Queueing Syst (2014) 78:225–254
DOI 10.1007/s11134-014-9399-5

Hitting time in Erlang loss systems with moving boundaries

Martin Nilsson

Received: 16 April 2013 / Revised: 24 February 2014 / Published online: 10 April 2014

Abstract When the boundary—the total number of servers— in an Erlang loss system is a function of time, customers may also be lost due to boundary variations. On condition that these customers are selected independently of their history, we solve for the hitting-time distribution and transient distribution of busy servers. We derive concise asymptotic expressions in the time domain for normal loads in the heavy-traffic limit, i.e., when the offered load ρ is high, and the number of servers scales as $\rho + O\left(\sqrt{\rho}\right)$. The solutions are computationally efficient, and simulations confirm the theoretical results.

Keywords First passing time · Spectral decomposition · Charlier polynomial · Hermite function · Diffusion · Colored noise

Mathematics Subject Classification Primary: 60K25 Queueing theory · 90B22 Queues and service · Secondary: 60J80 Branching processes

1 Introduction

1.1 The problem

We consider Erlang loss systems [5,16,37,45], which are queueing systems with Poisson arrivals of constant intensity λ, q servers with stationary i.i.d. exponential service durations, and no waiting, also described as *M/M/q/q* in Kendall notation. If a customer arrives when all servers are busy, the customer is lost. Erlang loss systems

M. Nilsson (✉)
SICS (Swedish Institute of Computer Science), POB 1263, 164 29 Kista, Sweden
e-mail: hitting.time@drnil.com

are birth–death processes [16, Sect. XVII.7], or in the terminology of [9, Sect. 4.3], immigration–death processes.

The *hitting time* or *first passing time* for an Erlang loss system is the time T from a given initial probability distribution $\boldsymbol{\pi}(0)$ of states until the first loss. The cumulative distribution function $F(T)$ or equivalently, the survivor function or tail distribution $S(T) = 1 - F(T)$, which depend on $\boldsymbol{\pi}$, q, and λ, and the mean service rate μ (defined as the reciprocal of the mean service duration) are exhaustive descriptions of the hitting time. For instance, the expected hitting time is $E[T] = \int_0^\infty S(T)\, dT$. Our goal is to find expressions for $S(T)$ valid for all $T \geq 0$.

Many papers have been published on the transient behavior of Erlang loss systems with *constant* boundaries. The problem of finding $F(T)$ has been solved in great generality for birth–death processes by Laplace and Stieltjes transforms [20–22]. The hitting time for Erlang loss systems specifically has been obtained directly by Laplace transforms [16, Sect. XIV.9], [37, Sect. 5.2]. The resulting Laplace transforms are usually difficult to invert analytically, but have been successfully inverted numerically [1]. Several authors have proposed approximate asymptotic solutions based on large deviations theory [30,32,42], singular perturbations [24,50], or diffusion (stochastic differential equations) [4,43]. An elegant exact method has been proposed, directly manipulating the master equations [48]. Hitting times have also been computed by combinations of approximate methods and simulation [40].

The hitting time to a *moving* boundary, i.e., when the boundary q is a function of time, is a classic problem in diffusion, e.g. [6,12], [35, Sect. 4.7], but has been little studied for Erlang loss systems. Unfortunately, diffusion methods do not easily carry over to Erlang loss systems, since the common diffusion assumption of white or uncorrelated noise is violated. Although there are techniques for handling correlated noise in diffusion systems, such schemes are involved [18], [38, Sect. S.10], [47, Sect. XVI.6] and do not generally offer analytic solutions.

1.2 Outline of the general method

We start out by expressing the transient distribution $\boldsymbol{\pi}$ of busy servers as a Markov process in a matrix formulation [9,21,27], but instead of using transform or diffusion methods, we elaborate a representation in terms of orthogonal polynomials. The ensuing reduction in complexity enables us to derive both exact and asymptotic concise formulas for $\boldsymbol{\pi}$ and $S(T)$.

The reduction proceeds in five principal steps: *first*, we observe that the losses due to boundary changes can be encoded by zero-padded matrix multiplications, allowing an expression of $S(T)$ as a product (6) of matrix exponentials. *Second*, we perform a spectral decomposition of the generator matrices, leading to (12). The benefit of spectral decomposition is that the eigenmodes decay independently, and their decay can be easily computed. *Third*, we use the Christoffel–Darboux identity [8,44] to simplify the matrix products, using expressions of the matrix elements in terms of orthogonal polynomials and their zeros (Theorem 1). The main effort lies in the *fourth* step, which uses asymptotic transitions from orthogonal polynomials to special functions in order to find asymptotic limits for the factors of the matrix products (Theorems 6, 8, 9). Most important here is that a boundary change leads to a perturbation of the eigenmodes that we can compute to *second-order* ($O\left(1/\rho\right)$) accuracy (Theorem 6). The *fifth* step finally integrates the sequence of matrix products into a differential equation for the spectral representation $\boldsymbol{\omega}$ of the transient distribution $\boldsymbol{\pi}$ (Theorem 10). This loses an order ($O\left(1/\sqrt{\rho}\right)$) of accuracy, but since the previous step left a second-order error, the final expression (3) of $S(T)$ remains asymptotically correct.

Instrumental in the derivation are recently shown asymptotic transitions [33] from Charlier polynomials [7,8,14,34,44], and their derivatives to the Hermite function $H_\nu(\cdot)$ [26, Sect. 10]. This function is related to the parabolic cylinder functions $D_\nu(\cdot)$ and $U(\cdot,\cdot)$ by

$$H_\nu\left(z\right) = e^{z^2/2}2^{-\nu/2}D_\nu\left(z\sqrt{2}\right) = e^{z^2/2}2^{-\nu/2}U(-\nu-1/2, z\sqrt{2}).$$

The Hermite function is an analytic extension of the Hermite polynomials $H_n\left(z\right)$, n integer, and satisfies many of the polynomials' convenient recursion rules—a fact we use extensively in the following.

Although a transition from Charlier polynomials to Hermite *polynomials* was given over seventy years ago [31,44], and many related works exist, e.g. [11,15,29,39] and the references therein, results pertaining to the more general transition to the Hermite *function* are rare. A pointwise transition for non-negative real ν is implicit in [10], but beyond that the present work requires uniform transitions for both the Charlier polynomials and their derivatives, together with error rates, as well as the convergence of Charlier polynomial zeros to Hermite function zeros shown in [33]. Interestingly, some expressions for the hitting time to constant boundaries for Ornstein–Uhlenbeck processes [2,36] also involve zeros of the Hermite function, but in these cases derived from the Laplace transform of the probability density function.

1.3 Assumptions

We define the offered load $\rho = \lambda/\mu$ and the scaled boundary distance

$$\sigma\left(\rho; t\right) = \frac{q\left(\rho; t\right) - \rho}{\sqrt{\rho}},$$

which can be seen intuitively as the distance to the boundary from the mean number of busy servers, in the unit of standard deviations. The factor $\sqrt{\rho}$ originates in the variability of Poisson traffic. The traffic intensity is ρ/q. We assume a simultaneous approach to infinity by ρ and $q = \rho + O\left(\sqrt{\rho}\right)$ often referred to as the Halfin–Whitt regime [17], but this has also been used previously [3,19]. The rationale is that under normal loads, σ converges when the offered load ρ grows large. An illustrative example of such a boundary q is

$$q_{\mathrm{EXP}}(t) = \lfloor \rho + \left(3e^{-t} + 1\right)\sqrt{\rho} \rfloor, \tag{1}$$

which is shown in Fig. 1. We will return to this example several times below.

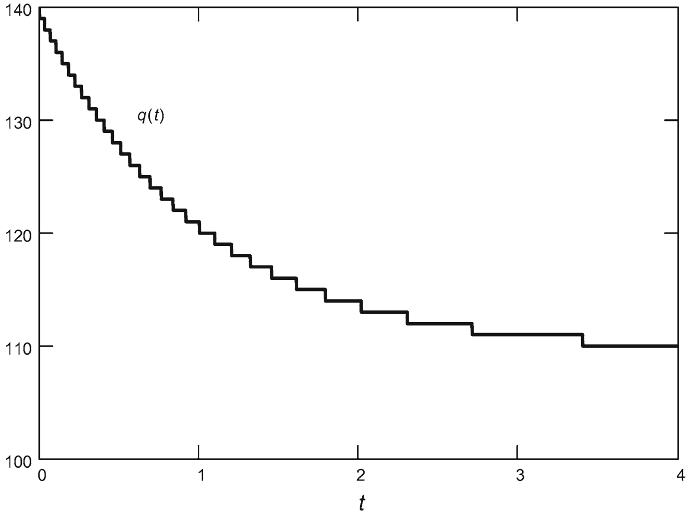

Fig. 1 In an Erlang loss system with moving boundaries, losses are caused by boundary changes (*vertical sections*) in addition to temporal decay (*horizontal sections*)–in this example, $q(t) = \lfloor \rho + \left(3e^{-t} + 1\right) \sqrt{\rho} \rfloor$ and $\rho = 100$

The boundary q is integer-valued and therefore piecewise constant. We assume that (**A1**) for fixed ρ, q has only a finite number of discontinuities in any finite interval. This is necessary for the existence of the finite sums and products used in the derivation. The timing of the jumps depends on ρ.

For the asymptotic case, we require (**A2**) that when ρ approaches infinity, the time intervals Δt between changes of q scale as $O\left(1/\sqrt{\rho}\right)$, and that $\sigma(\rho; t)$ converges to a limit $\sigma(t)$. We will often abbreviate $\sigma(t)$ as σ_t or σ for improved readability. Also, $\sigma(\rho; t)$ may be abbreviated as σ, unless there is a risk for confusion. The steps $\Delta\sigma$ in $\sigma(q; t)$ shrink as $O(1/\sqrt{\rho})$, while the number of steps in q grows as $O\left(\sqrt{\rho}\right)$ during any finite time interval. In the q_{EXP} example, $\sigma_t = 3e^{-t} + 1$.

We generally assume (**A3**) that σ_t is continuously differentiable in the intervals of interest, and that each boundary change $|\Delta q|$ is a unit change, $|\Delta q| = 1$, except in Sect. 4, where we relax this assumption in order to study isolated discontinuities in σ_t.

When customers are lost due to boundary changes, we assume (**A4**) that such customers are selected independently and uniformly at random, i.e., regardless of their history. This is necessary for describing the number of busy servers as a Markov process.

1.4 Overview of the main results

In the following, all vectors are column vectors, and are given in lower-case boldface. The notations **0** and **1** are used for vectors of zeros and ones, respectively. Matrices are written in upper-case boldface.

Let $\boldsymbol{\omega}^T = \boldsymbol{\pi}^T\mathbf{M}$ be the spectral representation of the transient distribution $\boldsymbol{\pi}$, where the raised T indicates transpose. The matrix $\mathbf{M}$ derives from the diagonalization of the Markov generator matrix, and is given in explicit form by Theorem 9. The main result is that asymptotically, $\boldsymbol{\omega}$ satisfies the matrix differential equation

$$\frac{\mathrm{d}\boldsymbol{\omega}^T}{\mathrm{d}t} = \boldsymbol{\omega}^T \left(\mathbf{C}\left|\frac{\mathrm{d}\sigma_t}{\mathrm{d}t}\right| - \mu\boldsymbol{\Lambda}\right) \tag{2}$$

during time intervals, including infinite, where σ_t is continuously differentiable. Here, $\boldsymbol{\Lambda} = \operatorname{diag}\left(\nu^j\right)$ is the diagonal matrix of positive zeros $\nu^j(\sigma)$, $j = 1, 2, \ldots$, of the expression $H_\nu(-\sigma/\sqrt{2})$ in ascending order, and $\mathbf{C}$ is an eigenmode perturbation matrix defined in Theorem 10. Intuitively speaking, the matrices $\boldsymbol{\Lambda}$ and $\mathbf{C}$ represent losses due to temporal decay and boundary changes, respectively.

The expansion of the survivor function is

$$S(T) = \boldsymbol{\omega}(T)^T\mathbf{M}^{-1}\mathbf{1}, \tag{3}$$

expressing that $S(T)$ equals the sum of the components of $\boldsymbol{\pi}(T)$ not including the absorbing state, i.e., lost customers. Theorem 10 also gives an exact formula for finite ρ, but in this case, $\boldsymbol{\omega}(t)$ is a sequence of matrix multiplications.

