

**Importance sampling for Markovian tandem queues using subsolutions
Exploring the possibilities**

Buijsrogge, Anne; de Boer, Pieter-Tjerk ; Scheinhardt, Werner R.W.

DOI

[10.1177/00375497211041351](https://doi.org/10.1177/00375497211041351)

Publication date

2021

Document Version

Final published version

Published in

Simulation

Citation (APA)

Buijsrogge, A., de Boer, P.-T., & Scheinhardt, W. R. W. (2021). Importance sampling for Markovian tandem queues using subsolutions: Exploring the possibilities. *Simulation*, 97(12), 849-866.
<https://doi.org/10.1177/00375497211041351>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Importance sampling for Markovian tandem queues using subsolutions: exploring the possibilities

Anne Buijsrogge¹ , Pieter-Tjerk de Boer² and
Werner R W Scheinhardt²

Abstract

We consider importance sampling simulation for estimating the probability of reaching large total number of customers in an $M|M|1$ tandem queue, during a busy cycle of the system. Our main result is a procedure for obtaining a family of asymptotically efficient changes of measure based on subsolutions. We explicitly show these families for two-node tandem queues and we find that there exist more asymptotically efficient changes of measure based on subsolutions than currently available in literature.

Keywords

$M|M|1$ tandem queue, rare event simulation, importance sampling, subsolutions

1. Introduction

In this paper, we explore possibilities for importance sampling in a Markovian tandem queue. With importance sampling, the underlying probability distributions are changed to speed up a simulation. This change of the probability distributions is also called a *change of measure*. When exploring the possibilities for a change of measure, we focus on the probability of total buffer overflow during a busy cycle of an $M|M|1$ tandem queue.

This problem was first studied by Parekh and Walrand,¹ where a *state-independent* change of measure is suggested based on heuristics. In that paper, the authors note that their change of measure performs poorly in practice in some cases. Afterwards, in Glasserman and Kou,² necessary and sufficient conditions have been determined for d -node $M|M|1$ tandem queues in order for the same change of measure to be *asymptotically efficient* (which means that the relative error of the estimator grows less than exponentially fast). These conditions have been extended in the work of de Boer³ for the case $d = 2$ and, in that paper, it also has been shown that the change of measure from Parekh and Walrand¹ is the only state-independent change of measure that may be asymptotically efficient when $d = 2$. The conclusion of these two papers is that the state-independent change of measure from Parekh and Walrand¹ for the d -node $M|M|1$ tandem queue is not asymptotically efficient for all input parameters.

To resolve this issue, in the work of Dupuis et al.,⁴ a *state-dependent* change of measure has been developed for the two-node $M|M|1$ tandem queue such that this change of measure is asymptotically efficient for all input parameters. This change of measure is derived using so-called *subsolutions*, which have been introduced (in the context of importance sampling) by Dupuis and Wang.⁵ Afterward, the work from Dupuis et al.⁴ has been extended to Jackson networks in the work of Dupuis and Wang,⁶ and to non-Markovian tandem queues in the work of Buijsrogge et al.⁷ For more details on subsolutions, see also Budhiraja and Dupuis's,⁸ Chapters 14 and 15.

Another approach has been developed by Blanchet,⁹ where an algorithm is presented that gives bounded relative error. This approach uses the time reversed process of Jackson networks. While the time reversed process is known for Jackson networks, this is not the case in other contexts, e.g., for non-Markovian tandem queues. Thus, even though a more efficient simulation approach might exist for the model in the current paper, the approach that we use here is generalizable to processes where the time

¹Delft University of Technology, The Netherlands

²University of Twente, The Netherlands

Corresponding author:

Anne Buijsrogge, Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4 2628CD, Delft, The Netherlands.

Email: a.buijsrogge@tudelft.nl

reversed process is unknown, see, for example, Buijsrogge et al.⁷ Therefore, we believe that the insights obtained in this paper can be useful for those types of processes.

In addition, in the work of Sezer,¹⁰ an approximation method has been developed for the two-node $M|M|1$ tandem queue using subsolutions. While the approximation of this probability could be very interesting in some cases, an unbiased simulation might be preferred in other cases or could be used to verify the results. Thus, in this paper, we focus on simulation.

Even though the problem that we consider has been studied previously in^{1-4,6,9,11} and asymptotic efficiency of changes of measure based on subsolutions has been shown in some of these works^{4,6,11} still there are “only” three such changes of measure known that have been proven to be asymptotically efficient for the two-node $M|M|1$ tandem queue: two when queue 2 is the bottleneck queue (i.e. $\mu_2 \leq \mu_1$, where μ_i is the service rate of queue i), and one when queue 1 is the bottleneck queue ($\mu_1 \leq \mu_2$), see Dupuis et al.⁴ and Dupuis and Wang.⁶ However, in the work of Buijsrogge et al.,⁷ we have shown that there are several possible changes of measure based on subsolutions yielding an asymptotically efficient estimator for $GI|GI|1$ tandem queues with bounded support. Since the exponential distribution clearly does not have bounded support, this raises some thoughts on whether there exist more asymptotically efficient changes of measure based on subsolutions for the two-node $M|M|1$ tandem queue.

In this paper, we give sufficient conditions for an asymptotically efficient change of measure based on subsolutions for the $M|M|1$ tandem queue. These conditions follow naturally when using the same method as in the works of Dupuis et al.⁴ and Dupuis and Wang,⁶ but have never been stated as such. We show how to find changes of measure satisfying these conditions and we believe the same could be done for other models. There is no numerical section as we do not aim for more efficient simulation schemes for the particular tandem model at hand (for which we could simply follow Blanchet⁹), but we aim for providing a procedure to obtain possible schemes based on subsolutions that surely lead to an asymptotically efficient estimator. For more complex models, we do have numerical results, see Buijsrogge et al.⁷ and Buijsrogge.¹²

Moreover, in this paper, we consider both $\mu_1 \leq \mu_2$ and $\mu_2 \leq \mu_1$. Even though this does not affect the probability of interest, since both queues are interchangeable,¹³ it is particularly interesting to see the possibilities for a change of measure when $\mu_1 \leq \mu_2$. This is in line with Buijsrogge et al.,⁷ where we have considered both queue 1 and queue 2 being the bottleneck queue (for a two-node $M|M|1$ tandem queue this is equivalent to $\mu_1 \leq \mu_2$ and $\mu_2 \leq \mu_1$, respectively). This is in slight contrast with most existing

literature on importance sampling for two-node *Markovian* tandem queues, see for example,^{2,3,4,11} where only the second queue being the bottleneck queue has been discussed (i.e., $\mu_2 \leq \mu_1$).

The contributions of this paper are two-fold. After summarizing the subsolution method for importance sampling and stating the results from Dupuis et al.⁴ and Dupuis and Wang⁶ in Section 2, our first contribution is in Section 3, where we state conditions for a change of measure for the d -node $M|M|1$ tandem queue based on subsolutions to give an asymptotically efficient estimator, and we prove that these conditions are sufficient for $d = 2$. The other contribution, in Section 4, is that we provide a whole *family* of changes of measure for the two-node $M|M|1$ tandem queue that satisfy these conditions and hence result in an asymptotically efficient estimator. We conclude this paper in Section 5.

2. Model and preliminaries

2.1. The model

In this paper, we consider a d -node $M|M|1$ tandem queue, with arrival rate λ and service rates μ_1, \dots, μ_d for queues 1, \dots , d , respectively, and we are interested in estimating the probability that the total number of customers in the system reaches some high level N during a busy cycle of the system. Since the case $d = 2$ has been studied mostly in literature, we mainly consider this case. However, it seems likely that all of the results can also be extended to $d > 2$ and so we will briefly comment on these extensions when necessary. Thus, we now let $d = 2$.

We consider the underlying embedded discrete time Markov chain and we assume without loss of generality $\lambda + \mu_1 + \mu_2 = 1$. Furthermore, we assume that $\lambda < \min\{\mu_1, \mu_2\}$, so that the system is stable. As in previous works,^{4,6,11} we let the state description be the number of customers in each queue, denoted by $\mathbf{Z}_i = (Z_{1,i}, Z_{2,i})$, where $Z_{j,i}$ is the number of customers in queue j after i transitions.

