

Asymptotic Approximations and Bottleneck Analysis in Product Form Queueing Networks with Large Populations

Charles Knessl and Charles Tier

Department of Mathematics, Statistics, and Computer Science
University of Illinois at Chicago
851 South Morgan St.
Chicago, IL 60607-7045

ABSTRACT

Asymptotic approximations are constructed for the performance measures of product form queueing networks consisting of single server, fixed rate nodes with large populations. The approximations are constructed by applying singular perturbation methods to the recursion equations of Mean Value Analysis. Networks with a single job class are studied first to illustrate the use of perturbation techniques. The leading term in the approximation is related to bottleneck analysis, but fails to be accurate if there is more than one bottleneck node. A uniform approximation is constructed which is valid for networks with many bottleneck nodes. The accuracy of the uniform approximation is demonstrated for both small and large population sizes. Next, multiclass networks are considered. The leading term in the asymptotic approximation is again related to bottleneck analysis but fails to be valid across “switching surfaces”. Across these the bottleneck nodes of the network change as a function of the fraction of jobs in the different job classes. A boundary layer correction is constructed near the switching surfaces which provides an asymptotic connection across the switching surfaces. Numerical examples are presented to demonstrate the accuracy of the results. We illustrate the asymptotic approach on some simple networks and indicate how to treat more complicated problems.

1 Introduction

Closed, product form queueing networks are widely used as models for the performance of computer and communications networks. For this class of networks, an exact expression for the stationary queue length distribution is known. Though the stationary distribution for product form networks has a simple analytic formula, calculation of performance measures, such as mean queue lengths, mean response times and throughputs, is computationally difficult for networks with large populations. To facilitate the calculation of these performance measures, a number of computational algorithms have been developed which greatly reduce the computational cost.

The computational algorithms can be divided into two general classes: (i) computing the normalization constant for the product form solution or (ii) computing the performance measures directly. Algorithms that compute the normalization constant include the Convolution algorithm by Buzen [?] and the RECAL algorithm developed by Conway and Georganas [?]. Algorithms that

This research was supported in part by DOE Grant DE-FG02-93ER25168

compute performance measures directly include Mean Value Analysis (MVA) discovered by Reiser and Lavenberg [?] and the distribution analysis by chain algorithm (DAC) by Souza e Silva and Lavenberg [?]. Monte Carlo methods were used to compute the performance measures by Ross, Tsang and Wang [?]. Even though these computational algorithms are improvements over direct calculation methods, each has disadvantages and often requires a large number of calculations for networks with large populations.

Another approach is to develop approximations for the normalization constant or the performance measures. Bounds on the performance measures are the simplest approximations to compute (see Zahorjan, Sevcik, Eager, and Galler [?]). Asymptotic bound analysis (ABA) provides an upper (lower) bound on the throughput (mean response time) based on the assumption that either no queues are formed or the bottleneck node has utilization equal to 1. Balanced job bounds (BJB) provide both upper and lower bounds on the throughput and mean response time. These bounds are obtained by replacing the original network by a balanced network with all nodes identical.

Asymptotic expansions of the normalization constant were obtained by McKenna and Mitra [?], [?]. They obtain approximations by first developing an integral representation to the normalization constant and then expanding the integral as an asymptotic series in inverse powers of a large parameter, the total number of jobs circulating in the system. Knessl and Tier [?, ?] and Mei and Tier [?] computed explicitly the leading term in the asymptotic expansion of the normalization constant when both the number of jobs and the number of nodes in the network are large. Their work is based on applying the ray method (see Keller [?]) to a scaled form of the Convolution algorithm. Kogan [?, ?] constructed asymptotic approximations to the normalization constant by using the inversion integral for the generating function of the normalization constant.

Asymptotic approximations have some advantages over computational algorithms. They yield results for very large networks without requiring much computing time, and they also lead to formulas that give qualitative insights into the behavior of the performance measures. Asymptotic formulas often clearly show how the system behaves in terms of the variables and/or parameters in the model.

Asymptotic approximations to the performance measures of closed, product form queueing networks with large populations can also be constructed using the MVA Algorithm. As described above, MVA directly computes mean queue lengths, mean response times and throughputs using a recursion in the number of jobs in the network. It relies on the key result that the mean queue lengths seen by arrivals are equal to the mean queue lengths in the same closed queueing network from which the arriving customer has been removed (arrival theorem). Unfortunately, direct application of the finite MVA recursion has a high computational cost if there are many jobs and job classes. An approximate version of MVA was developed by Schweitzer in [?] in which an iter-

ation is developed by guessing the performance measures for the full network and then using an approximation to compute the measures for a network with one less job. Finally, MVA is used to recompute the measures for the full network. Based on this iteration, approximations to the performance measures are obtained. Thus, an approximate MVA algorithm is used and the performance measures are not computed for all size networks. A refinement of this approximation, called the Linearizer, was given in Chandy and Neuse [?].

A systematic approach to constructing approximations for networks with a unique bottleneck was given by Schweitzer [?] and by Schweitzer, Serazzi and Broglia [?]. Here a perturbation scheme was presented based on MVA for networks with large populations which also agrees with the approximate MVA algorithm presented by Schweitzer [?]. Further work using the perturbation scheme was presented Balbo and Serazzi [?, ?]. The leading term in these approximation provides an asymptotic bottleneck analysis of the network and the higher order terms provide corrections to the standard bottleneck analysis. However, it was observed that the perturbation solution worked poorly when there were two or more bottleneck nodes. In multiclass networks, the bottleneck configuration can change as a function of the fraction of jobs in each class. The term *switching surfaces* of bottlenecks was used in Schweitzer, Serazzi and Broglia [?] to describe critical parameter values where the bottlenecks change. On and near these switching surfaces, the perturbation expansion develops a non-uniformity and is no longer valid.

Our goal is to re-examine the perturbation scheme of Schweitzer [?] and Schweitzer, Serazzi, and Broglia [?] and present a new analysis to correct the non-uniformities that develop at the switching surfaces or when there are multiple bottleneck nodes. To do so, we use a singular perturbation approach which relies on the construction of a boundary layer correction in the vicinity of these switching surfaces. In contrast to computational algorithms, our results yield explicit formulas for the performance measures.

The paper is organized as follows. In Section 2, we illustrate our approach for single class networks. We first present an asymptotic solution as in Schweitzer and Schweitzer, Serazzi, and Broglia [?, ?] and then construct a uniform approximation that is valid if there are multiple bottleneck nodes and for all values of the parameters. In addition, we show that our uniform approximation agrees with the asymptotic expansion of an exact solution given in Schweitzer [?]. Numerical results are presented to illustrate the accuracy of the approximation. Multiclass networks are considered in Section 3. Here we analyze a network with two nodes and two job classes, to illustrate the methods. In Section 4 we discuss the results and indicate how to treat more complicated networks.

2 Single Class Networks

We consider a product form queueing network with a single job class and K single server, fixed rate nodes. There are M jobs circulating in the network. The service time at node i is exponentially distributed with parameter μ_i . The visit ratio at node i is v_i where the visit ratio is equal to one at some distinguished node. The total demand at node i for an average job in the network is defined to be $D_i = v_i/\mu_i$.

The MVA algorithm is based on a recursion formed by adding one job to the network per iteration until the total number of jobs equals M . The following performance measures are computed as a function of the population size m :

$$\begin{aligned} T_i(m) &= \text{mean response time at node } i \\ U_i(m) &= \text{utilization of node } i \\ N_i(m) &= \text{mean queue length at node } i \\ X_i(m) &= \text{throughput at node } i \\ X(m) &= \text{system throughput measured at distinguished node} \end{aligned}$$

for $m = 1, \dots, M$. The full set of equations that comprise MVA are then given by:

I. Arrival Theorem

$$T_i(m) = \frac{1}{\mu_i}(1 + N_i(m-1))$$

II. Little's Law

$$X(m) = \frac{m}{\sum_{k=1}^K v_k T_k(m)}$$

III. Forced Flow Law

$$X_i(m) = v_i X(m)$$

IV. Update N_i and repeat I

$$N_i(m) = X_i(m) T_i(m).$$

The initial condition is

$$N_i(0) = 0$$

and the mean queue lengths must satisfy

$$\sum_{i=1}^K N_i(m) = m. \tag{2.1}$$

A more compact form of the iteration can be formed by not directly computing some of the performance measures, i.e. combine the above equations, to yield:

$$N_i(m) = X(m) D_i (1 + N_i(m-1)), \quad 1 \leq i \leq K \tag{2.2}$$

$$X(m) = \frac{m}{\sum_{k=1}^K D_k (1 + N_k(m-1))}, \quad 1 \leq m \leq M. \tag{2.3}$$

Of course, these equations could be simplified further by eliminating the throughput $X(m)$.

2.1 Perturbation Method of Schweitzer

We consider a network as above with M large. We expect that as the number of jobs in the network increases, and becomes very large, that one node or a small set of nodes becomes saturated first, i.e. utilization ≈ 1 . This set of nodes is called the *bottleneck set* (see Schweitzer, Serazzi, and Broglia[?]) and is identified as the nodes with the maximum load D_i . We assume that the bottleneck set consists only of node 1 and we consider the perturbation approach in Schweitzer, Serazzi, and Broglia [?]. We introduce the following scalings

$$\varepsilon = 1/M, \quad z = m/M = \varepsilon m, \quad h_i(z) = N_i(m)/M$$

into the MVA equations (??)-(??) to obtain the scaled recursion

$$h_i(z) - X(z)D_i h_i(z - \varepsilon) = \varepsilon X(z)D_i \quad (2.4)$$

$$X(z) = \frac{z}{\sum_i [D_i h_i(z - \varepsilon) + \varepsilon D_i]}. \quad (2.5)$$

The initial condition becomes

$$h_i(0) = 0 \quad (2.6)$$

and (??) is now

$$\sum_{i=1}^K h_i(z) = z. \quad (2.7)$$

The stability condition is that the utilization in each node be ≤ 1 so that

$$XD_i \leq 1, \quad \text{all } i. \quad (2.8)$$

We seek a regular perturbation expansion as in [?] in the form

$$\begin{aligned} h_i(z) &\sim h_i^0(z) + \varepsilon h_i^1(z) + \dots \\ X(z) &\sim X^0 + \varepsilon X^1 + \dots \end{aligned} \quad (2.9)$$

We substitute (??) into (??)-(??) to obtain to leading order

$$h_i^0(z)[1 - D_i X^0] = 0, \quad i = 1, \dots, K \quad (2.10)$$

and at the next order

$$h_i^1(z)[1 - D_i X^0] = D_i(X^0 + X^1 h_i^0 - X^0 \frac{dh_i^0}{dz}) \quad (2.11)$$

For each i , we can choose one of the following

$$h_i^0(z) = 0 \text{ or } X^0 = 1/D_i.$$

We can use bottleneck analysis to help determine the proper choice. Since node 1 is the unique bottleneck with

$$D_1 = \max_i D_i,$$

and taking into account (??), we choose $1 - D_1 X^0 = 0$. Hence, the leading order system throughput is

$$X^0 = 1/D_1. \quad (2.12)$$

For $i \neq 1$, we have $X D_i < 1$ and we choose

$$h_i^0(z) = 0$$

and using (??), we then have

$$h_1^0(z) = z.$$

Using the above in (??), we find that the $\mathcal{O}(\varepsilon)$ equation is

$$h_i^1 \left(1 - \frac{D_i}{D_1} \right) = X^1 D_i h_i^0 + \frac{D_i}{D_1} (1 - \delta_{i1})$$

where δ_{ij} is the Kronecker delta. We now must choose $X^1 \equiv 0$ and, for $i \neq 1$, we obtain

$$h_i^1 \left(1 - \frac{D_i}{D_1} \right) = D_i/D_1 = U_i$$

with the solution

$$h_i^1 = \frac{U_i}{1 - U_i}, \quad i \neq 1.$$

The term $h_1^1(z)$ is then obtained using (??). At this order the series truncates and no further corrections can be computed. We summarize the results.