In practice, truncating the system (2) to a small number of dimensions (eigenmodes) usually produces sufficient accuracy. If the initial distribution $\boldsymbol{\pi}(0)$ contains only the fundamental eigenmode, such as after a long period of constant boundary, and the boundary changes slowly, meaning $|\mathrm{d}\sigma/\mathrm{d}t| \lessapprox 2\mu\ \exp(-\sigma/\sqrt{2})$, then the survivor function can be written as

$$S\left(T\right) \approx \frac{G(\sigma_T)}{G(\sigma_0)} \exp\left(-\int_0^T \mu\nu^1\left(\sigma_t\right)\,\mathrm{d}t\right), \tag{4}$$

where

$$G(\sigma) = \frac{1}{\sqrt[4]{2\pi}} \exp\left(-\frac{\sigma^2}{4}\right)\sqrt{\frac{\mathrm{d}}{\mathrm{d}\sigma}\frac{1}{\nu^1\left(\sigma\right)}}.$$

The improvement by (4) over the naïve approximation $S(T) \approx \exp\Big(-\int_0^T \mu\nu^1\mathrm{d}t\Big)$, which is exact for a constant boundary and a fundamental-only initial state, is the factor $G(\sigma_T)/G(\sigma_0)$. The truncation error can be estimated by truncating (2) to a few different dimensions and comparing the results.

In 2001, the general problem of the first passage to a moving boundary with non-monotonic motion was described as still appearing relatively open [35, p. 120]. The hitting time of a Wiener process to a continuous, piecewise linear boundary was given by [49]. The hitting time of an Ornstein–Uhlenbeck process to a general moving boundary can be expressed as an integral equation [6], and many papers have appeared on the efficient numerical solution of this equation. In our case, although the Cauchy problem (2) does not in general admit a closed form for the unknown, it can be solved quickly and efficiently by standard numerical techniques. At this level of simplicity, (2) and (3) seem to constitute, if not the first, at least a rare asymptotic solution to a hitting-time problem with a general boundary.

1.5 Structure of the paper

In the next section, we introduce notation, review some well-known results, and decompose the probability distribution $\boldsymbol{\pi}$ into eigenmodes associated with the eigenvectors of a Markov generator matrix. In Sect. 3, we find explicit expressions and bounds for scalar products of the eigenvectors. Theorem 6 in this section is a highlight and a key to the compact formulation of the probability distribution. We use the expressions from Sect. 3 for mapping between eigenmodes and probability in Sect. 5. After preparing the ground in Sects. 3 and 5, the derivation of the eigenmode time evolution in Sect. 6 becomes straightforward. Section 4 analyzes discontinuities in σ_t, a special case of practical importance. Section 7 presents simulations confirming the theory. In Sect. 7.4, we briefly probe a variation of the classical Lee–Longton approximation [28], approximating an $M/G/q_t/q_t$ system by an $M/M/q_t/q_t$ system. Section 8 concludes the paper.

2 Preliminaries

2.1 Notation

The notation "$a \triangleq b$" is to be read as "a is defined b", and is used for highlighting the introduction of new symbols. The expression $e^{\mathbf{A}}$ denotes the matrix exponential $\sum_{k=0}^{\infty} \mathbf{A}^k/k!$.

For a constant boundary, the number of busy servers is a Markov process when the service duration has a stationary exponential distribution. Let the vector $\boldsymbol{\pi}^*(t)$, $t \geq 0$, be the probability distribution of numbers of busy servers as a function of t. The asterisk indicates the inclusion of an absorbing state for lost customers. The time evolution of a continuous-time Markov process with constant generator matrix $\mathbf{Q}^*$ can be written as [9, Sect. 4.5]

$$\boldsymbol{\pi}^{*T}(t) = \boldsymbol{\pi}^{*T}(0)\, e^{\mathbf{Q}^* t}.$$

Truncating $\boldsymbol{\pi}^*$ to $\boldsymbol{\pi}$ and $\mathbf{Q}^*$ to $\mathbf{Q}$ by removing the element, row, and column representing the absorbing state, the survivor function can then be obtained by the matrix formula

$$S(t) = \boldsymbol{\pi}^T(t)\, \mathbf{1} = \boldsymbol{\pi}^T(0)\; e^{t\mathbf{Q}}\mathbf{1}, \tag{5}$$

expressing the probability of no loss or absorption before, i.e., survival until epoch t.

For simplicity of notation, we introduce the convention that matrices in matrix products with incompatible dimensions are padded below or to the right with zero rows or columns as required, except for diagonal matrices that keep their diagonal elements and are merely truncated to size. For instance, with

$$\mathbf{U}_1 \triangleq \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad \mathbf{U}_2 \triangleq \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}, \quad \mathbf{D} \triangleq \begin{pmatrix} d_1 & 0 & \cdots \\ 0 & d_2 & \\ \vdots & & \ddots \end{pmatrix}$$

the product $\mathbf{U}_1\mathbf{D}\mathbf{U}_2$ would be interpreted as

$$\mathbf{U}_1\mathbf{D}\mathbf{U}_2 = \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}.$$

The boundary $q(t)$ is integer-valued, so necessarily piecewise constant. The number of busy servers is a Markov process for such a q, since passage to the absorbing state only depends on the current state and not on the history (**A4**). The boundary q corresponds to a piecewise constant generator matrix $\mathbf{Q}$, i.e., a sequence of constant matrices $\mathbf{Q}_1,\ \mathbf{Q}_2, \ldots \mathbf{Q}_n$ for the respective durations $\Delta t_1, \Delta t_2, \ldots \Delta t_n$, where $T = \sum_k \Delta t_k$. We then have the generalized formula

$$S(T) = \boldsymbol{\pi}^T(0)\ e^{\Delta t_1 \mathbf{Q}_1} e^{\Delta t_2 \mathbf{Q}_2} \ldots\ e^{\Delta t_n \mathbf{Q}_n} \mathbf{1}. \tag{6}$$

As it reads, (6) is rather awkward to compute, but the special structure of $\mathbf{Q}$ for a birth–death process can be used in order to radically simplify the equation.

2.2 Spectral decomposition for moving boundaries

For Erlang loss systems, the generator matrix $\mathbf{Q}$ can be written as [37, p. 88]

$$\mathbf{Q} = \mu \begin{pmatrix} -\rho & \rho & & & & \\ 1 & -\rho-1 & \rho & & & \\ & 2 & -\rho-2 & \rho & & \\ & & 3 & & & \\ & & & & \cdots & \rho \\ & & & & q & -\rho-q \end{pmatrix}.$$

The eigenvalues of the matrix $-\mathbf{Q}/\mu$ are precisely the zeros of the Charlier polynomials $C_{q+1}(\rho, x)$ [23,27], a family of orthogonal polynomials, here defined as the *Erdélyi-normalized* Charlier polynomials [14, pp. 226–227], [25,34],

$$C_n(\rho, x) \triangleq \sum_{k=0}^{n} \binom{n}{k} \binom{x}{k} k!(-\rho)^{-k}.$$

For compactness, we will write $c_n^\rho(x)$ or $c_n(x)$ or even just c_n for $C_n(\rho, x)$ when there is no risk for confusion. Another variant of Charlier polynomial we also use below is the *Szegő-normalized* orthonormal Charlier polynomials [44, p. 35],

$$s_k^\rho(x) \triangleq (-1)^k \sqrt{\frac{\rho^k}{k!}}\, c_k^\rho(x). \tag{7}$$

The zeros of the Charlier polynomial $c_q^\rho(x)$ are real, positive, and simple [23]. If the zeros are numbered $\nu_q^1 < \nu_q^2 < \ldots < \nu_q^q$, then the sequence of zeros $\left\{\nu_q^j\right\}_{q=j}^\infty$ is decreasing [8, Sect. 1.5], and the zeros of $c_q^\rho(x)$ and $c_{q+1}^\rho(x)$ mutually separate each other (the *separation property*), i.e., $\nu_{q+1}^j < \nu_q^j < \nu_{q+1}^{j+1}$ for $1 \le j \le q$ [*ibid.*]. The distance between adjacent zeros is greater than one [23],

$$\nu_q^{j+1} - \nu_q^j > 1. \tag{8}$$

A right eigenvector of $-\mathbf{Q}/\mu$ corresponding to zero ν_{q+1}^j, $j = 1, \ldots q+1$, is [27, Chap. 1]

$$\mathbf{w}_q^j \triangleq \left(c_0^\rho(\nu_{q+1}^j)\ \ c_1^\rho(\nu_{q+1}^j)\ \ldots\ c_q^\rho(\nu_{q+1}^j)\right)^T.$$

In the following, superscripts of vectors and matrices will often be used to indicate order, and subscripts to indicate degree of the corresponding Charlier polynomials, but may be dropped when there is no risk for misunderstanding. In the rare case where a power of a zero is intended, it will be enclosed in parentheses, e.g., $\left(\nu^1\right)^2$ for the square of the first zero ν^1.

The similarity transformation $\mathbf{DQD}^{-1}$ with $\mathbf{D} \triangleq \operatorname{diag}\left(\sqrt{\rho^k/k!}\right)$, $k = 0, \ldots q$, transforms $\mathbf{Q}$ to symmetric form, so it can be written as

$$\mathbf{DQD}^{-1} = -\mathbf{U}\mu\boldsymbol{\Lambda}\mathbf{U}^T,$$

where $\boldsymbol{\Lambda} \triangleq \operatorname{diag}\left(\nu^j\right)$ is the diagonal matrix of eigenvalues of $-\mathbf{Q}/\mu$, sorted in increasing order, and $\mathbf{U}$ is an orthonormal matrix of right eigenvectors of $\mathbf{DQD}^{-1}$. Obviously, $\mathbf{D}^{-1}\mathbf{U}$ is a matrix of right eigenvectors for $-\mathbf{Q}/\mu$. On the other hand, so is $\mathbf{W} \triangleq \left(\mathbf{w}^1\ \mathbf{w}^2 \ldots \mathbf{w}^{q+1}\right)$, and accordingly, we can choose $\mathbf{U}$ by normalizing the columns of $\mathbf{DW}$, i.e., by letting

$$\mathbf{u}_q^p \triangleq \frac{\mathbf{Dw}_q^p}{\left|\mathbf{Dw}_q^p\right|} \tag{9}$$

be the pth column vector of $\mathbf{U}$. The survivor function for a constant boundary in (5) now reads

$$S(t) = \boldsymbol{\pi}^T(0)\ \mathbf{D}^{-1}\mathbf{U}\, e^{-t\mu\boldsymbol{\Lambda}}\, \mathbf{U}^T\mathbf{D1}. \tag{10}$$

The busy server count probability distribution $\boldsymbol{\pi}(t)$ has $q+1$ components, or eigenmodes, that decay over time due to the losses. The zero ν^j determines how quickly mode j decays. The dominant zero ν^1 corresponds to the fundamental mode, with the slowest decay.

In case $\boldsymbol{\pi}(0)$ contains the fundamental mode only, then, since $\mathbf{u}^1 = \mathbf{D}\mathbf{w}^1 / \left|\mathbf{D}\mathbf{w}^1\right|$,

$$\boldsymbol{\pi}(0) = \frac{\mathbf{D}\mathbf{u}^1}{\sum \mathbf{D}\mathbf{u}^1} = \frac{\mathbf{D}^2\mathbf{w}^1}{\mathbf{1}^T\mathbf{D}^2\mathbf{w}^1},$$

so that

$$\boldsymbol{\pi}^T(0)\,\mathbf{D}^{-1}\mathbf{u}^1 = \frac{\mathbf{w}^{1T}\mathbf{D}^2}{\mathbf{1}^T\mathbf{D}^2\mathbf{w}^1}\mathbf{D}^{-1}\frac{\mathbf{D}\mathbf{w}^1}{\left|\mathbf{D}\mathbf{w}^1\right|} = \frac{\left|\mathbf{D}\mathbf{w}^1\right|}{\mathbf{1}^T\mathbf{D}^2\mathbf{w}^1}. \tag{11}$$

Suppose now that a time-variable boundary $q = q(t)$ in (6) is constant during time Δt_k, then changes, and is held constant again during time Δt_{k+1}. The loss caused by the boundary change is then expressed precisely by a matrix multiplication, which zeroes out the border probabilities. Define $\mathbf{M}_k \triangleq \mathbf{D}^{-1}\mathbf{U}_k$. The sequence (6) of step changes from boundary $q(0)$ to $q(T)$ translates into

$$S(T) = \boldsymbol{\pi}^T(0)\,\mathbf{M}_1\left[\prod_{k=1}^{n} e^{-\Delta t_k \mu \boldsymbol{\Lambda}_k}\mathbf{U}_k^T\mathbf{U}_{k+1}\right]\mathbf{M}_{n+1}^{-1}\mathbf{1}. \tag{12}$$

The intuition behind the right-hand side is that the factor $\mathbf{M}_1$ translates the prior probability distribution $\boldsymbol{\pi}(0)$ into eigenmodes $\boldsymbol{\omega}^T = \boldsymbol{\pi}^T(0)\,\mathbf{M}_1$; the factors $e^{-\Delta t_k \mu \boldsymbol{\Lambda}_k} = \operatorname{diag}\left\{\exp\left(-\Delta t_k \mu \nu_{q+1}^j\right)\right\}$, $j = 1, \ldots, q+1$, decay each mode temporally and independently; the products $\mathbf{U}_k^T\mathbf{U}_{k+1}$ express the perturbations of the eigenmodes due to the boundary change; the factor $\mathbf{U}_{n+1}{}^T\mathbf{D} = \mathbf{M}_{n+1}^{-1}$ projects modes back to state probabilities; and the final $\mathbf{1}$ sums them up. Specifically, the fundamental mode projects to probability through multiplication by $\mathbf{u}^{1T}\mathbf{D}$, where $\mathbf{u}^1$ is the first column of $\mathbf{U}_{n+1}$. In order to simplify this expression for the survivor function, we will inspect the factors $\exp(-\Delta t_k \mu \boldsymbol{\Lambda}_k)$ and $\mathbf{U}_k^T\mathbf{U}_{k+1}$ in Sect. 3, and $\mathbf{M}$ in Sect. 5.