We let \mathbf{v}_k denote the possible transitions and $\Theta(\mathbf{v}_k)$ their corresponding probabilities, i.e., $\mathbf{v}_0 = (1, 0)$ corresponds to an arrival to the first queue, and has probability $\Theta(\mathbf{v}_0) = \lambda$. Similarly, we have $\mathbf{v}_1 = (-1, 1)$ and $\mathbf{v}_2 = (0, -1)$, $\Theta(\mathbf{v}_1) = \mu_1$ and $\Theta(\mathbf{v}_2) = \mu_2$. The transitions \mathbf{v}_1 and \mathbf{v}_2 can only occur when $Z_{1,i} > 0$ and $Z_{2,i} > 0$, respectively. If $Z_{j,i} = 0$ for some queue j , thus, there are no customers in queue j , we add a self-loop transition with probability μ_j to make sure that the sum of all rates equals 1 (so that all rates are probabilities).

As in the work of Dupuis et al.,⁴ we let $\mathbf{X}_i = (\frac{1}{N})\mathbf{Z}_i$ be the scaled state description, which has the advantage that its elements are in $[0, 1]$ and therefore do not increase as N

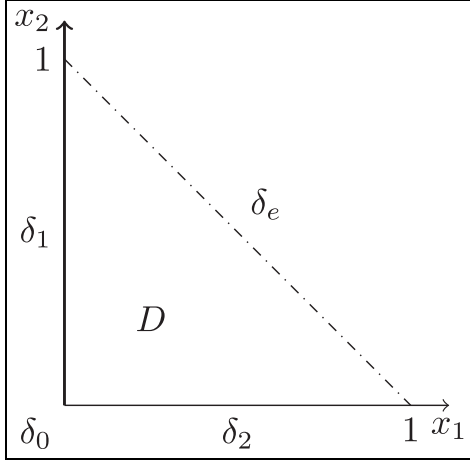


Figure 1. A sketch of the scaled state description of the event of interest.

increases. This allows us to make the following definitions:

$$\begin{aligned} D &= \{(x_1, x_2) : x_j > 0, x_1 + x_2 < 1\}, \\ \delta_1 &= \{(0, x_2) : 0 < x_2 < 1\}, \\ \delta_2 &= \{(x_1, 0) : 0 < x_1 < 1\}, \\ \delta_e &= \{(x_1, x_2) : x_j \geq 0, x_1 + x_2 = 1\}, \\ \delta_0 &= \{(0, 0)\}, \end{aligned}$$

and this is sketched in Figure 1.

Using these definitions, we can define the first time that the process hits level N in a busy cycle as

$$\tau_N = \inf\{i > 0 : X_i \in \delta_e, \mathbf{X}_k \notin \delta_0 \forall k = 1, \dots, i-1\}$$

and we set $\tau_N = \infty$ if the process hits δ_0 before δ_e . Therefore, our probability of interest can be written as

$$p_N = \mathbb{P}(\tau_N < \infty \mid \mathbf{X}_0 = (\frac{1}{N}, 0))$$

It is known that the asymptotic decay rate of this probability is given by

$$\begin{aligned} \gamma &\doteq -\lim_{N \rightarrow \infty} \frac{1}{N} \log p_N = -\log \left(\frac{\lambda}{\min\{\mu_1, \mu_2\}} \right) \\ &= \min\{\gamma_1, \gamma_2\}, \end{aligned} \quad (1)$$

where $\gamma_j = -\log(\lambda/\mu_j)$, see Glasserman and Kou² and Buijsrogge et al.¹⁴ The queue for which we have $\gamma_j = \gamma$ is called the *bottleneck queue*. For the $M|M|1$ tandem queue, the bottleneck queue is equivalent to the queue with the largest server utilization. In most papers on similar topics, see, for example, Glasserman and Kou,² de Boer,³ Dupuis et al.⁴ and de Boer and Scheinhardt,¹¹ it is assumed that $\mu_2 \leq \mu_1$, as for the probability of interest the queues are

interchangeable.¹³ In this paper, we both consider $\mu_2 \leq \mu_1$ and $\mu_1 \leq \mu_2$.

2.2. Importance sampling simulation

To estimate our probability of interest using simulation, we use importance sampling. In importance sampling, we perform our simulation under some new measure \mathbb{Q} . Under this new measure, we let $\bar{\Theta}(\mathbf{v}_k | \mathbf{x})$ denote the probability for transition \mathbf{v}_k given that we are in state \mathbf{x} . While doing the simulation, we keep track of the likelihood ratio $L(\mathcal{P})$ of a path $\mathcal{P} = (\mathbf{X}_i, i = 0, \dots, \tau_N)$:

$$L(\mathcal{P}) = \prod_{i=0}^{\tau_N-1} \frac{\Theta(\mathbf{Y}_i)}{\bar{\Theta}(\mathbf{Y}_i | \mathbf{X}_i)}, \quad (2)$$

where $\mathbf{Y}_i = (\mathbf{X}_{i+1} - \mathbf{X}_i)N$ if $\mathbf{X}_{i+1} \neq \mathbf{X}_i$ and $\mathbf{Y}_i = \mathbf{v}_k$ if $\mathbf{X}_{i+1} = \mathbf{X}_i$ and $\mathbf{X}_i \in \delta_k$, $k = 1, 2$, to include the self-loop transition when one of the queues is empty. Let $I(\mathcal{P}) = \mathbb{1}\{\tau_N < \infty\}$ indicate whether we have reached our event of interest during a busy cycle of the system or not. Then, under the new measure \mathbb{Q} , we have

$$p_N = \mathbb{E}^{\mathbb{Q}}[L(\mathcal{P})I(\mathcal{P})],$$

where $\mathbb{E}^{\mathbb{Q}}$ denotes the expectation under the new measure \mathbb{Q} . Thus, $L(\mathcal{P})I(\mathcal{P})$ is the estimator for our probability of interest.

As in the work of Dupuis et al.,⁴ we construct a subsolution $W(\mathbf{x})$ – formally defined in Section 2.3 – and we use this function to specify a change of measure in the following way:

$$\bar{\Theta}(\mathbf{v}_k | \mathbf{x}) = \Theta(\mathbf{v}_k) e^{-\langle DW(\mathbf{x}), \mathbf{v}_k \rangle} e^{\mathbb{H}(\mathbf{x}, DW(\mathbf{x}))}, \quad (3)$$

where $DW(\mathbf{x})$ denotes the gradient of $W(\mathbf{x})$ with respect to \mathbf{x} , and where:

$$\mathbb{H}(\mathbf{x}, DW(\mathbf{x})) = -\log \left(\sum_{k=0}^2 \Theta(\mathbf{v}_k) e^{-\langle DW(\mathbf{x}), \mathbf{v}_k \rangle} \right). \quad (4)$$

If we compare Equation (4) with the corresponding notation in the works of Dupuis et al.⁴ and de Boer and Scheinhardt,¹¹ a factor 2 is missing. However, we will also scale the function $W(\mathbf{x})$ accordingly, so that the change of measure remains the same.

As we are interested in finding a change of measure that gives an asymptotically efficient estimator for p_N , we need

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}^{\mathbb{Q}}[L(\mathcal{P})^2 I(\mathcal{P})] \leq -2\gamma,$$

or equivalently:

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}[L(\mathcal{P})I(\mathcal{P})] \leq -\gamma,$$

to hold.

In previous works,^{4,6,11} a subsolution is constructed (assuming $\mu_2 \leq \mu_1$ in the works of Dupuis et al.⁴ and de Boer and Scheinhardt¹¹) and asymptotic efficiency is proven for a specific choice of $W(\mathbf{x})$. In this paper, we provide conditions on $W(\mathbf{x})$ and prove that under these conditions, we get an asymptotically efficient estimator. Afterwards, we study the possibilities of the function $W(\mathbf{x})$, and the corresponding change of measure in Equation (3).

