_i \widetilde{h}_i^{(\alpha)} e^{-\alpha u} \mathbb{E}_{\alpha;i} \left[\frac{e^{-\widetilde{\tau}_+(u)\kappa(\alpha)}}{\widetilde{h}_{J_{\widetilde{\tau}_+(u)}}^{(\alpha)}}; \ \widetilde{\tau}_+(u) \leq T \right].$$

□

Proof of Theorem 7.1: Letting $\alpha = \gamma$ in Lemma 7.4 yields

$$\mathbb{P}_j(V_T < u, J_T = i) \approx \pi_i \widetilde{h}_i e^{-\gamma u} \mathbb{E}_{\gamma;i} \left[\frac{1}{\widetilde{h}_{J_{\widetilde{\tau}_+(u)}}}; \ \widetilde{\tau}_+(u) \leq T \right]$$

$$\approx \pi_i \widetilde{h}_i e^{-\gamma u} \lim_{u \to \infty} \mathbb{E}_{\gamma;i} \left[\frac{1}{\widetilde{h}_{J_{\widetilde{\tau}_+(u)}}} \right] \Phi\left(\frac{T - u/\kappa'(\gamma)}{\sqrt{u\omega_\gamma^2}} \right)$$

$$\approx \pi_i \widetilde{h}_i \widetilde{\zeta}^{(+)} \Delta_{\widetilde{h}}^{-1} e \cdot e^{-\gamma u} \Phi(y),$$

using Lemma 7.3 in the second step and the same estimate as in the proof of Corollary 4.9 in the third.   □

From Theorem 7.1, we immediately obtain:

Corollary 7.5:

$$\frac{\mathbb{P}(V_{yu} > u, J_{yu} = j)}{\mathbb{P}(V > u, J = j)} \to \begin{cases} 0 & y < 1/\kappa'(\gamma) \\ 1 & y > 1/\kappa'(\gamma) \end{cases}.$$

The implication of Corollary 7.5 is that if we are interested in valued in excess of u (say u is the buffer size of an ATM model), then $T = u/\kappa'(\gamma)$ plays a critical role as the time at which $\mathbb{P}(V_T > u)$ becomes of the same order of magnitude as for the steady state. We may be more ambitious and ask for bounds or approximations on the convergence rate in Corollary 7.5. Define α_y as the solution $> \gamma_0$ of $\kappa'(\alpha_y) = 1/y$, cf. Figure 4.1, and recall the definition of $C_+(\alpha)$ from Section 4.

Theorem 7.6: *Assume* $y > 1/max_{i \in E_+} r_i$ *and let* $\gamma_y = \alpha_y - y\kappa(\alpha_y)$. *Then*

$$\lim_{u \to \infty} \sup \frac{\mathbb{P}_j(V_{yu} > u, J_{yu} = i)}{e^{-\gamma_y u}} \leq \frac{\pi_i}{2} \widetilde{h}_i^{(\alpha_y)} C_+(\alpha_y). \tag{7.4}$$

Proof: Consider first the case $y < 1/\kappa'(\gamma)$. Then $\kappa(\alpha_y) > 0$ (see Figure 4.1). By Lemma 7.4,

$$\mathbb{P}_j(V_{yu} > u, J_{yu} = i) \approx \pi_i \widetilde{h}_i^{(\alpha_y)} e^{-\alpha_y u} \mathbb{E}_{\alpha_y;i} \left[\frac{e^{\widetilde{\tau}_+(u))\kappa(\alpha_y)}}{\widetilde{h}_{J_{\widetilde{\tau}_+(u)}}^{(\alpha_y)}}; \ \widetilde{\tau}_+(u) \leq yu \right] \tag{7.5}$$

$$\leq \pi_i \widetilde{h}_i^{(\alpha_y)} C_+(\alpha_y) e^{-\alpha_y u + yu\kappa(\alpha_y)} \mathbb{P}_{\alpha_y;i}(\widetilde{\tau}_+(u) \leq yu)$$

$$\to \pi_i \tilde{h}_i^{(\alpha_y)} C_+(\alpha_y) e^{-\gamma_y u} \cdot \frac{1}{2},$$

using Lemma 7.2 in the last step. □

Note that the condition $y > 1/max_{i \in E_+} r_i$ is no restriction: if $y < 1/max_{i \in E_+} r_i$, then $V_{yu} \le u$ so that $\mathbb{P}(V_{yu} > u) = 0$.

Remark 7.7: Heuristically, we can sharpen (7.4) to the approximation

$$\mathbb{P}_j(V_{yu} > u, J_{yu} = i) \approx \frac{\pi_i \tilde{h}_i^{(\alpha_y)} \tilde{\zeta}_+(\alpha_y) \Delta_{\tilde{h}^{(\alpha_y)}}^{-1} e}{\kappa(\alpha_y)\sqrt{2\pi y^3 \kappa''(\alpha_y)}} \cdot \frac{e^{-\gamma_y u}}{\sqrt{u}}. \tag{7.7}$$