3 Scalar products of eigenvectors

Central to simplifying (12) is the evaluation of the matrix product $\mathbf{U}_k^T\mathbf{U}_{k+1}$. For a decreasing boundary, its elements are the scalar products $\mathbf{u}_q^{pT}\mathbf{u}_{q-1}^r$ for $1 \le p \le q+1$ and $1 \le r \le q$. The following theorem provides an explicit formula for decreasing boundaries. Anticipating Sect. 4, where we drop assumption (**A4**), we prove a slightly more general theorem, giving the scalar products $\mathbf{u}_q^{pT}\mathbf{u}_{s-1}^r$ for $1 \le p \le q+1$, $s \le q$, and $1 \le r \le s$. The theorem can also be used for the increasing boundary case $s-1 > q$, via the identity $\mathbf{u}_q^{pT}\mathbf{u}_{s-1}^r = \mathbf{u}_{s-1}^{rT}\mathbf{u}_q^p$.

Theorem 1 (Eigenvector scalar products) *Let* ν_{q+1}^{p} *be the pth zero of the Charlier polynomial* $c_{q+1}^{\rho}(x)$. *Define* $w_q^j \triangleq \left(c_0^{\rho}(\nu_{q+1}^j)\ c_1^{\rho}(\nu_{q+1}^j)\ldots\ c_q^{\rho}(\nu_{q+1}^j)\right)^T$, $\mathbf{D} \triangleq$ *diag* $\left(\sqrt{\rho^k/k!}\right)$, *and* $\mathbf{u}_q^p \triangleq \mathbf{D}\mathbf{w}_q^p/\left|\mathbf{D}\mathbf{w}_q^p\right|$. *Then, for the decreasing boundary case* $s \leq q$,

$$\mathbf{u}_q^{pT}\mathbf{u}_{s-1}^r = \frac{1}{\nu_s^r - \nu_{q+1}^p}\sqrt{\frac{\rho^s q!}{\rho^q s!}}\frac{c_s(\nu_{q+1}^p)}{c_q(\nu_{q+1}^p)}\sqrt{-\frac{c_q(\nu_{q+1}^p)c_{s+1}\left(\nu_s^r\right)}{c'_{q+1}(\nu_{q+1}^p)c'_s\left(\nu_s^r\right)}}. \tag{13}$$

Proof According to the well-known Christoffel–Darboux identity [8, p. 23], [44, p. 43], for Erdélyi-normalized polynomials $c_q(x)$ and Szegő-normalized Charlier polynomials $s_q(x)$ (7),

$$\sum_{k=0}^{q} s_k(x)\, s_k(y) = \frac{\rho^{q+1}}{q!}\,\frac{c_{q+1}(x)\, c_q(y) - c_q(x)\, c_{q+1}(y)}{y-x}. \tag{14}$$

By this relation, and due to the fact that $c_k\left(\nu_k^r\right) = 0$ for all k, by (14),

$$\left(\mathbf{D}\mathbf{w}_q^p\right)^T\mathbf{D}\mathbf{w}_{s-1}^r = \sum_{k=0}^{s} s_k(\nu_{q+1}^p)s_k\left(\nu_s^r\right) = \frac{\rho^{s+1}}{s!}\,\frac{-c_s(\nu_{q+1}^p)c_{s+1}\left(\nu_s^r\right)}{\nu_s^r - \nu_{q+1}^p}.$$

Since $c_q\left(\nu_q\right) = 0$, by the limiting case $x \to y$ of the Christoffel–Darboux identity,

$$\left|\mathbf{D}\mathbf{w}_q^p\right|^2 = \sum_{k=0}^{q} s_k(\nu_{q+1}^p)^2 = -\frac{\rho^{q+1}}{q!}c_q(\nu_{q+1}^p)c'_{q+1}(\nu_{q+1}^p),$$

and since $s_s(\nu_s) = c_s(\nu_s) = 0$,

$$\left|\mathbf{D}\mathbf{w}_{s-1}^r\right|^2 = \sum_{k=0}^{s-1} s_k\left(\nu_s^r\right)^2 = \sum_{k=0}^{s} s_k\left(\nu_s^r\right)^2 = \frac{\rho^{s+1}}{s!}c_{s+1}\left(\nu_s^r\right)c'_s\left(\nu_s^r\right).$$

The theorem follows from substituting these expressions into

$$\mathbf{u}_q^{pT}\mathbf{u}_{s-1}^r = \frac{\left(\mathbf{D}\mathbf{w}_q^p\right)^T\mathbf{D}\mathbf{w}_{s-1}^r}{\left|\mathbf{D}\mathbf{w}_q^p\right|\left|\mathbf{D}\mathbf{w}_{s-1}^r\right|} = \frac{\sum_{k=0}^{s} s_k(\nu_{q+1}^p)s_k\left(\nu_s^r\right)}{\sqrt{\sum_{k=0}^{q}\left[s_k(\nu_{q+1}^p)\right]^2\sum_{k=0}^{s-1}\left[s_k\left(\nu_s^r\right)\right]^2}}.$$

□

This is an exact result, valid for all values of ρ. These formulas are suitable for numerical computations up to $\rho \approx 100$. For larger ρ, it can be advantageous to use asymptotic forms, i.e., the limiting values when ρ approaches infinity. In the following, we shall derive such asymptotic formulas for (13), and show how expression (12) can be further simplified for large ρ. This requires three theorems on the limiting behavior of Charlier polynomials, their derivatives, and their zeros, proved in [33]. Since the asymptotic forms are all intimately related to the Hermite function, we will refer to them as the *Hermite limits*.

Theorem 2 (Transition of Charlier polynomials) *For real σ, ν, and positive ρ,*

$$(2\rho)^{\nu/2} c^{\rho}_{\lceil \rho+\sigma\sqrt{\rho}\rceil}(\nu) = H_\nu\left(-\frac{\sigma}{\sqrt{2}}\right) + O\left(\frac{1}{\sqrt{\rho}}\right),$$

where $c^{\rho}_n(\nu)$ are the Charlier polynomials and H_ν is the Hermite function. The error bound $O(1/\sqrt{\rho})$ is uniform for ν and σ in any bounded interval, and is sharp in the sense that there are ν and σ such that the error is proportional to $1/\sqrt{\rho}$ for arbitrarily large ρ.

Theorem 3 (Transition of Charlier polynomial derivative) *For real σ, ν, and positive ρ,*

$$\frac{\partial}{\partial \nu}\left\{(2\rho)^{\nu/2} c^{\rho}_{\lceil \rho+\sigma\sqrt{\rho}\rceil}(\nu)\right\} = \frac{\partial}{\partial \nu} H_\nu\left(-\frac{\sigma}{\sqrt{2}}\right) + O\left(\frac{1}{\sqrt{\rho}}\right).$$

The error bound is uniform for ν and x in any bounded interval, and is sharp in the same sense as in Theorem 2.

Theorem 4 (Convergence of Charlier polynomial zeros) *For fixed real σ and positive $\rho \to \infty$, let $q \triangleq \lceil \rho + \sigma\sqrt{\rho}\rceil$. For a convergent sequence of zeros $\nu_q \to \nu$ such that $c^{\rho}_q\left(\nu_q\right) = 0$, the limit ν is a zero of the Hermite function, $H_\nu\left(-\sigma/\sqrt{2}\right) = 0$, satisfying $\nu = \nu_q + O\left(1/\sqrt{\rho}\right)$. Conversely, for a positive real zero ν of the Hermite function, there is a convergent sequence $\nu_q \to \nu$ of zeros of c^{ρ}_q, satisfying $\nu = \nu_q + O\left(1/\sqrt{\rho}\right)$.*

The zeros $x(\nu)$ of $H_\nu(x)$ for positive real ν are strictly monotonic and differentiable [13], implying that so are the positive real zeros $\nu(x)$. An important entity in the expression for the survivor function is the *zero difference* $\Delta\nu_q \triangleq \nu_{q+1} - \nu_q$ between same-order zeros of the Charlier polynomials of adjacent degrees. Since the zeros are decreasing with increasing q, this difference is always negative. The following theorem provides the asymptotics of $\Delta\nu_q$:

Theorem 5 (Charlier polynomial zero difference) *For fixed σ and $\rho \to \infty$ in such a way that $q = \rho + \sigma\sqrt{\rho}$ is integer, the zeros ν_q defined by $c^{\rho}_q\left(\nu_q\right) = 0$ converge to ν satisfying $H_\nu\left(-\sigma/\sqrt{2}\right) = 0$. The following asymptotic relation holds uniformly for bounded ν and σ:*

$$\Delta\nu_q = \frac{-1}{\sqrt{2\rho}}\,\frac{H_{\nu+1}\left(-\sigma/\sqrt{2}\right)}{\partial H_\nu\left(-\sigma/\sqrt{2}\right)/\partial\nu}\left[1 + O\left(\frac{1}{\sqrt{\rho}}\right)\right] = \frac{\partial\nu}{\partial\sigma}\cdot\left[\frac{1}{\sqrt{\rho}} + O\left(\frac{1}{\rho}\right)\right].$$

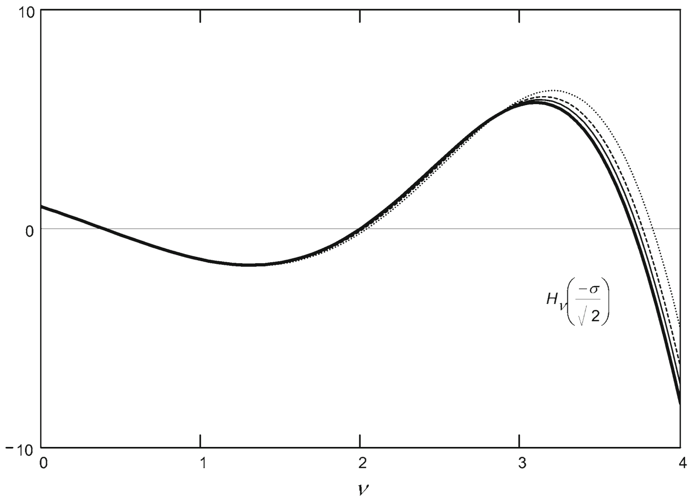

Fig. 2 There is a transition of Charlier polynomials $(2\rho)^{\nu/2}C_{\lceil\rho+\sigma\sqrt{\rho}\rceil}(\rho,\nu)$ (*thin lines*) to the Hermite function $H_\nu(-\sigma/\sqrt{2})$ (*thick line*). Here, $\sigma = 1$, and the different values of ρ are 100 (*dotted*), 400 (*dashed*), and 1,600 (*solid*)

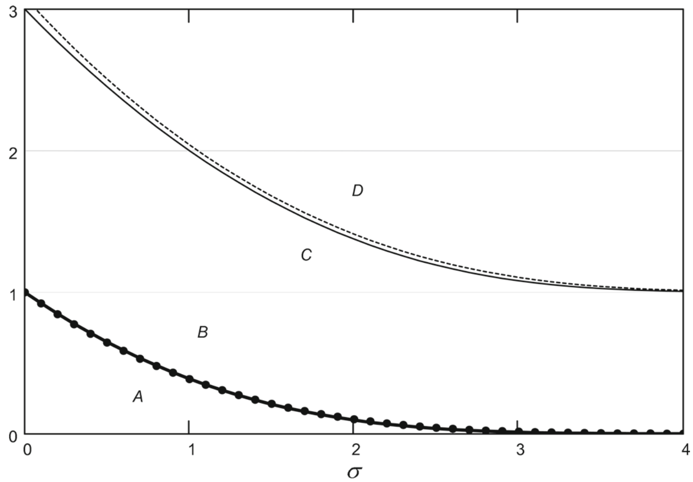

Fig. 3 *A*: The smallest positive zero in ν of Hermite function $H_\nu(-\sigma/\sqrt{2})$ (*lower, thick line*); *B*: Dominant zero of the corresponding Charlier polynomial (*lower, dotted line*) for $\rho = 100$; *C*: The second ero of Hermite function (*upper, thin line*) for $\rho = 100$; *D*: The second zero of the corresponding Charlier polynomial (*upper, dashed line*)

Figure 3 illustrates the distance between adjacent zeros as a function of σ, together with the Hermite limit given by Theorem 1. Before the proof of Theorem 5, we show some examples of the accuracy of the Hermite limit given by Theorems 2, 4, and 5 above.