2.3. Subsolution approach

In both Dupuis et al.⁴ and Dupuis and Wang,⁶ the change of measure for (two-node) tandem queues (and Jackson networks) has been studied. In those papers, the change of measure has been determined using subsolutions. We first briefly recap the ideas presented in those papers and we start with a formal definition of a subsolution, see also Dupuis et al.⁴

Definition 2.1. A function $W(\mathbf{x})$ is a classical subsolution if:

1. W is continuously differentiable,
2. $\mathbb{H}(\mathbf{x}, DW(\mathbf{x})) \geq 0$ for every $\mathbf{x} \in D \cup \delta_1 \cup \delta_2 \cup \delta_e \cup \delta_0$,
3. $W(\mathbf{x}) \leq 0$ for $\mathbf{x} \in \delta_e$.

In Dupuis et al.,⁴ there is an additional condition in this definition that needs to hold at the boundaries of the state space. Instead, we include the boundaries of the state space in $\mathbb{H}(\mathbf{x}, DW(\mathbf{x}))$, i.e., $\mathbb{H}(\mathbf{x}, DW(\mathbf{x}))$ differs along the boundaries (which is similar as in the work of Dupuis and Wang⁶).

It is known from Glasserman and Kou² and de Boer³ that for an asymptotically efficient change of measure it is not possible that $DW(\mathbf{x})$ is constant throughout the whole state space as in the work of Parekh and Walrand,¹ and in the work of Dupuis et al.⁴ this is “confirmed,” since the change of measure yielding an asymptotically efficient estimator differs from the change of measure in the work of Parekh and Walrand¹ along one of the boundaries of the state space and near the origin. Thus, to find an asymptotically efficient change of measure, we determine several (constant) changes of measure for various regions of the state space, in particular along the boundaries of the state space, and combine these so that we have a change of measure for the whole state space.

To determine such a change of measure that differs along various parts of the state space, in the work of Dupuis et al.,⁴ there are multiple – say r – affine functions $W_k^\delta(\mathbf{x})$, $k = 1, \dots, r$, considered such that for each of these functions $\mathbb{H}(\mathbf{x}, DW_k^\delta(\mathbf{x})) \geq 0$ for some part of the state space, so that all r functions cover the whole state space, and for at least one of these functions, we have

$W_k^\delta(\mathbf{x}) \leq 0$ for $\mathbf{x} \in \delta_e$. All these functions have the following form:

$$W_k^\delta(\mathbf{x}) = \langle \alpha_k, \mathbf{x} \rangle + c_k - d_k \delta, \quad (5)$$

where $\alpha_k = (\alpha_{k,1}, \alpha_{k,2})$, which we assume to be finite, specifies the gradient of $W_k^\delta(\mathbf{x})$ and both c_k and $d_k > 0$ are constants. Combining these functions to $W^\delta(\mathbf{x})$ by taking the minimum of these functions for all \mathbf{x} , then results in a piecewise affine function, that unfortunately is not continuously differentiable. That is, we have the following:

$$W^\delta(\mathbf{x}) = W_1^\delta(\mathbf{x}) \wedge \dots \wedge W_r^\delta(\mathbf{x}).$$

To satisfy the continuous differentiability, which is the first requirement for a function to be a classical subsolution, the functions $W_k^\delta(\mathbf{x})$ are combined into a (continuous) function $W^{\varepsilon, \delta}(\mathbf{x})$ by a similar mollification procedure as in the works of Dupuis et al.⁴ and Dupuis and Wang,⁵ where in this case we obtain the following:

$$W^{\varepsilon, \delta}(\mathbf{x}) = -\varepsilon \log \sum_{k=1}^r e^{-W_k^\delta(\mathbf{x})/\varepsilon}, \quad (6)$$

such that $W^{\varepsilon, \delta}(\mathbf{x})$ converges to $W^\delta(\mathbf{x})$ when $\varepsilon \rightarrow 0$. Throughout this paper, we make the following assumptions on ε and δ (that depend on N), which also have been made in previous works.^{4,6,11} For convenience, we do not write the explicit dependence of ε and δ on N .

Assumption 2.1. We choose ε and δ dependent on N , such that:

1. $\lim_{N \rightarrow \infty} \varepsilon = 0$;
2. $\lim_{N \rightarrow \infty} \delta = 0$;
3. $\lim_{N \rightarrow \infty} \varepsilon N = \infty$;
4. $\lim_{N \rightarrow \infty} \frac{\varepsilon}{\delta} = 0$.

The gradient of Equation (6) is then used as change of measure in Equation (3). It can be expressed as follows:

$$DW^{\varepsilon, \delta}(\mathbf{x}) = \sum_{k=1}^r \rho_k(\mathbf{x}) \alpha_k, \quad (7)$$

$$\text{where } \rho_k(\mathbf{x}) = \frac{e^{-W_k^\delta(\mathbf{x})/\varepsilon}}{\sum_{j=1}^r e^{-W_j^\delta(\mathbf{x})/\varepsilon}}.$$

The functions $\rho_k(\mathbf{x})$ are weight factors for the “influence” of each function $W_k^\delta(\mathbf{x})$, and so of each different “regional” change of measure, in the final change of measure. They can also be used to define a change of measure slightly different than in Equation (3) as follows, see also Dupuis et al.⁴ and de Boer and Scheinhardt¹¹:

$$\bar{\Theta}(\mathbf{v}_i | \mathbf{x}) = \Theta(\mathbf{v}_i) \sum_{k=1}^r \rho_k(\mathbf{x}) e^{-\langle \alpha_k, \mathbf{v}_i \rangle} e^{\mathbb{H}(\mathbf{x}, \alpha_k)}, \quad (8)$$

which we will refer to in this paper later on. In fact, we will also show that asymptotic efficiency for the change of measure in Equation (3) implies asymptotic efficiency for the change of measure in Equation (8), similar as in the work of de Boer and Scheinhardt.¹¹ We note that, from an implementation perspective, the change of measure in Equation (8) is preferred over the change of measure in Equation (3), see also Section 3.8.6 in the work of Dupuis et al.⁴

2.4. Existing changes of measure

Now that the general ideas of Dupuis et al.⁴ and Dupuis and Wang⁶ have been presented, we will show the different functions $W^\delta(\mathbf{x})$ that have been used in the works of Dupuis et al.⁴ and Dupuis and Wang⁶ to obtain an asymptotically efficient change of measure, as well as figures of the function $W^\delta(\mathbf{x})$ for both bottleneck queues j , in the following sections.

2.4.1. Change of measure from Dupuis et al.⁴ In the work of Dupuis et al.,⁴ queue 2 is always considered to be the bottleneck queue because for the probability of interest, the queues are interchangeable. In that paper, a subsolution $W^\delta(\mathbf{x})$ is determined by considering three functions $W_k^\delta(\mathbf{x})$, $k = 1, 2, 3$. The three functions from Dupuis et al.⁴ are as follows:

$$\begin{aligned} W_1^\delta(\mathbf{x}) &= -\gamma_2 x_1 - \gamma_2 x_2 + \gamma_2 - \delta, \\ W_2^\delta(\mathbf{x}) &= -\gamma_2 x_1 + \gamma_2 - 2\delta, \\ W_3^\delta(\mathbf{x}) &= \gamma_2 - 3\delta, \end{aligned} \quad (9)$$

and the function $W^\delta(\mathbf{x})$ is illustrated in Figure 2. We remark that also in the functions in Equation (9), we scaled the results from Dupuis et al.⁴ (except for the constant in front of δ) by a factor $1/2$, but this does not influence the resulting changes of measure in Equations (3) and (8).