Result 1 (Schweitzer) *For $M \gg 1$ and node 1 the unique bottleneck*

$$h_1(z) \sim z - \varepsilon \sum_{j=2}^K \frac{U_j}{1 - U_j}; \quad h_i(z) \sim \varepsilon \frac{U_i}{1 - U_i}, \quad i \neq 1$$

$$X(z) \sim 1/D_1.$$

In terms of the original variables:

$$N_1(m) \sim m - \sum_{j=2}^K \frac{U_j}{1 - U_j}; \quad N_i(m) \sim \frac{U_i}{1 - U_i}, \quad i \neq 1$$

$$X(m) \sim 1/D_1$$

where

$$U_i \equiv \frac{D_i}{D_1} = \text{utilization.}$$

This approximation is an improvement over asymptotic bound analysis (Zahorjan, Sevcik , Eager, and Galler [?]) since we obtain a correction in the expansion for the mean number in each node. We do not obtain a correction for the throughput.

We observe that if any $D_i \approx D_1, i \neq 1$, the above expansions breakdown due to the $1 - U_i$ terms in the denominators of the $\mathcal{O}(\varepsilon)$ terms. This will occur if the bottleneck set contains more than one node or if the loads of other nodes are “close” to the demand of the bottleneck node. This is important from a numerical point of view. An exact solution to the recursion (??)-(??) was obtained in Schweitzer [?]. The asymptotic solution described in Result 1 can be obtained by introducing our scaling to the result in Schweitzer [?] and expanding for $M \gg 1$ for the case of a unique bottleneck. To illustrate the breakdown in the expansions as the demand of one node

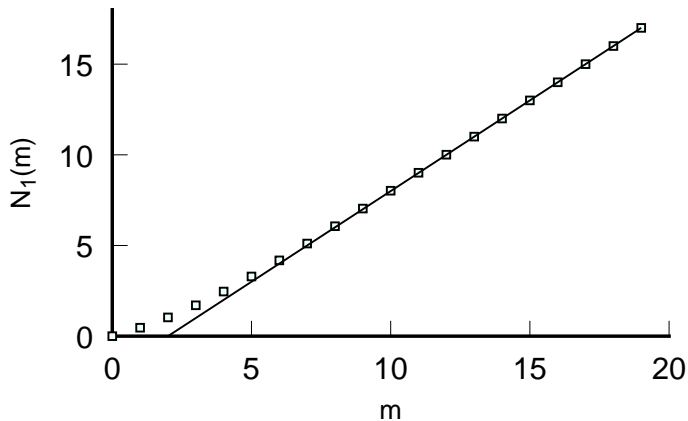


Figure 1: Graphs of $N_1(m)$ as a function of m . Exact MVA result = \square and approximation based on Result 1 = $—$ for a network with one class and 4 nodes with loads = 2.0, 1.0, 0.8, and 0.5.

approaches the demand of the bottleneck node, we present graphs of the mean number in the bottleneck node $N_1(m)$ as a function of m in Figures 1-3. The total population in a network with 4 nodes is $M = 20$ so that $\varepsilon = 0.05$. Node 1 is the bottleneck node for this network. In Figure 1, the nodes have loads given by [2.0, 1.0, 0.8, 0.5], respectively, and we see that the approximation is reasonably good for $m > 7$. However, if we increase the load in node 2 to 1.5 as in Figure 2, the approximation is only useful for $m > 15$. Finally, in Figure 3, we increase the demand in node 2 to 1.9 and find the approximation described in Result 1 in not useful. It is important to note that the demand in node 2 does not have to equal the demand in the bottleneck node 1 for the expansion to break down. For all three examples, the throughput $X = 0.5$ for all m , based on Result 1.

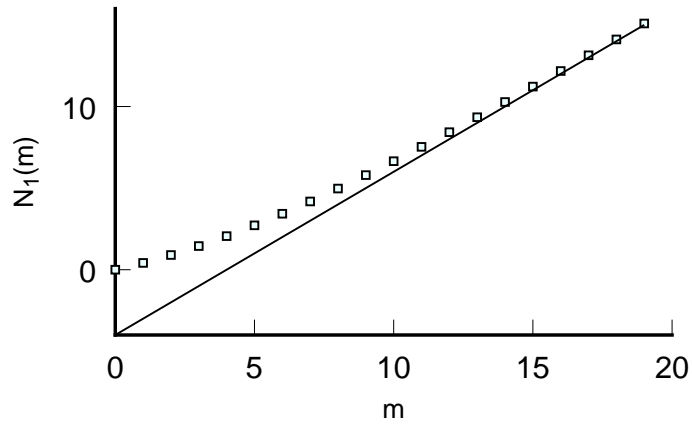


Figure 2: Graphs of $N_1(m)$ as a function of m . Exact MVA result = □ and approximation based on Result 1 = — for a network with one class and 4 nodes with loads = 2.0, 1.5, 0.8, and 0.5.

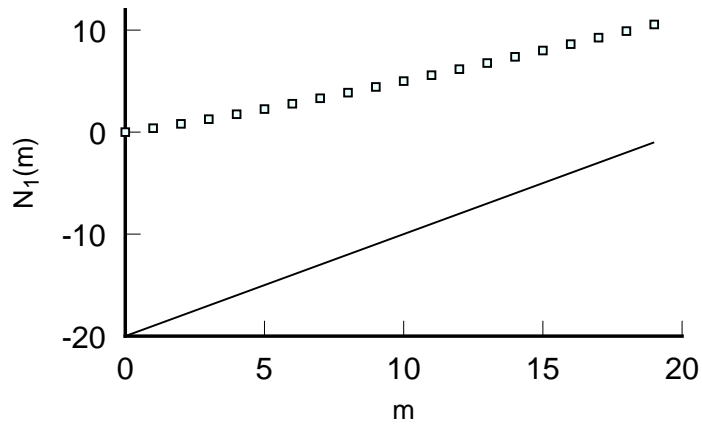


Figure 3: Graphs of $N_1(m)$ as a function of m . Exact MVA result = □ and approximation based on Result 1 = — for a network with one class and 4 nodes with loads = 2.0, 1.9, 0.8, and 0.5.

In order to obtain an improved approximation, we could develop approximations for the cases when there are two, three, or more nodes in the bottleneck set. However, each of these approximations would still break down if the demand of a non-bottleneck node was close to the demand of the nodes in the bottleneck set. A better approach is to develop a uniform expansion which is valid for all values of the demands and for any number of nodes in the bottleneck set.

2.2 The Uniform Approximation

We now develop a uniform approximation valid for any network with K nodes regardless of the values of the demands D_i , including networks with multiple bottlenecks. We introduce the following scaling:

$$\begin{aligned} D_i &\equiv D + \varepsilon a_i, \quad i = 1, 2, \dots, K \\ D &= \frac{1}{K} \sum_{k=1}^K D_k, \quad \sum_{i=1}^K a_i = 0. \end{aligned}$$

We use the single equation form of recursion (combining (??) and (??)) with the above scaling to obtain

$$\left\{ \sum_j [(D + \varepsilon a_j) h_j(z - \varepsilon) + \varepsilon(D + \varepsilon a_j)] \right\} h_i(z) - (D + \varepsilon a_i) z h_i(z - \varepsilon) = \varepsilon z D_i \quad (2.13)$$

with the conditions

$$h_i(0) = 0, \quad h_1 + \dots + h_K = z.$$

We assume an asymptotic solution of the form

$$h_i(z) \sim H_i(z) + \varepsilon G_i(z) + \dots \quad (2.14)$$

and substitute (??) into the recursion (??) to obtain the following equations at the first two orders in ε

$$zDH'_i(z) - za_iH_i(z) + [D(K-1) + \sum_{k=1}^K a_k H_k(z)]H_i(z) = zD, \quad i = 1, \dots, K \quad (2.15)$$

$$\begin{aligned} zDG'_i(z) + [(K-1)D - a_i z + \sum_{k=1}^K a_k H_k(z)]G_i(z) + \left(\sum_{k=1}^K a_k G_k(z) \right) H_i(z) = \\ za_i[1 - H'_i(z)] + \frac{z}{2}DH''_i(z) + H_i(z) \sum_{k=1}^K a_k H'_k(z). \end{aligned} \quad (2.16)$$

To solve (??), we first reduce the number of unknown functions to $K-1$ by using the fact that

$$H_K = z - \sum_{k=1}^{K-1} H_k, \quad a_K = -(a_1 + a_2 + \dots + a_{K-1})$$

to obtain the nonlinear system of ordinary differential equations

$$\begin{aligned}
zDH'_i(z) - za_iH_i(z) + [D(K-1) + \sum_{k=1}^{K-1} a_kH_k(z) \\
+ a_K(z - \sum_{k=1}^{K-1} H_k(z))]H_i(z) = zD, \quad i = 1, \dots, K-1
\end{aligned} \tag{2.17}$$

with the initial conditions

$$H_i(0) = 0.$$

To solve (??), we set

$$H_i = \frac{D}{a_K - a_i} + f_i(z), \quad f_i(0) = \frac{D}{a_i - a_K}$$

to obtain

$$\begin{aligned}
f'_i + \left[\frac{a_K - a_i}{D} + \frac{U(z)}{Dz} \right] f_i = \frac{U(z)}{z(a_i - a_K)}, \quad i = 1, \dots, K-1 \\
U(z) = \sum_{k=1}^{K-1} (a_k - a_K) f_k, \quad U(0) = D(K-1).
\end{aligned} \tag{2.18}$$

The construction of the solution to the system (??) is given in the Appendix. We use (??) for f_i and hence obtain $H_i(z)$, which gives, for $i = 1, \dots, K$

$$H_i(z) = \frac{R(z)}{S(z)} \tag{2.19}$$

where

$$\begin{aligned}
S(z) = \sum_{l=1}^K \frac{e^{za_l/D}}{\mathcal{P}_l}, \quad \mathcal{P}_l = \prod_{\substack{j=1 \\ j \neq l}}^K (a_l - a_j), \\
R(z) = \frac{ze^{za_i/D}}{\mathcal{P}_i} + D \sum_{\substack{l=1 \\ l \neq i}}^K \frac{1}{\mathcal{P}_l} \left[\frac{e^{za_l/D} - e^{za_i/D}}{a_l - a_i} \right]
\end{aligned} \tag{2.20}$$

Next we solve the linear system (??) to obtain the correction term $G_i(z)$. We first simplify (??) by introducing $S(z)$ (cf. (??)) to obtain the linear ODE system

$$\begin{aligned}
z \left[D \frac{d}{dz} - a_i + D \frac{S'}{S} \right] G_i + H_i \sum_{l=1}^K a_l G_l \\
= za_i - \frac{1}{2} z DH''_i - z D \frac{S'}{S} H'_i + DH_i \frac{S'}{S},
\end{aligned} \tag{2.21}$$

with the conditions

$$G_i(0) = 0, \quad \sum_{i=1}^K G_i(z) = 0.$$

Here we have also written equation (??), for the leading term $H_i(z)$, as

$$DH_i' - a_i H_i + \frac{S'}{S} DH_i = D.$$

We construct the solution to (??) by setting

$$G_i(z) = -1 + A_0(z)H_i + A_1(z)H_i' + A_2(z)H_i'' \quad (2.22)$$

and substituting into (??). We find, after a long calculation, that

$$A_0(z) = \frac{S'}{S}, \quad A_1(z) = K - z \frac{S'}{S}, \quad A_2(z) = -\frac{z}{2}. \quad (2.23)$$

Result 2 *A uniform approximation for a closed, single class queueing network consisting of K fixed rate nodes and $M \gg 1$ is given by*

$$h_i(z) \sim H_i(z) + \varepsilon \left(-1 + A_0(z)H_i + A_1(z)H_i' + A_2(z)H_i'' \right) + \dots \quad (2.24)$$

where $H_i(z)$ are defined in (??)-(??) and $A_j(z)$ are defined in (??). An approximation to the throughput $X(z)$ can then be obtained by substituting (??) into (??).

