INFINITE LEVEL-DEPENDENT QBD PROCESSES AND MATRIX-ANALYTIC SOLUTIONS FOR STOCHASTIC CHEMICAL KINETICS

TUĞRUL DAYAR,* *Bilkent University*

WERNER SANDMANN,** *Clausthal University of Technology*

DAVID SPIELER *** AND

VERENA WOLF,*** *Saarland University*

Abstract

Systems of stochastic chemical kinetics are modeled as infinite level-dependent quasi-birth-and-death (LDQBD) processes. For these systems, in contrast to many other applications, levels have an increasing number of states as the level number increases and the probability mass may reside arbitrarily far away from lower levels. Ideas from Lyapunov theory are combined with existing matrix-analytic formulations to obtain accurate approximations to the stationary probability distribution when the infinite LDQBD process is ergodic. Results of numerical experiments on a set of problems are provided.

Keywords: Stochastic chemical kinetics; level-dependent quasi-birth-and-death process; state space truncation; Lyapunov bound; matrix-analytic solution

2010 Mathematics Subject Classification: Primary 60J22
Secondary 60J28; 92C45; 92D25

1. Introduction

We consider systems of stochastic chemical kinetics [3], [15], [19], [21], [24], [25], [31] described by continuous-time Markov chains (CTMCs) with infinite multidimensional state spaces. A state is a vector whose components represent the numbers of molecules of each chemical species present in the system. State transitions correspond to chemical reactions. According to the reaction stoichiometry, transitions are represented by a state change vector and state-dependent transition rates, which are multivariate polynomials in the state variables. Assuming that the CTMC is ergodic, our goal is to approximate its stationary distribution.

If the state space can be partitioned into disjoint *levels*, consecutively numbered such that transitions are only possible between states belonging to the same or adjacent levels, then the CTMC is said to be a quasi-birth-and-death (QBD) process and its generator matrix is block tridiagonal [16, pp. 129–130], [20, pp. 81–83]. In the case of level-dependent QBD (LDQBD) processes, the transition rates may depend on the level. Applying suitable state space truncation and augmentation procedures [10], [17], [30], [32], the stationary distribution of infinite LDQBD processes can be approximated by matrix-analytic methods [1], [4], [9], [14].

In this paper we first show how systems of stochastic chemical kinetics can be modeled as infinite LDQBD processes. The state variables that take infinitely many values determine the level numbers, whereas those that take finitely many values correspond to control (or phase) and determine the number of states within each level. In particular, the mild condition, which states that the maximum-valued infinite state variable in any state does not change by more than one through any transition in these systems, enables the maximum value among the infinite state variables to be the level number. Since the state vector holds only a finite number of state variables, even when there are no finite state variables, the states within each level are well defined and finite in this representation. To the best of our knowledge, this is the first representation of levels in infinite LDQBD processes using multiple state variables.

For these systems, in contrast to many other applications, levels have an increasing number of states as the level number increases and the probability mass may reside arbitrarily far away from lower levels. These make the LDQBD model both interesting and challenging to work with. The difficulty in computing the stationary probability distribution by the method introduced in [4] lies in the need to determine a sufficiently high level number, denoted by `high` in this contribution, from level 0 up to which most of the probability mass is located. Finding `high` requires matrix-analytic computations and can be very time consuming, since a matrix recording conditional expected sojourn times in states at level `high` + 1 per unit sojourn in states at level `high` must be computed with a certain tolerance on accuracy.

With the help of a suitably chosen Lyapunov function [11], [29], we show that the value of `high` can be obtained easily for a specified percentage of the stationary probability mass, thereby circumventing the need to do extensive computations. For the specified percentage, it may very well be the case that the lower level number, denoted by `low` in this contribution, which bounds the stationary probability mass from below, is larger than 0. Hence, the proposed technique guarantees the existence of the specified percentage of the stationary probability mass between levels `low` and `high` inclusive, `low` being obtained as a byproduct of the analysis which yields `high`. In practice, the percentage of the mass that lies in the finite subset of states between levels `low` and `high` may be much larger since the levels taken into consideration depend on the tightness of the particular Lyapunov function used. We remark that the tolerance on the accuracy of the matrix of conditional expected sojourn times does not provide such a guarantee [4].

Having obtained the pair (`low`, `high`), an approximation to the matrix of conditional expected sojourn times at level `high` + 1 must be determined. As indicated in [4], this is a chore in itself and can be quite time consuming. Once this matrix is available, the matrices of conditional expected sojourn times at lower level numbers down to `low` + 1 are computed using an already existing recurrence. Thereafter, a linear matrix equation for the boundary level `low` is solved to obtain the subvector of the stationary distribution associated with level `low`. The remaining stationary subvectors up to level `high` are computed recursively by vector-matrix multiplications using the subvector obtained for the current level and the matrix of conditional expected sojourn times available for the next level. We remark that all nonzero blocks in the block tridiagonal form used in the respective computations are sparse with nonzero entries having well-defined values and row/column indices for a given ordering of states. Hence, each nonzero block can be generated as a sparse matrix whenever needed and used only once during the computational procedure. Furthermore, the accuracy of the computed solution can be assessed by the norm of the residual vector.

Results of numerical experiments we provide on a set of problems from the literature suggest, among a number of alternative approaches, that the zero matrix as an approximation to the matrix

of conditional expected sojourn times at level `high` $+ 1$ [1], [14] fares very well in terms of accuracy. The solution is obtained rapidly with a residual norm in the order of 10^{-16} when the stationary probability mass is located in lower level numbers. We remark that the largest linear system solved has as many equations and unknowns as the number of states in level `high`. Besides, we pay special attention to a stable implementation and exploit two ideas. The first is to use the Grassmann–Taksar–Heyman (GTH) idea of employing solely positive floating-point arithmetic [13] in solving linear systems which involve M-matrices [5]. The second is to drop all nonzeros less than a specified threshold in matrices obtained during intermediate computations. Numerical experiments suggest that the latter idea is generally as good as the former and results in a faster solver. Even when the GTH idea is not used (except when solving the boundary system of equations), dropping nonzeros less than 10^{-16} from matrices results in almost no loss of accuracy in the solution compared to the solution obtained without dropping; in almost all cases both solutions are accurate to a residual norm in the order of 10^{-16}.

In the next section we introduce systems of stochastic chemical kinetics using a finite set of transition classes and show how they can be modeled as infinite LDQBD processes. Therein we have a running example and two other examples from the literature, each different than the others. In Section 3 we show how the pair of levels (`low`, `high`), between which a specified percentage of the stationary probability mass lies, can be computed using suitably chosen Lyapunov functions. In Section 4 we review an existing matrix-analytic method for LDQBD processes and discuss how the stationary distribution can be computed accurately when (`low`, `high`) is available. In Section 5, results of numerical experiments are provided on the examples introduced before together with an assessment of the overall technique. In Section 6, we conclude.

Throughout the paper, vectors associated with states are row vectors, consistent with the conventional definition of state probability vectors; otherwise, all vectors are column vectors. We represent by e a column vector of 1s and by e_i the ith column of the identity matrix, I. The lengths of the vectors are determined by the context in which they are used.

2. Infinite LDQBD processes

Let $x = (x_1, x_2, \ldots, x_n)$ denote the state vector of a system of stochastic chemical kinetics with n components, where $x_i \in \mathcal{S}_i$ is the state and $\mathcal{S}_i \subseteq \mathbb{Z}_+$ is the state space of the ith component for $i = 1, 2, \ldots, n$. Furthermore, let $\times_{i=1}^{n} \mathcal{S}_i$ be the product state space, where '$\times$' is the Cartesian product operator, and let $\mathcal{S}$ be the state space of the associated irreducible time-homogeneous CTMC $\{X(t), t \geq 0\}$. We remark that $X(t) = (X_1(t), X_2(t), \ldots, X_n(t))$, $\Pr\{X(t) = x\} = \Pr\{X_1(t) = x_1, X_2(t) = x_2, \ldots, X_n(t) = x_n\}$, and that each state in $\mathcal{S}$ is reachable from every other state in $\mathcal{S}$. However, note that $\mathcal{S} \subseteq \times_{i=1}^{n} \mathcal{S}_i \subseteq \mathbb{Z}_+^{1 \times n}$, and while $\mathcal{S}$ may be equal to $\mathbb{Z}_+^{1 \times n}$ in some problems, it will be only a subset of $\mathbb{Z}_+^{1 \times n}$ in others. In any case, if there is at least one component among the n with a state space of $\mathbb{Z}_+$, then the underlying CTMC has an infinite state space. Let n_I denote the number of components with infinite state spaces. Without loss of generality, let $|\mathcal{S}_i| = \infty$ for $i = 1, 2, \ldots, n_I$ and $|\mathcal{S}_i| < \infty$ for $i = n_I + 1, n_I + 2, \ldots, n$. In other words, each of the first n_I variables of the state vector can take infinitely many values, whereas each of the last $n - n_I$ variables of the state vector can take only finitely many values; the latter can be perceived as related to controlling the system. In this contribution, systems of stochastic chemical kinetics for which $2 \leq n_I \leq n$ are considered.

It is convenient to represent the evolution of a system of stochastic chemical kinetics by a set of transition classes. A transition class ϕ is a pair (α, v), where $\alpha : \mathcal{S} \to \mathbb{R}_{\geq 0}$ is a function that determines the transition rate in each state and $v \in \mathbb{Z}^{1 \times n}$ is a state change vector that determines

TABLE 1: Transition classes of the gene expression model.

j	ϕ_j	$\alpha_j(x)$	$v^{(j)}$
1	$(\alpha_1, v^{(1)})$	λ	$e_1^\top$
2	$(\alpha_2, v^{(2)})$	μx_1	$e_2^\top$
3	$(\alpha_3, v^{(3)})$	$\delta_1 x_1$	$-e_1^\top$
4	$(\alpha_4, v^{(4)})$	$\delta_2 x_2$	$-e_2^\top$

the successor state of the transition. Thus, if $x \in S$ and $\alpha(x) > 0$, then there is a transition from state x with rate $\alpha(x)$ to state $x + v \geq 0$. It is assumed that v has at least one nonzero entry and that $\alpha(x)$ is a multivariate polynomial in the state variables x_i for $i = 1, 2, \ldots, n$.