2.4.2. Changes of measure from Dupuis and Wang.⁶ In the work of Dupuis and Wang,⁶ the work of Dupuis et al.⁴ is extended to Jackson networks and hence in the work of Dupuis and Wang,⁶ all queues being the bottleneck queue are considered, as in this paper. Not only the probability that the total number of customers in the system reaches some high level N during a busy cycle of the system is considered in the work of Dupuis and Wang,⁶ but the authors also consider buffer overflow in a single queue or in several queues at the same time. If we consider a two-node tandem queue with queue 2 being the bottleneck queue, we find that the following four functions are used in Dupuis and Wang⁶:

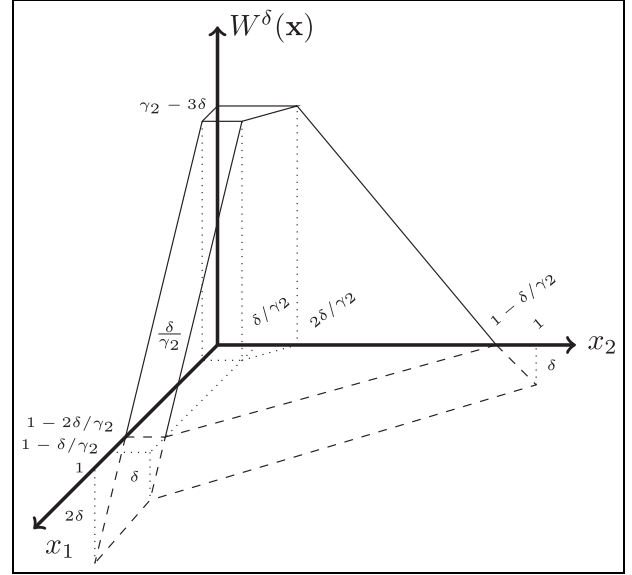


Figure 2. The function $W^\delta(\mathbf{x})$ from Equation (9). Queue 2 is the bottleneck queue (so $\gamma_2 \leq \gamma_1$).

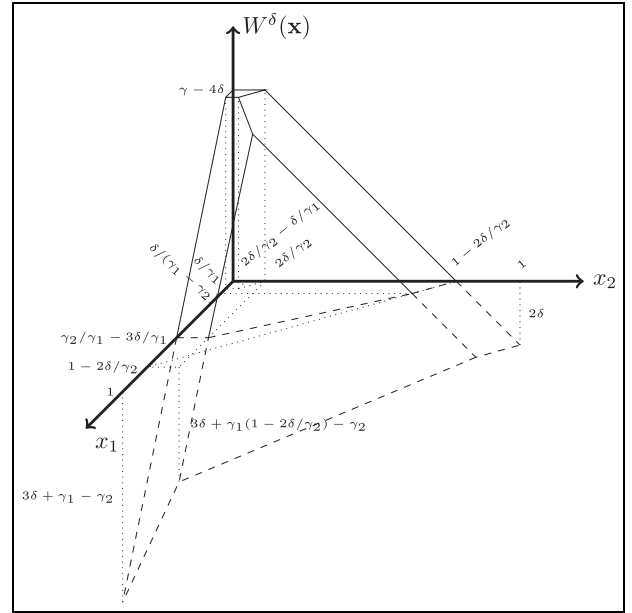


Figure 3. The function $W^\delta(\mathbf{x})$ from Equation (10). Queue 2 is the bottleneck queue (so $\gamma_2 \leq \gamma_1$).

$$\begin{aligned} W_1^\delta(\mathbf{x}) &= -\gamma_1 x_1 - \gamma_2 x_2 + \gamma_2 - \delta, \\ W_2^\delta(\mathbf{x}) &= -\gamma_2 x_1 - \gamma_2 x_2 + \gamma_2 - 2\delta, \\ W_3^\delta(\mathbf{x}) &= -\gamma_1 x_1 + \gamma_2 - 3\delta, \\ W_4^\delta(\mathbf{x}) &= \gamma_2 - 4\delta, \end{aligned} \quad (10)$$

and $W^\delta(\mathbf{x})$ is illustrated in Figure 3.

In the work of Dupuis and Wang,⁶ the authors do not explicitly mention by which constant δ is multiplied,

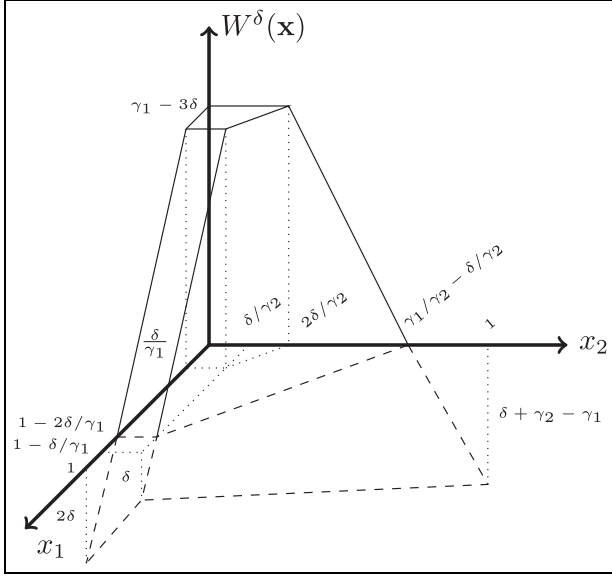


Figure 4. The function $W^\delta(\mathbf{x})$ from Equation(11). Queue 1 is the bottleneck queue (so $\gamma_1 \leq \gamma_2$).

though implicit requirements with a proof for asymptotic efficiency are given. For simplicity, we use $k\delta$ throughout this paper. It turns out that these values are sufficient for asymptotic efficiency, as we show later in this paper, but we will also see that they are by no means unique.

Since there is no limitation to queue 2 being the bottleneck queue in Dupuis and Wang,⁶ we also present the result from Dupuis and Wang⁶ when queue 1 is the bottleneck queue. The result from Dupuis and Wang,⁶ when queue 1 is the bottleneck queue, is different compared with when queue 2 is the bottleneck queue, even though for the probability of interest both queues are interchangeable. When queue 1 is the bottleneck queue, the following three functions are derived by Dupuis and Wang⁶:

$$\begin{aligned} W_1^\delta(\mathbf{x}) &= -\gamma_1 x_1 - \gamma_2 x_2 + \gamma_1 - \delta, \\ W_2^\delta(\mathbf{x}) &= -\gamma_1 x_1 + \gamma_1 - 2\delta, \\ W_3^\delta(\mathbf{x}) &= \gamma_1 - 3\delta, \end{aligned} \quad (11)$$

and the resulting function $W^\delta(\mathbf{x})$ is illustrated in Figure 4. These subsolutions are very similar to the ones in Equation (10), but with one function less. Also here, the multiplication factor of δ is not explicitly mentioned, but we use the values above which are sufficient for asymptotic efficiency, as we will show later in this paper.

2.4.3. Comparison of the existing changes of measure. In this section, we briefly comment on the similarities and differences of the changes of measure from Dupuis et al.⁴ and

Dupuis and Wang.⁶ We will do so by comparing the functions $W^\delta(\mathbf{x})$ for all different cases (see Figures 2–4).

When queue 2 is the bottleneck queue, we see from Figures 2 and 3 that along the x_2 -axis the function $W^\delta(\mathbf{x})$ is roughly the same. The only difference is that in Figure 3 the function, along the x_2 -axis, is δ lower. In particular, this part of the state space covers the most likely path. In all other parts of the state space, the function $W^\delta(\mathbf{x})$ in Figure 3 is slightly steeper than in Figure 2 since $\gamma_2 < \gamma_1$. These observations suggest that any change of measure based on some function $W^\delta(\mathbf{x})$ that somehow lies “in between” the functions in Figures 2 and 3 is also asymptotically efficient. In Section 4, we show that this is indeed the case.

When queue 1 is the bottleneck queue, there is not much to compare. However, since there are already two possibilities for the change of measure based on subsolutions to be asymptotically efficient, and even more to expect, when queue 2 is the bottleneck queue, also the case when queue 1 is the bottleneck queue is studied in Section 4.

3. Sufficient conditions for asymptotic efficiency

Similar to Dupuis et al.⁴ and Dupuis and Wang,⁶ the construction of the changes of measure in this paper is based on finding appropriate subsolutions $W^{e,\delta}(\mathbf{x})$. We start with a general proof for asymptotic efficiency for a change of measure based on the subsolution approach and the mollification procedure that is explained in Section 2.3. In our theorem, we provide sufficient conditions for the subsolution yielding an asymptotic efficient change of measure, which we use later for the derivation of the possibilities for the change of measure. Afterwards, we discuss some general observations with respect to some of the conditions in our theorem.

