The Role of Orthogonal Polynomials in Determining the Decay Rates of Multidimensional Queueing Processes

Peter Taylor

Department of Mathematics and Statistics,
The University of Melbourne
(with D Kroese, W Scheinhardt and A Motyer)
Quasi-birth-and-death processes

• A QBD process is a two-dimensional continuous-time Markov Chain \( \{(Y_t, J_t), t \geq 0\} \) on the state space \( \{0, 1, \ldots\} \times \{0, 1, \ldots, m\} \).

• The variable \( Y_t \) is called the level of the process at time \( t \) and the variable \( J_t \) is called the phase of the process at time \( t \).

• The parameter \( m \) may be either finite or infinite.

• State transitions are restricted to states in the same level or in the two adjacent levels (hence the name QBD).

• Transition intensities are assumed to be level-independent away from the boundary.
Our class of interest
Quasi-birth-and-death processes

A general QBD process has a block partitioned generator $Q$ with tri-diagonal structure

$$Q = \begin{pmatrix} \tilde{Q}_1 & Q_0 \\ Q_2 & Q_1 & Q_0 \\ & Q_2 & Q_1 & Q_0 \\ & & \ddots & \ddots & \ddots \end{pmatrix}.$$ 

$Q_0$, $Q_1$, $Q_2$ and $\tilde{Q}_1$ are $(m+1) \times (m+1)$ matrices.
In our class of interest, the blocks in the generator are themselves tridiagonal and homogeneous away from the boundary. Thus, we can write

\[ Q_0 = \begin{pmatrix} \tilde{c}_1 & c_0 \\ c_2 & c_1 & c_0 \\ c_2 & c_1 & c_0 \\ \vdots & \vdots & \vdots \end{pmatrix}, \]

\[ Q_1 = \begin{pmatrix} \tilde{b}_1 & b_0 \\ b_2 & b_1 & b_0 \\ b_2 & b_1 & b_0 \\ \vdots & \vdots & \vdots \end{pmatrix}, \]
\[ Q_2 = \begin{pmatrix} \tilde{a}_1 & a_0 \\ a_2 & a_1 & a_0 \\ a_2 & a_1 & a_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \]

and

\[ \tilde{Q}_1 = \begin{pmatrix} \bar{b}_1 & b_0 \\ b_2 & \hat{b}_1 & b_0 \\ b_2 & \hat{b}_1 & b_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \]
The Two-node Jackson Network

- Let $J_t$ denote the number of customers in the first queue at time $t$ (the phase), and
- $Y_t$ denote the number of customers in the second queue at time $t$ (the level).
Our Tools: 1

The Matrix-Geometric Property.

Assume that the QBD is positive recurrent and denote the limiting probabilities

\[ \pi_{k,j} := \lim_{t \to \infty} P(Y_t = k, J_t = j). \]

With

\[ \pi_k = (\pi_{k0}, \pi_{k1}, \ldots, \pi_{km}), \]

then

\[ \pi_k = \pi_0 R^k, \quad k \geq 0 \]

for both finite and infinite \( m \).
The \((m + 1) \times (m + 1)\) matrix \(R\) is the minimal non-negative solution to the equation

\[ Q_0 + R Q_1 + R^2 Q_2 = 0. \]

The stationary distribution \(\pi_0\) at level zero must satisfy

\[ \pi_0 (\tilde{Q}_1 + R Q_2) = 0. \]
Definition

If there exists a positive scalar $z$ and a positive row vector $w \in \ell^1$ such that

$$\lim_{k \to \infty} \frac{\pi_k}{z^k} = w$$

elementwise, then we say that the stationary distribution decays at rate $z$. 
For $m < \infty$, it follows immediately from the matrix-geometric property that

$$\lim_{k \to \infty} \frac{\sum_i \pi_{ki}}{(\text{sp}(R))^k} = \kappa,$$

where $\kappa$ is a constant. That is, the level process has tail decay rate $\text{sp}(R) < 1$.