Now, let the system of stochastic chemical kinetics under study be represented by $\Phi = \{\phi_1, \phi_2, \ldots, \phi_J\}$, a set of J transition classes $\phi_j = (\alpha_j, v^{(j)})$ with state change vectors $v^{(j)}$ for $j = 1, 2, \ldots, J$. We assume that the CTMC associated with the system under study is ergodic. In the next section, we show how this can be verified by a suitably chosen Lyapunov function. When X is ergodic, its infinite infinitesimal generator matrix, Q, up to an ordering of its states has off-diagonal and diagonal entries defined respectively by

$$Q(x, y) = \sum_{j \text{ such that } y = x + v^{(j)}} \alpha_j(x) \quad \text{and} \quad Q(x, x) = -\sum_{y \neq x} Q(x, y) \quad \text{for } x \in \mathcal{S}, \quad (1)$$

and a positive stationary probability distribution π such that $\pi Q = 0$ and $\pi e = 1$. The sum in the expression for $Q(x, y)$ indicates that there may be multiple state change vectors which yield the same successor state y given the current state x. Since J is finite, Q in (1) is sparse and has a finite number of nonzeros in each row and column, although $\sup_{x \in \mathcal{S}} |Q(x, x)|$ may be infinitely large due to the way in which $\alpha(x)$ is defined.

Example 1. (*Gene expression.*) Let us consider a system of stochastic chemical kinetics modeling the biological process associated with a gene expression [27] through the transition classes in Table 1. Here, $n = 2$, $x = (x_1, x_2)$, $\mathcal{S}_1 = \mathcal{S}_2 = \mathbb{Z}_+$, $n_I = 2$, $J = 4$, and $\lambda, \mu, \delta_1, \delta_2 \in \mathbb{R}_{>0}$. Furthermore, $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 = \mathbb{Z}_+^{1 \times 2}$. Observe that the states of the underlying infinite CTMC can be ordered as $(0, 0)$, $(0, 1)$, $(1, 1)$, $(1, 0)$, $(0, 2)$, $(1, 2)$, $(2, 2)$, $(2, 1)$, $(2, 0)$, $\ldots$, and these states can be partitioned into consecutive levels of subsets $\mathcal{S}^{(l)}$ for $l \in \mathbb{Z}_+$ in which states $(0, l)$, $(1, l)$, $\ldots$, (l, l), $\ldots$, $(l, 1)$, $(l, 0)$ belong to $\mathcal{S}^{(l)}$, the subset corresponding to level $l \in \mathbb{Z}_+$ and having $(2l + 1)$ states.

More formally, the subset of states corresponding to level $l \in \mathbb{Z}_+$ is defined as

$$\mathcal{S}^{(l)} = \{x \in \mathcal{S} \mid \max(x_1, x_2, \ldots, x_{n_I}) = l\}, \qquad \mathcal{S} = \bigcup_{l=0}^{\infty} \mathcal{S}^{(l)},$$

$$\mathcal{S}^{(l)} \cap \mathcal{S}^{(k)} = \varnothing \quad \text{for } l \neq k. \tag{2}$$

The justification for the maximum function follows from the observation that the maximum-valued state variable among $x_1, x_2, \ldots, x_{n_I}$ in any state $x \in \mathcal{S}$ does not change by more than one through any transition due to the particular form of the state change vectors in the transition classes. Since n is finite, even when $n = n_I$, the states within each level are well defined and finite in this representation. Now, we relate this to an infinite LDQBD process.

An infinite infinitesimal generator matrix of the block tridiagonal form

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & & & & \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & & & \\ & \ddots & \ddots & \ddots & & \\ & & Q_{l,l-1} & Q_{l,l} & Q_{l,l+1} & \\ & & & \ddots & \ddots & \ddots \end{pmatrix} \tag{3}$$

is said to correspond to an infinite LDQBD process. The nonzero blocks at level l are given by $Q_{l,l-1} \in \mathbb{R}_{\geq 0}^{|\mathcal{S}^{(l)}| \times |\mathcal{S}^{(l-1)}|}$, $Q_{l,l} \in \mathbb{R}^{|\mathcal{S}^{(l)}| \times |\mathcal{S}^{(l)}|}$, and $Q_{l,l+1} \in \mathbb{R}_{\geq 0}^{|\mathcal{S}^{(l)}| \times |\mathcal{S}^{(l+1)}|}$; (some of) their entries depend on the level number l. Note that only the diagonal entries in $Q_{l,l}$ are negative. Level 0 is the boundary level and has two nonzero blocks. There are an infinite number of levels, and transitions from level l are either to states within itself, to states in level $l - 1$, or to states in level $l + 1$. Clearly, the ordering of states within a level is only fixed up to a permutation.

In the following, we assume that Q in (3) is irreducible. This is a natural assumption since $\mathcal{S}$ is defined to include only states reachable from each other; however, it is also a consequence of our ergodicity assumption and implies that $Q_{l,l-1}$ and $Q_{l,l+1}$ are nonnegative rectangular matrices and $-Q_{l,l}$ is a nonsingular M-matrix [2, pp. 132–164], [23, pp. 45–46]. Recall that an M-matrix is any square matrix A of the form $A = \tau I - B$ with $\tau > 0$ and $B \geq 0$ for which $\tau \geq \rho(B)$. Here, $\rho(B)$ is the spectral radius of B, that is, the maximum magnitude of B's eigenvalues. The nonsingularity of $-Q_{l,l}$ follows from the fact that it has a positive diagonal, nonpositive off-diagonal entries, and nonnegative row sums with at least one positive row sum (this last property is a result of the irreducibility of Q). Being a nonsingular M-matrix, its inverse is nonnegative, that is, $-Q_{l,l}^{-1} \geq 0$ for $l \in \mathbb{Z}_+$.

Example 2. (*Example 1 continued: gene expression.*) For our particular example, with the ordering of states within levels as described, $Q_{0,0} = (-\lambda)$, $Q_{0,1} = \lambda e_3^\top$, and, for $l > 0$,

$$Q_{l,l-1} = \begin{array}{c} {\scriptstyle (0,l)} \\ {\scriptstyle \vdots} \\ {\scriptstyle (l-1,l)} \\ {\scriptstyle (l,l)} \\ {\scriptstyle (l,l-1)} \\ {\scriptstyle \vdots} \\ {\scriptstyle (l,0)} \end{array} \begin{pmatrix} l\delta_2 & & & & \\ & \ddots & & & \\ & & l\delta_2 & & \\ & & l\delta_1 & & \\ & & & \ddots & \\ & & & & l\delta_1 \end{pmatrix}_{(2l+1)\times(2l-1)},$$

$$Q_{l,l} = \begin{array}{c} {\scriptstyle (0,l)} \\ {\scriptstyle (1,l)} \\ {\scriptstyle \vdots} \\ {\scriptstyle (l-1,l)} \\ {\scriptstyle (l,l)} \\ {\scriptstyle (l,l-1)} \\ {\scriptstyle \vdots} \\ {\scriptstyle (l,1)} \\ {\scriptstyle (l,0)} \end{array} \begin{pmatrix} * & \lambda & & & & & \\ \delta_1 & * & \lambda & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & (l-1)\delta_1 & * & \lambda & & \\ & & & l\delta_1 & * & l\delta_2 & \\ & & & & l\mu & * & (l-1)\delta_2 \\ & & & & & \ddots & \ddots & \ddots \\ & & & & & & l\mu & * & \delta_2 \\ & & & & & & & l\mu & * \end{pmatrix}_{(2l+1)\times(2l+1)},$$

where the column headers for $Q_{l,l-1}$ are $(0,l-1) \cdots (l-1,l-1) \cdots (l-1,0)$.

$$Q_{l,l+1} = \begin{matrix} & \begin{matrix} (0,l+1) & (1,l+1) & \cdots & (l,l+1) & (l+1,l+1) & (l+1,l) & \cdots & (l+1,0) \end{matrix} \\ \begin{matrix} (0,l) \\ (1,l) \\ \vdots \\ (l,l) \\ \vdots \\ (l,0) \end{matrix} & \left(\begin{matrix} & \mu & & & & & & \\ & & \ddots & & & & & \\ & & & l\mu & & \lambda & & \\ & & & & & & \ddots & \\ & & & & & & & \lambda \end{matrix} \right)_{(2l+1)\times(2l+3)} \end{matrix}.$$

Now, we introduce two other examples from the literature.

Example 3. (*Toggle switch.*) Let us consider a system of stochastic chemical kinetics modeling the biological process associated with a toggle switch [8] through the transition classes in Table 2. Here, $n = 4$, $x = (x_1, x_2, x_3, x_4)$, $\mathcal{S}_1 = \mathcal{S}_2 = \mathbb{Z}_+$, $\mathcal{S}_3 = \mathcal{S}_4 = \{0, 1\}$, $n_I = 2$, $J = 8$, and $\lambda_1, \lambda_2, \delta_1, \delta_2, \beta_0, \beta_1 \in \mathbb{R}_{>0}$. Furthermore, $\mathcal{S} = \times_{i=1}^{4} \mathcal{S}_i$. The states of the underlying infinite CTMC can be ordered as $(0, 0, 0, 0)$, $(0, 0, 0, 1)$, $(0, 0, 1, 0)$, $(0, 0, 1, 1)$, $(0, 1, 0, 0)$, $(0, 1, 0, 1)$, $(0, 1, 1, 0)$, $(0, 1, 1, 1)$, $(1, 1, 0, 0)$, $(1, 1, 0, 1)$, $(1, 1, 1, 0)$, $(1, 1, 1, 1)$, $(1, 0, 0, 0)$, $(1, 0, 0, 1)$, $(1, 0, 1, 0)$, $(1, 0, 1, 1)$, ..., and can be partitioned into consecutive levels of subsets $\mathcal{S}^{(l)}$ for $l \in \mathbb{Z}_+$ in which states $(0, l, x_3, x_4)$, $(1, l, x_3, x_4)$, ..., (l, l, x_3, x_4), ..., $(l, 1, x_3, x_4)$, $(l, 0, x_3, x_4)$ belong to $\mathcal{S}^{(l)}$ having $8l + 4$ states. We remark that we again have a definition of levels as in (2).