Example 1 (Hermite limits) The transition of the Charlier polynomials to the Hermite function predicted by Theorem 2 is shown for $\rho = 100$, 400, and 1600 in Fig. 2. The

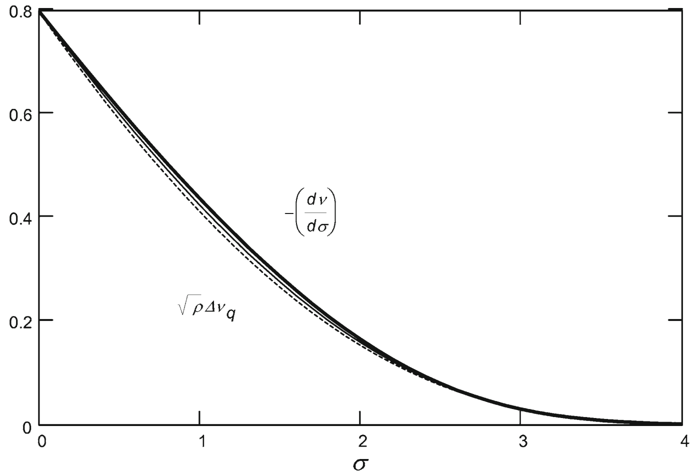

Fig. 4 Difference $\Delta\nu_q^1$ between dominant zeros for $\rho = 100$ (*dashed line*) and $\rho = 400$ (*thin line*) compared to Hermite limit $-d\nu^1/d\sigma$ (*thick line*)

first two zeros of the Charlier polynomials for $\rho = 100$ and the Hermite function are shown in Fig. 3. The zero difference $\Delta\nu_q$ is compared to the Hermite limit given by Theorem 5 in Fig. 4 for $\rho = 100$.

Proof When $\rho \to \infty$, the zeros ν_q converge to a limit ν_∞, since they are decreasing and positive, so the first part follows from Theorem 4. From this theorem it also follows that

$$\Delta\nu_q = \left(\nu_{q+1} - \nu_\infty\right) - \left(\nu_q - \nu_\infty\right) = O\left(\frac{1}{\sqrt{\rho}}\right).$$

In order to improve this bound, let $\sigma' \triangleq \sigma + 1/\sqrt{\rho}$, and define for arbitrary ν and σ,

$$y\left(\nu, \sigma\right) \triangleq (2\rho)^{\nu/2} c^{\rho}_{\lceil \rho + \sigma\sqrt{\rho}\rceil}(\nu). \tag{15}$$

By the definition of ν_q and ν_{q+1}, $y\left(\nu_q, \sigma\right) = y\left(\nu_{q+1}, \sigma'\right) = 0$. A zero of Charlier polynomial c^{ρ}_{q+1} cannot be a zero of polynomial c^{ρ}_{q}, implying that $y\left(\nu_{q+1}, \sigma\right) \neq 0$, so

$$\Delta\nu_q = \frac{y\left(\nu_{q+1}, \sigma\right) - y\left(\nu_{q+1}, \sigma'\right)}{\left[\dfrac{y\left(\nu_{q+1}, \sigma\right) - y\left(\nu_q, \sigma\right)}{\Delta\nu_q}\right]}. \tag{16}$$

Since y is an entire function of ν, the denominator in (16) can be Taylor expanded. In this expansion, the derivative $\partial y\left(\nu, \sigma\right)/\partial\nu$ is also entire, and by Theorem 3, converges uniformly to $\partial H_\nu\left(-\sigma/\sqrt{2}\right)/\partial\nu$ for bounded ν with rate $O\left(1/\sqrt{\rho}\right)$, so

$$\frac{y\left(\nu_{q+1},\sigma\right)-y\left(\nu_q,\sigma\right)}{\Delta\nu_q}=\frac{\partial H}{\partial\nu}\left(\nu_q,\sigma\right)+O\left(\Delta\nu_q\right), \tag{17}$$

where the error term is uniform for bounded σ. Although $y(\nu,\sigma)$ is an entire function of ν, it is not continuously differentiable with respect to σ, so the enumerator in (16) cannot be directly approximated by a derivative. Instead, we can use the backward recurrence rule for the Charlier polynomials [25],

$$c_q^\rho(\nu)-c_{q+1}^\rho(\nu)=\frac{\nu}{\rho}c_q^\rho\left(\nu-1\right),$$

implying that

$$y\left(\nu,\sigma\right)-y\left(\nu,\sigma'\right)=\frac{\nu\sqrt{2}}{\sqrt{\rho}}\,y\left(\nu-1,\sigma\right) \tag{18}$$

so for the enumerator, by Theorem 2, and substituting into (16),

$$\Delta\nu_q=\frac{\sqrt{2}}{\sqrt{\rho}}\,\frac{\nu_q\left[H_{\nu_q-1}\left(-\sigma/\sqrt{2}\right)+O\left(1/\sqrt{\rho}\right)\right]}{\partial H_{\nu_q}\left(-\sigma/\sqrt{2}\right)/\partial\nu+O\left(1/\sqrt{\rho}\right)}.$$

By the analyticity of the Hermite function,

$$H_{\nu_q-1}\left(-\sigma/\sqrt{2}\right)=H_{\nu-1}\left(-\sigma/\sqrt{2}\right)+O\left(\nu-\nu_q\right)$$

and similarly for $\partial H_\nu\left(-\sigma/\sqrt{2}\right)/\partial\nu$. Since $O\left(\nu-\nu_q\right)=O\left(1/\sqrt{\rho}\right)$ uniformly for bounded ν and $H_{\nu-1}\left(-\sigma/\sqrt{2}\right)\neq 0$ and $\partial H_\nu\left(-\sigma/\sqrt{2}\right)/\partial\nu\neq 0$ [13], and by the recurrence [26, p. 289]

$$H_{\nu+1}(x)-2xH_\nu(x)+2\nu H_{\nu-1}=0, \tag{19}$$

we obtain the first equality in Theorem 5. On the other hand, by differentiating the equation $H_\nu\left(-\sigma/\sqrt{2}\right)=0$, and using the rule $\partial H_\nu(z)/\partial z=2\nu H_{\nu-1}(z)$ [26, p. 289],

$$-\nu\sqrt{2}H_{\nu-1}\left(-\sigma/\sqrt{2}\right)+\frac{\partial H_\nu(-\sigma/\sqrt{2})}{\partial\nu}\frac{\partial\nu}{\partial\sigma}=0.$$

For brevity, with some abuse of notation, we have written $\partial H_\nu(-\sigma/\sqrt{2})/\partial\sigma$ for $\partial H_\nu(s)/\partial s$ at $s=-\sigma/\sqrt{2}$. After division by the factor $\partial H_\nu/\partial\nu$, we obtain the second equality, completing the proof. □

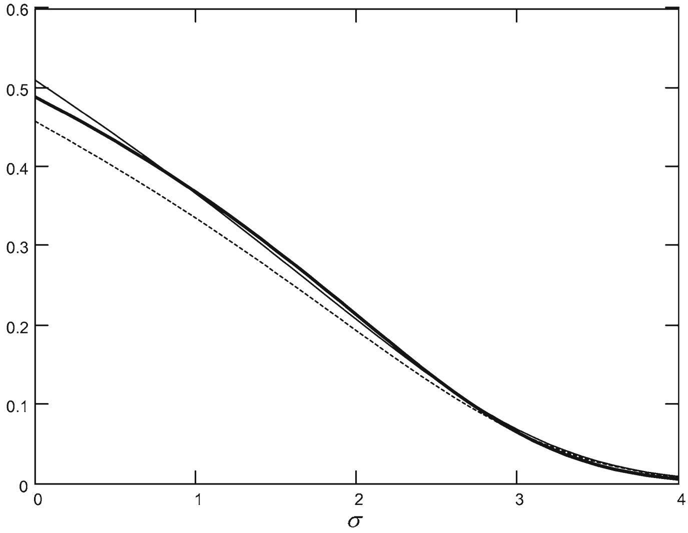

Fig. 5 Eigenvector scalar products $\mathbf{u}_q^{1T}\mathbf{u}_{q-1}^2\sqrt{\rho}$ (*thin, dashed*) and $-\mathbf{u}_q^{2T}\mathbf{u}_{q-1}^1\sqrt{\rho}$ (*thin, solid*) for $\rho = 100$, together with Hermite limit (*thick line*)

We now have asymptotic forms for Charlier polynomials $c(\nu)$ (Theorem 2), their derivatives $c'(\nu)$ (Theorem 3), zeros ν (Theorem 4), and zero differences $\Delta\nu$ (Theorem 5). In terms of these, we can express the asymptotics of the scalar product $\mathbf{u}_q^{pT}\mathbf{u}_{q-1}^r$. Theorem 6 computes the elements of the mode transition matrices. In particular, it shows that $\mathbf{u}_q^{pT}\mathbf{u}_{q-1}^p = 1$ with only a $O(1/\rho)$ error, i.e., *second-order* error.

Theorem 6 (Asymptotic eigenvector scalar products) *Let $\mathbf{u}_q^p$ be defined as in Theorem 1. When $p \neq r$,*

$$\mathbf{u}_q^{pT}\mathbf{u}_{q-1}^r = \frac{1}{\sqrt{\rho}\,(\nu^r - \nu^p)}\sqrt{\frac{d\nu^p}{d\sigma}\frac{d\nu^r}{d\sigma}} + O\left(\frac{1}{\rho}\right)$$

and when $p = r$, $\mathbf{u}_q^{pT}\mathbf{u}_{q-1}^p = 1 + O(1/\rho)$. The error terms are uniform for bounded σ.

Example 2 (Eigenvector scalar products for unit step) Figure 5 illustrates the scalar product $|\mathbf{u}_q^{pT}\mathbf{u}_{q-1}^r|\sqrt{\rho}$ for $p, r = 1, 2$, together with the Hermite limit given by Theorem 6 for $\rho = 100$.

Proof We proceed by computing the asymptotic limit of the expression $\mathbf{u}_q^{pT}\mathbf{u}_{q-1}^r$ by setting $q = s$ in (13) in Theorem 1.

Using the same definition of y (15) as in the proof of Theorem 5,

$$\frac{\partial y(\nu,\sigma)}{\partial \nu} = \frac{\partial}{\partial \nu}\left[(2\rho)^{\nu/2}\right] c_{q+1}(\nu) + (2\rho)^{\nu/2}\frac{\partial}{\partial \nu}\left[c_{q+1}(\nu)\right],$$

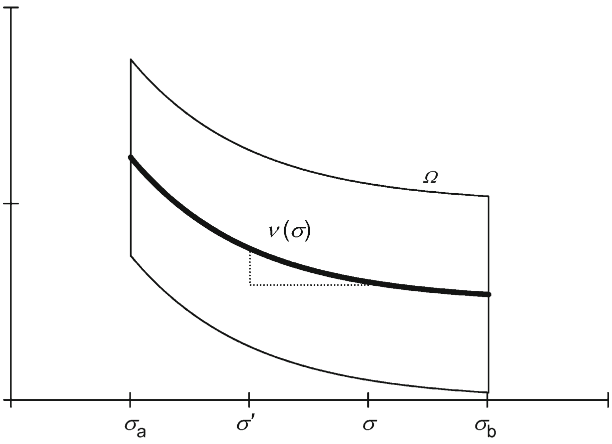

Fig. 6 There must be an open bounded set of σ and ν such that $\partial y/\partial \nu$ is non-zero

so for $\nu = \nu_{q+1}$,

$$\frac{c_q\left(\nu_{q+1}\right)}{c'_{q+1}\left(\nu_{q+1}\right)} = \frac{y\left(\nu_{q+1},\sigma\right) - y\left(\nu_{q+1},\sigma'\right)}{\frac{\partial y}{\partial \nu}\left(\nu_{q+1},\sigma'\right)}, \tag{20}$$

where $\sigma' \triangleq \sigma + 1/\sqrt{\rho}$.