For the class of processes considered here, $m = \infty$ so this result is not applicable.
Our Tools: 2

**Theorem** (Kroese, Scheinhardt, Taylor, 2004)

Consider an irreducible QBD process. If there exists a nonnegative vector $w \in \ell^1$ and a nonnegative number $z < 1$ such that

$$w R = zw,$$

and

$$w(\tilde{Q}_1 + R Q_2) = 0,$$

then the QBD process is positive recurrent with $\pi_0 = w$, and for all $k$,

$$\frac{\pi_k}{z^k} = w.$$
Our Tools: 3

**Theorem** (Ramaswami and Taylor, 1996)

Let \( q_n = -Q_1(n, n) \). If the complex variable \( z \) and the vector \( w = \{ w_n \} \) are such that \( |z| < 1 \) and \( \sum_n |w_n|q_n < \infty \), then

\[
\begin{align*}
    w \left( Q_0 + zQ_1 + z^2Q_2 \right) &= 0
\end{align*}
\]

implies that

\[
    w R = zw.
\]
Due to our assumptions about the structure of the generator $Q$, the condition $\sum_n |w_n| q_n < \infty$ is equivalent to $w \in \ell^1$ and the equation

$$w (Q_0 + zQ_1 + z^2Q_2) = 0$$

can be written as a homogeneous second order recurrence relation in the $w_n$ with coefficients that depend on $z$. 
Thus
\[ w_0 \tilde{\gamma}_1(z) + w_1 \gamma_2(z) = 0 \]
and, for \( n \geq 0 \),
\[ w_n \gamma_0(z) + w_{n+1} \gamma_1(z) + w_{n+2} \gamma_2(z) = 0, \]
where
\[ \tilde{\gamma}_1(z) = \tilde{c}_1 + \tilde{b}_1 z + \tilde{a}_1 z^2 \]
and, for \( i = 0, 1, 2, \)
\[ \gamma_i(z) = c_i + b_i z + a_i z^2. \]
Our Tools: 4

It is elementary to find conditions on $z$ such that $w \in \ell^1$. It is harder to find conditions on $z$ such that $w$ is nonnegative. We use the theory of orthogonal polynomials for this purpose. Introducing a new variable $x$, we can generalise our equations for $w_n$ to

\[
\begin{align*}
P_0(x; z) &= 1, \\
\gamma_2(z)P_1(x; z) &= x - \tilde{\gamma}_1(z), \\
\gamma_2(z)P_n(x; z) &= (x - \gamma_1(z))P_{n-1}(x; z) - \gamma_0(z)P_{n-2}(x; z), \quad n \geq 2.
\end{align*}
\]

When $x = 0$, the equations are the same as those for $w_n$, scaled so that $w_0 = 1$, and so $w_n = P_n(0; z)$. 
The $P_n(0; z)$ are positive for all $n$ if and only if the zeros of all the $P_n(x; z)$, considered as polynomials in $x$, are less than zero.

Let

$$T_n(x) = \left( \sqrt{\frac{\gamma_2(z)}{\gamma_0(z)}} \right)^n P_n \left( 2x \sqrt{\gamma_0(z) \gamma_2(z)} + \gamma_1(z); z \right).$$

The $T_n(x)$s satisfy the recursion which defines *perturbed Chebyshev polynomials*. The behaviour of their zeros has been well-studied, and we can translate this into information about the behaviour of the zeros of the $P_n(x)$s.
Let

\[ \tau(z) = \gamma_1(z) + 2\sqrt{\gamma_0(z)\gamma_2(z)}, \]
\[ \chi(z) = \tilde{\gamma}_1(z) + \frac{\gamma_0(z)\gamma_2(z)}{\tilde{\gamma}_1(z) - \gamma_1(z)}, \]
\[ \chi_1(z) = \begin{cases} \tau(z) & \text{if } (\tilde{\gamma}_1(z) - \gamma_1(z))^2 \leq \gamma_0(z)\gamma_2(z), \\ \chi(z) & \text{otherwise}. \end{cases} \]

For \( z > 0 \), \( P_n(x; z) \) is positive for all \( n \) if and only if \( \chi_1(z) \leq x \). Thus, the vector \( \mathbf{w} = (P_n(0; z)) \) is positive if and only if \( \chi_1(z) \leq 0 \).