Example 4. (*Exclusive switch.*) Let us consider a system of stochastic chemical kinetics modeling the biological process associated with an exclusive switch [18] through the transition classes in Table 3. Here, $n = 5$, $x = (x_1, x_2, x_3, x_4, x_5)$, $\mathcal{S}_1 = \mathcal{S}_2 = \mathbb{Z}_+$, $\mathcal{S}_3 = \mathcal{S}_4 = \mathcal{S}_5 = \{0, 1\}$, $n_I = 2$, $J = 10$, and $p, d, b, u \in \mathbb{R}_{>0}$. Furthermore, $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \{(0, 0, 1), (0, 1, 0), (1, 0, 0)\} \subset \times_{i=1}^{5} \mathcal{S}_i$. The states of the underlying infinite CTMC can be ordered as $(0, 0, 0, 0, 1)$, $(0, 0, 0, 1, 0)$, $(0, 0, 1, 0, 0)$, $(0, 1, 0, 0, 1)$, $(0, 1, 0, 1, 0)$, $(0, 1, 1, 0, 0)$, $(1, 1, 0, 0, 1)$, $(1, 1, 0, 1, 0)$, $(1, 1, 1, 0, 0)$, $(1, 0, 0, 0, 1)$, $(1, 0, 0, 1, 0)$, $(1, 0, 1, 0, 0)$, ..., and can be partitioned into consecutive levels of subsets $\mathcal{S}^{(l)}$ for $l \in \mathbb{Z}_+$ in which states $(0, l, x_3, x_4, x_5)$, $(1, l, x_3, x_4, x_5)$, ..., (l, l, x_3, x_4, x_5), ..., $(l, 1, x_3, x_4, x_5)$, $(l, 0, x_3, x_4, x_5)$

TABLE 2: Transition classes of the toggle switch model.

j	ϕ_j	$\alpha_j(x)$	$v^{(j)}$
1	$(\alpha_1, v^{(1)})$	$\lambda_1(1 - x_4)$	$e_1^\top$
2	$(\alpha_2, v^{(2)})$	$\delta_1 x_1$	$-e_1^\top$
3	$(\alpha_3, v^{(3)})$	$\beta_0 x_1(1 - x_3)$	$(-e_1 + e_3)^\top$
4	$(\alpha_4, v^{(4)})$	$\beta_1 x_3$	$(e_1 - e_3)^\top$
5	$(\alpha_5, v^{(5)})$	$\lambda_2(1 - x_3)$	$e_2^\top$
6	$(\alpha_6, v^{(6)})$	$\delta_2 x_2$	$-e_2^\top$
7	$(\alpha_7, v^{(7)})$	$\beta_0 x_2(1 - x_4)$	$(-e_2 + e_4)^\top$
8	$(\alpha_8, v^{(8)})$	$\beta_1 x_4$	$(e_2 - e_4)^\top$

TABLE 3: Transition classes of the exclusive switch model.

j	ϕ_j	$\alpha_j(x)$	$v^{(j)}$
1	$(\alpha_1, v^{(1)})$	$px_3(1 - x_4)(1 - x_5)$	$e_1^\top$
2	$(\alpha_2, v^{(2)})$	$px_3(1 - x_4)(1 - x_5)$	$e_2^\top$
3	$(\alpha_3, v^{(3)})$	dx_1	$-e_1^\top$
4	$(\alpha_4, v^{(4)})$	dx_2	$-e_2^\top$
5	$(\alpha_5, v^{(5)})$	$bx_1x_3(1 - x_4)(1 - x_5)$	$(-e_1 - e_3 + e_4)^\top$
6	$(\alpha_6, v^{(6)})$	$bx_2x_3(1 - x_4)(1 - x_5)$	$(-e_2 - e_3 + e_5)^\top$
7	$(\alpha_7, v^{(7)})$	$u(1 - x_3)x_4(1 - x_5)$	$(e_1 + e_3 - e_4)^\top$
8	$(\alpha_8, v^{(8)})$	$u(1 - x_3)(1 - x_4)x_5$	$(e_2 + e_3 - e_5)^\top$
9	$(\alpha_9, v^{(9)})$	$p(1 - x_3)x_4(1 - x_5)$	$e_1^\top$
10	$(\alpha_{10}, v^{(10)})$	$p(1 - x_3)(1 - x_4)x_5$	$e_2^\top$

belong to $\mathcal{S}^{(l)}$ having $6l + 3$ states. We remark that we again have a definition of levels as in (2). However, in this case $(x_3, x_4, x_5) \in \{(0, 0, 1), (0, 1, 0), (1, 0, 0)\} \subset \mathcal{S}_3 \times \mathcal{S}_4 \times \mathcal{S}_5$. That is, the values the last three state variables take are mutually exclusive. Note that the state change vectors in this model are not unique.

We remark that all three examples are different from each other in that they respectively have two, four, and five variables in the state vector only, the first two of which take values in the set of nonnegative integers. The first example does not have any state variables for control, whereas the second and third examples have two and three state variables for control, respectively. The second and the third examples are different from each other in that the two control variables in the former represent four distinct possibilities, whereas the three control variables in the latter represent only three distinct possibilities. Hence, the state space of the exclusive switch model is a proper subset of its product state space.

We have not specified the blocks in the block tridiagonal form corresponding to the second and third examples due to space limitations. However, observe that all nonzero blocks are sparse with nonzero entries having well-defined values and row/column indices for a given ordering of states. Hence, each block can be generated as a sparse matrix whenever needed.

3. Lyapunov bounds

The LDQBD process $\{X(t), \ t \geq 0\}$ is ergodic if and only if there exists a function $g \colon \mathcal{S} \to \mathbb{R}_{\geq 0}$, called a Lyapunov function, that guarantees the ergodicity of X as well as a finite attractor set $\mathcal{C} \subset \mathcal{S}$ [29]. More precisely, g and $\mathcal{C}$ meet the conditions

(i) $(\mathrm{d}/\mathrm{d}t) \, \mathrm{E}[g(X(t)) \mid X(t) = x] \leq -\gamma$ for all $x \in \mathcal{S} \setminus \mathcal{C}$,

(ii) $(\mathrm{d}/\mathrm{d}t) \, \mathrm{E}[g(X(t)) \mid X(t) = x] < \infty$ for all $x \in \mathcal{C}$, and

(iii) $\{x \in \mathcal{S} \mid g(x) \leq r\}$ is finite for all $r < \infty$

for some $\gamma > 0$.

The value $d(x) = (\mathrm{d}/\mathrm{d}t)\,\mathrm{E}[g(X(t)) \mid X(t) = x]$ is called the drift in state x and is given by

$$d(x) = \sum_{j=1}^{J} \alpha_j(x)(g(x + v^{(j)}) - g(x)). \tag{4}$$

Observe that, with a slight abuse of the functional notation introduced earlier, $g(x)$ and $d(x)$ for all $x \in \mathcal{S}$ can be written in the form of row vectors $g \in \mathbb{R}^{1 \times |\mathcal{S}|}$ and $d \in \mathbb{R}^{1 \times |\mathcal{S}|}$, respectively. Then (4) can be expressed in matrix-vector form as

$$d^\top = Q g^\top, \tag{5}$$

assuming the same ordering of states for the vectors g and d as in Q.

Now, we derive a nontrivial upper bound on the stationary probability mass that resides outside the set $\mathcal{C}$ by a scaling of g with $c + \gamma$, where $c = \sup_{x \in \mathcal{S}} d(x)$ and γ is as in condition (i). Note that conditions (i) and (ii) imply that $c < \infty$. We define $g^*(x) = g(x)/(c + \gamma)$ and $d^*(x) = (\mathrm{d}/\mathrm{d}t)\,\mathrm{E}[g^*(X(t)) \mid X(t) = x]$. This yields

$$d^*(x) = \frac{d(x)}{c + \gamma} \leq \frac{c}{c + \gamma} - \chi_{\bar{\mathcal{C}}}(x) \tag{6}$$

for $x \in \mathcal{S}$, where $\chi_{\bar{\mathcal{C}}}(x) = 1$ if $x \in \mathcal{S} \setminus \mathcal{C}$ and 0 if $x \in \mathcal{C}$. Since X is ergodic and Q has a finite number of nonzeros in each row and column, it follows from (5) that $\pi d^\top = \pi Q g^\top = 0$ and $\pi (d^*)^\top = \pi Q (g^*)^\top = 0$. Together with (6) we get

$$\sum_{x \in \mathcal{S} \setminus \mathcal{C}} \pi(x) \leq \frac{c}{c + \gamma}.$$

We remark that $c > 0$. This is so because π is positive, d contains at least one negative entry due to condition (i), and $\pi d^\top = 0$. These together imply at least one positive entry in d. Since $\gamma > 0$ and $c > 0$, this gives a nontrivial upper bound on the stationary probability of being in $\mathcal{S} \setminus \mathcal{C}$, and, therefore, the stationary probability mass inside $\mathcal{C}$ must be at least $1 - c/(c + \gamma)$.

Given a Lyapunov function g that satisfies condition (iii) and $c = \sup_{x \in \mathcal{S}} d(x) < \infty$, an upper bound $\varepsilon \in (0, 1)$ can be a priori specified as $\varepsilon = c/(c + \gamma)$. Then $\gamma = c/\varepsilon - c$, and this yields the set

$$\mathcal{C} = \{x \in \mathcal{S} \mid d^*(x) > \varepsilon - 1\},$$

because the condition $d^*(x) > \varepsilon - 1$ is equivalent to $d(x) > -\gamma$. If $\mathcal{C}$ is finite, the aforementioned conditions above are fulfilled and $\sum_{x \in \mathcal{C}} \pi(x) \geq 1 - \varepsilon$.

Here, we are interested in determining the pair of level numbers, (low, high), of the LDQBD process such that the states between levels low and high inclusively encompass $\mathcal{C}$. Thus, we have

$$\texttt{low} = \min\{l \in \mathbb{Z}_+ \mid \mathcal{S}^{(l)} \cap \mathcal{C} \neq \varnothing\} \quad \text{and} \quad \texttt{high} = \max\{l \in \mathbb{Z}_+ \mid \mathcal{S}^{(l)} \cap \mathcal{C} \neq \varnothing\},$$

and the finite set $\bigcup_{\texttt{low} \leq l \leq \texttt{high}} \mathcal{S}^{(l)}$ contains at least $1 - \varepsilon$ of the stationary probability mass.

Example 5. (*Example 1 continued: gene expression.*) We first consider the function

$$g(x_1, x_2) = \max(x_1, x_2).$$

A few moments of reflection reveal that it does not guarantee the ergodicity of the LDQBD process, because, when $x_1 = x_2$, from (4), the drift is $d(x_1, x_2) = \lambda + \mu x_1$, and $d(x_1, x_2)$ increases as x_1 increases.

The function $g(x_1, x_2) = x_1 + x_2$ works only if $\mu < \delta_1$, because from (4) the drift is $d(x_1, x_2) = \lambda + \mu x_1 - \delta_1 x_1 - \delta_2 x_2$. If $\mu \geq \delta_1$ then $d(x_1, x_2)$ is positive for infinitely many states.

On the other hand, the function

$$g(x_1, x_2) = \frac{2\mu x_1}{\delta_1} + x_2$$

is suitable, since from (4) the drift is

$$d(x_1, x_2) = \frac{2\lambda\mu}{\delta_1} - \mu x_1 - \delta_2 x_2,$$

and $d(x_1, x_2)$ decreases as x_1 and x_2 increase.

In the following, we use two subscripts to indicate the example number and the parameter set, respectively, and report the values of c and γ with two decimal digits of precision after the decimal point.