3.1. Main result

Theorem 3.1. Consider a two-node $M|M|1$ tandem queue. Let $W_k^\delta(\mathbf{x})$ be affine functions for all $k = 1, \dots, r$, as in Equation (5), and let the classical subsolution $W^{e,\delta}(\mathbf{x})$ be constructed using Equation (6). Then, using the gradient of the function $W^{e,\delta}(\mathbf{x})$ as change of measure in Equations (3) and (8) results in an asymptotically efficient estimator if there exist functions $f(N)$, $g(N)$, and $h(N)$ with $\lim_{N \rightarrow \infty} f(N) = \lim_{N \rightarrow \infty} g(N) = \lim_{N \rightarrow \infty} h(N) = 0$, such that:

1. $\sum_{k=1}^r \rho_k(\mathbf{x}) \mathbb{H}(\mathbf{x}, \boldsymbol{\alpha}_k) \geq f(N)$;
2. $W^{e,\delta}(\mathbf{x}) \leq g(N)$, for all $\mathbf{x} \in \delta_e$;
3. $W^{e,\delta}(\mathbf{0}) \geq \gamma + h(N)$.

Proof. We start by showing that under the above conditions the change of measure in Equation (3) is asymptotically efficient, after which it follows that also the change of measure in Equation (8) is asymptotically efficient using a similar argument as in Theorem 2 from de Boer and Scheinhardt.¹¹ \square

From Equations (2) and (3), it follows that the likelihood ratio of a path $L(\mathcal{P})$ is given by the following:

$$\log L(\mathcal{P}) = N \sum_{j=0}^{\tau_N-1} \langle DW^{e,\delta}(\mathbf{X}_j), \mathbf{X}_{j+1} - \mathbf{X}_j \rangle - \sum_{j=0}^{\tau_N-1} \mathbb{H}(\mathbf{X}_j, DW^{e,\delta}(\mathbf{X}_j)).$$

We find, using Equation (7), that for all states \mathbf{x} we have the following:

$$\mathbb{H}(\mathbf{x}, DW^{e,\delta}(\mathbf{x})) \geq \sum_{k=1}^r \rho_k(\mathbf{x}) \mathbb{H}(\mathbf{x}, \boldsymbol{\alpha}_k),$$

due to concavity of $\mathbb{H}(\mathbf{x}, DW^{e,\delta}(\mathbf{x}))$ in its second argument, see Proposition 3.2 in the work of Dupuis et al.⁴ Combining the two expressions above, we arrive at the following:

$$\log L(\mathcal{P}) \leq N \sum_{j=0}^{\tau_N-1} \langle DW^{e,\delta}(\mathbf{X}_j), \mathbf{X}_{j+1} - \mathbf{X}_j \rangle - \sum_{j=0}^{\tau_N-1} \sum_{k=1}^r \rho_k(\mathbf{X}_j) \mathbb{H}(\mathbf{X}_j, \boldsymbol{\alpha}_k) \quad (12)$$

$$\leq N \sum_{j=0}^{\tau_N-1} \langle DW^{e,\delta}(\mathbf{X}_j), \mathbf{X}_{j+1} - \mathbf{X}_j \rangle - f(N)\tau_N, \quad (13)$$

where the last inequality follows from Condition 1.

Similar to Lemma 2 in the work of de Boer and Scheinhardt,¹¹ also when using r regions, we can obtain the following bound. The idea of this Lemma in de Boer and Scheinhardt¹¹ is to replace the summation in Equation (13) by $W^{e,\delta}(\mathbf{X}_{\tau_N}) - W^{e,\delta}(\mathbf{0})$ and to give an upper bound on the error that is introduced. Thus, since $\boldsymbol{\alpha}_k$ is finite by construction, and hence $|\boldsymbol{\alpha}_k| \leq c$ for some $0 \leq c < \infty$, we find the following:

$$\left| N \sum_{j=0}^{\tau_N-1} \langle DW^{e,\delta}(\mathbf{X}_j), \mathbf{X}_{j+1} - \mathbf{X}_j \rangle - N(W^{e,\delta}(\mathbf{X}_{\tau_N}) - W^{e,\delta}(\mathbf{0})) \right| \leq \frac{5c^2}{\varepsilon N} \tau_N. \quad (14)$$

Next, we follow similar steps as in Theorem 1 of the same paper. By combining Equations (13) and (14) we have the following:

$$\begin{aligned} \log L(\mathcal{P}) &\leq N(W^{e,\delta}(\mathbf{X}_{\tau_N}) - W^{e,\delta}(\mathbf{0})) + \frac{5c^2}{\varepsilon N} \tau_N - f(N)\tau_N \\ &\leq (g(N) - h(N) - \gamma)N + \left(\frac{5c^2}{\varepsilon N} - f(N) \right) \tau_N, \end{aligned}$$

where the second inequality follows from Conditions 2 and 3 when $\mathbf{X}_{\tau_N} \in \delta_e$, and thus, we find, as in de Boer and Scheinhardt,¹¹ the following:

$$\begin{aligned} &\frac{1}{N} \log \mathbb{E}[L(\mathcal{P})I(\mathcal{P})] \\ &= \frac{1}{N} \log(\mathbb{E}[L(\mathcal{P})|I(\mathcal{P}) = 1] \cdot \mathbb{P}(I(\mathcal{P}) = 1)) \\ &\leq \frac{1}{N} \log \left(\mathbb{E} \left[e^{(g(N)-h(N)-\gamma)N + \left(\frac{5c^2}{\varepsilon N} - f(N) \right) \tau_N} \mid \tau_N < \infty \right] p_N \right) \\ &= g(N) - h(N) - \gamma + \frac{1}{N} \log \mathbb{E} \left[e^{\left(\frac{5c^2}{\varepsilon N} - f(N) \right) \tau_N} \mid \tau_N < \infty \right] \\ &\quad + \frac{1}{N} \log p_N. \end{aligned} \quad (15)$$

To conclude the proof, we need Lemma 3 from de Boer and Scheinhardt,¹¹ which states that for any sequence $\theta_N \geq 0$ such that $\lim_{N \rightarrow \infty} \theta_N = 0$, we have the following:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E} \left[e^{\theta_N \tau_N} \mid \tau_N < \infty \right] = 0. \quad (16)$$

Thus, taking limits in Equation (15) gives the following:

$$\begin{aligned} &\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}[L(\mathcal{P})I(\mathcal{P})] \\ &\leq \limsup_{N \rightarrow \infty} \left(g(N) - h(N) - \gamma \right. \\ &\quad \left. + \frac{1}{N} \log \mathbb{E} \left[e^{\left(\frac{5c^2}{\varepsilon N} - f(N) \right) \tau_N} \mid \tau_N < \infty \right] + \frac{1}{N} \log p_N \right) \\ &= -2\gamma, \end{aligned} \quad (17)$$

where the last equation follows using Equation (1), $\lim_{N \rightarrow \infty} g(N) = \lim_{N \rightarrow \infty} h(N) = 0$ and since $\left(\frac{5c^2}{\varepsilon N} - f(N) \right) \rightarrow 0$ when $N \rightarrow \infty$, we can apply Equation (16) to the fourth term of Equation (17). This concludes the proof for the change of measure in Equation (3).

For the change of measure in Equation (8), we note that similar to de Boer and Scheinhardt,¹¹ we have the following:

$$\begin{aligned}
\log L(\mathcal{P}) &= \log \prod_{j=0}^{\tau_N-1} \frac{1}{\sum_{k=1}^r \rho_k(\mathbf{X}_j) e^{-\langle \alpha_k, \mathbf{X}_{j+1} - \mathbf{X}_j \rangle} e^{\mathbb{H}(\mathbf{X}_j, \alpha_k)}} \\
&= - \sum_{j=0}^{\tau_N-1} \log \left(\sum_{k=1}^r \rho_k(\mathbf{X}_j) e^{-\langle \alpha_k, \mathbf{X}_{j+1} - \mathbf{X}_j \rangle} e^{\mathbb{H}(\mathbf{X}_j, \alpha_k)} \right) \\
&\leq \sum_{j=0}^{\tau_N-1} \sum_{k=1}^r \rho_k(\mathbf{X}_j) \langle \alpha_k, \mathbf{X}_{j+1} - \mathbf{X}_j \rangle \\
&\quad - \sum_{k=1}^r \rho_k(\mathbf{X}_j) \mathbb{H}(\mathbf{X}_j, \alpha_k) \\
&= \sum_{j=0}^{\tau_N-1} \langle DW^{e,\delta}(\mathbf{X}_j), \mathbf{X}_{j+1} - \mathbf{X}_j \rangle \\
&\quad - \sum_{k=1}^r \rho_k(\mathbf{X}_j) \mathbb{H}(\mathbf{X}_j, \alpha_k),
\end{aligned}$$

where the inequality follows by concavity of the logarithm and the last equality follows by definition of $DW^{e,\delta}(\mathbf{x})$, see Equation (7). Thus, we have the same bound as in Equation (12) and so we also find that the change of measure in Equation (8) is asymptotically efficient. \square