Writing $\tilde{\tau}_+(u) \approx yu + u^{1/2} \omega_{\alpha_y} V$, where V is normal $(0,1)$ under $\mathbb{P}_{\alpha_y}$, we get heuristically that

$$\mathbb{E}_{\alpha_y}\left[e^{\tilde{\tau}_+(u)\kappa(\alpha_y)}; \tilde{\tau}_+(u) \le yu\right] \quad \approx e^{yu\kappa(\alpha_y)} \mathbb{E}_{\alpha_y}\left[e^{\kappa(\alpha_y) u^{1/2} \omega_{\alpha_y} V}; V \le 0\right]$$

$$= e^{yu\kappa(\alpha_y)} \int_0^\infty e^{-\kappa(\alpha_y) u^{1/2} \omega_{\alpha_y} x} \varphi(x) dx$$

$$= e^{yu\kappa(\alpha_y)} \frac{1}{\kappa(\alpha_y) u^{1/2} \omega_{\alpha_y}} \int_0^\infty e^{-z} \varphi(z/(u^{1/2} \omega_{\alpha_y})) dz$$

$$= e^{yu\kappa(\alpha_y)} \frac{1}{\kappa(\alpha_y) u^{1/2} \omega_{\alpha_y}} \int_0^\infty e^{-z} \cdot \frac{1}{\sqrt{2\pi}} dz$$

$$= e^{yu\kappa(\alpha_y)} \frac{1}{\kappa(\alpha_y)\sqrt{2\pi u \omega_{\alpha_y}^2}}.$$

Inserting these estimates in (7.5) and noting that $\omega_{\alpha_y}^2 = y^3 \kappa''(\alpha_y)$, (7.6) follows.

The main difficulty in making the proof precise is that one needs a sharpened version of the CLT for $\tau_+(u)$ (basically a local CLT with remainder term). However, also (7.7) needs a more rigorous proof. □

If y is larger than the critical value $1/\kappa'(\gamma)$, we can get a bound on the deviation from the steady-state value:

Theorem 7.8: *Assume $y > 1/\kappa'(\gamma)$ and let $\alpha_y - y\kappa(\alpha_y)$. Then*

$$0 \le \mathbb{P}(V > u, J = i) - \mathbb{P}\pi(V_{yu} > u, J_{yu} = i) \le \pi_i \tilde{h}_i^{(\alpha_y)} C_+(\alpha_y) e^{-\gamma_y u}. \tag{7.8}$$

Proof: Let $\psi_j(u) = \mathbb{P}_j(M_{\tilde{S}} > u)$. Since $\kappa(\gamma_y) < 0$, (4.4) yields

$$\psi_j(u) \le C_+(\alpha_y) \tilde{h}_j^{(\alpha_y)} e^{-\gamma_y u}.$$

Hence by Proposition 2.1

$$\mathbb{P}(V > u, J = i) - \mathbb{P}_{\pi}(V_{yu} > u, J_{yu} = i)$$

$$= \pi_i \mathbb{P}_i(M_{\widetilde{S}} > u) - \sum_{j \in E} \pi_j \frac{\pi_i}{\pi_j} \mathbb{P}_i(M_{\widetilde{S}}(yu) > u, \widetilde{J}_{yu} = j)$$

$$= \pi_i \sum_{j \in E} \{\mathbb{P}_i(M_{\widetilde{S}} > u, \widetilde{J}_{yu} = j) - \mathbb{P}_i(M_{\widetilde{S}}(yu) > u, \widetilde{J}_{yu} = j)\}$$

$$= \pi_i \sum_{j \in E} \mathbb{P}_i(M_{\widetilde{S}} > u, M_{\widetilde{S}}(yu) \le u, \widetilde{J}_{yu} = j) \qquad (7.9)$$

$$= \pi_i \sum_{j \in E} \mathbb{P}_i(M_{\widetilde{S}} > u, M_{\widetilde{S}}(yu) \le u, \widetilde{S}_{yu} \le u, J_{yu} = j)$$

$$= \pi_i \sum_{j \in E} \mathbb{E}_i\left[\psi_j(u - \widetilde{S}_{yu}); M_{\widetilde{S}}(yu) \le u, \widetilde{S}_{yu} \le u, \widetilde{J}_{yu} = j\right]$$

$$\le \pi_i \sum_{j \in E} C_+(\gamma_0) e^{-\alpha_y u} \widetilde{h}_j^{(\alpha_y)} \mathbb{E}_i\left[e^{\alpha_y \widetilde{S}_{yu}}; \widetilde{J}_{yu} = j\right]$$

$$= \pi_i \widetilde{h}_i^{(\alpha_y)} C_+(\alpha_y) e^{-\alpha_y u} e^{\kappa(\gamma_y) yu}.$$

Non-negativity follows from (7.9).   □

The form of the above results originate from classical collective risk theory, a setting which is mathematically equivalent to the $M/G/1$ queue. Thus Theorem 7.1 was proved in that framework by Segerdahl [36], whereas Theorem 7.6 goes back to Gerber [21] (in the setting of [21], Theorem 7.6 takes the form of an exact inequality) and (7.6) to Arfwedson [3]. The present proof appears to use less information than is inherent in the definition of α_y, γ_y. However, as in [21] this definition will produce the maximal γ_y for which the argument works. The idea behind the choice of α_y is essentially the saddlepoint method, to make $\mathbb{E}_{\alpha_{y'}} \cdot \widetilde{\tau}_+(u) \approx T = yu$.

Here is an estimate of the rate of convergence to the steady state which is different from (7.8) by fixing u and letting only $T \to \infty$.