Parameter set 1. Let $\lambda = \mu = \delta_2 = 0.05$ and $\delta_1 = 0.005$. Then

$$d_{1,1}(x_1, x_2) = 1 - 0.05x_1 - 0.05x_2,$$

implying that $d_{1,1}(x_1, x_2) \leq 1$ with equality attained at $(0, 0)$. Hence, $c_{1,1} = 1$. When $\varepsilon = 0.1$, we obtain $\gamma_{1,1} = 9$, and $\mathcal{C}_{1,1} = \{(x_1, x_2) \mid d_{1,1}(x_1, x_2) \geq -9\}$ for which $(\texttt{low}, \texttt{high}) = (0, 200)$.

Parameter set 2. Now, let us consider the same model with the parameters $\lambda = 60$, $\mu = \delta_2 = 0.01$, and $\delta_1 = 0.2$. Even though the Lyapunov function $g(x_1, x_2) = 2\mu x_1/\delta_1 + x_2$ would work, we use the function

$$g_{1,2}(x_1, x_2) = (x_1 - 300)^2 + (x_2 - 300)^2.$$

The reason is that $g_{1,2}(x)$ measures the squared distance of x to $x^* = (300, 300)$, which is the equilibrium point of the mean field of the underlying CTMC. That is, we have $g_{1,2}(x) = \|x - x^*\|_2^2$, the squared Euclidean norm of $x - x^*$.

More precisely, from Propositions 2.2 and 2.3 of [7], we derive

$$\frac{d}{dt} \mathrm{E}[X_1(t)] = \lambda - \delta_1 \mathrm{E}[X_1(t)],$$

$$\frac{d}{dt} \mathrm{E}[X_2(t)] = \mu \mathrm{E}[X_1(t)] - \delta_2 \mathrm{E}[X_2(t)].$$

Setting the left-hand side of the system of differential equations to 0 yields the equilibrium point $x^* = (\lambda/\delta_1, \lambda\mu/\delta_1\delta_2)$.

For parameter set 2, $x^* = (300, 300)$, and, from (4), the drift with respect to $g_{1,2}(x_1, x_2)$ is given by

$$d_{1,2}(x_1, x_2) = -0.4x_1^2 + 0.02x_1x_2 + 234.21x_1 - 0.02x_2^2 + 6.01x_2 - 35\,940.$$

Using Geobound (see http://alma.cs.uni-saarland.de/?page_id=74), we obtain $c_{1,2} = 126.03$. Furthermore, we remark that the drift is positive in only finitely many states. When $\varepsilon = 0.05$, we obtain $\gamma_{1,2} = 2394.58$. Thus, the set

$$\mathcal{C}_{1,2} = \{(x_1, x_2) \mid d_{1,2}(x_1, x_2) \geq -2394.58\}$$

contains at least 95% of the stationary probability mass. Using the level definition in (2), we obtain $(\text{low}, \text{high}) = (221, 657)$.

Example 6. (*Example 3 continued: toggle switch.*) We consider the parameters $\lambda_1 = \lambda_2 = 6$, $\delta_1 = \delta_2 = 0.4$, $\beta_0 = 1$, and $\beta_1 = 0.5$. Similar to the gene expression model, we then estimate the location of attractor points by computing the equilibrium point $x^* = (x_1^*, x_2^*, x_3^*, x_4^*)$ of the mean-field equation as

$$0 = 6(1 - x_4^*) - 0.4x_1^* - x_1(1 - x_3^*) + 0.5x_3^*,$$
$$0 = 6(1 - x_3^*) - 0.4x_2^* - x_2^*(1 - x_4^*) + 0.5x_4^*,$$
$$0 = x_1^*(1 - x_3^*) - 0.5x_3^*,$$
$$0 = x_2^*(1 - x_4^*) - 0.5x_4^*.$$

This yields $x_1^* = x_2^* = \frac{5}{2}$ and $x_3^* = x_4^* = \frac{5}{6}$. We remark that this is only a rough estimate of $E[X(t)]$ when t approaches ∞ [7]. We again choose the Lyapunov function as $g_{2,1}(x) = \|x - x^*\|_2^2$. With $g_{2,1}(x)$, from (4) we get the drift function

$$d_{2,1}(x_1, x_2, x_3, x_4) = 2x_1^2x_3 - 2.8x_1^2 - \tfrac{13}{3}x_1x_3 - 12x_1x_4 + \tfrac{296}{15}x_1 + 2x_2^2x_4 - 2.8x_2^2$$
$$- \tfrac{13}{3}x_2x_4 - 12x_2x_3 + \tfrac{296}{15}x_2 + \tfrac{67}{3}x_3 + \tfrac{67}{3}x_4 - 48.$$

Using Geobound, we obtain the global maximum $c_{2,1} = 53.79$. Note that $d_{2,1}(x)$ is positive for only a finite number of states, because, for each positive term in $d_{2,1}(x)$, there is a negative term with higher order. This can be easily verified by considering all four possibilities for (x_3, x_4). For $\varepsilon = 0.05$, we compute $\gamma_{2,1} = 1021.92$. This yields a set $\mathcal{C}_{2,1}$ with $(\text{low}, \text{high}) = (0, 46)$. Note that the Lyapunov function $g_{2,1}(x)$ yields tighter bounds than, for instance, $g(x) = \|x\|_2^2$ or $g(x) = x_1 + x_2 + x_3 + x_4$. The reason is that, in the latter case we measure the distance from the origin, but the states with high stationary probability are also located around $(x_1, x_2) = (14, 0)$ and $(x_1, x_2) = (0, 14)$. For example, $g(x) = x_1 + x_2 + x_3 + x_4$ yields the pair $(\text{low}, \text{high}) = (0, 599)$.

Example 7. (*Example 4 continued: exclusive switch.*) *Parameter set 1.* We choose the parameters $p = 0.05$, $d = 0.005$, $b = 0.01$, and $u = 0.008$. For the Lyapunov function $g_{3,1}(x) = \|x\|_2^2$, we get the drift

$$d_{3,1}(x_1, x_2, x_3, x_4, x_5) = -0.02x_1^2x_3 - 0.01x_1^2 + 0.11x_1x_3 + 0.116x_1x_4 + 0.005x_1$$
$$- 0.02x_2^2x_3 - 0.01x_2^2 + 0.11x_2x_3 + 0.116x_2x_5 + 0.005x_2$$
$$+ 0.042x_3 + 0.058.$$

By checking all valid combinations of $x_3, x_4, x_5 \in \{0, 1\}$, it is possible to verify that the drift is positive for only finitely many states. We compute the global maximum $c_{3,1} = 0.42$ using Geobound. For $\varepsilon = 0.05$, we get $\gamma_{3,1} = 8.07$ and $(\text{low}, \text{high}) = (0, 35)$.

Parameter set 2. Now we choose $p = 0.5$, $d = 0.005$, $b = 0.00001$, and $u = 0.008$, and consider the Lyapunov function

$$g_{3,2}(x_1, x_2, x_3, x_4, x_5) = \left\| (100, 80) - (x_1, x_2) \right\|_2^2 \left\| (80, 100) - (x_1, x_2) \right\|_2^2.$$

The corresponding drift has as its maximum $c_{3,2} = 1.55 \times 10^5$, again computed using Geobound. For $\varepsilon = 0.05$, we obtain $\gamma_{3,2} = 2.94 \times 10^6$ and $(\texttt{low}, \texttt{high}) = (1, 206)$, while $\varepsilon = 0.1$ yields $\gamma_{3,2} = 1.39 \times 10^7$ and $(\texttt{low}, \texttt{high}) = (13, 189)$. Since the number of states per level increases with the level number, the tight bounds of $(13, 189)$ are computationally more attractive. Furthermore, experiments we report later show that including levels beyond 189 does not improve accuracy in this example.

4. Matrix-analytic computations

Our objective is to compute the stationary probability distribution

$$\pi = (\pi^{(0)}, \pi^{(1)}, \dots, \pi^{(l)}, \dots)$$

of the ergodic infinite LDQBD process partitioned conformally with its underlying infinitesimal generator matrix Q in (3). Thus, we seek the unique positive solution to $\pi Q = 0$ subject to $\pi e = 1$.

In our setting, we are given the pair $(\texttt{low}, \texttt{high})$, and the task amounts to obtaining the matrix subsequence $\tilde{R}_l \in \mathbb{R}_{\geq 0}^{|\mathscr{S}^{(l)}| \times |\mathscr{S}^{(l+1)}|}$ that approximates R_l for $l \in \{\texttt{low}, \texttt{low} + 1, \dots, \texttt{high}\}$. Here, the entry $R_l(x, y)$ is the expected sojourn time in state $y \in \mathscr{S}^{(l+1)}$ at level $l + 1$ per unit sojourn in state $x \in \mathscr{S}^{(l)}$ at level l before returning to level l, given the process started in state x at level l. We remark that R_l (and, therefore, $\tilde{R}_l$) is nonnegative and rectangular. Hence, the behavior of its largest singular value, denoted by $\sigma_1(R_l)$ for $l \in \{\texttt{low}, \texttt{low} + 1, \dots, \texttt{high}\}$ is likely to provide insight regarding the behavior of the stationary probability distribution of the system under study. Recall that the singular values of a real rectangular matrix C are the square roots of the eigenvalues of $C^\top C$, which is symmetric positive definite and, therefore, has positive real eigenvalues (see, for instance, [12, pp. 410–411, 427]).

Given the approximations $\tilde{R}_l$, we can obtain approximations $\tilde{\pi}^{(l)} \in \mathbb{R}_{\geq 0}^{1 \times |\mathscr{S}^{(l)}|}$ of $\pi^{(l)}$ by using the boundary condition

$$\tilde{\pi}^{(\texttt{low})}(Q_{\texttt{low},\texttt{low}} + \tilde{R}_{\texttt{low}} Q_{\texttt{low}+1,\texttt{low}}) = 0 \qquad (7)$$

and the recurrence

$$\tilde{\pi}^{(l+1)} = \tilde{\pi}^{(l)} \tilde{R}_l$$

subject to $(\tilde{\pi}^{(\texttt{low})}, \tilde{\pi}^{(\texttt{low}+1)}, \dots, \tilde{\pi}^{(\texttt{high})})e = 1$ [4]. We assume that $\tilde{\pi}^{(l)} = 0$ for $l \notin \{\texttt{low}, \texttt{low} + 1, \dots, \texttt{high}\}$ since the stationary probability mass of these levels is at most ε.