The formula (??) is also valid for networks in which some or all of the nodes have the same demands, i.e. $a_i = a_j$ for some i and j . For such networks, a new formula for $h_i(z)$ can be obtained in the limit as $a_i \rightarrow a_j$, etc. For example, if all the nodes have the same demand (i.e. a balanced network) then we set

$$a_j = a, \quad j = 1, \dots, K.$$

We find that

$$S(z) \rightarrow \frac{1}{(K-1)!} e^{za/D} \frac{z^{K-1}}{D^{K-1}}, \quad R(z) \rightarrow \frac{1}{K!} e^{za/D} \frac{z^K}{D^{K-1}}$$

so that

$$H_i(z) \rightarrow \frac{z}{K}, \quad G_i(z) \rightarrow 0.$$

and our result (??) reduces to the exact solution for a balanced network, namely,

$$h_i(z) = \frac{z}{K}$$

We illustrate in Figure 4 the uniform approximation for the parameters in Figure 3, in which the approximation in Result 1 was the worst. The uniform result is accurate for all values of $m \leq M$ (see Table 1). In Figure 5, we graph the throughput using the uniform expansion in Result 2 along with the bounds on the throughput using ABA and BJB (see Zahorjan, Sevcik, Eager, and Galler [?]). The uniform result is clearly an improvement over the non-uniform expansion, as well as over ABA and BJB methods, and is accurate for all $m \leq M$.

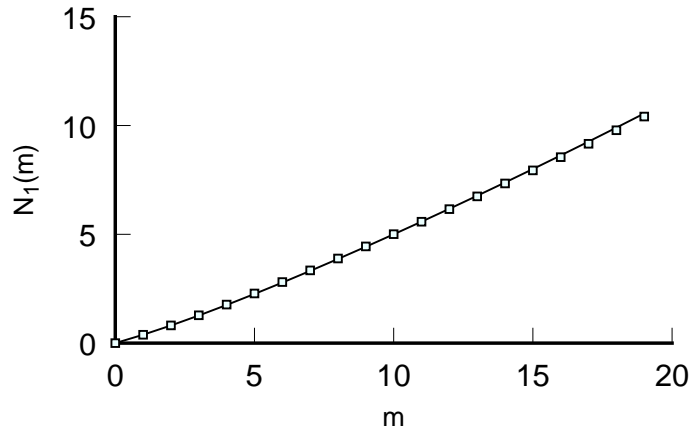


Figure 4: Graphs of $N_1(m)$ as a function of m . Exact MVA result = \square and approximation based on Result 2 = $—$ for the network with one class and 4 nodes in Figure 3.

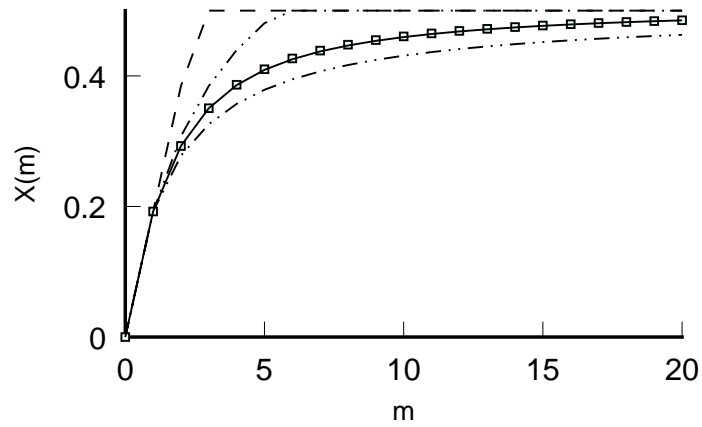


Figure 5: Graphs of $X(m)$ as a function of m using the exact MVA result = \square , asymptotic approximation based on Result 2 = $—$, ABA = $- - -$ and BJB = $- \cdot -$ for the network with one class and 4 nodes in Figure 3.

Table 1: $N_1(m)$ for the example in Figure 4

| m | Exact | Asymptotic |
|-----|---------|------------|
| 0 | 0 | 0 |
| 1 | .3846 | .3839 |
| 2 | .8103 | .8139 |
| 3 | 1.2690 | 1.2793 |
| 4 | 1.7537 | 1.7708 |
| 5 | 2.2592 | 2.2813 |
| 6 | 2.7818 | 2.8062 |
| 7 | 3.3187 | 3.3426 |
| 8 | 3.8682 | 3.8889 |
| 9 | 4.4291 | 4.4440 |
| 10 | 5.0005 | 5.0074 |
| 11 | 5.5819 | 5.5786 |
| 12 | 6.1728 | 6.1574 |
| 13 | 6.7728 | 6.7437 |
| 14 | 7.3818 | 7.3372 |
| 15 | 7.9995 | 7.9379 |
| 16 | 8.6257 | 8.5458 |
| 17 | 9.2603 | 9.1608 |
| 18 | 9.9032 | 9.7829 |
| 19 | 10.5542 | 10.4120 |

3 Multiple Class Networks

We now consider a multiclass, closed product form queueing network consisting of K single server, fixed-rate nodes. There are R job classes (chains) with M_r jobs in class r . The total number of jobs in the network is

$$M = \sum_r M_r$$

and the loads are defined to be

$$D_{ri} = \text{load of class } r \text{ jobs at node } i.$$

The performance measures in the compact form of MVA are functions of the population vector $\vec{m} = (m_1, m_2, \dots, m_R)$ which we define as

$$\begin{aligned} N_i(\vec{m}) &= \text{mean queue length at node } i \\ X_r(\vec{m}) &= \text{system throughput - class } r. \end{aligned}$$

The MVA algorithm is now given by

$$\begin{aligned}
X_r(\vec{m}) &= \frac{m_r}{\sum_{k=1}^K D_{rk} [1 + N_k(\vec{m} - \vec{e}_r)]}, \quad r = 1, \dots, R \\
N_k(\vec{m}) &= \sum_{r=1}^R D_{rk} X_r(\vec{m}) [1 + N_k(\vec{m} - \vec{e}_r)], \quad k = 1, \dots, K,
\end{aligned} \tag{3.25}$$

with the initial conditions

$$N_k(\vec{0}) = 0.$$

Here \vec{e}_r is a vector of dimension R with 1 in the r th component and zeros elsewhere.

We introduce the following scaling into (??)

$$z_r = \varepsilon m_r, \quad h_k(\vec{z}) = \varepsilon N_k(\vec{m}); \quad \varepsilon = 1/M \tag{3.26}$$

to obtain the scaled equations

$$\begin{aligned}
X_r(\vec{z}) &= \frac{z_r}{\sum_{k=1}^K D_{rk} [\varepsilon + h_k(\vec{z} - \varepsilon \vec{e}_r)]}, \quad r = 1, \dots, R \\
h_k(\vec{z}) &= \sum_{r=1}^R D_{rk} X_r(\vec{z}) [\varepsilon + h_k(\vec{z} - \varepsilon \vec{e}_r)], \quad k = 1, \dots, K.
\end{aligned} \tag{3.27}$$

The initial condition is again

$$h_k(\vec{0}) = 0, \quad k = 1, \dots, K. \tag{3.28}$$

We fix $z_1 + z_2 + \dots + z_R = 1$ and seek an asymptotic solution of the form

$$h_k(\vec{z}) \sim h_k^0(\vec{z}) + \varepsilon h_k^1(\vec{z}) + \dots \tag{3.29}$$

$$X_r(\vec{z}) \sim X_r^0 + \varepsilon X_r^1 + \dots$$

The equation for the leading term is given by

$$(1 - U_k^0) h_k^0 = 0, \quad k = 1, \dots, K \tag{3.30}$$

where the utilization U_k^0 is defined by

$$U_k^0 = \sum_{r=1}^R X_r^0 D_{rk}. \tag{3.31}$$

Clearly, there are numerous possible solutions to (??). As has been shown in [?]-[?], the choice of solution as well as the bottleneck set depends on the value of \vec{z} . The points where the bottleneck

set changes are called switching surfaces in Schweitzer [?]. The analysis of (??)-(??) away from the switching surfaces was given in [?]-[?]. However, their approximations, defined by (??), fail to be valid at or near the switching surfaces. Our goal is to provide a new approximation which is valid near the switching surfaces and is consistent with the expansions derived in [?]-[?] valid away from these surfaces. To simplify our presentation, we first consider the case $R = 2$ and $K = 2$. This will illustrate the types of asymptotic behavior that arise.

3.1 MVA Algorithm - $R = 2$ and $K = 2$

For the case when $R = 2$ and $K = 2$, the MVA algorithm (??)-(??) reduces to the system of equations

$$\begin{aligned} X_1 &= \frac{m_1}{D_{11}[1 + N_1(m_1 - 1, m_2)] + D_{12}[1 + N_2(m_1 - 1, m_2)]} \\ X_2 &= \frac{m_2}{D_{21}[1 + N_1(m_1, m_2 - 1)] + D_{22}[1 + N_2(m_1, m_2 - 1)]} \end{aligned} \quad (3.32)$$

$$N_1(m_1, m_2) = X_1 D_{11}[1 + N_1(m_1 - 1, m_2)] + X_2 D_{21}[1 + N_1(m_1, m_2 - 1)]$$

$$N_2(m_1, m_2) = X_1 D_{12}[1 + N_2(m_1 - 1, m_2)] + X_2 D_{22}[1 + N_2(m_1, m_2 - 1)]$$

with the initial conditions

$$N_1(0, 0) = 0, \quad N_2(0, 0) = 0.$$

In addition, the total population in the network must satisfy

$$N_1(m_1, m_2) + N_2(m_1, m_2) = m_1 + m_2.$$

We are interested in the behavior for $M \gg 1$ so we introduce the following scalings into (??)