Theorem 7.9: Let $\gamma_0 > 0$ satisfy $\kappa'(\gamma_0) = 0$ and let $\delta = e^{\kappa(\gamma_0)}$. Then

$$0 \le \mathbb{P}(V > u, J = i) - \mathbb{P}_{\pi}(V_T > u, J_T = i) \le \pi_i \widetilde{h}_i^{(\gamma_0)} C_+(\gamma_0) e^{-\gamma_0 u} \cdot \delta^T. \qquad (7.10)$$

Proof: Replace α_y by γ_0 and yu by T in the proof of Theorem 7.8.   □

The $GI/G/1$ version of Theorem 7.9 is due to Borokov [13]. We conjecture that the condition of stationary initial conditions for $\{J_t\}$ is not critical for the rates in Theorem 7.8, 7.9 and that (cf. standard relaxation time results for simple queues, e.g. [5], pp. 95, 262-262) the correct rate of convergence in Theorem 7.9 is $\delta^T/T^{3/2}$.

Note that (7.10) can be seen as a limiting case of (7.4) (except for the constant $1/2$ there). Indeed, if we write $T = yu$ with u fixed, we have $y \to \infty$ which implies $\alpha_y \downarrow \gamma_0$ and

$$e^{-\gamma_0 u} \cdot \delta^T = e^{-\gamma_0 u + yu\kappa(\gamma_0)} \approx e^{-\gamma_y u}.$$

8. Some Exact Formulas

We first compute the Laplace transform of the busy period.

Theorem 8.1: *For $i \in E_+$, $j \in E_-$ and $\alpha \geq \kappa(\gamma_0)$,*

$$\mathbb{E}_i\left[e^{-\alpha P_i}; J_{P_i} = j\right] = \frac{h_i^{(\theta)}}{h_j^{(\theta)}}\alpha_{ij}^{(+-)}(\theta). \tag{8.1}$$

where $\theta = \theta(\alpha)$ is the unique solution $\leq \gamma_0$ of $\kappa(\theta) = \alpha$. If $\alpha < \kappa(\gamma_0)$, then

$$\mathbb{E}_i\left[e^{-\alpha P_i}; J_{P_i} = j\right] = \infty.$$

Proof: Since $\kappa(\theta)$ increases monotonically from $\kappa(\gamma_0)$ to ∞ as θ decreases from γ_0 to $-\infty$, the existence and uniqueness of θ is clear, and we get

$$\mathbb{E}_i\left[e^{-\alpha P_i}; J_{P_i} = j\right] = \frac{h_i^{(\theta)}}{h_j^{(\theta)}}\mathbb{E}_{\theta;i}\left[e^{-\theta S_{P_i} + \kappa(\theta)P_i - \alpha P_i}; J_{P_i} = j\right]$$

$$= \frac{h_i^{(\theta)}}{h_j^{(\theta)}}\,\mathbb{P}_{\theta;i}(J_{P_i} = j) = \frac{h_i^{(\theta)}}{h_j^{(\theta)}}\alpha_{ij}^{(+-)}(\theta)$$

using $S_{P_i} = 0$. If $\alpha = \kappa(\gamma_0) - \epsilon$, then similarly

$$\mathbb{E}_i\left[e^{-\alpha P_i}; J_{P_i} = j\right] = \frac{h_i^{(\gamma_0)}}{h_j^{(\gamma_0)}}\,\mathbb{E}_{\gamma_0;i}\left[e^{\epsilon P_i}; J_{P_i} = j\right].$$

However, some easy irreducibility arguments show that since $\kappa'(\gamma_0) = 0$, we have $\mathbb{E}_{\gamma_0;i}P_i = \infty$

(otherwise, the $\mathbb{P}_{\gamma_0;\cdot}$-process would be positive recurrent) and hence $\mathbb{E}_{\gamma_0;i}\left[e^{\epsilon P_i}; J_{P_i} = j\right] = \infty.$ □

Now consider the time $\tau_-(u)$ for the system to empty starting from $V_0 = u > 0$.

Corollary 8.2: *If α,θ are as in Theorem 8.1, then for $i,j \in E_-$*

$$\mathbb{E}_i\left[e^{-\alpha\tau_-(u)}; J_{\tau_-(u)} = j\right] = \begin{cases} \dfrac{h_i^{(\theta)}}{h_j^{(\theta)}}e^{\theta u}e_i'U^{(-)}(\theta)u\, e_j & i \in E_- \\[2em] \dfrac{h_i^{(\theta)}}{h_j^{(\theta)}}e^{\theta u}\alpha_i^{(+-)}(\theta)e^{U^{(-)}(\theta)u}e_j & \in E_+. \end{cases} \tag{8.2}$$

If $\alpha < \kappa(\gamma_0)$, then $\mathbb{E}_i\left[e^{-\alpha\tau_-(u)}; J_{\tau_-(u)} = j\right] = \infty.$

Proof: As in the proof of Theorem 8.1, it follows from $S_{\tau_-(u)} = -u$ that

$$\mathbb{E}_i\left[e^{-\alpha\tau_-(u)}; J_{\tau_-(u)} = j\right] = \frac{h_i^{(\theta)}}{h_j^{(\theta)}}\mathbb{E}_{\theta;i}\left[e^{-\theta S_{\tau_-(u)} + \kappa(\theta)\tau_-(u) - \alpha\tau_-(u)}; J_{\tau_-(u)} = j\right]$$

$$= \frac{h_i^{(\theta)}}{h_j^{(\theta)}} e^{\theta u} \mathbb{P}_{\theta;i}(J_{\tau_-(u)} = j).$$

That $\mathbb{P}_{\theta;i}(J_{\tau_-(u)} = j)$ has the asserted form is obvious. The case $\alpha < \kappa(\gamma_0)$ is treated as above.□

The difficulty in applying Theorem 8.a and Corollary 8.2 is that the explicit form of $\theta = \kappa^{-1}(\alpha)$ is complicated, and that the matrices $\alpha_i^{(-+)}(\theta)$ and $U^{(-)}(\theta)$ do not appear to reduce (except for the case $\theta = \gamma$, cf. Lemma 4.8); see however, the next section for a simple example. However, the less ambitious goal of computing the mean of $\tau_-(u)$ is attainable by a direct argument. Note that for $u = 0$, this has already been carried out in Section 3 by computing the matrix $\bar{P}$.