When the infinite LDQBD process is ergodic, the sequence of R_l matrices for $l > 0$ is given by

$$R_l = \sum_{k=0}^{\infty} U_l^{(k)} \prod_{m=0}^{k-1} D_{l+2^{k-m}}^{(k-1-m)}, \qquad (8)$$

where

$$U_l^{(0)} = Q_{l,l+1}(-Q_{l+1,l+1})^{-1}, \qquad D_{l+2}^{(0)} = Q_{l+2,l+1}(-Q_{l+1,l+1})^{-1}, \qquad (9)$$

and, for $k > 0$,

$$U_l^{(k)} = U_l^{(k-1)} U_{l+2^{k-1}}^{(k-1)} (I - U_{l+2^k}^{(k-1)} D_{l+3.2^{k-1}}^{(k-1)} - D_{l+2^k}^{(k-1)} U_{l+2^{k-1}}^{(k-1)})^{-1},$$

$$D_{l+2^{k+1}}^{(k)} = D_{l+2^{k+1}}^{(k-1)} D_{l+3.2^{k-1}}^{(k-1)} (I - U_{l+2^k}^{(k-1)} D_{l+3.2^{k-1}}^{(k-1)} - D_{l+2^k}^{(k-1)} U_{l+2^{k-1}}^{(k-1)})^{-1} \qquad (10)$$

are pairs of nonnegative matrices [4] defined recursively. The pair in (9) for $k = 0$ supplies the boundary conditions with $U_l^{(0)} \in \mathbb{R}_{\geq 0}^{|\mathcal{S}^{(l)}| \times |\mathcal{S}^{(l+1)}|}$, $D_{l+2}^{(0)} \in \mathbb{R}_{\geq 0}^{|\mathcal{S}^{(l+2)}| \times |\mathcal{S}^{(l+1)}|}$, and the pair in (10) for $k > 0$ satisfies $U_l^{(k)} \in \mathbb{R}_{\geq 0}^{|\mathcal{S}^{(l)}| \times |\mathcal{S}^{(l+2^k)}|}$, $D_{l+2^{k+1}}^{(k)} \in \mathbb{R}_{\geq 0}^{|\mathcal{S}^{(l+2^{k+1})}| \times |\mathcal{S}^{(l+2^k)}|}$. Note that the inverse matrices in (9) and (10) are the same, and each can be obtained once and used in the other [28].

The kth term in (8) calculates the expected sojourn times at level $l + 1$ for sample paths that go to levels between $l - 1 + 2^k$ and $l - 1 + 2^{k+1}$ [22]. Hence, if the infinite sum in (8) is truncated using $K \geq 0$ as the upper limit, as in

$$R_l^{(K)} = \sum_{k=0}^{K} U_l^{(k)} \prod_{m=0}^{k-1} D_{l+2^{k-m}}^{(k-1-m)}, \qquad (11)$$

then only sample paths below level $l - 1 + 2^{K+1}$ are considered. Note that, for $K = 0$, $R_l^{(0)} = U_l^{(0)}$, which is nonzero and nonnegative. Furthermore, $R_l^{(K+1)} \geq R_l^{(K)}$ since $U_l^{(k)} \geq 0$ and $D_{l+2^{k-m}}^{(k-1-m)} \geq 0$ for $0 \leq k \leq K$, $0 \leq m < k$. This also follows from the physical interpretations of $R_l^{(K+1)}$ and $R_l^{(\overline{K})}$.

As pointed out in [4], rather than (8), we can equivalently use

$$\tilde{R}_l = Q_{l,l+1}(-Q_{l+1,l+1} - \tilde{R}_{l+1} Q_{l+2,l+1})^{-1} \quad \text{for } \texttt{low} \leq l < \texttt{high} \qquad (12)$$

if there is an accurate $\tilde{R}_{\texttt{high}}$ to start with.

There are various approaches for obtaining $\tilde{R}_{\texttt{high}}$ given $\texttt{high}$.

(A1) As in [4], it is possible to try to compute a value for K in (11) by considering an increasing number of terms until the maximum entry in absolute value in the difference between two consecutive approximations is less than a specified tolerance, say 10^{-10}. That is, K is chosen as the smallest positive integer which satisfies $\max(R_{\texttt{high}}^{(K)} - R_{\texttt{high}}^{(K-1)}) < 10^{-10}$, and then $\tilde{R}_{\texttt{high}}$ is set to $R_{\texttt{high}}^{(K)}$.

(A2) A simpler approach is to limit the number of terms in (11) to five or six, that is, to use $R_{\texttt{high}}^{(4)}$ or $R_{\texttt{high}}^{(5)}$ as $\tilde{R}_{\texttt{high}}$ [4]. In either case, the maximum entry in absolute value in the difference between the last two approximations (that is, $\max(R_{\texttt{high}}^{(4)} - R_{\texttt{high}}^{(3)})$ or $\max(R_{\texttt{high}}^{(5)} - R_{\texttt{high}}^{(4)})$, respectively) can be used as an indication of the error in the approximation, although it does not provide any guarantees.

(A3) An alternative discussed in [1] and [14] is to let $\tilde{R}_{\texttt{high}}$ be the zero matrix. The existence of $\tilde{R}_{\texttt{high}-1}$ in (12) is guaranteed by the fact that $-Q_{\texttt{high,high}}$ is a nonsingular M-matrix. However, it is not clear how much approximation error is incurred with this approach.

(A4) One variation on (A3) is to add $Q_{\texttt{high,high}+1}e$ to the last column of $Q_{\texttt{high,high}}$ during the computation of $\tilde{R}_{\texttt{high}-1}$ in (12). However, this is known not to be optimal in some cases [32].

In any case, the important thing is to have a measure against which these approaches can be compared. To this end, $\|\tilde{\pi} Q\|_\infty$, the infinity norm of the residual vector corresponding to the approximate stationary distribution $\tilde{\pi}$ obtained with the respective approach, emerges as a convenient measure.

Now, we turn to implementation issues and remark that the computation of $\tilde{R}_l$ in (12) and $\tilde{\pi}_{\texttt{low}}$ in (7) are prone to inaccuracies. The inaccuracy of the former stems from the possibility that matrices whose inverses appear in (12) are close to being singular in machine precision. Note that in exact arithmetic these matrices should be nonsingular. The inaccuracy of the latter is somewhat different and more serious in that $(Q_{\texttt{low},\texttt{low}} + \tilde{R}_{\texttt{low}}Q_{\texttt{low}+1,\texttt{low}})$ in (7) must be singular in exact arithmetic for $\tilde{\pi}_{\texttt{low}}$ to be nonnegative since the right-hand side is 0. Now we have the possibility that a matrix, which must be singular, is nonsingular in machine precision.

The remedy for both of these situations is to employ the GTH idea of using only positive floating-point arithmetic as discussed in [5]. The fix for the former kind of problem is straightforward. The matrix to be inverted can be factorized using the GTH idea since its inverse should be nonnegative. Then a linear system with right-hand side I should be solved for the matrix to be inverted. The fix for the latter kind of problem when $\texttt{low} = 0$ is similar. However, for the case $\texttt{low} > 0$, it is not so obvious; but, we have observed that not augmenting $Q_{\texttt{low},\texttt{low}}$ in any way (similar to $Q_{\texttt{high},\texttt{high}}$) yields the best accuracy.

5. Numerical results

All numerical experiments are carried out on an Intel® Core™2 Duo T7700 at 2.4 GHz with 2 GB main memory, except those with the second parameter set of the gene expression model which are carried out on an Intel Core i7 at 2.8 GHz with 8 GB main memory since they cannot be solved with 2 GB. Approaches (A1), (A2), (A3), and (A4) are implemented in MATLAB® and the code is available online at http://www.cs.bilkent.edu.tr/~tugrul/software.html.

In the tables, the first two columns give the values of ε and the corresponding $(\texttt{low}, \texttt{high})$ pair for the particular Lyapunov function devised in Section 3. We remark that $(1 - \varepsilon) \times 100$ is a lower bound on the percentage of the stationary probability mass that lies between levels $\texttt{low}$ and $\texttt{high}$ inclusive. Column 'Approach' indicates the approach used in the solution procedure. A threshold of 1e−16 beside the name of the approach indicates that nonzero entries less than 10^{-16} in all intermediate matrices are dropped. Column 'T_{setup}' lists the time in seconds to compute $\tilde{R}_{\texttt{high}}$ and column 'T_{solve}' lists the time in seconds to solve the problem using the method discussed in the previous section with the particular approach. The sum of the values in columns T_{setup} and T_{solve} is the total solution time. Column 'Res' gives the infinity norm of the residual vector (i.e. Res $= \|\tilde{\pi} Q\|_\infty$). Note that the value in column T_{setup} can be larger than 0 only for approaches (A1) and (A2), and T_{solve} includes the time to compute the value in column Res. Finally, the last column gives Tol, which is equal to $\max(R_{\texttt{high}}^{(4)} - R_{\texttt{high}}^{(3)})$ for (A2). Since (A2) does not yield competitive total solution times compared to (A3), we have not experimented with (A1) at all. It is expected for the total solution time with (A1) to be larger than that of (A2) since its T_{setup} is almost always costlier. The results with (A4) are the same as those for (A3) and, therefore, are not reported. Since solving (12) using the GTH idea has not improved accuracy of the solutions, we do not report results with GTH.

5.1. Gene expression

Parameter set 1. With a lower bound of 90% on the stationary probability mass, $(\texttt{low}, \texttt{high}) = (0, 200)$ and a residual norm in the order of 10^{-17} is obtained in about 13 seconds using (A3) (see Table 4). The LDQBD process restricted to levels $(0, 200)$ has 40 401 states. Observe that the marginal stationary probability across levels is maximized at level 10 (see Figure 1(a)), whereas the marginal stationary probability across the infinite state variables is maximized at level 9 (see Figure 1(b)). The maximum singular value of the matrices of conditional expected sojourn times appears at level 0 in Figure 1(c). We remark that there is

TABLE 4: Results for the gene expression model with parameter set 1: $(\lambda, \mu, \delta_1, \delta_2) = (0.05, 0.05, 0.005, 0.05)$ and $g_{1,1}(x_1, x_2) = 2\mu x_1/\delta_1 + x_2$.