Remark 3.1. In Equation (17), we see that we end up with some term $g(N) - h(N) - \gamma$, which goes to $-\gamma$ as $N \rightarrow \infty$. This term arises from bounding $W^{e,\delta}(\mathbf{X}_{\tau_N}) - W^{e,\delta}(\mathbf{0})$ on a path that leads to the overflow level. However, we cannot have $W^{e,\delta}(\mathbf{X}_{\tau_N}) - W^{e,\delta}(\mathbf{0}) < -\gamma$ when $N \rightarrow \infty$, since this would contradict with Jensen's inequality:

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}[L(\mathcal{P})I(\mathcal{P})] \geq -2\gamma.$$

That is, it is impossible to obtain a tighter bound. As a result, we find that for an asymptotically efficient change of measure we need $W^{e,\delta}(\mathbf{X}_{\tau_N}) - W^{e,\delta}(\mathbf{0}) \rightarrow -\gamma$ when $N \rightarrow \infty$ on at least one path that leads to reaching the overflow level, e.g., the most likely path, see also Dupuis and Wang.⁶ Thus, on such a path, we need $\lim_{N \rightarrow \infty} W^{e,\delta}(\mathbf{x}) = 0$ for $\mathbf{x} \in \delta_e$ and $\lim_{N \rightarrow \infty} W^{e,\delta}(\mathbf{0}) = \gamma$.

Remark 3.2. In the sequel, we use a slightly stronger condition than Condition 1 of Theorem 3.1, namely that for each k separately, we have $\rho_k(\mathbf{x}) \mathbb{H}(\mathbf{x}, \alpha_k) \geq f_k(N)$, where $\lim_{N \rightarrow \infty} f_k(N) = 0$.

Remark 3.3. It seems likely that Theorem 3.1 can also be extended to a d -node $M|M|1$ tandem queue (and Jackson networks), with the same sufficient conditions as in the current statement for $d = 2$. To do so, observe that in the

proof of Theorem 3.1, Lemma 3 from de Boer and Scheinhardt¹¹ is the only part that restricts to $d = 2$ (and hence to tandem queues). Thus, one could either extend this result to $d > 2$ (and Jackson networks), or use similar techniques as in the works of Dupuis et al.⁴ and Dupuis and Wang⁶ in order to show that the theorem holds in a more general setting.

3.2. General observations

Now that we have shown under which conditions we obtain an asymptotically efficient change of measure based on subsolutions, it remains to find α_k , c_k and d_k for all $k = 1, \dots, r$ such that Conditions 1, 2, and 3 of Theorem 3.1 are satisfied.

In this section, we make some general observations with respect to Conditions 1 and 3 of Theorem 3.1 when considering a two-node $M|M|1$ tandem queue, that are used later to construct the possibilities for the change of measure. These observations are independent of the bottleneck queue.

3.2.1. Observations with respect to Condition 1 of Theorem 3.1. We recall that the first condition is $\sum_{k=1}^r \rho_k(\mathbf{x}) \mathbb{H}(\mathbf{x}, \alpha_k) \geq f(N)$ for some $f(N)$ with $\lim_{N \rightarrow \infty} f(N) = 0$. In this section, we state some observations with respect to $\mathbb{H}(\mathbf{x}, \alpha)$ for some general α , independent of k , after which we present some observations with respect to $\rho_k(\mathbf{x})$.

By considering all possibilities for $\mathbb{1}\{x_j > 0\}$, $j = 1, 2$, in a busy cycle of the system, we find the following from Equation (4)

$$\begin{aligned}
\mathbb{H}(\mathbf{x}, \alpha) &= \\
&\begin{cases} -\log(\lambda e^{-\alpha_1} + \mu_1 e^{\alpha_1 - \alpha_2} + \mu_2 e^{\alpha_2}) & \text{if } x_1 > 0, x_2 > 0, \\ -\log(\lambda e^{-\alpha_1} + \mu_1 e^{\alpha_1 - \alpha_2} + \mu_2) & \text{if } x_1 > 0, x_2 = 0, \\ -\log(\lambda e^{-\alpha_1} + \mu_1 + \mu_2 e^{\alpha_2}) & \text{if } x_1 = 0, x_2 > 0. \end{cases}
\end{aligned} \tag{18}$$

We start by finding solutions to $\mathbb{H}(\mathbf{x}, \alpha) \geq 0$, for particular parts of the state space, or equivalently to the following:

$$\lambda e^{-\alpha_1} + \mu_1 e^{(\alpha_1 - \alpha_2) \mathbb{1}\{x_1 > 0\}} + \mu_2 e^{\alpha_2 \mathbb{1}\{x_2 > 0\}} \leq 1. \tag{19}$$

In Lemma 3.2, we consider $x_j > 0$ for $j = 1, 2$. We fix either α_1 or α_2 , and we show for which values of α_2 or α_1 Equation (19) holds. In Lemma 3.3, we state two relations considering Equation (19), at one of the boundaries (either $x_1 = 0$ or $x_2 = 0$) that will be used several times later in this paper. In Sections 4.1–4.4, we use the results of these Lemmas to lower bound $\sum_{k=1}^r \rho_k(\mathbf{x}) \mathbb{H}(\mathbf{x}, \alpha_k)$, such that Condition 1 of Theorem 3.1 is satisfied.

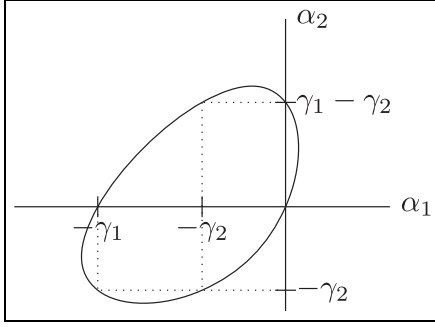


Figure 5. Sketch of the level set for which $\mathbb{H}(\mathbf{x}, \boldsymbol{\alpha}) = 0$ (i.e., Equation (19) holds with equality) for $\mathbf{x} > 0$, when queue 2 is the bottleneck queue (and so $\gamma_2 \leq \gamma_1$).

Lemma 3.2. Suppose $x_j > 0$ for $j = 1, 2$. We prove the following statements:

- If $\alpha_2 = -\gamma_2$, then Equation (19) holds iff $\alpha_1 \in [-\max_j \gamma_j, -\min_j \gamma_j]$;
- If $\alpha_1 = -\gamma_1$, then Equation (19) holds iff $\alpha_2 \in [-\gamma_2, 0]$;
- If $\alpha_1 = -\gamma_2$, then Equation (19) holds iff $\alpha_2 \in [-\gamma_2, \gamma_1 - \gamma_2]$;
- If $\alpha_1 = 0$, then Equation (19) holds iff $\alpha_2 \in [0, \gamma_1 - \gamma_2]$.

Proof. For the first statement, let $\alpha_2 = -\gamma_2$ and $x_j > 0$ for $j = 1, 2$. Then, Equation (19) reduces to $\lambda e^{-\alpha_1} + (\mu_1 \mu_2 / \lambda) e^{\alpha_1} + \lambda \leq 1$ which holds if and only if:

$$(1 - \mu_1 - \mu_2)^2 x^2 - (\mu_1 + \mu_2)(1 - \mu_1 - \mu_2)x + \mu_1 \mu_2 \leq 0.$$

where we let $x = e^{-\alpha_1}$. Using elementary calculus we find that $\alpha_1 \in [-\max_j \gamma_j, -\min_j \gamma_j]$. The other statements follow similarly. \square

As a result of Lemma 3.2, we can sketch the level set for all α_1, α_2 such that $\mathbb{H}(\mathbf{x}, \boldsymbol{\alpha}) = 0$ for all $\mathbf{x} > 0$, using the concavity of the function shown in the work of Dupuis et al.⁴ see Figures 5 and 6.