Define $M^{(-)}(u): E_- \times E_-$ as the matrix with $ij$the element $\mathbb{E}_i[\tau_-(u); J_{\tau_-(u)} = j]$ and $M^{(+)}(u): E_+ \times E_-$ as the matrix with ijth element $\mathbb{E}_i[\tau_-(u); J_{\tau_-(u)} = j]$. Note that it is easily seen that asymptotically

$$\mathbb{E}\tau_-(u) \approx \frac{u}{|\sum_{i \in E} \pi_i r_i|}. \tag{8.3}$$

The following result gives an exact expression incorporating also information on the $J_{\tau_-(u)}$ component.

Proposition 8.3:

$$M^{(-)}(u) = \int_0^u e^{U^{(-)}x} M^{(-)\prime}(0) e^{U^{(-)}(u-x)} dx$$

$$M^{(+)}(u) = \bar{P} e^{U^{(-)}u} + \alpha^{(+-)} M^{(-)}(u) \tag{8.4}$$

Here $M^{(-)\prime}(0) = \Delta_{|r|}^{-1} + T^{(-+)}\bar{P}$.

Proof: Consider first $M^{(-)}(u)$. We decompose $\tau_-(u+v)$ as $\tau_-(u) + \tau_-^*(v)$, where $\tau_-^*(v)$ has the same distribution as $\tau_-(v)$, only started from $J_{\tau_-(u)}$ rather than i. Given $J_{\tau_-(u)} = k$, the condition probability of $J_{\tau_-(u+v)} = j$ is $e_k' e^{U^{(-)}v} e_j$, and the conditional expectation of $\tau_-^*(v)I(J_{\tau_-(u+v)} = j$ is $e_k' M(v)e_j$. Since the $\mathbb{P}_i$-distribution of $J_{\tau_-(u)}$ is $e_i' e^{U^{(-)}u}$, we obtain

$$M(u+v) = M(u)e^{Uv} + e^{Uu}M(v).$$

This leads to $M^{(-)\prime}(u) = M^{(-)}(u)U + e^{Uu}M^{(-)\prime}(0)$ and the solution of this differential equation subject to the obvious boundary condition $M^{(-)}(0) = 0$ is indeed (8.4). Furthermore,

$$\mathbb{E}_i\left[\tau - \left(\frac{du}{|r_i|}\right); J_{\tau_-(du)} = j\right] = \left(1 - \frac{\lambda_i du}{|r_i|}\right)\frac{du}{|r_i|}\delta_{ij} + \sum_{k \in E_+}\frac{\lambda_{ik}du}{|r_i|}\mathbb{E}_k[P_k; J_{\tau_-(u)} = j]$$

$$= \frac{du}{|r_i|}\delta_{ij} + \sum_{k \in E_+}\frac{\lambda_{ik}du}{|r_i|}\mathbb{E}_k[P_k; J_{j_k} = j],$$

which immediately leads ot the asserted expression for $M^{(-)\prime}(0)$. Finally for $M^{(+)}(u)$ and $i \in E_+$, we decompose $\tau_-(u)$ as $P_i + \tau_-^*(u)$, where $\tau_-^*(u)$ has the same distribution as $\tau_-(u)$ stated from J_{P_i}. Given $J_{P_i} = k$, the conditional probability of $J_{\tau_-(u)} = j$ is $e_k' e^{U^{(-)}u} e_j$, and the conditional expectation of $\tau_-^*(u)I(J_{\tau_-(u)} = j$ is $e_k' M(u)e_j$. From this the asserted formula

easily follows.   □

Corollary 8.4: *In terms of the diagonal form (2.5),*

$$M^{(-)}(u) = \sum_{i \in E_-} u e^{s_i u} \frac{f_{ii}}{s_i} r_i I_i + \sum_{i,j \in E_-, i \neq j} (e^{s_i u} - e^{s_i u}) \frac{f_{ij}}{s_i - s_j} r_i I_j,$$

where $f_{ij} = I_i M^{(-)'}(0) r_j$.

Proof: By straightforward integration.   □

9. An Example

Consider a two-state fluid flow model with $E_+ = \{1\}, E_- = \{2\}$,

$$\Lambda = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \ r = \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \text{ i.e., } K[s] = \begin{pmatrix} -1+s & 1 \\ 1 & -1-2s \end{pmatrix}.$$

In particular, $\pi = (1/2 \ \ 1/2)$ and the mean drift is

$$\kappa'(0) = \pi_i r_1 + \pi_2 r_2 = -\frac{1}{2}.$$

Note that any two-state model is automatically time reversible so that $\tilde{h} = h$ etc.