The condition \( \chi_1(z) \leq 0 \) turns out to be expressible in terms of inequalities involving polynomials in \( z \) of degree, at most, four.
Summary

• We use elementary techniques and the theory of orthogonal polynomials to derive the set of values of $z \in [0, 1]$ for which

$$w(Q_0 + zQ_1 + z^2Q_2) = 0$$

has a solution $w \in \ell^1$.

• This gives us the set of $w \in \ell^1$ and $z \in [0, 1]$ for which $wR = zw$.

• Provided that $w(\tilde{Q}_1 + zQ_2) = 0$, then $\pi_0 = w$, and for all $k$,

$$\frac{\pi_k}{z^k} = w.$$ 

Thus the decay rate is $z$. 
Summary

We have a lot of flexibility in modifying \( \tilde{Q}_1 \) and so, for a large range of \( w \) and \( z \), it is reasonable to think that we can do so in order that

\[
w(\tilde{Q}_1 + zQ_2) = 0
\]

is satisfied. Indeed, for a number of examples we have shown how to modify \( \tilde{Q}_1 \) so that this equation is satisfied for all \( w \) and \( z \) compatible with

\[
w R = zw.
\]

Thus we have shown that, by altering the transition rates at level zero, we can obtain any decay rate in the calculated range.
The two-node Jackson network

The stationary distribution has the product form

$$\pi(n_1, n_2) = (1 - \rho_1)(1 - \rho_2)\rho_1^{n_1} \rho_2^{n_2}, \quad n_1, n_2 \geq 0,$$

where $\rho_1$ and $\rho_2$ are solutions to the well-known traffic equations. The stability condition is $\rho_1 < 1, \rho_2 < 1$.

The decay rate at queue 2 is always $\rho_2$. However, the ranges of possible decay rates, obtainable by varying the transition rates when queue 2 is empty can be non-trivial for different parameter values.
The two-node Jackson network

<table>
<thead>
<tr>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$p$</th>
<th>$q$</th>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>Possible decay rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1.5</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.667</td>
<td>0.500</td>
<td>[0.477, 0.750)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1.5</td>
<td>1</td>
<td>0</td>
<td>0.500</td>
<td>0.667</td>
<td>[0.667, 1)</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1.5</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0.667</td>
<td>0.500</td>
<td>{0.500}</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1.5</td>
<td>0</td>
<td>1</td>
<td>0.500</td>
<td>0.667</td>
<td>{0.667}</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.1</td>
<td>0.8</td>
<td>0.978</td>
<td>0.598</td>
<td>[0.597, 0.600)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.8</td>
<td>0.1</td>
<td>0.598</td>
<td>0.978</td>
<td>[0.978, 1)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.833</td>
<td>0.833</td>
<td>[0.833, 0.900)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>0.5</td>
<td>0.5</td>
<td>0.200</td>
<td>0.200</td>
<td>[0.200, 0.600)</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>0.4</td>
<td>0.9</td>
<td>0.859</td>
<td>0.563</td>
<td>[0.558, 0.600)</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>15</td>
<td>10</td>
<td>0.9</td>
<td>0.4</td>
<td>0.563</td>
<td>0.859</td>
<td>[0.859, 1)</td>
</tr>
</tbody>
</table>
Finite Truncations

We can extend our orthogonal polynomial analysis to show that the limiting value of the decay rate at queue 2 when queue 1 is truncated to size $m$ and then $m$ is allowed to approach infinity is always the infimum of the interval of possible decay rates.

We see from our table that this is not always $\rho_2$. In these cases, there is a ‘discontinuity at infinity’ with respect to the parameter $m$. 