Suppose that σ belongs to a bounded interval and let $\nu(\sigma)$ be the kth smallest positive real zero of $H_\nu\left(-\sigma/\sqrt{2}\right)$ (cf. Fig. 3 for an illustration of the smallest and second smallest zero). Since $\partial y\left(\nu_q,\sigma\right)/\partial\nu$ approaches $\partial H_{\nu(\sigma)}\left(-\sigma/\sqrt{2}\right)/\partial\nu \neq 0$ uniformly for bounded ν when $\rho \to \infty$ by Theorem 3, there must be an open bounded set Ω

$$\Omega = \{(\sigma,\nu) \mid \sigma_a < \sigma < \sigma_b,\ \ |\nu - \nu(\sigma)| < \varepsilon\}$$

for some $\varepsilon > 0$ (Fig. 6), such that $\partial y(\nu,\sigma)/\partial\nu$ does not vanish on Ω. This enables us to introduce

$$z(\nu,\sigma) \triangleq \frac{\frac{\partial^2 y}{\partial \nu^2}(\nu,\sigma)}{\frac{\partial y}{\partial \nu}(\nu,\sigma)},$$

which is bounded and analytic in ν on Ω. Combining (17) and (20),

$$\frac{c_q\left(\nu_{q+1}\right)}{c'_{q+1}\left(\nu_{q+1}\right)} = \Delta\nu_q\left\{1 + \frac{\Delta\nu_q}{2} z\left(\nu_q,\sigma\right) + O\left(\frac{1}{\rho}\right)\right\}\frac{\frac{\partial y}{\partial \nu}\left(\nu_q,\sigma\right)}{\frac{\partial y}{\partial \nu}\left(\nu_{q+1},\sigma'\right)}, \tag{21}$$

where the error term is uniform on Ω. We can write

$$\frac{\partial y}{\partial \nu}\left(\nu_{q+1},\sigma'\right)=\left[\frac{\partial y}{\partial \nu}\left(\nu_{q+1},\sigma'\right)-\frac{\partial y}{\partial \nu}\left(\nu_{q},\sigma'\right)\right]+\left[\frac{\partial y}{\partial \nu}\left(\nu_{q},\sigma'\right)-\frac{\partial y}{\partial \nu}\left(\nu_{q},\sigma\right)\right]+\frac{\partial y}{\partial \nu}\left(\nu_{q},\sigma\right).$$

For the first bracket, by the analyticity of y in ν,

$$\frac{\partial y}{\partial \nu}\left(\nu_{q+1},\sigma'\right)-\frac{\partial y}{\partial \nu}\left(\nu_{q},\sigma'\right)=O\left(\frac{1}{\sqrt{\rho}}\right),$$

where the error terms are uniform on Ω. For the second bracket, the function $\partial y\left(\nu,\sigma\right)/\partial\nu$ cannot be differentiated with respect to σ, but by differentiating (18), and again using the analyticity of y in ν,

$$\frac{\partial y}{\partial \nu}\left(\nu_{q},\sigma'\right)-\frac{\partial y}{\partial \nu}\left(\nu_{q},\sigma\right)=O\left(\frac{1}{\sqrt{\rho}}\right).$$

Accordingly, the quotient in (21) is

$$\frac{\frac{\partial y}{\partial \nu}\left(\nu_{q},\sigma\right)}{\frac{\partial y}{\partial \nu}\left(\nu_{q+1},\sigma'\right)}=\frac{\frac{\partial y}{\partial \nu}\left(\nu_{q},\sigma\right)}{\frac{\partial y}{\partial \nu}\left(\nu_{q},\sigma\right)+O\left(1/\sqrt{\rho}\right)}=1+O\left(\frac{1}{\sqrt{\rho}}\right),$$

where the error terms are uniform on Ω.

On the other hand, similar to the derivation of (17) in the proof of Theorem 5,

$$\Delta\nu_q=\frac{-y\left(\nu_q,\sigma'\right)}{\frac{y\left(\nu_q,\sigma'\right)}{-\Delta\nu_q}}=-\frac{y\left(\nu_q,\sigma'\right)-y\left(\nu_q,\sigma\right)}{\frac{\partial y}{\partial \nu}\left(\nu_{q+1},\sigma'\right)-\frac{\Delta\nu_q}{2}\frac{\partial^2 y}{\partial \nu^2}\left(\nu_{q+1},\sigma'\right)+O\left(1/\rho\right)}$$

and

$$\frac{c_{q+1}\left(\nu_q\right)}{c_q'\left(\nu_q\right)}=\frac{c_{q+1}\left(\nu_q\right)-c_q\left(\nu_q\right)}{c_q'\left(\nu_q\right)}=\frac{y\left(\nu_q,\sigma'\right)-y\left(\nu_q,\sigma\right)}{\frac{\partial y}{\partial \nu}\left(\nu_q,\sigma\right)}$$

so

$$\frac{c_{q+1}\left(\nu_q\right)}{c_q'\left(\nu_q\right)}=-\Delta\nu_q\left\{1-\frac{\Delta\nu_q}{2}z\left(\nu_{q+1},\sigma'\right)+O\left(\frac{1}{\rho}\right)\right\}\frac{\frac{\partial y}{\partial \nu}\left(\nu_{q+1},\sigma'\right)}{\frac{\partial y}{\partial \nu}\left(\nu_q,\sigma\right)},\tag{22}$$

where all error terms are uniform on Ω. Multiplying (21) and (22),

$$-\frac{c_q\left(\nu_{q+1}^p\right)c_{q+1}\left(\nu_q^r\right)}{c_{q+1}'\left(\nu_{q+1}^p\right)c_q'\left(\nu_q^r\right)}=\Delta\nu_q^p\Delta\nu_q^r\left[1+O\left(\frac{1}{\sqrt{\rho}}\right)\right]R,\tag{23}$$

where

$$R \triangleq \left\{1 + \frac{\Delta \nu_q^p}{2} z\left(\nu_q^p, \sigma\right)\right\} \left\{1 - \frac{\Delta \nu_q^r}{2} z\left(\nu_{q+1}^r, \sigma'\right)\right\} + O\left(\frac{1}{\rho}\right) \tag{24}$$

uniformly on Ω. By Theorems 1 and 5, the theorem is proved when $p \neq r$. When $p = r$,

$$R = 1 + \frac{\Delta \nu_q}{2}\left[z\left(\nu_q, \sigma\right) - z\left(\nu_{q+1}, \sigma'\right)\right] + O\left(\frac{1}{\rho}\right),$$

where the middle term can be split into the two parts

$$z\left(\nu_q, \sigma\right) - z\left(\nu_{q+1}, \sigma'\right) = \left[z\left(\nu_q, \sigma\right) - z\left(\nu_q, \sigma'\right)\right] + \left[z\left(\nu_q, \sigma'\right) - z\left(\nu_{q+1}, \sigma'\right)\right].$$

For the second bracket, by the analyticity of z in ν,

$$z\left(\nu_q, \sigma'\right) - z\left(\nu_{q+1}, \sigma'\right) = -\Delta \nu_q \frac{\partial z}{\partial \nu}\left(\nu_q, \sigma'\right) + O\left(\frac{1}{\rho}\right) = O\left(\frac{1}{\sqrt{\rho}}\right).$$

For the first bracket, the function $z(\nu, \sigma)$ cannot be differentiated with respect to σ, but by differentiating (18) twice, and using the analyticity of y in ν,

$$\frac{\partial^2}{\partial \nu^2} y(\nu, \sigma) - \frac{\partial^2}{\partial \nu^2} y\left(\nu, \sigma'\right) = O\left(\frac{1}{\sqrt{\rho}}\right)$$

so

$$z\left(\nu, \sigma'\right) = \frac{\frac{\partial^2 y}{\partial \nu^2}(\nu, \sigma) + O\left(1/\sqrt{\rho}\right)}{\frac{\partial y}{\partial \nu}(\nu, \sigma) + O\left(1/\sqrt{\rho}\right)} = z(\nu, \sigma) + O\left(1/\sqrt{\rho}\right),$$

implying that $z(\nu, \sigma) - z\left(\nu, \sigma'\right) = O\left(1/\sqrt{\rho}\right)$ uniformly on Ω. Since $p = r$, the factor $1 + O(\sqrt{\rho})$ cancels out in the product (23), resulting in $\mathbf{u}_q^{pT} \mathbf{u}_{q-1}^p = 1 + O(1/\rho)$ uniformly on Ω. □

4 Asymptotic discontinuities

Theorem 6 does not cover discontinuities in σ_t. Since this is an important case in practice, in this section, we will analyze this situation while temporarily suspending assumption (**A3**).

Theorem 7 (Asymptotic eigenvector scalar products for large steps) *Assume that* σ *has a discontinuity at* t*, and define* $q \triangleq q(t-)$*,* $s \triangleq q(t+) + 1$*,* $\sigma_{t-} \triangleq$

$\lim_{\rho\to\infty}(q-\rho)/\sqrt{\rho}$, *and* $\sigma_{t+} \triangleq \lim_{\rho\to\infty}(s-\rho)/\sqrt{\rho}$. *Under the same conditions as in Theorem 1,*

$$\mathbf{u}_q^{pT}\mathbf{u}_{s-1}^r = \frac{e^{(\sigma_{t-}^2-\sigma_{t+}^2)/4}}{\nu^p-\nu^r}\sqrt{2\,\frac{d\nu^r}{d\sigma}\frac{d\nu^p}{d\sigma}}\,\frac{H_{\nu^p}\left(-\sigma_{t+}/\sqrt{2}\right)}{H_{\nu^p+1}\left(-\sigma_{t-}/\sqrt{2}\right)}\left[1+O\left(\frac{1}{\sqrt{\rho}}\right)\right], \tag{25}$$

where ν^p *is the* p*th smallest zero of* $H_\nu\left(-\sigma_{t-}/\sqrt{2}\right)$, *and* ν^r *is the* r*th smallest zero of* $H_\nu\left(-\sigma_{t+}/\sqrt{2}\right)$. *The derivatives* $d\nu^p/d\sigma$ *and* $d\nu^r/d\sigma$ *are to be taken at* σ_{t-} *and* σ_{t+}, *respectively.*

Example 3 (Eigenvector scalar products for large steps) As an example, consider an instantaneous boundary reduction from $\sigma = 4$ to $\sigma = 1$,

$$q_{\text{STEP}}(t) \triangleq \begin{cases} \lfloor \rho + 4\sqrt{\rho} \rfloor & \text{if } t \le 0 \\ \lfloor \rho + \sqrt{\rho} \rfloor & \text{if } t > 0 \end{cases}. \tag{26}$$

Let $t = 0$ and $t = 0+$ denote the epochs just before and after the step, respectively. Then, by (12),

$$\boldsymbol{\omega}^T(0+) = \boldsymbol{\omega}^T(0)\mathbf{U}^T(q_0)\,\mathbf{U}(q_0+\Delta q),$$

where $q_0 = \rho + 4\sqrt{\rho}$ and $\Delta q = -3\sqrt{\rho}$. If the fundamental mode is the only mode initially present in $\boldsymbol{\omega}(0) = (\omega_1, 0, \ldots)^T$, we obtain the component-wise relations

$$\omega_k(0+) = \omega_1(0)\mathbf{u}^{1T}(q_0)\,\mathbf{u}^k(q_0+\Delta q). \tag{27}$$

The Charlier and Hermite limit mode transition matrices $\mathbf{T} \triangleq \mathbf{U}^T(q_0)\,\mathbf{U}(q_0+\Delta q)$ for $\rho = 100$ become

$$\mathbf{T}_C = \begin{pmatrix} .8589 & .2124 & .1333 & \\ -.4976 & .4887 & .2043 & \\ -.0311 & -.7837 & .0537 & \\ & & & \ddots \end{pmatrix} \mathbf{T}_H = \begin{pmatrix} .8378 & .2210 & .1393 & \\ -.5310 & .4496 & .1934 & \\ -.0133 & -.7904 & .0208 & \\ & & & \ddots \end{pmatrix},$$

respectively.