The characteristic polynomial of $K[s]$ is $z^2 + z(s+2) + s - 2s^2$. Thus, $\kappa(s)$ is the largest root,

$$\kappa(s) = \frac{-s - 2 + \sqrt{9s^2 + 4}}{2}.$$

From this it is readily seen that at the positive right eigenvector $h^{(\theta)}$ of $K[\theta]$ corresponding to the eigenvalue $\kappa(\theta)$ can be taken as

$$h^{(\theta)} = c \begin{pmatrix} 1 \\ 1 - \theta + \kappa(\theta) \end{pmatrix} = c \begin{pmatrix} 1 \\ \dfrac{-3\theta + \sqrt{9\theta^2 + 4}}{2} \end{pmatrix} \tag{9.5}$$

and that the solution $\gamma > 0$ of $\kappa(\gamma) = 0$ is $\gamma = 1/2$. In particular,

$$h = h^{(\gamma)} = c \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}$$

and $C_+(\theta) = C_-(\theta) = 1/c$. Thus in this case, Corollary 4.7 is exact and takes the form

$$\mathbb{P}(V > u, J = i) = \begin{cases} \frac{1}{2} e^{-u/2} & i = 1 \\ \frac{1}{4} e^{-u/2} & i = 2 \end{cases} ; \tag{9.6}$$

from this we obtain

$$\mathbb{P}(V = 0, J = i) = \pi_i - \mathbb{P}(V > 0, J = i) = \begin{cases} 0 & i = 1 \\ \frac{1}{4} & i = 2 \end{cases}.$$

It is also easily checked that the constants in (9.6) coincide with those of Corollary 4.9, so that the approximations there are exact as well.

Theorem 3.2 yields the mean busy period $\mathbb{E}_1 P_1$ as

$$\bar{P} = \bar{p}_{12} = \frac{1 + 1/2}{1 - 1/2} = 3.$$

Alternatively, the same result follows by Theorem 3.3 which takes the form

$$0 = \bar{p}_{12} + 1 - \frac{\bar{p}_{12}}{2} + \frac{\bar{p}_{12}}{2} + \frac{\bar{p}_{12}}{2} + \frac{1}{2}.$$

Since the idle period is exponential with mean 1, we get $\mathbb{E}_1 C_1 = \mathbb{E}_1 C_1^* = 3 + 1 = 4$.

In Sections 5-6,

$$D_C = \frac{1}{2} \cdot \frac{1}{c}, \; D_1^* = D_C \cdot c = \frac{1}{2}.$$

Thus Theorem 5.1 yields the asymptotic tail of the cycle maximum $M_V(P_1)$ as

$$\mathbb{P}_1(M_V(P_1) > u) \approx \frac{1}{2} e^{-u/2}.$$

Note, incidentally, that for this simple example the exact distribution of $M_V(P_1)$ is easily obtained: a sample path inspection shows that it is the same as the distribution of the cycle maximum of the virtual waiting time of a $M/M/1$ queue with service rate $1/r_1 = 1$ and arrival rate $1/r_2 = 1/2$, so that ([9], Corollary 2.1 or Theorem 3.1)

$$\mathbb{P}_1(M_V(P_1) > u) = \frac{e^{-u/2}}{2(1 - e^{-u/2}/2)}.$$

By Corollary 5.3, the time $\tau_V(u)$ of the first occurrence of the rare event $\{V_t \geq u\}$ is asymptotically exponential,

$$\mathbb{P}(e^{-u/2} \tau_V(u) > t) \to e^{-8x}.$$

Also, Corollary 5.4 takes the form

$$\frac{1}{2} M_V(T) - \log T + \log 8 \xrightarrow{\mathcal{D}} H,$$

which shows that the maximal value $M_V(T)$ of $\{V_t\}$ in the time interval is of the order of magnitude $2 \log T$; more precisely

$$\mathbb{P}(M_V(T) \leq 2 \log T + x) \approx e^{-e^{-x/2}/8}.$$

Finally, the changed measure $\mathbb{P}_{\gamma;\,.}$ occurring in the conditioned limit theorems and the quick simulation schemes is described by the changed intensity matrix

$$\Lambda(\gamma) = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ 2 & -2 \end{pmatrix}.$$

I.e., as expected, the sojourns in state 1 are longer and those in state 2 shorter.

In Section 7, we get first

$$\kappa'(\alpha) = -\frac{1}{2} + \frac{9\alpha}{2\sqrt{9\alpha^2 + 4}}, \ \kappa''(\alpha) = \frac{18\alpha}{(9\alpha^2 + 4)^{3/2}}.$$

In particular, $\kappa'(\gamma) = 2/5$ so that the critical time for $\mathbb{P}(V_T > u, J_T = i)$ to reach its steady-state value $\mathbb{P}(V > u, J = i)$ is approximately $5u/2$ when u is large. For $T = 5u/2 + O(\sqrt{u})$, Theorem 7.1 yields the approximation

$$\mathbb{P}(V_T > u, J_t = i) \approx \begin{cases} \frac{1}{2} e^{-u/2} \Phi\left(\dfrac{T - 5u/2}{3\sqrt{2u}} \right) & i = 1 \\[3mm] \frac{1}{4} e^{-u/2} \Phi\left(\dfrac{T - 5u/2}{3\sqrt{2u}} \right) & i = 2 \end{cases}$$

(calculating ω_γ^2 as 18). To exemplify (7.4) and (7.6), let $y = 2$. Then $\kappa'(\alpha_y) = 1/y$ yields

$$\alpha_y = \frac{4}{3\sqrt{5}}, \ \gamma_y = \alpha_y = y\kappa(\alpha_y) = 2 - \frac{2\sqrt{5}}{3}, \ \boldsymbol{h}^{(\alpha_y)} = c\begin{pmatrix} 1 \\ \dfrac{1}{\sqrt{5}} \end{pmatrix}.$$