ε	(low, high)	Approach	T_{setup} (s)	T_{solve} (s)	Res	Tol
1e−1	(0, 200)	(A2)	286	12	6e−17	2e−3
		(A2), 1e−16	19	5	6e−17	2e−3
		(A3)	0	13	6e−17	
		(A3), 1e−16	0	9	6e−17	

(a) $\|\tilde{\pi}^{(l)}\|_1$, $\max(\tilde{\pi}^{(l)})$, and $\min(\tilde{\pi}^{(l)})$ across levels

(b) $\tilde{\pi}_{(x_1, x_2)}$ across population sizes
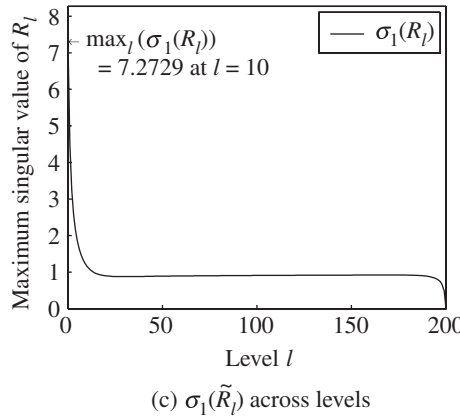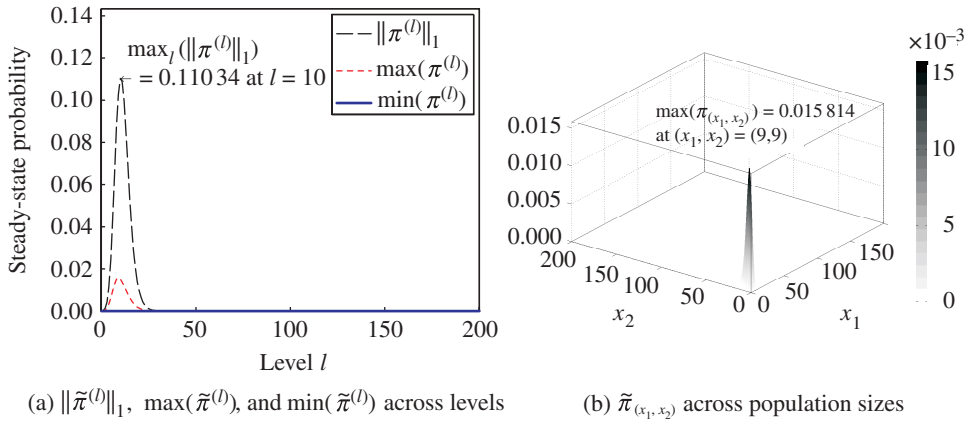

(c) $\sigma_1(\tilde{R}_l)$ across levels

FIGURE 1: Plots for the gene expression model with parameter set 1: $(\lambda, \mu, \delta_1, \delta_2) = (0.05, 0.05, 0.005, 0.05)$ and (low, high) = (0, 200).

no need to consider a smaller ε since the solution already has a residual norm in the order of 10^{-17} for $\varepsilon = 0.1$

Parameter set 2. With a lower bound of 95% on the stationary probability mass, (low, high) = (221, 657) and a residual norm in the order of 10^{-13} is obtained in about 41 minutes using (A3) (see Table 5). The LDQBD process restricted to levels (221, 657) has 384 123 states. Observe that the marginal stationary probability across levels is maximized at level 308

TABLE 5: Results for the gene expression model with parameter set 2: $(\lambda, \mu, \delta_1, \delta_2) = (60, 0.01, 0.2, 0.01)$ and $g_{1,2}(x_1, x_2) = (x_1 - 300)^2 + (x_2 - 300)^2$.

ε	(low, high)	Approach	T_{setup} (s)	T_{solve} (s)	Res	Tol
1e−1	(245, 552)	(A2)	2859	1078	3e−8	2e−4
		(A2), 1e−16	236	1202	3e−8	2e−4
		(A3)	0	1059	3e−8	
		(A3), 1e−16	0	1098	3e−8	
5e−2	(221, 657)	(A2)	3687	2463	2e−13	1e−4
		(A2), 1e−16	350	2557	2e−13	1e−4
		(A3)	0	2442	2e−13	
		(A3), 1e−16	0	2484	2e−13	

(a) $\|\tilde{\pi}^{(l)}\|_1$, $\max(\tilde{\pi}^{(l)})$, and $\min(\tilde{\pi}^{(l)})$ across levels

(b) $\tilde{\pi}_{(x_1, x_2)}$ across infinite state variables
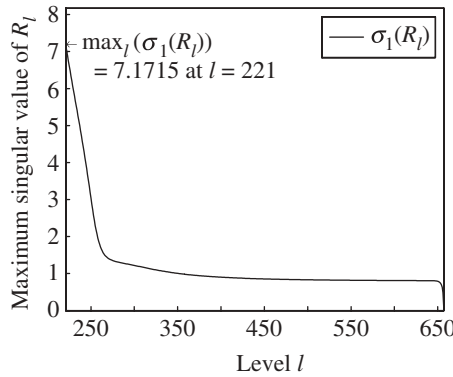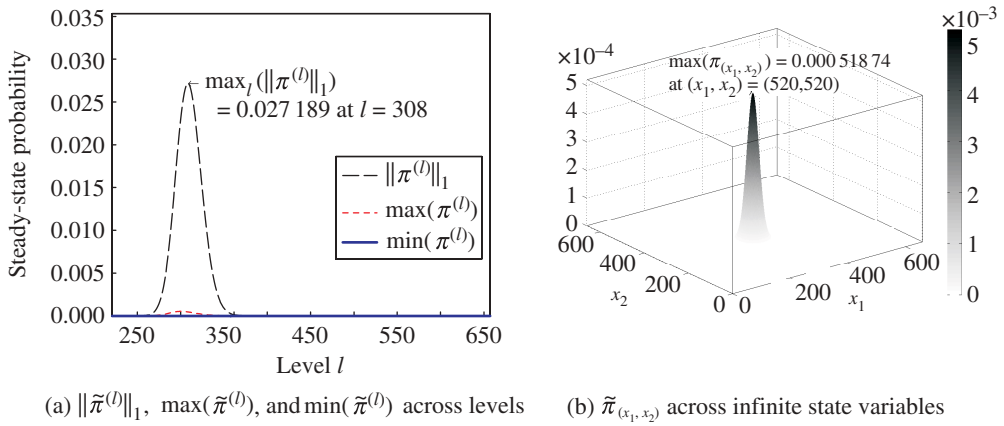

(c) $\sigma_1(\tilde{R}_l)$ across levels

FIGURE 2: Plots for the gene expression model with parameter set 2: $(\lambda, \mu, \delta_1, \delta_2) = (60, 0.01, 0.2, 0.01)$ and (low, high) = (221, 657).

(see Figure 2(a)), whereas the marginal stationary probability across the infinite state variables is maximized at level 520 (see Figure 2(b)). The maximum singular value of the matrices of conditional expected sojourn times appears at level 221 in Figure 2(c). In order to see the effect

of using `low` $= 0$ on accuracy without changing the value of `high`, we have carried out an additional experiment on the same platform. The results show that a residual norm in the order of 10^{-14} is obtained with $(\texttt{low}, \texttt{high}) = (0, 552)$ using (A3). This indicates that accuracy improves considerably if lower levels are taken into account.

5.2. Toggle switch

With a lower bound of 99% on the stationary probability mass, $(\texttt{low}, \texttt{high}) = (0, 91)$ and a residual norm in the order of 10^{-16} is obtained in 9 seconds using (A3) (see Table 6). The LDQBD process restricted to levels $(0, 91)$ has 33 856 states. In this problem, there is a slight loss of accuracy when entries less than 10^{-16} in intermediate matrices are dropped. Observe that the marginal stationary probability across levels is maximized at level 3 (see Figure 3(a)), whereas the marginal stationary probability across the infinite state variables is maximized at level 1 (see Figure 3(b)). The maximum singular value of the matrices of conditional expected sojourn times appears at level 0 in Figure 3(c).

5.3. Exclusive switch

Parameter set 1. With a lower bound of 99% on the stationary probability mass, $(\texttt{low}, \texttt{high}) = (0, 71)$ and a residual norm in the order of 10^{-18} is obtained in 2 seconds using (A3) (see Table 7). The LDQBD process restricted to levels $(0, 71)$ has 15 552 states. Observe that the marginal stationary probability across levels is maximized at level 9 (see Figure 4(a)), whereas the marginal stationary probability across the infinite state variables is maximized at level 10 (see Figure 4(b)). The maximum singular value of the matrices of conditional expected sojourn times appears at level 0 in Figure 4(c).

Parameter set 2. With a lower bound of 90% on the stationary probability mass, $(\texttt{low}, \texttt{high}) = (13, 189)$ and a residual norm in the order of 10^{-17} is obtained in about 1.5 minutes using (A3) (see Table 8). The LDQBD process restricted to levels $(13, 189)$ has 103 293 states. Observe that the marginal stationary probability across levels is maximized at level 102 (see Figure 5(a)), whereas the marginal stationary probability across the infinite state variables is maximized at level 110 (see Figure 5(b)). The maximum singular value of the matrices of conditional expected sojourn times appears at level 27 in Figure 5(c).

TABLE 6: Results for the toggle switch model with $(\lambda_1, \lambda_2, \delta_1, \delta_2, \beta_0, \beta_1) = (6, 6, 0.4, 0.4, 1, 0.5)$ and $g_{2,1}(x_1, x_2, x_3, x_4) = (x_1 - \frac{5}{2})^2 + (x_2 - \frac{5}{2})^2 + (x_3 - \frac{5}{6})^2 + (x_4 - \frac{5}{6})^2$.

ε	(`low`, `high`)	Approach	T_{setup} (s)	T_{solve} (s)	Res	Tol
1e$-$1	(0, 35)	(A2)	52	0	3e$-$6	4e$-$8
		(A2), 1e$-$16	10	1	3e$-$6	4e$-$8
		(A3)	0	0	3e$-$6	
		(A3), 1e$-$16	0	1	3e$-$6	
5e$-$2	(0, 46)	(A2)	88	1	4e$-$11	1e$-$9
		(A2), 1e$-$16	13	3	4e$-$11	1e$-$9
		(A3)	0	1	5e$-$11	
		(A3), 1e$-$16	0	3	5e$-$11	
1e$-$2	(0, 91)	(A2)	675	9	8e$-$17	8e$-$14
		(A2), 1e$-$16	63	28	1e$-$15	8e$-$14
		(A3)	0	9	8e$-$17	
		(A3), 1e$-$16	0	27	1e$-$15	

(a) $\|\tilde{\pi}^{(l)}\|_1$, $\max(\tilde{\pi}^{(l)})$, and $\min(\tilde{\pi}^{(l)})$ across levels