In the following lemma, we consider Equation (19) at one of the boundaries given a certain choice for either α_1 or α_2 . The first equation considers $x_1 > 0$, $x_2 = 0$ and $\alpha_1 = -\gamma_1$, and the second equation considers $x_1 = 0$, $x_2 > 0$ and $\alpha_2 = -\gamma_2$.

Lemma 3.3. $\mu_1 + \lambda e^{-\alpha_2} + \mu_2 \leq 1$ iff $\alpha_2 \geq 0$, and $\lambda e^{-\alpha_1} + \mu_1 + \lambda \leq 1$ iff $\alpha_1 \geq -\gamma_2$.

Proof. The statements follow directly by elementary calculus, combined with $\lambda + \mu_1 + \mu_2 = 1$ and $\lambda > 0$. \square

We conclude this section with a remark on $\rho_k(\mathbf{x})$. Using Equation (7) we find the following:

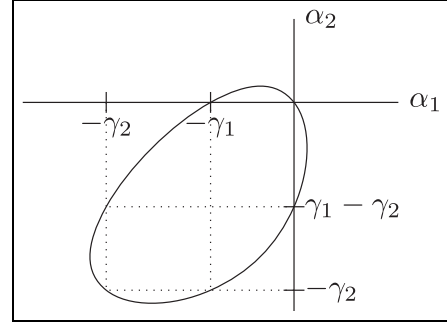


Figure 6. Similar to Figure 5, but when queue 1 is the bottleneck queue (and so $\gamma_1 \leq \gamma_2$).

$$\rho_k(\mathbf{x}) = \frac{e^{-W_k^\delta(\mathbf{x})/\varepsilon}}{\sum_{i=1}^r e^{-W_i^\delta(\mathbf{x})/\varepsilon}} \leq \frac{e^{-W_k^\delta(\mathbf{x})/\varepsilon}}{e^{-W_\ell^\delta(\mathbf{x})/\varepsilon}} = e^{(W_\ell^\delta(\mathbf{x}) - W_k^\delta(\mathbf{x}))/\varepsilon}, \quad (20)$$

for any ℓ , where the inequality follows trivially.

3.2.2. Observations with respect to Condition 3 of Theorem 3.1. In Remark 3.1, it is noted that we need $\bar{h}(N) \leq W^{\varepsilon, \delta}(\mathbf{0}) - \gamma \leq h(N)$ such that $\lim_{N \rightarrow \infty} h(N) = \lim_{N \rightarrow \infty} \bar{h}(N) = 0$. This is satisfied when $W_k^\delta(\mathbf{0}) - \gamma \geq \bar{h}_k(N)$, such that $\lim_{N \rightarrow \infty} \bar{h}_k(N) = 0$ for all k and $W_k^\delta(\mathbf{0}) - \gamma \leq h_k(N)$, such that $\lim_{N \rightarrow \infty} h_k(N) = 0$ for some k , see Equation (6). By Equation (5), we have $W_k^\delta(\mathbf{0}) = c_k - d_k \delta$. Therefore, we find the following:

$$c_k = \gamma \quad \text{for all } k,$$

as a sufficient condition to satisfy Condition 3 since $\delta \rightarrow 0$ when $N \rightarrow \infty$ by Assumption 2.1.

As a result of the observations above, to construct the possibilities for the change of measure, it remains to find α_k and d_k for all $k = 1, \dots, r$ that satisfy Conditions 1, 2, and 3 of Theorem 3.1. This is the topic of the next section.

4. Construction of the subsolution $W^{\varepsilon, \delta}(\mathbf{x})$

In this section, we construct possible subsolutions, based on the approach mentioned in Section 2.3, that satisfy the conditions in Theorem 3.1 and thus yield an asymptotically efficient estimator. It may not be clear at first sight that the method below results in subsolutions that satisfy all these conditions, since the construction is partly based on intuition. However, we conclude all sections by showing that the conditions are indeed satisfied.

For the two-node $M|M|1$ tandem queue, we consider both possibilities for the bottleneck queue, i.e., we consider both queue 1 and queue 2 as bottleneck queue. In addition, we focus on a maximum of four different regions. More

regions may or may not be possible. However, this is undesirable from a practical point of view and does not contribute to an easier implementation of the change of measure.

We start using three regions and queue 2 being the bottleneck queue, since this case has been studied most in literature. Afterwards, we consider three regions and queue 1 the bottleneck queue. We conclude this section with four regions, for which we again consider both queue 2 and queue 1 as the bottleneck, respectively. For brevity, when considering four regions, we will only state the result of the construction (which is similar to the construction when having three regions) and show that indeed the conditions in Theorem 3.1 are satisfied.

4.1. Three regions and queue 2 bottleneck

In this section, our starting point is to consider three functions $W_k^\delta(\mathbf{x})$, $k = 1, 2, 3$, and we let queue 2 be the bottleneck queue, thus $\gamma_2 \leq \gamma_1$. As the boundaries turned out to be crucial in designing an asymptotically efficient change of measure, we consider the following regions: (i) $x_2 > 0$, which also covers $x_1 = 0$; (ii) $x_1 > 0$, which also covers $x_2 = 0$; and (iii) $x_1 \geq 0$ and $x_2 \geq 0$, so that we have covered the whole state space. All regions overlap in the sense that they all cover the case in which both $x_1 > 0$ and $x_2 > 0$. However, by construction of the continuously differentiable subsolution, in that part of the state space the function $W_k^\delta(\mathbf{x})$ of only one of the three regions will be used since we use the minimum of the functions $W_k^\delta(\mathbf{x})$. Clearly, the third region covers the whole state space, but it is important to note that there is no non-trivial solution that satisfies Condition 1 from Theorem 3.1 for the whole state space. The most important part of region three is that it covers $x_1 = x_2 = 0$.

It turns out that the zero change of measure can be used when both $x_1 = x_2 = 0$. This may seem strange at first sight, since simulating under the zero change of measure is the same as simulating the original system. However, using this as a starting point for this part of the state space turns out to be a good choice (which is the same choice as in the works of Dupuis et al.⁴ and Dupuis and Wang⁶). In particular, the condition in Equation (19) for this part of the state space is equivalent to $\alpha_{3,1} \geq 0$. The zero change of measure gives us $\alpha_3 = (\alpha_{3,1}, \alpha_{3,2}) = (0, 0)$ and hence satisfies the condition mentioned.

The ordering of the regions that we assign can be found in Table 1, the reasons for this ordering will become clear later in this section.

4.1.1. Finding α_1 . To find α_1 , we start with Condition 2, which is $W_k^\delta(\mathbf{x}) \leq g(N)$ for all $\mathbf{x} \in \delta_e$, where $\lim_{N \rightarrow \infty} g(N) = 0$. In Remark 3.1, it is noted that for the most likely path equality should hold. Trivially, we have $W_k^\delta(\mathbf{x}) \leq W_1^\delta(\mathbf{x})$, with equality on the most likely path when $N \rightarrow \infty$, since $W_1^\delta(\mathbf{x})$ covers the most likely path. As

Table 1. Overview of proposed regions for the case $r = 3$.

Region	k
$x_2 > 0$	1
$x_1 > 0$	2
$x_1 \geq 0, x_2 \geq 0$	3

a result, taking into account Section 3.2.2 and recalling that $\alpha_k = (\alpha_{k,1}, \alpha_{k,2})$ for all k , we must have $\alpha_{1,2} = -\gamma_2$.