Proof The proof consists of taking the asymptotic limit of each factor on the right-hand side of (13). By Theorem 4,

$$\frac{1}{\nu_s^r - \nu_{q+1}^p} = \frac{1}{\nu^r - \nu^p}\left[1+O\left(\frac{1}{\sqrt{\rho}}\right)\right]. \tag{28}$$

It follows easily from Stirling's approximation, taking the logarithm, and then Taylor expanding that

$$\frac{\rho^q e^{-\rho}}{q!} = \frac{1}{\sqrt{2\pi\rho}} \exp\left(-\frac{\sigma^2}{2}\right)\left[1 + O\left(\frac{1}{\sqrt{\rho}}\right)\right], \tag{29}$$

so

$$\sqrt{\frac{\rho^s q!}{\rho^q s!}} = e^{(\sigma_{t-}^2 - \sigma_{t+}{}^2)/4}\left[1 + O\left(\frac{1}{\sqrt{\rho}}\right)\right]. \tag{30}$$

Since

$$c_q\left(\nu_{q+1}\right) = c_q\left(\nu_{q+1}\right) - c_q\left(\nu_q\right) = \Delta\nu_q \cdot \frac{\partial}{\partial\nu} c_q\left(\nu_q\right) + O\left(\left(\Delta\nu_q\right)^2\right),$$

then, by Theorems 2, 3, 5, and the recurrence relation (19),

$$\frac{c_s\left(\nu_{q+1}\right)}{c_q\left(\nu_{q+1}\right)} = \frac{-\sqrt{2}}{\Delta\nu_q} \frac{H_\nu\left(-\sigma_{t+}/\sqrt{2}\right)}{H_{\nu+1}\left(-\sigma_{t-}/\sqrt{2}\right)}\left[1 + O\left(\frac{1}{\sqrt{\rho}}\right)\right]. \tag{31}$$

By (21) and (22) in the proof of Theorem 6,

$$\sqrt{-\frac{c_q\left(\nu_{q+1}^p\right) c_{s+1}\left(\nu_s^r\right)}{c_{q+1}'\left(\nu_{q+1}^p\right) c_s'\left(\nu_s^r\right)}} = \sqrt{\Delta\nu_q^p \Delta\nu_s^r}\left[1 + O\left(\frac{1}{\sqrt{\rho}}\right)\right]. \tag{32}$$

Substituting (28), (30), (31), and (32) into (13) in Theorem 1 completes the proof. □

5 Conversion between eigenmodes and probability

Converting between the busy-server probability distribution $\boldsymbol{\pi}$ and the eigenmode $\boldsymbol{\omega}$ is important, because the initial condition and the survivor function are expressed in terms of probability, while the time evolution is more easily computed in terms of eigenmodes. First, we consider the translation of the mode vector to the survival function.

We define the pth *mode conversion factor*

$$G_p \triangleq e^{-\rho/2} \frac{\mathbf{w}_q^{pT} \mathbf{D}^2 \mathbf{1}}{\left|\mathbf{w}_q^{pT} \mathbf{D}\right|}. \tag{33}$$

We will use the theorems from Sect. 3 for simplifying this expression.

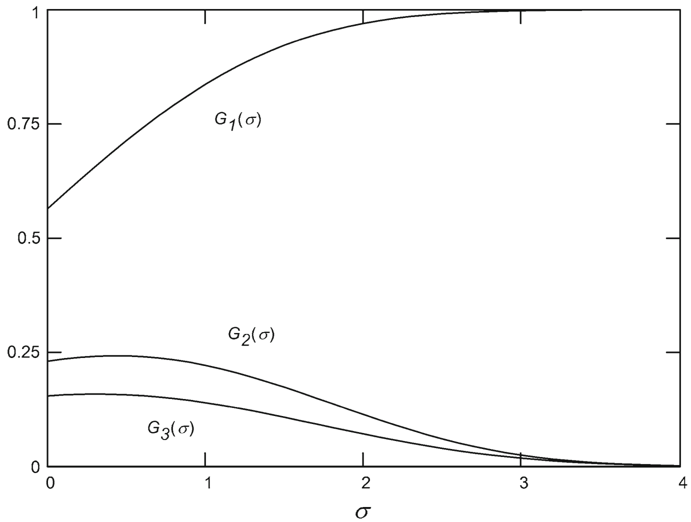

Fig. 7 Hermite limit of conversion factors $G_k(\sigma)$. For $k = 1$ and $k > 1$, $G_k(\sigma)$ approaches 1 and 0, respectively

Theorem 8 (Translation of eigenmodes to survivor function) *The translation* $\boldsymbol{\omega}^T \mathbf{M}^{-1}\mathbf{1}$ *of the mode vector to the survivor function can be written as*

$$S(t) = \boldsymbol{\omega}^T \mathbf{g}\, e^{\rho/2},$$

where $\mathbf{g} \triangleq \mathbf{M}^{-1}\mathbf{1}\, e^{-\rho/2} = \mathbf{M}^T\mathbf{D^2 1}\, e^{-\rho/2} = \left(G_1, G_2, \ldots G_q\right)^T$ *is a vector of mode conversion factors. It satisfies* $|\mathbf{g}| \leq 1$*, and*

$$G_p(\sigma) = \frac{1}{\sqrt[4]{2\pi}} \exp\left(-\frac{\sigma^2}{4}\right) \sqrt{\frac{d}{d\sigma}\frac{1}{\nu^p}} \cdot \left[1 + O\left(\frac{1}{\sqrt{\rho}}\right)\right],$$

where ν^p *is the pth smallest positive zero of the Hermite function* $H_\nu\left(-\sigma/\sqrt{2}\right)$.

Example 4 (Mode conversion factors) Figure 7 illustrates the asymptotic conversion factors $G_p(\sigma)$ for $p = 1, 2, 3$ and $\rho = 100$.

Proof Since $\mathbf{g} = \mathbf{M}^{-1}\mathbf{1}\, e^{-\rho/2} = \mathbf{U}^T\mathbf{D1}\, e^{-\rho/2} = \mathbf{M}^T\mathbf{D^2 1}\, e^{-\rho/2}$,

$$|\mathbf{g}| = \sqrt{\mathbf{1}^T\mathbf{D}^T\mathbf{U} \cdot \mathbf{U}^T\mathbf{D1}}\, e^{-\rho/2} \leq e^{-\rho/2}\sqrt{\mathbf{1}^T\mathbf{D}^2\mathbf{1}} = e^{-\rho/2}\sqrt{\sum_{k=0}^{\infty}\frac{\rho^k}{k!}} = 1.$$

For a single-mode $\boldsymbol{\omega} = \left(0, \ldots, 0, \omega_p, 0, \ldots 0\right)^T$, by (9), the projection becomes

$$\boldsymbol{\omega}^T\mathbf{U}^T\mathbf{D1} = \omega_p \mathbf{u}_q^{pT}\mathbf{D1} = \omega_p G_p.$$

By the Christoffel–Darboux identity (14), tacitly understanding superscript p on $\mathbf{w}_q$, ν_{q+1}, and $\Delta\nu_q$,

$$\frac{\mathbf{w}_q^T \mathbf{D}^2 \mathbf{1}}{\left|\mathbf{w}_q^T \mathbf{D}\right|} = \frac{\sum_{k=0}^{q} s_k(0)\, s_k\left(\nu_{q+1}\right)}{\sqrt{\sum_{k=0}^{q} s_k\left(\nu_{q+1}\right)^2}} = \frac{1}{\nu_{q+1}}\sqrt{-\frac{\rho^{q+1}}{q!}\frac{c_q\left(\nu_{q+1}\right)}{c'_{q+1}\left(\nu_{q+1}\right)}}. \tag{34}$$

By (21), $c_q\left(\nu_{q+1}\right)/c'_{q+1}\left(\nu_{q+1}\right) = \Delta\nu_q\left[1 + O\left(1/\sqrt{\rho}\right)\right]$, so by (29),

$$e^{-\rho/2}\frac{\mathbf{w}_q^T \mathbf{D}^2 \mathbf{1}}{\left|\mathbf{w}_q^T \mathbf{D}\right|} = \frac{1}{\sqrt[4]{2\pi}}\exp\left(-\frac{\sigma^2}{4}\right)\frac{\sqrt{-\Delta\nu_q^p\sqrt{\rho}}}{\nu_{q+1}^p}\left[1 + O\left(\frac{1}{\sqrt{\rho}}\right)\right].$$

Application of Theorems 4 and 5 completes the proof. □

We will now consider the reverse translation $\mathbf{M}$ from probability distribution $\boldsymbol{\pi}$ to mode vector $\boldsymbol{\omega}^T = \boldsymbol{\pi}^T\mathbf{M}$.

Theorem 9 (Translation of probability distribution to eigenmodes) *The probability distribution* $\boldsymbol{\pi}$ *translates to the mode vector* $\boldsymbol{\omega}$ *by* $\boldsymbol{\omega}^T = \boldsymbol{\pi}^T\mathbf{M}$ *where the* j*th element* $(j = 0, 1, \ldots q)$ *of the* p*th column* $\mathbf{m}^p$ *of* $\mathbf{M}$ *is*

$$m_j^p = \sqrt{\frac{q!}{\rho^{q+1}}}\frac{c_j\left(\nu_{q+1}\right)}{\sqrt{-c_q\left(\nu_{q+1}\right)c'_{q+1}\left(\nu_{q+1}\right)}}. \tag{35}$$

Let $z(j) \triangleq (j-\rho)/\sqrt{\rho}$, *and let* $E_\pi[\cdot]$ *denote the expectation with respect to* $\boldsymbol{\pi}$. *Then the* p*th component of* $\boldsymbol{\omega}$ *is*

$$\omega^p = \frac{e^{-\rho/2}}{G_p(\sigma)}\frac{E_\pi\left[H_{\nu^p}\left(-z/\sqrt{2}\right)\right]}{\partial H_{\nu^p}\left(-\sigma/\sqrt{2}\right)/\partial\nu^p}\left[1 + O\left(\frac{1}{\sqrt{\rho}}\right)\right], \tag{36}$$

where ν^p *is the* p*th smallest positive zero of the Hermite function* $H_{\nu^p}\left(-\sigma/\sqrt{2}\right)$. *In the special case where* $\boldsymbol{\pi}$ *contains only the fundamental mode, then*

$$\omega^1 = \frac{e^{-\rho/2}}{G_1(\sigma)}\left[1 + O\left(\frac{1}{\sqrt{\rho}}\right)\right]. \tag{37}$$

Proof The pth mode equals the projection of $\boldsymbol{\pi}$ on the pth column $\mathbf{m}^p$ of $\mathbf{M}$, $\boldsymbol{\pi}^T\mathbf{m}^p = \boldsymbol{\pi}^T\mathbf{w}^p/|\mathbf{D}\mathbf{w}^p|$. The jth component of $\mathbf{w}^p$ is $c_j\left(\nu_{q+1}\right)$, and if we tacitly understand superscript p on $\mathbf{w}_q$ and ν_{q+1}, again by the Christoffel–Darboux identity (14), then

$$\left|\mathbf{D}\mathbf{w}_q\right| = \sqrt{\frac{-\rho^{q+1}}{q!}c_q\left(\nu_{q+1}\right)c'_{q+1}\left(\nu_{q+1}\right)}$$

so the jth component m_j^p of $\mathbf{m}^p$ is (35). By (34) in the proof of Theorem 8,

$$m_j^p = \frac{e^{-\rho/2}}{G_p(\sigma)} \frac{c_j\left(\nu_{q+1}\right)}{c'_{q+1}\left(\nu_{q+1}\right)} \left[1 + O\left(\frac{1}{\sqrt{\rho}}\right)\right].$$

By Theorems 2 and 3, (36) follows. Equation (37) follows from (11) and (33). □

6 Time evolution of the mode vector

We are now ready to present the general exact and asymptotic solutions to the hitting-time problem, formulated as a recursion and a differential equation, respectively. In essence, this theorem carries out a temporal integration of the product (12), and expresses it as a differential equation.