$$\begin{aligned} \varepsilon &= 1/M, \quad M = M_1 + M_2 \\ z_i &= n_i/M = \varepsilon n_i \\ h_i(z_1, z_2) &= N_i(m_1, m_2)/M \end{aligned}$$

which leads to the scaled MVA recursion

$$\begin{aligned}
h_1(z_1, z_2) &= X_1 D_{11}[\varepsilon + h_1(z_1 - \varepsilon, z_2)] + X_2 D_{21}[\varepsilon + h_1(z_1, z_2 - \varepsilon)] \\
h_2(z_1, z_2) &= X_1 D_{12}[\varepsilon + h_2(z_1 - \varepsilon, z_2)] + X_2 D_{22}[\varepsilon + h_2(z_1, z_2 - \varepsilon)] \\
X_1 &= \frac{z_1}{D_{11}[\varepsilon + h_1(z_1 - \varepsilon, z_2)] + D_{12}[\varepsilon + h_2(z_1 - \varepsilon, z_2)]} \\
X_2 &= \frac{z_2}{D_{21}[\varepsilon + h_1(z_1, z_2 - \varepsilon)] + D_{22}[\varepsilon + h_2(z_1, z_2 - \varepsilon)]}.
\end{aligned} \tag{3.33}$$

The initial conditions and normalization conditions are now

$$h_1(0, 0) = 0, \quad h_2(0, 0) = 0,$$

$$h_1(z_1, z_2) + h_2(z_1, z_2) = z_1 + z_2.$$

The recursion is defined on the domain $0 \leq z_1 + z_2 \leq 1$. As was done in [?]-[?], we restrict our analysis to the line $z_1 + z_2 = 1$, i.e. total population = $M \gg 1$ so that

$$h_1 + h_2 = 1. \tag{3.34}$$

We assume an asymptotic solution for $\varepsilon \ll 1$ in the form

$$h_j \sim h_j^0 + \varepsilon h_j^1 + \dots, \quad j = 1, 2, \tag{3.35}$$

$$X_r \sim X_r^0 + \varepsilon X_r^1 + \dots, \quad r = 1, 2$$

and substitute into (??) to obtain to leading order (cf. (??))

$$(U_1^0 - 1)h_1^0 = 0 \tag{3.36}$$

$$(U_2^0 - 1)h_2^0 = 0$$

and to the next order ($\mathcal{O}(\varepsilon)$)

$$(U_1^0 - 1)h_1^1 = X_1^0 D_{11} \partial_{z_1} h_1^0 + X_2^0 D_{21} \partial_{z_2} h_1^0 - U_1^1 h_1^0 - U_1^0 \tag{3.37}$$

$$(U_2^0 - 1)h_2^1 = X_1^0 D_{12} \partial_{z_1} h_2^0 + X_2^0 D_{22} \partial_{z_2} h_2^0 - U_2^1 h_2^0 - U_2^0.$$

Here U_j^i is the i th term in the expansion of the utilization at node j and is defined by

$$U_j^i \equiv D_{1j}X_1^i + D_{2j}X_2^i, \quad j = 1, 2. \quad (3.38)$$

where $U_j^0 \leq 1$. We also need to expand the equations in (??) for the throughput. We obtain to leading order

$$X_1^0 = \frac{z_1}{D_{11}h_1^0 + D_{12}h_2^0}, \quad X_2^0 = \frac{z_2}{D_{21}h_1^0 + D_{22}h_2^0} \quad (3.39)$$

and to $\mathcal{O}(\varepsilon)$, after using (??), we find

$$X_1^1 = -z_1 \frac{D_{11} + (D_{11} - D_{12})(h_1^1 - \partial_{z_1} h_1^0)}{(D_{11}h_1^0 + D_{12}h_2^0)^2} \quad (3.40)$$

$$X_2^1 = -z_2 \frac{D_{21} + (D_{21} - D_{22})(h_1^1 - \partial_{z_2} h_1^0)}{(D_{21}h_1^0 + D_{22}h_2^0)^2}.$$

There are three possible solutions of the leading order equations (??):

1. **Node 1 in Bottleneck Set:** $h_2^0 = 0$; $U_1^0 = 1$
2. **Node 2 in Bottleneck Set:** $h_1^0 = 0$; $U_2^0 = 1$
3. **Nodes 1 and 2 in Bottleneck Set:** $U_1^0 = 1$; $U_2^0 = 1$.

The nodes in the bottleneck set, and hence the form of the leading term in the expansion, depend on the values of z_1 and z_2 , i.e. the fraction of jobs in each of the job classes. We now consider each of the cases separately.

Case 1: In this case, node 1 is the only node in the bottleneck set and we choose $h_2^0 = 0$; $U_1^0 = 1$. Using (??), we find that $h_1^0 = 1$. We now use the throughput equations in (??), taking into account the leading terms for h_j^0 , to obtain

$$X_1^0 = \frac{z_1}{D_{11}}; \quad X_2^0 = \frac{z_2}{D_{21}}. \quad (3.41)$$

We require that the utilization at the non-bottleneck node 2 be < 1 , hence

$$U_2^0 = \frac{D_{12}}{D_{11}}z_1 + \frac{D_{22}}{D_{21}}z_2 < 1 \quad (3.42)$$

which leads to the following restriction on the value of z_1

$$z_1 > z_1^* \equiv \frac{D_{11}(D_{22} - D_{21})}{\Delta}, \quad \Delta \equiv D_{11}D_{22} - D_{12}D_{21} > 0 \quad (3.43)$$

$$z_1 < z_1^*, \quad \Delta < 0. \quad (3.44)$$

At the point $z_1 = z_1^*$, $U_2^0 = 1$ so that node 2 is now also a bottleneck node. We refer to z_1^* as a *switching point* (Schweitzer [?]) since the elements in the bottleneck set change at this point. The

most interesting case is when the switching point lies in the interval (0,1) which occurs under the conditions

$$D_{11} > D_{12}, \quad D_{22} > D_{21} \quad \text{if } \Delta > 0 \quad (3.45)$$

$$D_{11} < D_{12}, \quad D_{22} < D_{21} \quad \text{if } \Delta < 0. \quad (3.46)$$

When the switching point lies outside of the interval [0,1], it plays no role in the asymptotics.

The correction terms, i.e. next order in ε , can be easily obtained using *Maple*. Using the above results in (??), we obtain

$$U_1^1 = -1, \quad h_2^1 = \frac{U_2^0}{1 - U_2^0} = -1 + \frac{D_{11}D_{21}}{\Delta(z_1 - z_1^*)} \quad (3.47)$$

and using (??)

$$h_1^1 = -h_2^1. \quad (3.48)$$

We can then use (??) to obtain the next term in the throughput. To summarize, we have found that

$$\begin{aligned} h_2 &\sim \varepsilon \left[-1 + \frac{D_{11}D_{21}}{\Delta(z_1 - z_1^*)} \right] + \varepsilon^2 \frac{D_{11}D_{21}(\Delta z_1^2 + 2D_{12}D_{21}z_1 - D_{11}D_{22})}{\Delta^2(z_1 - z_1^*)^3} \\ h_1 &= 1 - h_2 \\ X_1 &\sim \frac{z_1}{D_{11}} \left(1 - \varepsilon \frac{z_1 - 1}{z_1 - z_1^*} \right), \quad X_2 \sim \frac{z_2}{D_{21}} \left(1 - \varepsilon \frac{z_1}{z_1 - z_1^*} \right) \\ U_1 &= 1 + \mathcal{O}(\varepsilon^2), \quad U_2 \sim U_2^0 = 1 - \frac{\Delta}{D_{11}D_{21}}(z_1 - z_1^*) \end{aligned}$$

which holds for $z_1 \lesseqgtr z_1^*$ when $\Delta \lesseqgtr 0$.

Case 2: The analysis when node 2 is the only bottleneck is the same as in case 1 so we merely summarize the results:

$$\begin{aligned} h_1 &\sim \varepsilon \left[-1 - \frac{D_{12}D_{22}}{\Delta(z_1 - z_1^{**})} \right] - \varepsilon^2 \frac{D_{12}D_{22}(\Delta z_1^2 - 2D_{11}D_{22}z_1 + D_{12}D_{21})}{\Delta^2(z_1 - z_1^{**})^3} \\ h_2 &= 1 - h_1 \\ X_1 &\sim \frac{z_1}{D_{12}} \left(1 + \varepsilon \frac{1 - z_1}{z_1 - z_1^{**}} \right), \quad X_2 \sim \frac{z_2}{D_{22}} \left(1 - \varepsilon \frac{z_1}{z_1 - z_1^{**}} \right) \\ U_2 &= 1 + \mathcal{O}(\varepsilon^2), \quad U_1 \sim U_1^0 = 1 - \frac{\Delta}{D_{12}D_{22}}(z_1^{**} - z_1) \end{aligned}$$

which holds for

$$z_1 < z_1^{**} \equiv \frac{D_{12}(D_{22} - D_{21})}{\Delta}, \quad \text{for } \Delta > 0 \quad (3.49)$$

$$z_1 > z_1^{**} \quad \text{for } \Delta < 0 \quad (3.50)$$

The switching points z_1^* and z_1^{**} satisfies the conditions

$$0 < z_1^{**} < z_1^* < 1 \quad \text{if } \Delta > 0, \quad D_{11} > D_{12}, \quad D_{22} > D_{21}, \quad (3.51)$$

$$0 < z_1^* < z_1^{**} < 1 \quad \text{if } \Delta < 0, \quad D_{11} < D_{12}, \quad D_{22} < D_{21}. \quad (3.52)$$

We also observe that when $\Delta = 0$ only node 1 or node 2 is in the bottleneck set depending upon which node has the largest values of the loads D_{ri} .

We will assume that (??) holds, i.e. the switching points satisfy (??), and will only summarize the final results when (??) holds, since the analysis is similar.

Case 3: When both node 1 and node 2 are contained in the bottleneck set, we have

$$U_1^0 \equiv D_{11}X_1^0 + D_{21}X_2^0 = 1 \quad (3.53)$$

$$U_2^0 \equiv D_{12}X_1^0 + D_{22}X_2^0 = 1.$$

Solving (??), we find that the limiting throughputs are given by

$$X_1^0 = \frac{D_{22} - D_{21}}{\Delta}, \quad X_2^0 = \frac{D_{11} - D_{12}}{\Delta}. \quad (3.54)$$

both of which are positive when either condition (??) or (??) holds. We can then use the equations for h_j in (??) to obtain if $\Delta > 0$

$$h_1^0 = \frac{z_1 - z_1^{**}}{z_1^* - z_1^{**}}, \quad h_2^0 = \frac{z_1^* - z_1}{z_1^* - z_1^{**}}. \quad (3.55)$$

Summary: We summarize the leading term in the asymptotic expansions (??).

$\Delta > 0, D_{11} > D_{12}, D_{22} > D_{21}$:

$$h_1^0 = \begin{cases} 0, & 0 < z_1 < z_1^{**} \\ \frac{z_1 - z_1^{**}}{z_1^* - z_1^{**}}, & z_1^{**} < z_1 < z_1^* \\ 1, & z_1^* < z_1 < 1 \end{cases} \quad X_1^0 = \begin{cases} \frac{z_1}{D_{12}}, & 0 < z_1 < z_1^{**} \\ \frac{D_{22} - D_{21}}{\Delta}, & z_1^{**} < z_1 < z_1^* \\ \frac{z_1}{D_{11}}, & z_1^* < z_1 < 1. \end{cases} \quad h_2^0 = 1 - h_1^0.$$

$\Delta < 0, D_{11} < D_{12}, D_{22} < D_{21}$:

$$h_1^0 = \begin{cases} 1, & 0 < z_1 < z_1^* \\ \frac{z_1^{**} - z_1}{z_1^{**} - z_1^*}, & z_1^* < z_1 < z_1^{**} \\ 0, & z_1^{**} < z_1 < 1, \end{cases} \quad X_1^0 = \begin{cases} \frac{z_1}{D_{11}}, & 0 < z_1 < z_1^* \\ \frac{D_{22} - D_{21}}{\Delta}, & z_1^* < z_1 < z_1^{**} \\ \frac{z_1}{D_{12}}, & z_1^{**} < z_1 < 1. \end{cases} \quad h_2^0 = 1 - h_1^0.$$

We observe that the leading term agrees with the result obtained by Asymptotic Bound Analysis described above, and the second term provides a correction to it. However, the next term in the asymptotic expansion for h_j has a singularity as z_1 approaches a switching point (z_1^{**} or z_1^*). Hence, the asymptotic expansion in the form (??) is no longer valid in the neighborhood of the a switching point. This behavior was previously pointed out in [?]-[?]. In singular perturbation problems, such a singularity indicates the need for a boundary layer correction, i.e. a new expansion form, at or near the switching point. This new analysis is described below.