(b) $\tilde{\pi}_{(x_1, x_2)}$ across infinite state variables

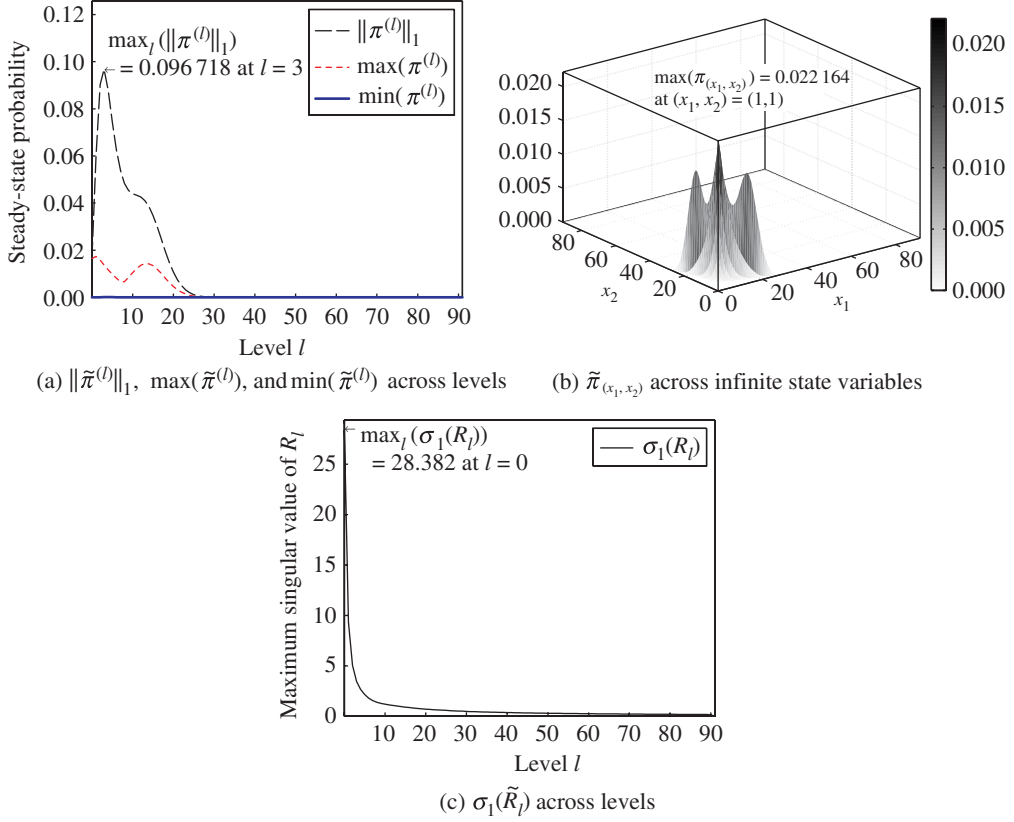(c) $\sigma_1(\tilde{R}_l)$ across levels

FIGURE 3: Plots for the toggle switch model with $(\lambda_1, \lambda_2, \delta_1, \delta_2, \beta_0, \beta_1) = (6, 6, 0.4, 0.4, 1, 0.5)$ and $(\text{low}, \text{high}) = (0, 91)$.

TABLE 7: Results for the exclusive switch model with parameter set 1: $(p, d, b, u) = (0.05, 0.005, 0.01, 0.008)$ and $g_{3,1}(x_1, x_2, x_3, x_4, x_5) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2$.

ε	$(\text{low}, \text{high})$	Approach	T_{setup} (s)	T_{solve} (s)	$\|\tilde{\pi} Q\|_\infty$	Tol
1e$-$1	$(0, 26)$	(A2)	13	0	2e$-$7	3e$-$9
		(A2), 1e$-$16	6	0	2e$-$7	3e$-$9
		(A3)	0	0	2e$-$7	
		(A3), 1e$-$16	0	0	2e$-$7	
5e$-$2	$(0, 35)$	(A2)	22	0	7e$-$12	8e$-$11
		(A2), 1e$-$16	6	0	7e$-$12	8e$-$11
		(A3)	0	0	7e$-$12	
		(A3), 1e$-$16	0	0	7e$-$12	
1e$-$2	$(0, 71)$	(A2)	111	2	3e$-$18	5e$-$15
		(A2), 1e$-$16	14	7	2e$-$18	5e$-$15
		(A3)	0	2	3e$-$18	
		(A3), 1e$-$16	0	7	2e$-$18	

(a) $\|\tilde{\pi}^{(l)}\|_1$, $\max(\tilde{\pi}^{(l)})$, and $\min(\tilde{\pi}^{(l)})$ across levels   (b) $\tilde{\pi}_{(x_1, x_2)}$ across infinite state variables

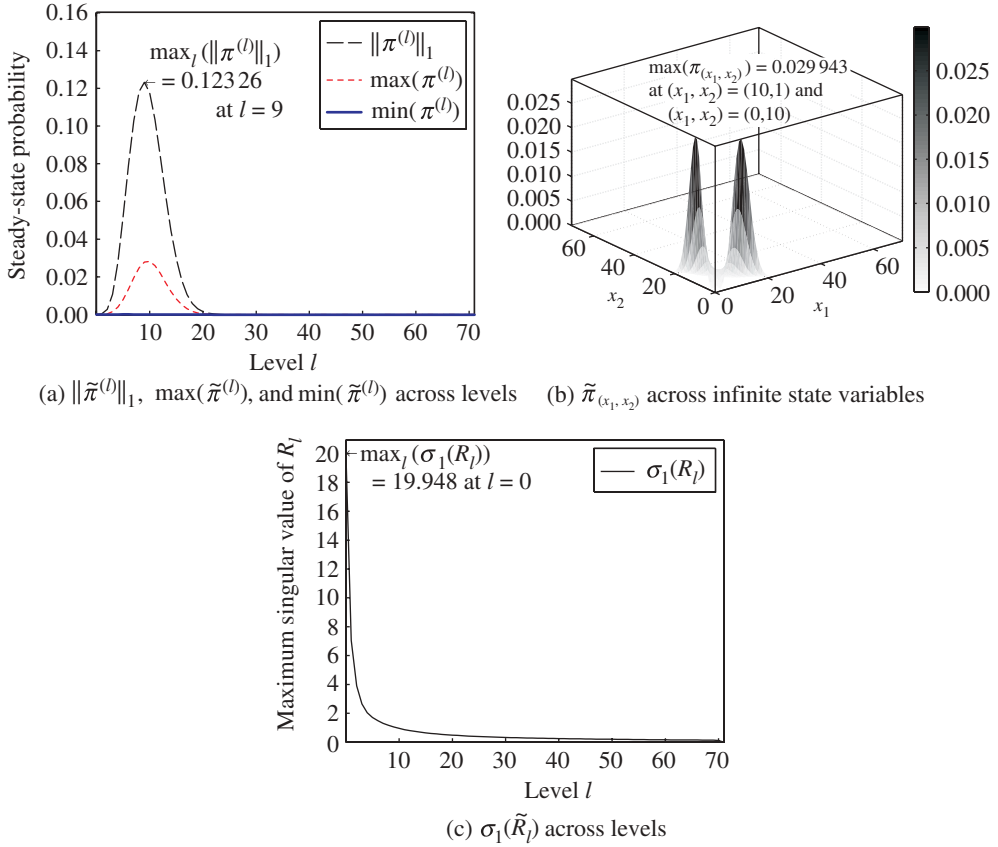(c) $\sigma_1(\tilde{R}_l)$ across levels

FIGURE 4: Plots for the exclusive switch model with parameter set 1: $(p, d, b, u) = (0.05, 0.005, 0.01, 0.008)$ and $(\texttt{low}, \texttt{high}) = (0, 71)$.

TABLE 8: Results for the exclusive switch model with parameter set 2: $(p, d, b, u) = (0.5, 0.005, 0.000\,01, 0.008)$ and $g_{3,2}(x_1, x_2, x_3, x_4, x_5) = ((x_1 - 100)^2 + (x_2 - 80)^2)((x_1 - 80)^2 + (x_2 - 100)^2)$.

ε	$(\texttt{low}, \texttt{high})$	Approach	T_{setup} (s)	T_{solve} (s)	$\|\tilde{\pi} Q\|_\infty$	Tol
1e−1	(13, 189)	(A2)	2 047	83	5e−18	6e−6
		(A2), 1e−16	122	354	5e−18	6e−6
		(A3)	0	87	1e−17	
		(A3), 1e−16	0	305	1e−17	
5e−2	(1, 206)	(A2)	2 915	145	3e−17	2e−6
		(A2), 1e−16	144	534	4e−17	2e−6
		(A3)	0	140	3e−17	
		(A3), 1e−16	0	471	4e−17	

5.4. Discussion

With the proposed approach, five problems are solved. In two problems, $\texttt{low} > 0$ is used. In problems where the stationary probability mass is located at lower level numbers (such as
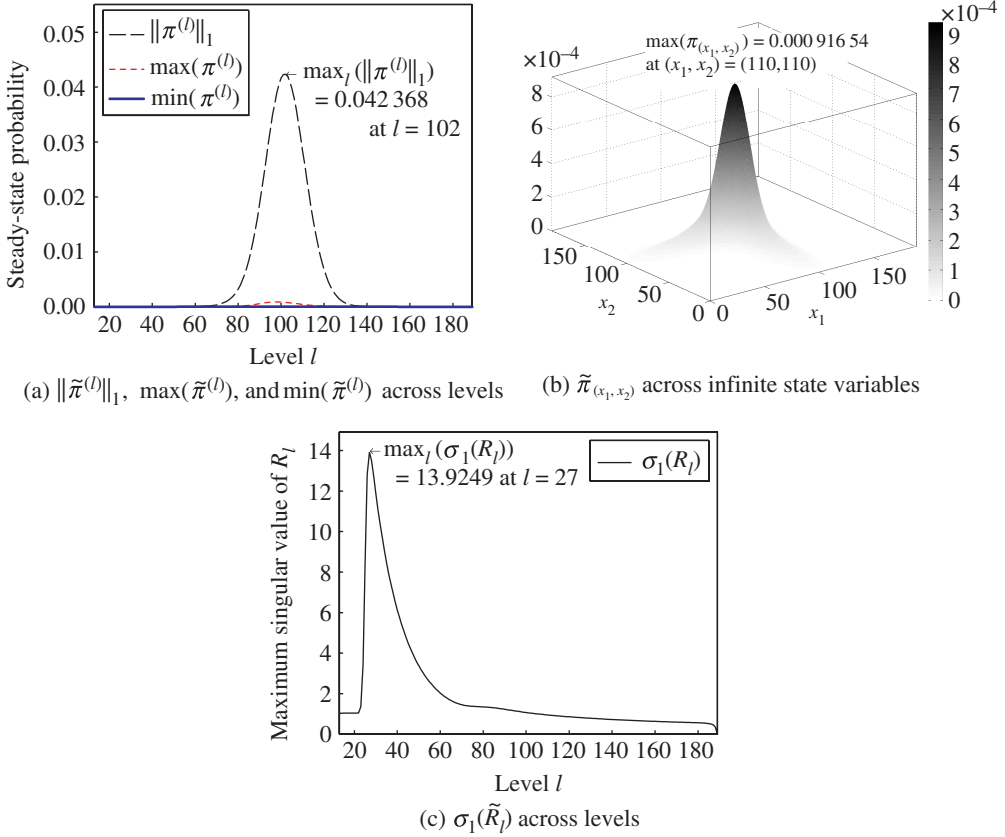
(a) $\|\tilde{\pi}^{(l)}\|_1$, $\max(\tilde{\pi}^{(l)})$, and $\min(\tilde{\pi}^{(l)})$ across levels