Thus (7.4) yields the approximate upper bounds

$$\frac{1}{4} e^{-(2 - 2\sqrt{5}/3)u} \ (i = 1), \quad \frac{1}{4\sqrt{5}} e^{-(2 - 2\sqrt{5}/3)u} \ (i = 2)$$

for $\mathbb{P}(V_{2u} > u, J_{2u} = i)$ and (7.6) the approximations

$$\frac{9}{8(2\sqrt{5} - 3)\sqrt{\pi}} \cdot \frac{e^{-(2 - 2\sqrt{5}/3)u}}{\sqrt{u}} \ (i = 1), \quad \frac{9}{8(10 - 3\sqrt{5})\sqrt{\pi}} \cdot \frac{e^{-(2 - 2\sqrt{5}/3)u}}{\sqrt{u}} \ (i = 2).$$

For Theorem 7.9, $\kappa'(\gamma_0) = 0$ yields

$$\gamma_0 = \frac{1}{2\sqrt{2}} \text{ so that } \kappa(\gamma_0) = \frac{4}{3\sqrt{2}} - 1, \ \boldsymbol{h}^{(\gamma_0)} = c\begin{pmatrix} 1 \\ \dfrac{1}{\sqrt{2}} \end{pmatrix},$$

and an upper bound on the rate at which $\mathbb{P}_{\boldsymbol{\pi}}(V_T > u, J_T = i)$ approaches the steady-state value as $T \to \infty$ is

$$\frac{1}{2} e^{-u/3\sqrt{2}} \cdot \left(e^{4/3\sqrt{2}} - 1 \right)^T \ (i = 1), \quad \frac{1}{2\sqrt{2}} e^{-u/3\sqrt{2}} \cdot \left(e^{4/3\sqrt{2}} - 1 \right)^T \ (i = 2).$$

Finally consider the results of Section 8. For a given α the solution $\theta = \theta(\alpha)$ of $\kappa(\theta) = \alpha$ is seen after some algebra to be one of the roots of $2\theta^2 - \theta(\alpha + 1) - \alpha^2 - 2\alpha$, i.e.

$$\theta = \frac{\alpha + 1 - \sqrt{9\alpha^2 + 18\alpha + 1}}{4}$$

(the sign of $\sqrt{\cdot}$ is $-$ because $\theta \to -\infty$ as $\alpha \to \infty$). Using the first expression for $\boldsymbol{h}^{(\theta)}$ in (9.5) yields

$$\frac{h_1^{(\theta)}}{h_2^{(\theta)}} = \frac{4}{3 + 3\alpha + \sqrt{9\alpha^2 + 18\alpha + 1}}. \tag{9.7}$$

Since $\boldsymbol{U}^{(-)}(\theta) = 0$ for $\theta \le \gamma_0$, it follows that

$$\mathbb{E}_1 e^{-\alpha \tau_-(u)} = \frac{h_1^{(\theta)}}{h_2^{(\theta)}} exp\left\{ -\left(\frac{\sqrt{9\alpha^2 + 18\alpha + 1} - 1 - \alpha}{4} \right)u \right\},$$

$$\mathbb{E}_2 e^{-\alpha \tau_-(u)} = exp\left\{ -\left(\frac{\sqrt{9\alpha^2 + 18\alpha + 1} - 1 - \alpha}{4} \right)u \right\};$$

in particular, letting $u = 0$ yields the Laplace transform of the busy period as (9.7). In Proposition 8.3, $M^{(-)'} = 1/2 + 1/2 \cdot 3 = 2$, and we get $M^{(-)}(u) = \mathbb{E}_2 \tau_-(u) = 2u$ (thus, in this case (8.3) is exact) and $M^{(+)}(u) = \mathbb{E}_1 \tau_-(u) = 3 + 2u$.

10. Appendix: Miscellaneous Proofs

Proof of Theorem 4.2: The strict convexity follows from [27] by noting that $e^{t\kappa(s)}$ may be viewed as the Perron-Frobenious root of $\widehat{F}_t[s]$ for some fixed t. For s large, $s\Delta_r$ dominates Λ so that $\kappa(s) \approx s \ max_{i \in E_+} r_i$ and $\kappa'(s) \approx max_{i \in E_+} r_i$; the behavior at $-\infty$ is obtained in a similar way.

Differentiating $\nu^{(s)} K[s] = \kappa(s)\nu^{(s)}$ w.r.t. s and omitting the argument s for brevity yields

$$\kappa'\nu + \kappa\nu' = \nu'K + \nu\Delta_r, \tag{10.1}$$

$$\kappa''\nu + 2\kappa'\nu' + \kappa\nu'' = \nu''K + 2\nu'\Delta_r. \tag{10.2}$$

From $\nu e \equiv 1$ it follows that $\nu'e \equiv \nu''e \equiv \ldots \equiv 0$. Thus letting $s = 0$ in (10.1) and multiplying by e, we get

$$\kappa'(0) + 0 = 0 + \pi\Delta_r e = \sum_{i \in E} \pi_i r_i;$$

that this is $lim_{t\to\infty} S_t/t$ follows by the law of large numbers for Markov processes, and the proof of (4.1) is complete.