3.2 Approximation near a Switching Point

As we have seen in the previous section, the “outer” solution in the form (??) is no longer valid in the neighborhood of a switching point. To find the proper asymptotic behavior near z_1^{**} , we introduce the following new scaling

$$z_1 = z_1^{**} + \sqrt{\varepsilon}\eta, \quad z_2 = 1 - z_1^{**} - \sqrt{\varepsilon}\eta, \quad (3.56)$$

where $z_1 + z_2 = 1$. In addition, we seek a boundary layer solution of the form

$$\begin{aligned} h_1 &= \sqrt{\varepsilon}g(\eta; \varepsilon) = \sqrt{\varepsilon}[g_0 + \sqrt{\varepsilon}g_1 + \dots] \\ h_2 &= 1 - \sqrt{\varepsilon}g(\eta; \varepsilon) = 1 - \sqrt{\varepsilon}[g_0 + \sqrt{\varepsilon}g_1 + \dots] \end{aligned} \quad (3.57)$$

$$\begin{aligned} X_1 &= X_1^0 + \sqrt{\varepsilon}X_1^{(1)} + \dots \\ X_2 &= X_2^0 + \sqrt{\varepsilon}X_2^{(1)} + \dots \end{aligned}$$

We use the scaling (??) and the solution form (??) in (??) and expand for $\varepsilon \ll 1$. The expansion of the equation for h_1 leads to, at $\mathcal{O}(\sqrt{\varepsilon})$,

$$[D_{21}X_2^0 - D_{11}X_1^0]g'_0 + [D_{21}X_2^{(1)} + D_{11}X_1^{(1)}]g_0 + 1 = 0 \quad (3.58)$$

and at $\mathcal{O}(\varepsilon)$

$$\begin{aligned} &[D_{21}X_2^{(1)} - D_{11}X_1^{(1)}]g'_1 + [D_{11}X_1^{(1)} + D_{21}X_2^{(1)}]g_1 \\ &+ D_{11}X_1^{(1)} + D_{21}X_2^{(1)} + [D_{11}X_1^{(2)} + D_{21}X_2^{(2)}]g_0 \\ &+ [-D_{11}X_1^{(1)} + D_{21}X_2^{(1)}]g'_0 + \frac{1}{2}g''_0 = 0. \end{aligned} \quad (3.59)$$

Here we have used the fact the $U_1^0 = 1$ and (??) to simplify the equations. We next expand the throughput equations in (??) to obtain, for class 1,

$$\begin{aligned} D_{12}X_1^0 &= z_1^{**} \\ D_{12}X_1^{(1)} + (D_{11} - D_{12})g_0X_1^0 &= \eta \\ D_{12}X_1^{(2)} + (D_{11} - D_{12})g_0X_1^{(1)} + [D_{11} + (D_{11} - D_{12})(g_1 - g'_0)]X_1^0 &= 0 \end{aligned} \quad (3.60)$$

and for class 2

$$\begin{aligned} D_{22}X_2^0 &= 1 - z_1^{**} \\ D_{22}X_2^{(1)} + (D_{21} - D_{22})g_0X_2^0 &= -\eta \\ D_{22}X_2^{(2)} + (D_{21} - D_{22})g_0X_2^{(1)} + [D_{21} + (D_{21} - D_{22})(g_1 + g'_0)]X_2^0 &= 0. \end{aligned} \quad (3.61)$$

We can solve for $X_1^{(1)}$ and $X_2^{(1)}$ in terms of X_1^0 and X_2^0 , which are known, and the unknown function g_0 . We then substitute these results into (??) to obtain the leading order equation

$$Ag'_0 + B\eta g_0 + Cg_0^2 + 1 = 0 \quad (3.62)$$

where

$$\begin{aligned} A &\equiv D_{21}X_2^0 - D_{11}X_1^0, \\ B &\equiv \frac{D_{11}}{D_{12}} - \frac{D_{21}}{D_{22}} = \frac{\Delta}{D_{12}D_{22}} \\ C &\equiv D_{11}X_1^0\left(1 - \frac{D_{11}}{D_{12}}\right) + D_{21}X_2^0\left(1 - \frac{D_{21}}{D_{22}}\right). \end{aligned} \quad (3.63)$$

The equation for g_1 is obtained from (??), by first solving for the throughput correction terms as functions of g_0 and X_j^0 in (??)-(??). It is a linear first-order differential equation given by

$$\begin{aligned} Ag'_1 + (B\eta + 2Cg_0)g_1 + B\eta + (2C - 1)g_0 + \frac{1}{2}g_0'' + 2E_1g_0g_0' + \\ E_2\eta g_0^2 + E_3g_0^3 + B'\eta g_0' = 0. \end{aligned} \quad (3.64)$$

where

$$\begin{aligned} E_1 &\equiv A + \frac{D_{11}^2}{D_{12}}X_1^0 - \frac{D_{21}^2}{D_{22}}X_2^0, \quad B' \equiv -\frac{D_{11}}{D_{12}} - \frac{D_{21}}{D_{22}} \\ E_2 &\equiv \frac{D_{11}}{D_{12}}\left(1 - \frac{D_{11}}{D_{12}}\right) - \frac{D_{21}}{D_{22}}\left(1 - \frac{D_{21}}{D_{22}}\right) \\ E_3 &\equiv D_{11}X_1^0\left(1 - \frac{D_{11}}{D_{12}}\right)^2 + D_{21}X_2^0\left(1 - \frac{D_{21}}{D_{22}}\right)^2. \end{aligned} \quad (3.65)$$

The boundary layer solution $g(\eta; \varepsilon)$ must also match (i.e. agree with) the solution valid away from the switching point in the limits as $\eta \rightarrow \pm\infty$. This can be written symbolically as

$$\begin{aligned} \sqrt{\varepsilon}g(\eta; \varepsilon)|_{\eta \rightarrow \infty} &\sim h_1(z_1)|_{z_1 \downarrow z_1^{**}} \\ \sqrt{\varepsilon}g(\eta; \varepsilon)|_{\eta \rightarrow -\infty} &\sim h_1(z_1)|_{z_1 \uparrow z_1^{**}}. \end{aligned}$$

We will concentrate on the case when $\Delta < 0$ and when (??) holds. In this case, the switching points are ordered as in (??) and $A > 0$, $B < 0$, $C < 0$ in (??). To find the solution of (??), we convert the Riccati equation to a linear second order equation by introducing

$$g_0 = -\frac{B\eta}{2C} + \frac{AF'}{CF}$$

and find that F satisfies the parabolic cylinder equation Abramowitz and Stegun [?]

$$F_{\omega\omega} + \left[-\frac{1}{4}\omega^2 + \frac{1}{2} - \frac{C}{AB}\right]F = 0, \quad \eta = \left|\frac{A}{B}\right|^{1/2}\omega. \quad (3.66)$$

The general solution of (??) is

$$F(\omega) = c_1 U\left(\frac{C}{AB} - \frac{1}{2}, \omega\right) + c_2 V\left(\frac{C}{AB} - \frac{1}{2}, \omega\right) \quad (3.67)$$

where $U(a, x)$ and $V(a, x)$ are parabolic cylinder functions described in [?]. We substitute $F(\omega)$ into the formula for g_0 and fix the arbitrary constants by applying the matching condition as $\eta \rightarrow \infty$, which to leading order becomes

$$\sqrt{\varepsilon} g_0 \sim \varepsilon h_1^1 \sim -\frac{D_{12} D_{22} \varepsilon}{\Delta(z_1 - z_1^{**})}, \quad z_1 \downarrow z_1^{**}.$$

To satisfy the matching conditions, we must choose $c_2 = 0$ so that

$$g_0 = \frac{1}{\sqrt{-AB}} \frac{U\left(\frac{C}{AB} + \frac{1}{2}, \omega\right)}{U\left(\frac{C}{AB} - \frac{1}{2}, \omega\right)}. \quad (3.68)$$

After a tedious calculation, we find that the solution of the linear equation (??) for the correction term, which satisfies the appropriate matching conditions, is given by

$$g_1(\eta) = \alpha_1 + \alpha_2 \eta g_0(\eta) + \alpha_3 g_0^2(\eta) + \frac{1}{AF^2(\eta)} \int_{\eta}^{\infty} F^2(u) (\nu_4 u^2 g_0(u) + \nu_5 g_0^3(u)) du \quad (3.69)$$

where

$$F(\eta) = U\left(\frac{C}{AB} - \frac{1}{2}, \sqrt{\frac{-B}{A}} \eta\right) \quad (3.70)$$

and

$$\begin{aligned} \alpha_1 &= \left(\frac{2C}{AB^2} - \frac{1}{B}\right) \nu_1 - \frac{1}{AB} \nu_2 + \frac{2}{AB^2} \nu_3 \\ \alpha_2 &= \frac{2C}{AB} \nu_1 - \frac{1}{A} \nu_2 + \frac{2}{AB} \nu_3 \\ \alpha_3 &= \frac{2C^2}{AB^2} \nu_1 - \frac{C}{AB} \nu_2 + \left(\frac{1}{B} + \frac{2C}{AB^2}\right) \nu_3 \\ \nu_1 &= B + \frac{B}{2A^2} - \frac{B'}{A} \\ \nu_2 &= 2C - 1 - \frac{B}{2A} + \frac{C}{A^2} - \frac{2E_1}{A} \\ \nu_3 &= E_2 + \frac{3BC}{2A^2} - \frac{1}{A}(B'C + 2E_1 B) \\ \nu_4 &= \frac{B^2}{2A^2} - \frac{BB'}{A} \\ \nu_5 &= E_3 + \frac{C^2}{A^2} - \frac{2E_1 C}{A}. \end{aligned} \quad (3.71)$$

The constants $A, B, B', C, E_1, E_2,$ and E_3 are defined in (??) and (??). It can be verified that the boundary layer solution matches to two orders to the outer expansion, in the limit as $\eta \rightarrow -\infty$,

i.e. $z_1 < z_1^{**}$ and $\eta \rightarrow \infty$, i.e. $z_1 > z_1^{**}$. A similar analysis is needed near the switching point z_1^* which is analogous to the boundary layer analysis near z_1^{**} . The asymptotic solution consists of the solution valid away from the switching points and the boundary layer solutions near z_1^* and z_1^{**} . The main results are summarized below.