(b) $\tilde{\pi}_{(x_1, x_2)}$ across infinite state variables

(c) $\sigma_1(\tilde{R}_l)$ across levels

FIGURE 5: Plots for the exclusive switch model with parameter set 2: $(p, d, b, u) = (0.5, 0.005, 0.000\,01, 0.008)$ and $(\texttt{low}, \texttt{high}) = (13, 189)$.

around level 3 for the toggle switch model and around level 9 for the first parameter set of the exclusive switch model), the solution is obtained within some number of seconds with approach (A3). The solution time depends on the value of $\texttt{high}$ and the sizes of the blocks close to level $\texttt{high}$, and, thus, the number of states residing between levels 0 and $\texttt{high}$. On the other hand, when the stationary probability mass is located relatively far away from lower levels as in the gene expression model or the second parameter set of the exclusive switch model (these three problems have their stationary probability masses located around levels 96, 308, and 102, respectively), a solution takes longer. Note that in these three problems, even though a much larger value of ε is used compared to those of the other two problems for similar residual norms, much larger sets of states end up being analyzed. The solution time of the second problem is about 40 minutes with (A3) and much larger than that of the first problem, which is about 3 minutes with (A3), due to the necessity to consider the much larger blocks at level numbers between 221 and 657. We note that the value of $\max(R_{\texttt{high}}^{(4)} - R_{\texttt{high}}^{(3)})$ for the (A2) approach is not indicative of the quality of the obtained solution. On the other hand, the level number where the largest singular values of the approximate matrices of conditional expected sojourn times is maximized, generally turns out to be close to where the marginal stationary probability across levels for the same problem is maximized. Finally, it is observed clearly that (A3) provides

the best overall approach from a computational perspective to handle the approximation to the matrix of conditional expected sojourn times at level high $+ 1$, and that dropping nonzero entries less than 10^{-16} from matrices does not have an adverse effect on accuracy, but is likely to help with memory requirements.

Recently a method that computes componentwise bounds on the stationary probability distribution of Markov population models was proposed [6]. The work in this paper is complementary to that in several ways. Although both techniques use a suitable Lyapunov function to locate a finite subset of states where a specified percentage of the stationary probability mass resides, the technique proposed in this paper for the LDQBD model requires a larger finite subset of states with which to work than the technique in [6] for the same percentage. This clearly implies larger memory requirements. However, it results in a rapid solution with a sufficiently small residual norm when a relatively large percentage of the probability mass is located in relatively small level numbers. Note that the largest linear system solved has as many equations and unknowns as the number of states in level high. On the contrary, the computation of bounds with comparable accuracy with the bounding technique in [6] would require a finite subset of states with at least comparable size to those that lie between levels low and high to be explored and a system of linear equations with coefficient matrix incident on this relatively larger subset to be solved for multiple right-hand sides. Although there is no duplication of work when the coefficient matrix is relatively small to be factorized and a direct solution method employed, the nonzero entries of the same coefficient matrix will be used over and over again when the finite subset is relatively large to warrant the use of an iterative method in the solution procedure [26, pp. 121–230].

6. Conclusion

We have shown how stochastic chemical kinetic systems can be modeled as infinite LDQBD processes and how their stationary distribution can be approximated by matrix-analytic methods. The major novelties of our contribution are the use of LDQBD processes for stochastic chemical kinetics, the way of defining the levels, and the application of Lyapunov functions for determining truncation levels such that the maximum approximation error due to state space truncation can be prescribed.

The level definition is a key issue in LDQBD modeling. In many applications of QBD processes, such as queueing models for computer and communication systems, the level of a state is usually defined through only one single component of the multidimensional state space in that the value of that component determines the level number. Often the notion of a QBD process is even restricted to two-dimensional state spaces. But actually the only essential feature to render matrix-analytic methods possible is the block tridiagonality of the generator matrix. We define the levels as subsets of the state space such that the level number of a state is determined by a (simple) function of the values of all components of the state variable. This enables us to consider infinite state spaces of potentially arbitrary dimension.

In order to cope with the infinite state space with regard to numerically approximating the stationary distribution, Lyapunov functions were obtained such that the state space can be truncated accordingly. The corresponding bounds provided by these Lyapunov functions guarantee that at most a prespecified (small) probability mass lies in the truncated parts of the state space. Once the state space has been truncated, matrix-analytic methods can be applied. Numerical results demonstrated that our approach yields accurate, computationally efficient, and numerically stable approximations to the stationary distribution.

A couple of interesting issues for further research arise. In our examples, the maximum of the state variables that may take infinitely many values defines the level number. This may not work in general and other definitions may be reasonable as well. In any case, it is clear that the level can be defined much more flexibly than usually done in previous QBD applications. Our approach similarly applies to Markovian queueing networks. In many regards these are often even simpler than systems of stochastic chemical kinetics. In particular, except for cases with batch arrivals, batch services, or generalized queueing concepts like *G*-networks, state transitions in the underlying CTMC typically correspond to arrivals, departures, or moves of single customers. Therefore, the definition of the level as the maximum of state variables seems to be quite generally applicable in the queueing context.

One crucial point that deserves further investigation is the derivation of suitable Lyapunov functions. It is not difficult to find Lyapunov functions that guarantee the ergodicity of the LDQBD process, but the corresponding bounds on the stationary distribution may be loose. It is common that the derivation of Lyapunov functions is mainly done by hand or by using insights of the specific model under consideration. However, this particularly hampers the application by practitioners. Clearly, a systematic scheme for obtaining Lyapunov functions that provide tight bounds for a given model would be highly desirable. First steps towards automated computation of the required bounds are made with the Geobound tool (see http://alma.cs.uni-saarland.de/?page_id=74), but more work needs to be done.

Acknowledgements

References

[1] BAUMANN, H. AND SANDMANN, W. (2010). Numerical solution of level dependent quasi-birth-and-death processes. *Procedia Comput. Sci.* **1,** 1561–1569.

[2] BERMAN, A. AND PLEMMONS, R. J. (1994). *Nonnegative Matrices in the Mathematical Sciences*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

[3] BHARUCHA-REID, A. T. (1960). *Elements of the Theory of Markov Processes and Their Applications*. McGraw-Hill, New York.

[4] BRIGHT, L. AND TAYLOR, P. G. (1995). Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes. *Commun. Statist. Stoch. Models* **11,** 497–525.

[5] DAYAR, T. AND STEWART, W. J. (1996). On the effects of using the Grassmann–Taksar–Heyman method in iterative aggregation–disaggregation. *SIAM J. Sci. Comput.* **17,** 287–303.

[6] DAYAR, T., HERMANNS, H., SPIELER, D. AND WOLF, V. (2011). Bounding the equilibrium distribution of Markov population models. To appear in *Numer. Linear Algebra Appl.*

[7] ENGBLOM, S. (2006). Computing the moments of high dimensional solutions of the master equation. *Appl. Math. Comput.* **180,** 498–515.

[8] GARDNER, T. S., CANTOR, C. R. AND COLLINS, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. *Nature* **403,** 339–342.

[9] GAVER, D. P., JACOBS, P. A. AND LATOUCHE, G. (1984). Finite birth-and-death models in randomly changing environments. *Adv. Appl. Prob.* **16,** 715–731.

[10] GIBSON, S. AND SENETA, E. (1987). Augmented truncation of infinite stochastic matrices. *J. Appl. Prob.* **24,** 600–608.

[11] GLYNN, P. W. AND ZEEVI, A. (2008). Bounding stationary expectations of Markov processes. In *Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz* (Inst. Math. Statist. Collect. **4**), Institute of Mathematical Statistics, Beachwood, OH, pp. 195–214.

[12] GOLUB, G. H. AND VAN LOAN, C. F. (1996). *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD.

[13] GRASSMANN, W. K., TAKSAR, M. I. AND HEYMAN, D. P. (1985). Regenerative analysis and steady state distributions for Markov chains. *Operat. Res.* **33,** 1107–1116.

[14] HANSCHKE, T. (1999). A matrix continued fraction algorithm for the multiserver repeated order queue. *Math. Comput. Modelling* **30,** 159–170.

[15] KURTZ, T. G. (1972). The relationship between stochastic and deterministic models for chemical reactions. *J. Chemical Phys.* **57,** 2976–2978.

[16] LATOUCHE, G. AND RAMASWAMI, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

[17] LATOUCHE, G. AND TAYLOR, P. (2002). Truncation and augmentation of level-independent QBD processes. *Stoch. Process. Appl.* **99,** 53–80.

[18] LOINGER, A., LIPSHTAT, A., BALABAN, N. Q. AND BIHAM, O. (2007). Stochastic simulations of genetic switch systems. *Phys. Rev. E* **75,** 021904.

[19] MCQUARRIE, D. A. (1967). Stochastic approach to chemical kinetics. *J. Appl. Prob.* **4,** 413–478.

[20] NEUTS, M. F. (1981). *Matrix-Geometric Solutions in Stochastic Models*. Johns Hopkins University Press, Baltimore, MD.

[21] OPPENHEIM, I., SHULER, K. E. AND WEISS, G. H. (1969). Stochastic and deterministic formulation of chemical rate equations. *J. Chemical Phys.* **50,** 460–466.

[22] RAMASWAMI, V. AND TAYLOR, P. G. (1996). Some properties of the rate operators in level dependent quasi-birth-and-death processes with a countable number of phases. *Commun. Statist. Stoch. Models* **12,** 143–164.

[23] SENETA, E. (1981). *Nonnegative Matrices and Markov Chains*, 2nd edn. Springer, New York.

[24] SINGER, K. (1953). Application of the theory of stochastic processes to the study of irreproducible chemical reactions and nucleation processes. *J. R. Statist. Soc. B* **15,** 92–106.

[25] STAFF, P. J. (1970). A stochastic development of the reversible Michaelis-Menten mechanism. *J. Theoret. Biol.* **27,** 221–232.

[26] STEWART, W. J. (1994). *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press.

[27] THATTAI, M. AND VAN OUDENAARDEN, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Nat. Acad. Sci. USA* **98,** 8614–8619.

[28] THORNE, J. (1997). An investigation of algorithms for calculating the fundamental matrices in level dependent quasi birth death processes. Doctoral Thesis, The University of Adelaide.

[29] TWEEDIE, R. L. (1975). Sufficient conditions for regularity, recurrence and ergodicity of Markov processes. *Math. Proc. Camb. Phil. Soc.* **78,** 125–136.

[30] TWEEDIE, R. L. (1998). Truncation approximation of invariant measures for Markov chains. *J. Appl. Prob.* **35,** 517–536.

[31] VAN KAMPEN, N. G. (1992). *Stochastic Processes in Physics and Chemistry*. North-Holland.

[32] ZHAO, Y. Q. AND LIU, D. (1996). The censored Markov chain and the best augmentation. *J. Appl. Prob.* **33,** 623–629.