Theorem 10 (Time evolution of the mode vector) *The change of the mode vector* $\boldsymbol{\omega}$ *over time intervals* Δt_k *where the boundary is constant, but ends with a unit step, can be expressed as*

$$\boldsymbol{\omega}^T\left(t + \Delta t_k\right) = \boldsymbol{\omega}^T(t) e^{-\Delta t_k \mu \boldsymbol{\Lambda}_k} \mathbf{U}_k^T \mathbf{U}_{k+1}. \tag{38}$$

The initial mode vector $\boldsymbol{\omega}^T(0) = \boldsymbol{\pi}^T \mathbf{M}(0)$ *is given by Theorem 9, the transient distribution* $\boldsymbol{\pi}^T(t) = \boldsymbol{\omega}^T(t)\mathbf{M}^{-1}(t)$*, and thereby the projection to the survivor function* $S(t) = \boldsymbol{\pi}^T(t)\mathbf{1}$ *by Theorem 8, and the mode transition* $\mathbf{U}_k^T \mathbf{U}_{k+1}$ *by Theorem 1.*

For the asymptotic case, assume that $\sigma(t) = \lim_{\rho\to\infty} \sigma(\rho; t)$ *is continuously differentiable. In the limit as* ρ *approaches infinity, and the time intervals between boundary changes* $\Delta t_k = O\left(1/\sqrt{\rho}\right)$ *approach zero, (38) becomes the differential equation*

$$\frac{d\boldsymbol{\omega}^T}{dt} = \boldsymbol{\omega}^T \left(\mathbf{C}\left|\frac{d\sigma}{dt}\right| - \mu \boldsymbol{\Lambda}\right), \tag{39}$$

where $\boldsymbol{\Lambda} = diag\left(\nu^j\right)$ *is the diagonal matrix of positive zeros in* ν *of the Hermite function* $H_\nu(\sigma/\sqrt{2})$*, and the skew-symmetric matrix* $\mathbf{C} \triangleq \lim_{\rho\to\infty}\left[\sqrt{\rho}\left(\mathbf{U}_k^T \mathbf{U}_{k+1} - \mathbf{I}\right)\right]$ *is given by Theorem 6.*

Proof Equation (38) is an immediate consequence of Eq. (12). For large ρ, the exponential $\exp\left(-\Delta t_k \mu \boldsymbol{\Lambda}_k\right)$ can be expanded into

$$e^{-\Delta t_k \mu \boldsymbol{\Lambda}_k} = \mathbf{I} - \Delta t_k \mu \boldsymbol{\Lambda}_k + O\left(\Delta t_k^2\right).$$

Inserting this into (38), subtracting $\boldsymbol{\omega}^T$ on both sides and dividing by Δt,

$$\frac{\boldsymbol{\omega}^T\left(t + \Delta t_k\right) - \boldsymbol{\omega}^T(t)}{\Delta t_k} = \boldsymbol{\omega}^T \left(\frac{\mathbf{U}_k^T \mathbf{U}_{k+1} - \mathbf{I}}{\Delta t_k} - \mu \boldsymbol{\Lambda}_k \mathbf{U}_k^T \mathbf{U}_{k+1}\right) + O(\Delta t_k).$$

By Theorem 6, $\mathbf{U}_k^T \mathbf{U}_{k+1} = \mathbf{I} + \mathbf{C}/\sqrt{\rho} + O\left(1/\rho\right)$. By (**A3**), $|\Delta\sigma| = 1/\sqrt{\rho}$, so

$$\frac{\boldsymbol{\omega}^T\left(t + \Delta t_k\right) - \boldsymbol{\omega}^T\left(t\right)}{\Delta t_k} = \boldsymbol{\omega}^T \left(\mathbf{C}\left|\frac{\Delta\sigma}{\Delta t_k}\right| - \mu\boldsymbol{\Lambda}_k\right) + O(\Delta t_k).$$

In the limit, when $\rho \to \infty$ and by (**A2**), $\Delta t_k \to 0$, this converges to (39). □

If the system (39) is truncated to one dimension, it has the exact solution

$$\boldsymbol{\omega}^T\left(T\right) = \boldsymbol{\omega}_0^T \exp\left(-\int\limits_0^T \mu\boldsymbol{\Lambda}\mathrm{d}t\right).$$

This solution does have a truncation error, but is a good approximation when the initial distribution $\boldsymbol{\pi}(0)$ contains only the fundamental mode, and $|\mathrm{d}\sigma/\mathrm{d}t|$ is small. We can quantify the meaning of "small" by requiring that the term $\mathbf{C}\,|\mathrm{d}\sigma/\mathrm{d}t|$ in (39) does not significantly affect the fundamental mode, or

$$\left|\boldsymbol{\omega}^T\mathbf{c}^1\frac{\mathrm{d}\sigma}{\mathrm{d}t}\right| \ll \omega_1\mu\nu^1,$$

where $\mathbf{c}^1$ is the first column of $\mathbf{C}$. Given that the fundamental mode dominates, i.e., $|\omega_1| \gg \sum_{j\neq 1}|\omega_j|$, the condition of a small disturbance can be formulated as

$$\left|\frac{d\sigma}{dt}\right| \lessapprox \frac{\mu\nu^1}{|\mathbf{c}^1|_\infty} \lessapprox 2\mu\,\exp\left(-\frac{\sigma}{\sqrt{2}}\right),$$

where we used Theorem 6 and some computation. The survivor function becomes

$$S(T) \approx \frac{G_1(\sigma_T)}{G_1(\sigma_0)}\exp\left(-\int\limits_0^T \mu\nu^1\mathrm{d}t\right). \tag{40}$$

The factor $G_1(\sigma_T)/G_1(\sigma_0)$ describes losses due to boundary changes, while the factor $\exp\left(-\int_0^T \mu\nu^1\mathrm{d}t\right)$ describes the fundamental's temporal decay. The first factor can be comfortably computed by Theorem 8 and the second by Theorem 4.

By combining Theorems 6 and 7, we can handle σ which have discontinuities and are only piecewise continuously differentiable.

7 Simulations

In order to illustrate and verify the theoretical results, we compare these with discrete-event simulations. In all cases, the simulation uses $\rho = 1600$ and $\mu = 1$. The experimental survivor functions are computed by sorting the hitting times from 400 simulation runs. We assume that the initial distribution is fundamental mode only.

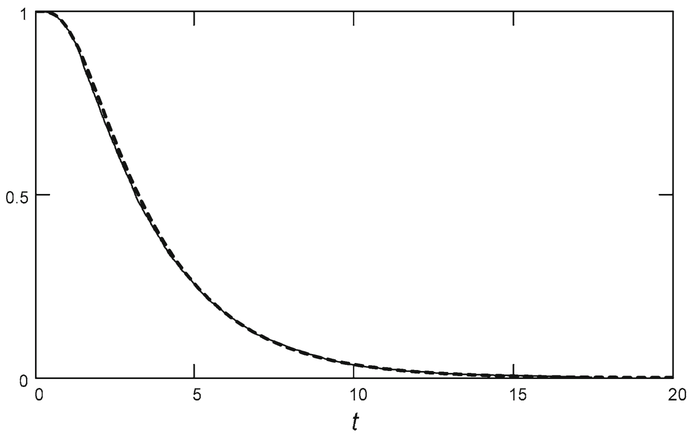

Fig. 8 Survivor function for the exponentially decreasing boundary q_{EXP} in (1). The theoretical prediction based on the fundamental mode only is shown as the *thick*, *dashed line*. The simulation result is shown as the *thin*, *solid line*

7.1 Exponential change

We first consider a system with the exponentially decreasing boundary q_{EXP} in (1). Simulation results are compared to the theoretical survivor function in Fig. 8. The service time is exponentially distributed. The theoretical survivor function $S(t)$ is computed by (40), i.e., by approximation with the fundamental mode only. The agreement is good, since the boundary changes slowly.

7.2 Step change

The second simulation uses the boundary q_{STEP} in (26) from example 3. Figure 9 displays the result of a simulation as the thin, solid line, and the theoretical value based on (40) as the dotted line. Here, there are some discrepancies due to the steepness of the boundary change.

If we instead use the lowest *two* modes from the component-wise relations (27) in example 3, then we have the survival probability

$$S(T) \approx \boldsymbol{\omega}^T(0+)\exp\left(-\int_0^T \mu\boldsymbol{\Lambda}\mathrm{d}t\right)\begin{pmatrix} G_1(\sigma_T) \\ G_2(\sigma_T) \end{pmatrix}.$$

Here, the first factor is a vector $\boldsymbol{\omega}$ holding the initial state of two modes *after* the step; the diagonal matrix $\boldsymbol{\Lambda}$ is composed of the first and second eigenvalues λ_1 and λ_2, respectively; and the third factor is a vector of two mode conversion factors. The result is shown as the thick, dashed line in Fig. 9, demonstrating the accuracy.

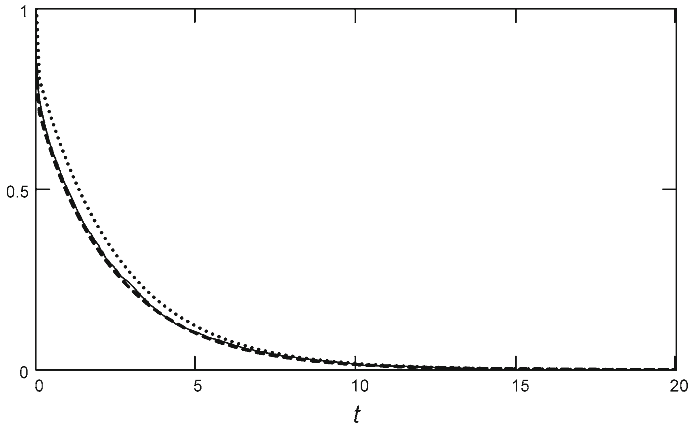

Fig. 9 Survivor function for the step boundary q_{STEP} in (26). The theoretical prediction based on the fundamental mode only, and on the two lowest modes are shown by the *dotted line* and the *dashed line*, respectively. The simulation result is shown as the *thin*, *solid line*

7.3 Rapid change

We finally consider a system with the rapidly decreasing boundary

$$q_{\mathrm{RAPID}}(t) \triangleq \lfloor \rho + \left(3e^{-100t} + 1\right)\sqrt{\rho} \rfloor. \tag{41}$$

The qualitative behavior for this boundary in simulation is the same as for q_{STEP} in (26). Truncating (39) to two modes and numerical integration produces the survivor function as shown in Fig. 10.

7.4 Non-exponential service duration

When service durations are non-exponentially distributed, the number of busy servers is not generally a Markov process, even for a constant boundary. Still, it is well known that the probability distribution of the number of busy servers upon arrival of a new customer depends only on the arrival rate and the mean service time, and is otherwise independent of the distribution of the service duration when q is *constant* [45].

Unfortunately, this does not hold for moving boundaries, since such a system is non-stationary. However, similar to approximations of other queueing systems with non-exponential service duration [46, Sect. 4.4], we have tested a variation of the Lee-Longton approximation of an $M/G/q$ system by an $M/M/q$ system [28], by mapping to an Erlang loss system having the same ρ, but modifying μ to *retain the second moment of the service time*,

$$\mu' = \mu\sqrt{\frac{2}{1 + C_v^2}}, \tag{42}$$

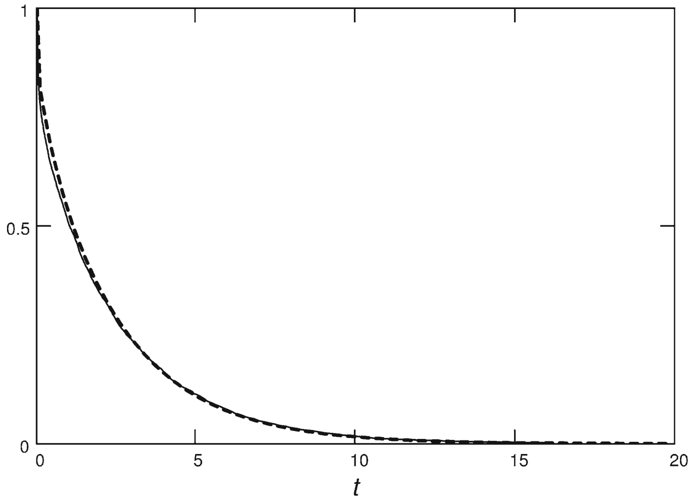

Fig. 10 Survivor function for the rapidly decreasing boundary q_{RAPID} in (41). The theoretical prediction based on (39) truncated to two modes is shown as the *thick*, *dashed line*. The simulation result is shown as the *thin*, *solid line*

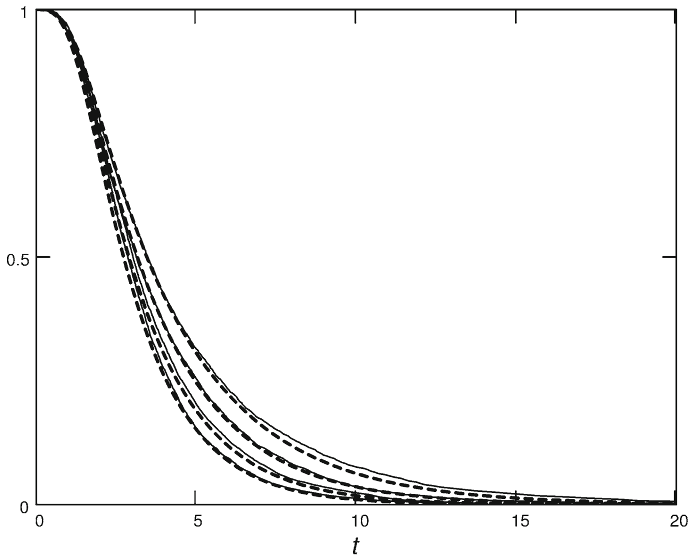

Fig. 11 The survivor function for non-exponential service distributions (*thin*, *solid lines*) may be approximated by exponential service distributions with a modified μ (*thick*, *dashed lines*), provided that the coefficient of variation is small. The boundary is q_{EXP} in (1). From *bottom* to *top* survivor functions for constant, uniform, exponential, and gamma service time distributions

where C_v is the coefficient of variation of the service time. Figure 11 shows simulation results for q_{EXP} above, with the following service time distributions, all with unit mean: constant ($C_v^2 = 0$), uniform ($C_v^2 = 1/3$), exponential ($C_v^2 = 1$), and gamma ($\Gamma(1/2, 2)$; $C_v^2 = 2$). Simulations are shown as solid thin lines, while dashed thick lines represent theoretical values. The theoretical prediction uses the asymptotic formula (39) truncated to the two lowest modes. Simulations indicate that the approximation is useful for $C_v^2 \leq 2$. For larger C_v, the behavior starts to depend substantially on the third moment, as has been observed for $M/G/q$ systems [41,46]. For hyperexponential distributions $H_2(p, \lambda_1, \lambda_2)$ with $C_v^2 \leq 4$, the systems still behave like $M/M/q_t/q_t$ systems, although the third moment influences the shape, and (42) no longer applies.