Result 3 Consider a closed, two class queueing network with two fixed rate nodes (i.e. $R = 2$ and $K = 2$). We assume that the parameters satisfy (??) and that on $z_1 + z_2 = 1$, the switching points z_1^* and z_1^{**} are defined by (??) and (??), respectively. The asymptotic expansions for the performance measures for $M \gg 1$ are:

z_1 away from the switching points:

$$h_1 \sim \begin{cases} 1 - \varepsilon \left[-1 + \frac{D_{11}D_{21}}{\Delta(z_1 - z_1^*)} \right] & 0 < z_1 < z_1^* \\ \frac{z_1^{**} - z_1}{z_1^{**} - z_1^*}, & z_1^* < z_1 < z_1^{**} \\ \varepsilon \left[-1 + \frac{D_{11}D_{21}}{\Delta(z_1 - z_1^*)} \right], & z_1^{**} < z_1 < 1, \end{cases} \quad h_2 = 1 - h_1,$$

$$X_1 \sim \begin{cases} \frac{z_1}{D_{11}} \left(1 - \varepsilon \frac{z_1 - 1}{z_1 - z_1^*} \right), & 0 < z_1 < z_1^* \\ \frac{D_{22} - D_{21}}{\Delta}, & z_1^* < z_1 < z_1^{**} \\ \frac{z_1}{D_{12}} \left(1 + \varepsilon \frac{1 - z_1}{z_1 - z_1^{**}} \right), & z_1^{**} < z_1 < 1, \end{cases}$$

$$X_2 \sim \begin{cases} \frac{z_2}{D_{21}} \left(1 - \varepsilon \frac{z_1}{z_1 - z_1^*} \right), & 0 < z_1 < z_1^* \\ \frac{D_{11} - D_{12}}{\Delta}, & z_1^* < z_1 < z_1^{**} \\ \frac{z_2}{D_{22}} \left(1 - \varepsilon \frac{z_1}{z_1 - z_1^{**}} \right), & z_1^{**} < z_1 < 1. \end{cases}$$

The asymptotic results for U_1 and U_2 are described in Cases 1-3 above and can be obtained using (??). Near the switching point z_1^{**} , the asymptotic expansion is given by the boundary layer solution:

$$\begin{aligned} h_1 &= \sqrt{\varepsilon}[g_0(\eta) + \sqrt{\varepsilon}g_1(\eta) + \dots], \quad z_1 = z_1^{**} + \sqrt{\varepsilon}\eta \\ h_2 &= 1 - h_1 \\ X_1 &= X_1^0 + \sqrt{\varepsilon}X_1^{(1)} + \dots \\ X_2 &= X_2^0 + \sqrt{\varepsilon}X_2^{(1)} + \dots \end{aligned}$$

where $g_0(\eta)$ is given by (??) and $g_1(\eta)$ is given by (??)-(??). The formulas for X_1^0 , $X_1^{(1)}$, X_2^0 , and $X_2^{(1)}$ are obtained by solving (??) and (??) in terms of g_0 and g_1 .

When (??) holds, the location of the switching points changes but the asymptotic analysis is completely analogous to that above.

3.3 Special Case

We now investigate a specific example in which the loads are given by the following matrix

$$(D_{ri}) = \begin{pmatrix} 1 & 2 \\ 7 & 4 \end{pmatrix}.$$

This example was presented in [?] where the expansions away from the switching points were given. We note that $\Delta < 0$ in this example and conditions (??) are satisfied.

The leading term in our asymptotic approximation and the locations of the switching points are summarized below:

1. **Node 1 is bottleneck** - $h_2^0 = 0$ and $U_1^0 = 1$ which leads to $h_1^0 = 1$ and $X_1^0 = z_1$. The constraint $U_2^0 < 1$ satisfied if

$$z_1 < 0.3 = z_1^*.$$

2. **Node 2 is bottleneck** - $h_1^0 = 0$ and $U_2^0 = 1$ which leads to $h_2^0 = 1$ and $X_1^0 = \frac{z_1}{2}$. The constraint $U_1^0 < 1$ satisfied if

$$z_1 > 0.6 = z_1^{**}$$

3. **Nodes 1 and 2 are bottlenecks** - $U_1^0 = U_2^0 = 1$ which implies $X_1^0 = \frac{3}{10}$ and $X_2^0 = \frac{1}{10}$. These lead to

$$h_1^0 = 2 - \frac{10}{3}z_1$$

which satisfies $0 \leq h_1^0 \leq 1$ if $0.3 \leq z_1 \leq 0.6$. The perturbation expansion valid away from the switch points, including correction terms, is given by

$$h_1(z_1) \sim \begin{cases} 1 - \varepsilon \frac{10z_1 + 4}{3 - 10z_1} - 70\varepsilon^2 \frac{10z_1^2 - 28z_1 + 4}{(3 - 10z_1)^3}, & 0 \leq z_1 < .3 \\ 2 - \frac{10}{3}z_1 - \varepsilon \frac{1}{3} \frac{50z_1^2 - 155z_1 + 63}{(10z_1 - 3)(5z_1 - 3)}, & .3 < z_1 < .6 \\ \varepsilon \frac{5z_1 - 7}{3 - 5z_1} - 20\varepsilon^2 \frac{5z_1^2 + 4z_1 - 7}{(3 - 5z_1)^3}, & .6 < z_1 \leq 1. \end{cases} \quad (3.72)$$

In Figure 6, we display graphs of the exact MVA recursion and the $\mathcal{O}(1)$ terms in the approximation (??) which is valid away from the switching points. The $\mathcal{O}(1)$ terms corresponds to the result obtained using Asymptotic Bound Analysis.

We can clearly see in Figure 7 that the approximation (??) fails near the switching points, as a singularity occurs in the correction terms. Thus the higher order terms do not prove particularly useful when $z_1 \approx 0.6$ and $z_1 \approx 0.3$. This illustrates the need for the boundary layer analysis.

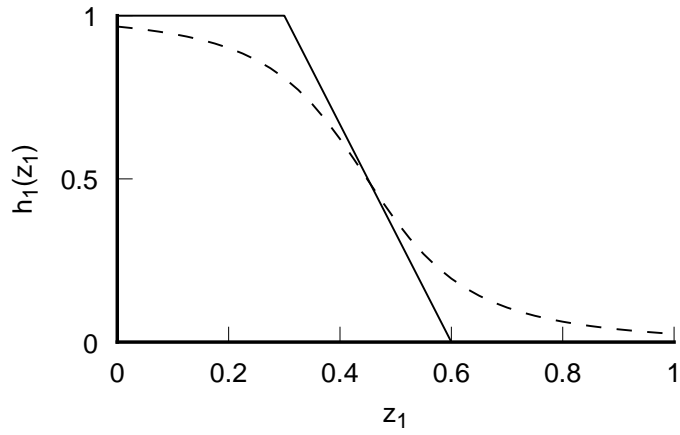


Figure 6: Graphs of h_1 as a function of z_1 for the special case with $M = 40$. Exact MVA result = $--$ and the $\mathcal{O}(1)$ approximation in (3.72) = $—$.

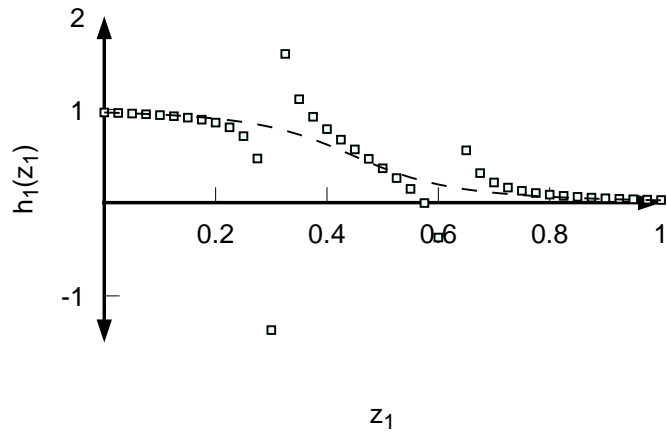


Figure 7: Graphs of h_1 as a function of z_1 for the special case with $M = 40$. Exact MVA result = $--$ and the approximation in (3.72) including $\mathcal{O}(\varepsilon)$ terms = \square .

We now implement the boundary layer analysis described in the previous section near the switching point $z_1 = .6$. We introduce the stretched variable

$$\eta = \frac{z_1 - .6}{\sqrt{\varepsilon}}, \quad z_1 = .6 + \sqrt{\varepsilon}\eta \quad (3.73)$$

and the boundary layer function (??)

$$h_1 = \sqrt{\varepsilon}g(\eta) = \sqrt{\varepsilon} [g_0(\eta) + \sqrt{\varepsilon}g_1(\eta) + \dots]. \quad (3.74)$$

The boundary layer equation (??) becomes

$$\frac{2}{5}g_0' - \frac{5}{4}\eta g_0 - \frac{3}{8}g_0^2 + 1 = 0$$

with the matching conditions:

$$h_1 \sim \sqrt{\varepsilon}\frac{4}{5\eta}, \quad z_1 > .6 \Rightarrow \eta \rightarrow \infty, \quad g_0 \sim \frac{4}{5\eta}$$

$$h_1 \sim \sqrt{\varepsilon}\frac{-10\eta}{3}, \quad z_1 < .6 \Rightarrow \eta \rightarrow -\infty, \quad g_0 \sim \frac{-10\eta}{3}.$$

The solution (??) is then given by

$$g_0(\eta) = \frac{\sqrt{2}U\left(\frac{5}{4}, \frac{5\sqrt{2}}{4}\eta\right)}{U\left(\frac{1}{4}, \frac{5\sqrt{2}}{4}\eta\right)} \quad (3.75)$$

where $U(a, x)$ is a parabolic cylinder function. The next term in the boundary layer solution is given by (??), which for this example becomes

$$g_1(\eta) = \frac{5}{16} + \frac{5}{64}\eta g_0(\eta) + \frac{41}{128}g_0^2(\eta) + \frac{5}{2U^2\left(\frac{1}{4}, \frac{5\sqrt{2}}{4}\eta\right)} \int_{\eta}^{\infty} U^2\left(\frac{1}{4}, \frac{5\sqrt{2}}{4}u\right) \left(-\frac{275}{128}u^2 g_0(u) + \frac{21}{256}g_0^3(u)\right) du \quad (3.76)$$

where g_0 is now given by (??).

A boundary layer solution is also needed near the switching point $z_1^* = 0.3$. Here we introduce the stretched variable

$$\xi = \frac{z_1 - 0.3}{\sqrt{\varepsilon}}$$

and use the equation for h_2 to simplify the analysis. Using the boundary layer expansion $h_2 = 1 - h_1 = \sqrt{\varepsilon}(\hat{g}_0(\xi) + \sqrt{\varepsilon}\hat{g}_1(\xi) \dots)$, we derive the leading order equation

$$-\frac{7}{5}\hat{g}_0' + 10\xi\hat{g}_0 - 3\hat{g}_0^2 + 7 = 0.$$

The solution satisfying the appropriate matching conditions is

$$\hat{g}_0(\xi) = \sqrt{\frac{7}{2}} \frac{U(2, -5\sqrt{\frac{2}{7}}\xi)}{U(1, -5\sqrt{\frac{2}{7}}\xi)}. \quad (3.77)$$

so that $h_1 \sim 1 - \sqrt{\varepsilon}\hat{g}_0(\xi)$. A correction term of the form (??) can also be computed, but we do not include it.