For (4.2), we get similarly from (10.2)

$$\kappa''(0) = 2\nu'(0)\Delta_r e.$$

Letting $s = 0$ in (10.1) and using $\pi(\Lambda - e\pi) = -\pi$ yields

$$-\kappa'(0)\pi(\Lambda - e\pi) = \kappa'(0)\pi = \nu'(0)\Lambda + \pi\Delta_r = \nu'(0)(\Lambda - e\pi) + \pi\Delta_r,$$

$$\nu' = -\kappa'(0)\pi - \pi\Delta_r D,$$

and the proof of (4.2) is complete. The identification in terms of the limiting variance can be obtained by appealing to discrete-time results of Keilson & Wishart [29], [30] and the method of discrete skeletons.   □

Proof of Proposition 4.6: Let θ be fixed and define

$$L_T = \frac{h_{J_T}^{(\theta)}}{h_{J_0}^{(\theta)}} exp\{\theta S_T - T\kappa(\theta)\}.$$

Then inspecting the explicit form of L_T shows that $\{L_T\}$ is a multiplicative functional of $\{(J_t, S_t)\}$. Since it is also a martingale w.r.t. $\mathbb{P}_i$ (Proposition 4.3), it follows by general Markov process theory ([31]) that there exist governing probabilities $\widetilde{\mathbb{P}}_{\theta;i}$, w.r.t. which $\{(J_t, S_t)\}$ is a time-homogeneous Markov process and

$$\widetilde{\mathbb{P}}_{\theta;i}(G) = \mathbb{E}_i[L_T; G], \ G \in \mathcal{F}_T.$$

Also, $\{1/L_T\}$ is a multiplicative functional and a martingale w.r.t. $\widetilde{\mathbb{P}}_{\theta;i}$, from which it follows in particular that

$$\widetilde{\mathbb{E}}_{\theta;i}\left[\frac{1}{L_T} \mid \mathcal{F}_\tau\right] = \frac{1}{L_\tau} \text{ on } \{\tau \le T\}. \tag{10.3}$$

To see that $\widetilde{\mathbb{P}}_{\theta;i} = \mathbb{P}_{\theta;i}$, it suffices to consider the one-dimensional transition function or equivalently the matrix m.g.f., i.e to show that

$$\left(\widetilde{\mathbb{E}}_{\theta;i}\left[e^{\alpha S_T}; J_T = j\right]\right)_{i,j \in E} = \left(\mathbb{E}_{\theta;i}\left[e^{\alpha S_T}; J_T = j\right]\right)_{i,j \in E}. \tag{10.4}$$

But the l.h.s. is

$$\left(\mathbb{E}_i\left[L_T e^{\alpha S_T}; J_T = j\right]\right) = \left(e^{-T\kappa(\theta)}\frac{h_j^{(\theta)}}{h_i^{(\theta)}}\mathbb{E}_i\left[e^{(\alpha+\theta)S_T}; J_T = j\right]\right)$$

$$= e^{-T\kappa(\theta)}\mathbf{\Delta}_{h(\theta)}^{-1}\widehat{\mathbf{F}}[\alpha + \theta]\mathbf{\Delta}_{h(\theta)},$$

which is the same as the r.h.s., cf. the proof of Proposition 4.5.

It follows that for $G \in \mathcal{F}_T$,

$$\mathbb{P}_i(G) = \mathbb{E}_{\theta;i}\left[\frac{1}{L_T}; G\right].$$

In particular, if $G \in \mathcal{F}_\tau, G \subseteq \{\tau \ge T\}$, we have $G \in \mathcal{F}_T$ and get

$$\mathbb{P}_i(G) = \mathbb{E}_{\theta;i}\mathbb{E}_{\theta;i}\left[\frac{I(G)}{L_T} \mid \mathcal{F}_\tau\right]$$

$$= \mathbb{E}_{\theta;i}\left[I(G)\mathbb{E}_{\theta;i}\left[\frac{1}{L_T} \mid \mathcal{F}_\tau\right]\right]$$

$$= \mathbb{E}_{\theta;i}\left[\frac{1}{L_\tau}; G\right],$$

using (10.3). Now consider a general $G \in \mathcal{F}_\tau$. Then $G_T = G \cap \{\tau \le T\}$ satisfies $G_T \in \mathcal{F}_\tau$, $G_T \subseteq \{\tau \le T\}$. Thus, according to what has just been proved, (4.3) holds with G replaced by G_T. Letting $T \uparrow \infty$ and using monotone convergence then shows that (4.3) holds for G as well. □

References

[1] Anantharam, V., How large delays build up in a $GI/G/1$ queue, *Queueing Systems* (1989).

[2] Anick, D., Mitra, D., Sondhi, M.M., Stochastic theory of a data-handling system with multiple sources, *Bell System Tech. J.* **61**, (1982), 1871-1894.

[3] Arfwedson, G., Research in collective risk theory. The case of equal risk sums, *Skand.*

Aktuar Tidskr. **38**, (1955), 53-100.

[4]   Asmussen, S., Conditioned limit theorems relating a random walk to its associate, with applications to risk reserve processes and the $GI/G/1$ queue, *Adv. Appl. Probab.* **14**, (1982), 143-170.

[5]   Asmussen, S., *Applied Probability and Queues*, John Wiley & Sons, Chichester, New York 1987.

[6]   Asmussen, S., Risk theory in a Markovian environment, *Scand. Act. J.* **1989**, (1989), 69-100.

[7]   Asmussen, S., Ladder heights and the Markov-modulated $M/G/1$ queue, *Stoch. Proc. Appl.* **37**, (1991), 313-326.

[8]   Asmussen, S., Stationary distributions for fluid flow models with or without Brownian noise, *Stochastic Models* (under revision), (1992).