8 Conclusions

Given an $M/M/q_t/q_t$ Erlang loss system with constant arrival intensity λ, constant service rate μ, and a variable number of servers $q = \rho + \sigma(t)\sqrt{\rho}$, where $\rho = \lambda/\mu$, we derived the transient distribution of busy servers and the cumulative distribution of the hitting time. By using spectral decomposition and special properties of orthogonal polynomials, we found computationally efficient formulas valid for arbitrary $t \geq 0$, both for finite ρ and for the asymptotic case. If the offered load is large, e.g., $\rho \geq 100$, then the asymptotic formulas typically differ by a relative error $O(1/\sqrt{\rho})$ from the exact values. For non-exponential service durations, we found the classical Lee-Longton approximation [28] useful.

A well-established technique for solving queueing problems asymptotically is translation into stochastic differential equations [4,43], but these become difficult to solve for general boundaries, which imply colored noise. Our results suggest that translation in the reverse direction could instead be a fruitful approach for solving some difficult stochastic differential equations involving moving boundaries and/or colored noise.

Acknowledgments This research was funded by the European Union FP7 research project THE, “The Hand Embodied,” under Grant agreement 248587. The author thanks the anonymous reviewers for comments, and Dr. Henrik Jörntell of Lund University, Department of Experimental Medical Science for support.

References

1. Abate, J., Whitt, W.: Calculating transient characteristics of the Erlang loss model by numerical transform inversion. Stoch. Models **3**, 663–680 (1998). doi:10.1080/15326349808807494
2. Alili, L., Patie, P., Pedersen, J.L.: Representations of the first hitting time density of an Ornstein–Uhlenbeck process. Stoch. Models **21**(4), 967–980 (2005). doi:10.1080/15326340500294702
3. Borovkov, A.A.: Stochastic Processes in Queueing Theory. Springer, Berlin (1976)
4. Borovkov, A.A.: Asymptotic Methods in Queueing Theory. Wiley, New York (1984)
5. Brockmeyer, E., Halstrøm, H.L., Jensen, A.: The life and works of A K Erlang: solution of some problems in the theory of the probabilities of significance in automatic telephone exchanges. Trans. Dan. Acad. Tech. Sci. **2**, 138–155 (1948)
6. Buonocore, A., Nobile, A.G., Ricciardi, L.M.: A new integral equation for the evaluation of first-passage-time probability densities. Adv. Appl. Prob. 19(4), 784–800 (1987). URL http://www.jstor.org/stable/1427102
7. Charlier, C.V.L.: Über die Darstellung willkürlicher Funktionen. Ark. mat. astron. fys. 2 (1905).
8. Chihara, T.S.: An Introduction to Orthogonal Polynomials. No. 13 in Mathematics and its applications. Gordon and Breach, New York (1978)

9. Cox, D.R., Miller, H.D.: The Theory of Stochastic Processes. Methuen, London (1965)
10. Dominici, D.: Asymptotic analysis of the Askey scheme I: from Krawchouk to Charlier. Cent. Eur. J. Math. **5**(2), 280–304 (2007). doi:10.2478/s11533-006-0041-6
11. Dunster, T.M.: Uniform asymptotic expansions for Charlier polynomials. J. Approx. Theory. **112**, 93–133 (2001). doi:10.1006/jath.2001.3595
12. Durbin, J.: The first-passage density of a continuous Gaussian process to a general boundary. J. Appl. Prob. 22(1), 99–122 (1985). URL http://www.jstor.org/stable/3213751
13. Elbert, A., Muldoon, M.E.: Inequalities and monotonicity properties for zeros of Hermite functions. Proc. R. Soc. Edinb., Sect. A **129**, 57–75 (1999). doi:10.1017/S0308210500027463
14. Erdélyi, A., Magnus, W., Oberhettinger, F., Tricomi, F.G.: Higher Transcendental Functions. McGraw-Hill, New York (1953)
15. Feinsilver, P.J.: Special Functions, Probability Semigroups, and Hamiltonian Flows Lecture Notes in Mathematics, vol. 696. Springer, Berlin (1978)
16. Feller, W.: An Introduction to Probability Theory and its Applications, vol. 1, 3rd edn. Wiley, New York (1968)
17. Halfin, S., Whitt, W.: Heavy-traffic limits for queues with many exponential servers. Oper. Res. 29(3), 567–588 (1981). URL http://www.jstor.org/stable/170115
18. Hänggi, P., Jung, P.: Colored noise in dynamical systems. Adv. Chem. Phys. **89**, 239–326 (1995). doi:10.1002/9780470141489.ch4
19. Jagerman, D.: Some properties of the Erlang loss function. Bell Syst. Tech. J. **53**(3), 525–551 (1974). doi:10.1002/j.1538-7305.1974.tb02756.x
20. Karlin, S., McGregor, J.: The classification of birth and death processes. Trans. Am. Math. Soc. 86(2), 366–400 (1957). URL http://www.jstor.org/stable/1993021
21. Karlin, S., McGregor, J.: Many server queueing processes with Poisson input and exponential service times. Pac. J. Math. 8(1), 87–118 (1958). URL http://projecteuclid.org/euclid.pjm/1103040247
22. Karlin, S., McGregor, J.L.: The differential equations of birth-and-death processes, and the Stieltjes moment problem. Trans. Am. Math. Soc. 85(2), pp. 489–546 (1957). URL http://www.jstor.org/stable/1992942
23. Kijima, M.: On the largest negative eigenvalue of the infinitesimal generator associated with $M/M/n/n$ queues. Oper. Res. Lett. **9**, 59–64 (1990). doi:10.1016/0167-6377(90)90041-3
24. Knessl, C.: On the transient behavior of the M/M/m/m loss model. Stoch. Models **6**(4), 749–776 (1990). doi:10.1080/15326349908807172
25. Koekoek, R., Lesky, P.A., Swarttouw, R.F.: Hypergeometric Orthogonal Polynomials and Their q-Analogues. Springer, Berlin (2010)
26. Lebedev, N.: Special Functions and their Applications. Dover Publications, New York (1972)
27. Ledermann, W., Reuter, G.E.H.: Spectral theory for the differential equations of simple birth and death processes. Philos. Trans. R. Soc. London 246(914), pp. 321–369 (1954). URL http://www.jstor.org/stable/91569
28. Lee, A.M., Longton, P.A.: Queueing processes associated with airline passenger check-in. Oper. Res. Q. 10(1), 56–71 (1959). URL http://www.jstor.org/stable/3007312
29. Maejima, M., van Assche, W.: Probabilistic proofs of asymptotic formulas for some classical polynomials. Math. Proc. Camb. Philos. Soc. **97**, 499–510 (1985). doi:10.1017/S0305004100063088
30. Mandjes, M., Ridder, A.: A large deviations approach to the transient of the Erlang loss model. Perform. Eval. **43**, 181–198 (2001). doi:10.1016/S0166-5316(00)00050-X
31. Meixner, J.: Erzeugende Funktionen der Charlierschen Polynome. Math. Z. **44**(1), 531–535 (1939). doi:10.1007/BF01210670
32. Mitra, D., Weiss, A.: The transient behavior in Erlang's model for large trunk groups and various traffic conditions. In: Bonatti, M. (ed.) Teletraffic Science for New Cost-Effective Systems, Networks, and Services, ITC-12, pp. 1367–1374. Elsevier-Science, Amsterdam (1988)
33. Nilsson, M.: On the transition of Charlier polynomials to the Hermite function. arXiv:1202.2557 [math.CA] (2013). URL http://arxiv.org/abs/1202.2557
34. Olver, F.W., Lozier, D.W., Boisvert, R.F., Clark, C.W. (eds.): NIST Handbook of Mathematical Functions. Cambridge University Press, Cambridge (2010). URL http://dlmf.nist.gov
35. Redner, S.: A Guide to First-Passage Processes. Cambridge University Press, Cambridge (2001)
36. Ricciardi, L.M., Sato, S.: First-passage-time density and moments of the Ornstein-Uhlenbeck process. J. Appl. Prob. 25(1), 43–57 (1988).URL http://www.jstor.org/stable/3214232

37. Riordan, J.: Stochastic Service Systems. The SIAM Series in Applied Mathematics. SIAM, New York (1962)
38. Risken, H.: The Fokker–Planck Equation: Methods of Solution and Applications, 2nd edn. Springer, Berlin (1989)
39. Ronveaux, A., Zarzo, A., Area, I., Godoy, E.: Transverse limits in the Askey tableau. J. Comput. Appl. Math. **99**, 327–335 (1998). doi:10.1016/S0377-0427(98)00167-8
40. Ross, S.M., Seshadri, S.: Hitting time in an Erlang loss system. Prob. Eng. Inf. Sci. **16**, 167–184 (2002). doi:10.1017/S0269964802162036
41. Shin, Y.W., Moon, D.H.: Sensitivity and approximation of M/G/c queue: numerical experiments. In: Proceedings of the 8th International Symposium Operations Research and its Application (ISORA'09), pp. 140–147. Zhangjiajie, China (2009).
42. Shwartz, A., Weiss, A.: Large Deviations for Performance Analysis: Queues, Communications, and Computing. Chapman & Hall, New York (1994)
43. Srikant, R., Whitt, W.: Simulation runlengths to estimate blocking probabilities. ASCM Trans. Comput. Simul. **6**, 7–52 (1996). doi:10.1145/229493.229496
44. Szegő, G.: Orthogonal Polynomials, vol. XXIII, 4th edn. AMS Colloquium Publication, Providence (1975)
45. Takacs, L.: On Erlang's formula. Ann. Math. Stat. 40(1), 71–78 (1969). URL http://www.jstor.org/stable/2239199
46. Tijms, H.C.: Stochastic Modelling and Analysis A Computational Approach. Wiley & Sons, Chichester (1986)
47. Van Kampen, N.G.: Stochastic Processes in Physics and Chemistry, 2nd edn. North-Holland, Amsterdam (1992)
48. Virtamo, J., Aalto, S.: Calculation of time-dependent blocking probabilities. In: B. Goldstein, A. Koucheryavy, M. Shneps-Shneppe (eds.) Proceedings of the ITC Sponsored St. Petersburg Regional International Teletraffic Seminar Teletraffic Theory as a Base for QoS: Monitoring, Evaluation, Decisions, pp. 365–375. St. Petersburg (1998). URL http://www.netlab.hut.fi/tutkimus/cost257/publ/transient
49. Wang, L., Pötzlberger, K.: Boundary crossing probability for Brownian motion and general boundaries. J. Appl. Prob. 34, 54–65 (1997). URL http://www.jstor.org/stable/3215174
50. Xie, S., Knessl, C.: On the transient behavior of the Erlang loss model: heavy usage asymptotics. SIAM J. Appl. Math. 53(2), 555–599 (1993). URL http://www.jstor.org/stable/2102350