The complete asymptotic solution consists of the solution valid away from the switching points given by (??), and the boundary layer solutions near the switching points 0.6 and 0.3, given by (??) and (??), respectively. The solution is summarized as follows:

$$h_1(z_1) \sim \begin{cases} 1 - \varepsilon \frac{10z_1 + 4}{3 - 10z_1} - 70\varepsilon^2 \frac{10z_1^2 - 28z_1 + 4}{(3 - 10z_1)^3}, & 0 \leq z_1 < .3 \\ 1 - \sqrt{\frac{7\varepsilon}{2}} \frac{U(2, -5\sqrt{\frac{2}{7}}\xi)}{U(1, -5\sqrt{\frac{2}{7}}\xi)}, & z_1 - 0.3 = \sqrt{\varepsilon}\xi = \mathcal{O}(\sqrt{\varepsilon}) \\ 2 - \frac{10}{3}z_1 - \varepsilon \frac{1}{3} \frac{50z_1^2 - 155z_1 + 63}{(10z_1 - 3)(5z_1 - 3)}, & .3 < z_1 < .6 \\ \frac{\sqrt{2\varepsilon}U(\frac{5}{4}, \frac{5\sqrt{2}}{4}\eta)}{U(\frac{1}{4}, \frac{5\sqrt{2}}{4}\eta)}, & z_1 - 0.6 = \sqrt{\varepsilon}\eta = \mathcal{O}(\sqrt{\varepsilon}) \\ \varepsilon \frac{5z_1 - 7}{3 - 5z_1} - 20\varepsilon^2 \frac{5z_1^2 + 4z_1 - 7}{(3 - 5z_1)^3}, & .6 < z_1 \leq 1. \end{cases} \quad (3.78)$$

We only give the leading term in each boundary layer region.

We present a graph of the asymptotic results in Figure 8 and demonstrate the numerical accuracy of our approximation in Table 2. The graph illustrates a smooth transition between the asymptotic solutions valid away from the switching points and the boundary layer solutions valid near the switching points. This smooth transition is expected since the different asymptotic solutions satisfy the matching conditions above. In Table 2, the data labeled ‘‘Outer’’ represents the approximation valid away from the switch points including $\mathcal{O}(\varepsilon)$ terms. The data labeled ‘‘Layer’’ was computed using the leading term in the boundary layer solution near z_1^* for $.15 \leq z_1 \leq .40$ and the leading term near z_1^{**} for $.50 \leq z_1 \leq .80$.

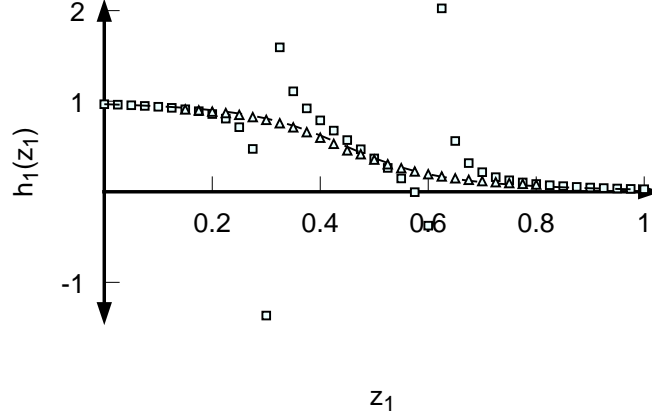


Figure 8: Asymptotic approximations to h_1 as a function of z_1 for the special case with $M = 40$. The exact solutions = —, the approximation in (3.72) including $\mathcal{O}(\varepsilon)$ terms = \square , and the leading term in the boundary layer solutions in (3.78) = \triangle .

3.4 Approximation at a Switching Surface

In the beginning of this section, we gave the scaled MVA equations for a general R class and K node network (cf. (??)-(??)). The dimension of the MVA recursion is equal to the number of job classes, since the performance measures are functions of the population vector $\vec{m} = (m_1, m_2, \dots, m_R)$. If there are more than two job classes, the analysis can become considerably more complicated. However, our methods are still applicable, as we will now show with an example.

As we discussed above, the leading term in the perturbation expansion (??) yields the same result as Asymptotic Bound Analysis. However, the correction terms will develop singularities at *switching surfaces* where the bottleneck set for the network changes. To illustrate how to construct the leading term near a switching surface, we consider a 3 class, 2 node network with the switching surface separating regions with a single node in the bottleneck set and regions with both nodes in the bottleneck set. This type of analysis applies to a switching surface in any size network.

We write the scaled recursion (??) for this network, $R = 3$ and $K = 2$, as

$$h_1(\vec{z}) = \varepsilon \sum_{j=1}^3 X_j D_{j1} + \sum_{j=1}^3 X_j D_{j1} h_1(\vec{z} - \varepsilon \vec{e}_j), \quad \vec{z} = (z_1, z_2, z_3) \quad (3.79)$$

and

$$X_r = \frac{z_r}{\varepsilon \sum_{k=1}^2 D_{rk} + D_{r1} h_1(\vec{z} - \varepsilon \vec{e}_r) + D_{r2} (1 - h_1(\vec{z} - \varepsilon \vec{e}_r))}, \quad r = 1, 2, 3. \quad (3.80)$$

Here we have used the fact that $h_1 + h_2 = 1$ and $z_1 + z_2 + z_3 = 1$.

Table 2: $h_1(z_1)$ for the example in Figure 8

| z_1 | Exact | Outer | Layer | |
|-------|-------|----------|-------|------------|
| 0 | .9666 | .9666 | | |
| .050 | .9570 | .9550 | | |
| .100 | .9441 | .9375 | | |
| .150 | .9263 | .9083 | .9090 | |
| .200 | .9009 | .8500 | .8850 | |
| .250 | .8638 | .6750 | .8489 | |
| .300 | .8090 | ∞ | .7937 | z_1^* |
| .350 | .7294 | 1.0316 | .7110 | |
| .400 | .6221 | .7416 | .5966 | |
| .450 | .4961 | .5725 | — | |
| .500 | .3724 | .3166 | .3548 | |
| .550 | .2699 | .0716 | .2601 | |
| .600 | .1944 | ∞ | .1938 | z_1^{**} |
| .650 | .1419 | .3750 | .1484 | |
| .700 | .1057 | .1750 | .1173 | |
| .750 | .0805 | .1083 | .0954 | |
| .800 | .0624 | .0750 | .0796 | |
| .850 | .0491 | .0550 | | |
| .900 | .0390 | .0416 | | |
| .950 | .0312 | .0321 | | |
| 1. | .0250 | .0250 | | |

We consider the switching surface S that separates region R_1 , in which node 2 is the unique bottleneck, and region R_{12} , in which both nodes 1 and 2 are bottlenecks. The location of this switching surface is defined using the leading terms in the approximation in R_1 , namely

$$h_1^0 = 0, \quad X_r^0 = \frac{z_r}{D_{r2}}, \quad U_2^0 = 1 \quad (3.81)$$

and the extra condition that at S

$$U_1^0 = \sum_{r=1}^3 X_r^0 D_{r1} = 1. \quad (3.82)$$

This leads to the following equation for the switching surface

$$\left(\frac{D_{11}}{D_{12}} - \frac{D_{31}}{D_{32}} \right) z_1 + \left(\frac{D_{21}}{D_{22}} - \frac{D_{31}}{D_{32}} \right) z_2 = \left(1 - \frac{D_{31}}{D_{32}} \right). \quad (3.83)$$

We define

$$f(\xi) = \frac{\left(1 - \frac{D_{31}}{D_{32}} \right) - \left(\frac{D_{11}}{D_{12}} - \frac{D_{31}}{D_{32}} \right) \xi}{\left(\frac{D_{21}}{D_{22}} - \frac{D_{31}}{D_{32}} \right)} \quad (3.84)$$

and introduce the following change of variables in the neighborhood of the switching surface

$$z_1 = \xi + \sqrt{\varepsilon}\eta, \quad z_2 = f(\xi) - \sqrt{\varepsilon}\eta. \quad (3.85)$$

The boundary layer function is given by

$$h_1(z_1, z_2) = \sqrt{\varepsilon}g(\eta, \xi) = \sqrt{\varepsilon}(g_0 + \sqrt{\varepsilon}g_1 + \dots). \quad (3.86)$$

We introduce (??)-(??) into (??) and (??) and expand for $\varepsilon \ll 1$, taking into account (??) and (??), to obtain to leading order

$$0 = \sum_{j=1}^3 D_{j1}X_j^0 + \left(\sum_{j=1}^3 D_{j1}X_j^1 \right) g_0 + A(\xi)g_{0,\eta} \quad (3.87)$$

where $A(\xi)$ is defined by

$$A(\xi) = \frac{1}{1+f'} [-f'X_1^0D_{11} + X_2^0D_{21} + (f' - 1)X_3^0D_{31}]. \quad (3.88)$$

To get an explicit form of the boundary layer equation, we expand the throughput as $X_j \sim X_j^0 + \sqrt{\varepsilon}X_j^1 + \dots$ and use (??) to obtain

$$D_{12}X_1^0 = \xi, \quad D_{22}X_2^0 = f(\xi), \quad D_{32}X_3^0 = 1 - \xi - f(\xi) \quad (3.89)$$

and

$$\begin{aligned} X_1^0(D_{11} - D_{12})g_0 + D_{12}X_1^1 &= \eta, \\ X_2^0(D_{21} - D_{22})g_0 + D_{22}X_2^1 &= -\eta, \\ X_3^0(D_{31} - D_{32})g_0 + D_{32}X_3^1 &= 0. \end{aligned}$$

In view of (??) and (??), we see that $A(\xi)$ is a linear function of ξ . The values of X_r^1 are obtained from the above as functions of X_r^0 and g_0 which, when substituted into (??), yields the Riccati equation

$$A(\xi)g_{0,\eta} + [B\eta + C(\xi)g_0]g_0 + 1 = 0 \quad (3.90)$$

where $A(\xi)$ is defined in (??) and

$$\begin{aligned} B &= \frac{D_{11}}{D_{12}} - \frac{D_{21}}{D_{22}} \\ C &= D_{11}X_1^0 \left(1 - \frac{D_{11}}{D_{12}} \right) + D_{21}X_2^0 \left(1 - \frac{D_{21}}{D_{22}} \right) + D_{31}X_3^0 \left(1 - \frac{D_{31}}{D_{32}} \right). \end{aligned}$$

We note the similarity between (??) and (??). The next step would be to match the boundary layer solution to the solutions in R_1 and R_{12} , which we illustrated for the network with two classes. Boundary layer analyses are also needed near any other switching surfaces.

4 Conclusion

To summarize, we have developed an asymptotic approach for analyzing the non-linear recurrence equations that arise in the MVA algorithm(s). This extends and improves upon previous work (cf. [?]-[?]) in that we treat the case of multiple bottleneck (and “near” bottleneck) nodes in single class networks, and also the switching surfaces that arise in multiclass networks.

The asymptotic formulas we obtained lead to accurate numerical approximations for the performance measures, even for moderate values of the total network population size (e.g. $M = 10$). For multiclass networks, we have developed the basic approach necessary to treat the vicinities of the switching surfaces, though we have not given a general formula that would apply to a network with an arbitrary number of nodes, customer classes and bottleneck nodes near the switching surface.

To treat the general case we would need to first locate all switching surfaces, introduce appropriate scales near the surfaces (such as (??)) and derive the differential equations that apply on these scales. It would seem that as long as crossing the switching surface involves adding (or subtracting) a single node to the bottleneck set, we invariably obtain a Ricatti equation of the type (??). This is explicitly solvable in terms of parabolic cylinder functions, whose numerical values are easily obtained using standard programs. If the bottleneck set changes by more than one (i.e. at intersections of two or more switching surfaces), then the local analysis will likely involve solving a non-linear system of ODEs.