[9]   Asmussen, S., Perry, D., On cycle maxima, first passage problems and extreme value theory for queues, *Stochastic Models* **8**, (1992), 421-458.

[10]   Asmussen, S., Rolski, T., Risk theory in a periodic environment: Lundberg's inequality and the Cramér-Lundberg approximations, *Math. Oper. Res.* (1993/4).

[11]   Barlow, M.T., Rogers, L.C.G., Williams, D., Wiener-Hopf factorization for matrices, *Seminaire de Probabilités* **XIV**, Springer Lecture Notes in Mathematics **784**, (1980), 324-331.

[12]   Bellman, R., On a generalization of a fundamental identity of Wald, *Proc. Camb. Philos. Soc.* **53**, (1957), 257-259.

[13]   Borovkov, A.A., *Stochastic Processes in Queueing Theory*, Springer-Verlag, New York 1976.

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

[15]   Chang, C.-S., Heidelberger, P., Juneja, S., Shahabuddin, P., Effective bandwidth and fast simulation of ATM intree networks, *Performance Evaluation* (1993).

[16]   Çinlar, E., Markov additive processes. II, *Z. Wahrscheinlichkeitsth. verw. Geb.* **24**, (1972), 93-121.

[17]   Cottrell, M., Fort, J.C., Malgouyres, C., Large deviations and rare events in the study of stochastic algorithms, *IEEE Trans. on Auto. Contr.* **AC-28**, (1983), 907-920.

[18]   van Doorn, E., Regterschot, G.J.K., Conditional PASTA, *Oper. Res. Letters* **7**, (1988), 229-232.

[19]   Gaver, D.P., Lehoczky, J.P., Channels that cooperatively service a data stream and voice messages, *IEEE Trans. Com.* **30**, (1982), 1153-1161.

[20]   Gaver, D.P., Lehoczky, J.P., Performance evaluation of voice/data queueing systems, *Appl. Prob. — Comp. Sci.: The Interface* Vol. I, (ed. by T.J. Ott), (1982), 329-346.

[21]   Gerber, H.U., Martingales in risk theory, *Mitt. Ver. Schweiz. Vers. Math.* **73**, (1973), 205-216.

[22]   Gnedenko, B.V., Kovalenko, I.N., *Introduction to Queueing Theory* (2nd ed.), Birkhäuser, Basel (1989).

[23]   Graham, A., *Kronecker Products and Matrix Calculus with Application*, Ellis Horwood, Chichester, New York (1981).

[24]   Iglehart, D.L., Extreme values in the $GI/G/1$ queue, *Ann. Math. Statist.* **43**, (1972), 627-635.

[25]   Kennedy, J., Williams, D., Probabilistic factorization of a quadratic matrix polynomial, *Math. Proc. Camb. Phil. Soc.* **107**, (1990), 591-600.

[26]   Kesidis, G., Walrand, J., Quick simulation of ATM buffers, *Proc. of the IEEE CDC'92 Conf.*, Tucson, Arizona, (1992), 1018-1019.

[27]   Kingman, J.F.C., A convexity property of positive matrices, *Quart. J. Math. Oxford* **12**, (1961), 283-284.

[28]   Karandikar, R.L., Kulkarni, V.G., Second-order fluid flow model of a data-buffer in

random environment, *submitted.*

[29] Keilson, J., Wishart, D.M.G., A central limit theorem for processes defined on a finite Markov chain, *Proc. Camb. Philos. Soc.* **61**, (1967), 173-190.

[30] Keilson, J., Wishart, D.M.G., Addenda to processes defined on a finite Markov chain. *Proc. Camb. Philos. Soc.* **63**, (1967), 187-193.

[31] Kunita, H., Absolute continuity of Markov processes, *Seminaire de Probabilités* **X**, Lecture Notes in Mathematics **511**, (1976), Springer-Verlag, 44-77.

[32] Miller, H.D., A generalization of Wald's identity with applications to random walks, *Ann. Math. Statist.* **32**, (1957), 268-285.

[33] Neuts, M.F., *Matrix-Geometric Solutions in Stochastic Models*, Johns Hopkins University Press, Baltimore 1981.

[34] Parekh, S., Walrand, J., Quick simulation of excessive backlogs in networks of queues. In *Stoch. Diff. Syst., Stoch. Contr. The. and Appl.*, IMA Vol. **10**, (ed. by W. Fleming and P.L. Lions), Springer-Verlag, New York (1988), 439-472.

[35] Rogers, C.L.G., Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains, *submitted.*

[36] Segerdahl, C.-O., When does ruin occur in the collective theory of risk? *Skand. Aktaur Tidsskr.* **1955**, (1955), 22-36.

[37] Stoyan, D., *Comparison Methods for Queues and Other Stochastic Models* (ed. by D.J. Daley), John Wiley & Sons, New York 1983.

[38] Stern, T.E., Elwalid, A.I., Analysis of separable Markov-modulated rate models for information-handling systems, *Adv. Appl. Probab.* **23**, (1991), 105-1991.

[39] Takács, L., *Introduction to the Theory of Queues*, Oxford University Press, New York 1962.

[40] Takács, L., Application of Ballot theorems in the theory of queues, *Proc. Symp. on Congestion Theory* (ed. by W.L.Smith & W.E. Wilkinson), University of North Carolina Press, Chapel Hill 1965.

[41] Tweedie, M.C.K., Generalization of Wald's fundamental identity of sequential analysis to Markov chains, *Proc. Camb. Philos. Soc.* **56**, (1960), 205-214.