We have not attempted to treat the most general case, but we believe that the basic approach will apply to networks with many nodes and job classes.

Appendix

We derive the solution of (??) by first solving the first order differential equation for f_i in terms of U to obtain

$$f_i(z) = \frac{D}{a_i - a_K} \int_0^z \exp[-(\frac{a_K - a_i}{D})(z - s)] G'(s) e^{G(s) - G(z)} ds, \quad i = 1, \dots, K - 1 \quad (\text{A.1})$$

where

$$G(z) = \int^z \frac{U(s)}{Ds} ds. \quad (\text{A.2})$$

An integral equation for $U(z)$ can be obtained by multiplying (??) by $a_i - a_K$ and then summing from $i = 1, \dots, K - 1$ to obtain

$$U(z) = D \int_0^z G'(s) e^{G(s) - G(z)} \sum_{i=1}^{K-1} \exp[-(\frac{a_K - a_i}{D})(z - s)] ds. \quad (\text{A.3})$$

We define

$$F(z) = \frac{U(z)}{z} \exp \left[\int_1^z \frac{U(\tau)}{D\tau} d\tau \right] \quad (\text{A.4})$$

with which (??) becomes

$$zF(z) = \int_0^z F(s) H(z-s) ds \quad (\text{A.5})$$

where

$$H(z) = \sum_{i=1}^{K-1} \exp \left[-\frac{a_K - a_i}{D} z \right].$$

We now define the Laplace transform $\hat{F}(\alpha) = \int_0^\infty e^{-\alpha z} F(z) dz$ and transform (??) into

$$-\hat{F}'(\alpha) = \hat{F}(\alpha) \sum_{i=1}^{K-1} \left(\alpha + \frac{a_K - a_i}{D} \right)^{-1}. \quad (\text{A.6})$$

The solution of (??) is

$$\hat{F}(\alpha) = c \prod_{i=1}^{K-1} \left(\alpha + \frac{a_K - a_i}{D} \right)^{-1}. \quad (\text{A.7})$$

To determine c , we use (??) and note that for small z

$$F(z) \sim U(0) z^{K-2} \exp \left[\int_0^1 \frac{U(0) - U(\tau)}{D\tau} d\tau \right]$$

or, in terms of Laplace transforms,

$$\hat{F}(\alpha) \sim D(K-1)! \alpha^{-(K-1)} \exp \left[\int_0^1 \frac{U(0) - U(\tau)}{D\tau} d\tau \right], \alpha \rightarrow +\infty. \quad (\text{A.8})$$

Here we have used $U(0) = D(K-1)$. From (??), we see that as $\alpha \rightarrow +\infty$,

$$\hat{F}(\alpha) \sim c \alpha^{-(K-1)}$$

so that we choose

$$c = D(K-1)! \exp \left[\int_0^1 \frac{U(0) - U(\tau)}{D\tau} d\tau \right]. \quad (\text{A.9})$$

We use (??) to solve for $U(z)$ in terms of $F(z)$ to obtain

$$U(z) = \frac{DzF(z)}{\int_0^z F(\tau) d\tau}. \quad (\text{A.10})$$

Inverting (??), we get

$$F(z) = c \sum_{i=1}^{K-1} \beta_i e^{(a_i - a_K)z/D} \quad (\text{A.11})$$

with

$$\beta_j = \prod_{i=1, i \neq j}^{K-1} \left\{ \frac{D}{a_j - a_i} \right\}. \quad (\text{A.12})$$

We find $U(z)$ using (??) and f_i using (??) to obtain

$$f_i(z) = \frac{D e^{(a_i - a_K)z/D}}{a_i - a_K} \sum_{l=1}^{K-1} \frac{\beta_l}{a_l - a_i} \left(e^{(a_l - a_i)z/D} - 1 \right) \bigg/ \sum_{l=1}^{K-1} \frac{\beta_l}{a_l - a_K} \left(e^{(a_l - a_K)z/D} - 1 \right). \quad (\text{A.13})$$

We then obtain H_i for $i = 1, 2, \dots, K-1$ from $H_i = \frac{D}{a_K - a_i} + f_i(z)$.

We now write the result for H_i in a more symmetric form, which will apply for all $1 \leq i \leq K$. Let us set

$$\mathcal{B}_j = \prod_{p=1, p \neq j}^{K-1} (a_j - a_p); \quad \mathcal{P}_l = (a_l - a_K) \mathcal{B}_l, \quad l < K; \quad \mathcal{P}_K = \mathcal{B}_K.$$

For $K \geq 2$, we can easily show that

$$\sum_{l=1}^K \frac{1}{\mathcal{P}_l} = \frac{1}{\mathcal{P}_1} + \frac{1}{\mathcal{P}_2} + \dots + \frac{1}{\mathcal{P}_K} = 0 \quad (\text{A.14})$$

and thus

$$\begin{aligned} & \sum_{l=1}^{K-1} \frac{1}{\mathcal{P}_l} (e^{a_l z/D} - e^{a_K z/D}) \\ &= \sum_{l=1}^K \frac{1}{\mathcal{P}_l} (e^{a_l z/D} - e^{a_K z/D}) = \sum_{l=1}^K \frac{1}{\mathcal{P}_l} e^{a_l z/D} \equiv S(z). \end{aligned}$$

Since $\beta_j = D^{K-2}/\mathcal{B}_j$, we obtain from (??)

$$\begin{aligned} H_i &= \frac{D}{a_K - a_i} + \frac{1}{S(z)} \frac{D}{a_i - a_K} \sum_{l=1}^{K-1} \frac{1}{\mathcal{B}_l} \frac{e^{a_l z/D} - e^{a_i z/D}}{a_l - a_i} \\ &= \frac{D}{a_i - a_K} \frac{1}{S(z)} \left\{ \sum_{l=1}^{K-1} \frac{1}{\mathcal{B}_l} \frac{e^{a_l z/D} - e^{a_i z/D}}{a_l - a_i} - \sum_{l=1}^K \frac{1}{\mathcal{P}_l} \right\}. \end{aligned}$$

Now, $\mathcal{B}_l^{-1} = (a_l - a_K) \mathcal{P}_l^{-1}$ and

$$\begin{aligned} & \frac{a_l - a_K}{a_l - a_i} \left(e^{a_l z/D} - e^{a_i z/D} \right) - e^{a_l z/D} \\ &= w \frac{a_i - a_K}{a_l - a_i} \left(e^{a_l z/D} - e^{a_i z/D} \right) - e^{a_l z/D} \end{aligned}$$

so that

$$H_i = \frac{D}{S(z)} \sum_{l=1}^{K-1} \frac{1}{\mathcal{P}_l} \frac{e^{a_l z/D} - e^{a_i z/D}}{a_l - a_i} + \frac{D}{a_K - a_i} \frac{1}{S(z)} \left\{ e^{a_i z/D} \sum_{l=1}^{K-1} \frac{1}{\mathcal{P}_l} + e^{a_K z/D} \frac{1}{\mathcal{P}_K} \right\}. \quad (\text{A.15})$$

From (??) we have $\sum_{l=1}^{K-1} \mathcal{P}_l^{-1} = -\mathcal{P}_K^{-1}$, which when used in (??) gives the expression (??). This equation also applies for $i = K$. In (??), the term with $l = i$ is understood to be evaluated by L'Hopital's rule.

REFERENCES

- [1] J. BUZEN, Computational algorithms for closed queueing networks with exponential servers, *Comm. ACM* 16, 1973, 527-531.
- [2] A.E. CONWAY AND N.D. GEORGANAS, RECAL – A new efficient algorithm for the exact analysis of multiple-chain closed queueing networks, *J. ACM* 33, 1986, 768-791.
- [3] M. REISER AND S.S. LAVENBERG, Mean-value analysis of closed multichain queueing networks, *J. ACM* 27, 1980, 313-322.
- [4] E. SOUZA E SILVA AND S.S. LAVENBERG, Calculating joint queue-length distributions in product-form queueing networks, *J. ACM* 36, 1989, 194-207.
- [5] K.W. ROSS, D.H.K. TSANG, AND J. WANG, Monte Carlo Summation and Integration Applied to Multiclass Queueing Networks, *J. ACM* 41, 1994, 1110-1135.
- [6] J. ZAHORJAN, K.C. SEVCIK, D.L. EAGER, AND B.I. GALLER, Balanced job bound analysis of queueing networks, *Commun. of ACM*, 25, 1982, 134-141.
- [7] J. MCKENNA AND D. MITRA, Integral representations and asymptotic expansions for closed Markovian queueing networks: normal usage, *Bell Syst. Tech. J.* 61, 1982, 661-683.
- [8] J. MCKENNA AND D. MITRA, Asymptotic expansions and integral representations of moments of queue lengths in closed Markovian networks, *J. ACM* 31, 1984, 346-360.
- [9] C. KNESSL AND C. TIER, Asymptotic expansions for large closed queueing networks, *J. ACM* 37, 1990, 144-174.
- [10] C. KNESSL AND C. TIER, Asymptotic expansions for large closed queueing networks with multiple job classes, *IEEE Trans. Comp.* 41, 1992, 480-488.
- [11] J.D. MEI AND C. TIER, Asymptotic approximations for a queueing network with multiple classes, *SIAM J. Appl. Math.* 54, 1994, 1147-1180.
- [12] J.B. KELLER, Rays, waves and asymptotics, *Bull. Amer. Math. Soc.* 44, 1978, 727-750.

- [13] Y. KOGAN, Another approach to asymptotic expansions for large closed queueing networks, *Oper. Res. Lett.* 11, 1992, 317-321.
- [14] A. BIRMAN AND Y. KOGAN, Asymptotic evaluation of closed queueing networks with many stations, *Commun. Statistics: Stoch. Models* 8, 1992, 543-563.
- [15] P. SCHWEITZER, Approximate analysis of multiclass closed networks of queues, *Proc. International Conference on Stochastic Control and Optimization*, 1979.
- [16] K.M. CHANDY AND D. NEUSE, Linearizer: A heuristic algorithm for queueing network models of computer systems, *Comm. ACM* 25, 1982, 126-134.
- [17] P. SCHWEITZER, A fixed-point approximation to product-form networks with large populations, presented at Second ORSA Telecommunications Conference, Boca Raton, Florida, 1992.
- [18] P. SCHWEITZER, G. SERAZZI AND M. BROGLIA, A survey of bottleneck analysis in closed queues, in *Performance Evaluation of Computer and Communications Systems*, L. Donatiello and R. Nelson, eds., Lecture Notes in Computer Science, No. 729, Springer-Verlag, Berlin, 1993, 491-508.
- [19] G. BALBO AND G. SERAZZI, Asymptotic analysis of multiclass closed queueing networks: common bottleneck, *Perf. Eval. J.* 26, 1996, 51-72.
- [20] G. BALBO AND G. SERAZZI, Asymptotic analysis of multiclass closed queueing networks: multiple bottlenecks, preprint, 1994.
- [21] P. SCHWEITZER, A nonlinear vector finite difference scheme, *SIAM Rev.* 23, 1981, 528-532.
- [22] M. ABRAMOWITZ AND I.E. STEGUN, *Handbook of Mathematical Functions*, Dover Publications, Inc., New York, 1984.