Using Condition 1, we can now determine $\alpha_{1,1}$. We only consider the term of the summation from this condition that involves $k = 1$, i.e., $\rho_1(\mathbf{x})\mathbb{H}(\mathbf{x}, \alpha_1)$ should be non-negative for large enough N . As $W_1^\delta(\mathbf{x})$ is not designed for $x_2 = 0$, the intuition is that the weight factor $\rho_1(\mathbf{x})$ tends to 0 for large enough N for all states \mathbf{x} such that $x_2 = 0$, see also Equation (28). Thus, using Equation (18) for $\alpha = \alpha_1$, we see that for all \mathbf{x} such that $x_2 > 0$, we can only satisfy Condition 1, if we have the following:

$$\lambda e^{-\alpha_{1,1}} + \frac{\mu_1 \mu_2}{\lambda} e^{\alpha_{1,1}} + \lambda \leq 1, \quad (21)$$

$$\lambda e^{-\alpha_{1,1}} + \mu_1 + \lambda \leq 1, \quad (22)$$

where we used $\alpha_{1,2} = -\gamma_2$. As a result of Lemma 3.2, the first bullet, we find that Equation (21) is satisfied when $\alpha_{1,1} \in [-\gamma_1, -\gamma_2]$. Using Lemma 3.3, we find that $\alpha_{1,1} \geq -\gamma_2$ is necessary to satisfy Equation (22), and hence, we must have $\alpha_{1,1} = -\gamma_2$. Therefore, we need the following:

$$\alpha_1 = (-\gamma_2, -\gamma_2),$$

to get an asymptotically efficient change of measure based on Theorem 3.1. For future reference, we remark that as a result of this condition on α_1 we find the following, using Equation (18),

$$\rho_1(\mathbf{x})\mathbb{H}(\mathbf{x}, \alpha_1) = \begin{cases} 0 & \text{if } x_1 > 0, \quad x_2 > 0, \\ -\rho_1(\mathbf{x}) \log(\mu_1 + 2\mu_2) & \text{if } x_1 > 0, \quad x_2 = 0, \\ 0 & \text{if } x_1 = 0, \quad x_2 > 0, \end{cases} \quad (23)$$

since $\lambda + \mu_1 + \mu_2 = 1$.

4.1.2. Finding α_2 . Using the underlying idea of the construction of the subsolution – i.e., the idea to construct several functions, each for different parts of the state space, that are combined through mollification to obtain a classical subsolution – we determine conditions on α_2 and d_k for $k = 1, 2, 3$. For some parts of the state space, it is known which function $W_k^\delta(\mathbf{x})$ has to be the minimum of the three functions or cannot be the minimum of the three.

For example, when $x_1 = 0$, it follows that $W_2^\delta(\mathbf{x})$ cannot be the minimum function, since it is designed for $x_1 > 0$. As a result, we find that for some parts of the state space, some weight factors $\rho_k(\mathbf{x})$ must tend to 0 as $N \rightarrow \infty$.

To start with, we consider the origin of the state space, i.e., $x_1 = x_2 = 0$. Here, we want $W_3^\delta(\mathbf{x})$ to be the minimum function, since this is the only function that is designed for this part of the state space. Thus, we need both $W_3^\delta(\mathbf{x}) < W_1^\delta(\mathbf{x})$ and $W_3^\delta(\mathbf{x}) < W_2^\delta(\mathbf{x})$ for $\mathbf{x} = \mathbf{0}$. Trivially, these inequalities result in the following:

$$d_1 < d_3 \quad \text{and} \quad d_2 < d_3. \quad (24)$$

Second, we consider the boundary $x_2 = 0$ (and so $x_1 > 0$). At this part of the state space, we want $W_2^\delta(\mathbf{x})$ to be the minimum function. Thus, we want to have both:

$$W_2^\delta(\mathbf{x}) < W_1^\delta(\mathbf{x}) \Leftrightarrow (\alpha_{2,1} - \alpha_{1,1})x_1 < (d_2 - d_1)\delta, \quad (25)$$

$$W_2^\delta(\mathbf{x}) < W_3^\delta(\mathbf{x}) \Leftrightarrow \alpha_{2,1} x_1 < (d_2 - d_3)\delta, \quad (26)$$

for all \mathbf{x} such that $x_2 = 0$. Clearly, the first inequality holds for all x_1 whenever:

$$\alpha_{2,1} \leq \alpha_{1,1} = -\gamma_2 \quad \text{and} \quad d_1 < d_2. \quad (27)$$

As a result, we immediately have, for all \mathbf{x} such that $x_2 = 0$, using Equation (20) for $k = 1$ and $\ell = 2$, the following:

$$\rho_1(\mathbf{x}) \leq e^{((\gamma_2 + \alpha_{2,1})x_1 + (d_1 - d_2)\delta)/\varepsilon} \leq e^{(d_1 - d_2)\delta/\varepsilon}, \quad (28)$$

for all $\alpha_{2,1} \leq -\gamma_2$ and since $d_1 < d_2$, the right-hand side tends to 0 as N tends to infinity. This intuitively implies that the weight factor $\rho_1(\mathbf{x})$ for \mathbf{x} such that $x_2 = 0$ tends to 0 when $N \rightarrow \infty$, as suggested in Section 4.1.1, and thus, the change of measure that is designed for $x_2 > 0$ hardly has any influence when $x_2 = 0$.

The second inequality, Equation (26), is satisfied for all $x_1 > \frac{(d_3 - d_2)\delta}{-\alpha_{2,1}}$, which is positive because $\alpha_{2,1} \leq -\gamma_2 < 0$, see Equation (27), and $d_2 < d_3$. Thus, $W_2^\delta(\mathbf{x})$ is the minimum function for all $x_1 > \frac{(d_3 - d_2)\delta}{-\alpha_{2,1}}$, and note that the right-hand side of this inequality tends to zero as $N \rightarrow \infty$. For all other (very small) x_1 , the function $W_3^\delta(\mathbf{x})$ is the minimum function. It turns out in the sequel that this is not a problem for the resulting change of measure to be asymptotically efficient, since the function $W_3^\delta(\mathbf{x})$ can be used throughout the whole state space. More importantly, $W_1^\delta(\mathbf{x})$ is *not* the minimum function for all \mathbf{x} such that $x_1 > 0$ and $x_2 = 0$ whenever Equation (27) holds.

Combining all conditions on d_k , see Equations (24) and (27), we find $d_1 < d_2 < d_3$. Here, we see that the choice $d_k = k$, as in the work of Dupuis et al.,⁴ satisfies all requirements that we have imposed until now to get an

asymptotically efficient change of measure based on Theorem 3.1, but it is by no means unique.

Next, we consider the boundary $x_1 = 0$ (and so $x_2 > 0$). Here, we want $W_1^\delta(\mathbf{x})$ to be the minimum function, since all other functions are not designed for this part of the state space. Thus, for all \mathbf{x} such that $x_1 = 0$, we want to have both:

$$W_1^\delta(\mathbf{x}) < W_2^\delta(\mathbf{x}) \Leftrightarrow (\alpha_{1,2} - \alpha_{2,2})x_2 < (d_1 - d_2)\delta, \quad (29)$$

$$W_1^\delta(\mathbf{x}) < W_3^\delta(\mathbf{x}) \Leftrightarrow \alpha_{1,2} x_2 < (d_1 - d_3)\delta. \quad (30)$$

The first inequality is equivalent to the following:

$$(\alpha_{2,2} + \gamma_2)x_2 > (d_2 - d_1)\delta.$$

which unfortunately holds only for $x_2 > (d_2 - d_1)\delta/(\alpha_{2,2} + \gamma_2)$, provided that $\alpha_{2,2} > -\gamma_2$. Since we do *not* want $W_2^\delta(\mathbf{x})$ to be the minimum function for \mathbf{x} such that $x_1 = 0$ and $x_2 \leq (d_2 - d_1)\delta/(\alpha_{2,2} + \gamma_2)$, we need $W_3^\delta(\mathbf{x}) < W_2^\delta(\mathbf{x})$ for those states \mathbf{x} . This condition is equivalent to the following:

$$\alpha_{2,2}x_2 > (d_2 - d_3)\delta, \quad (31)$$

and if $\alpha_{2,2} \geq 0$ this inequality holds for all x_2 , so in particular for $x_2 \leq (d_2 - d_1)\delta/(\alpha_{2,2} + \gamma_2)$. So we need the following:

$$\alpha_{2,2} \geq 0. \quad (32)$$

As a result, using Equation (20) for $k = 2$ and $\ell = 3$, we find for all \mathbf{x} such that $x_1 = 0$:

$$\rho_2(\mathbf{x}) \leq e^{(-\alpha_{2,2}x_2 + (d_2 - d_3)\delta)/\varepsilon} \leq e^{(d_2 - d_3)\delta/\varepsilon}. \quad (33)$$

Since $d_2 < d_3$, the right-hand side tends to 0 as $N \rightarrow \infty$ and this implies that the weight factor $\rho_2(\mathbf{x})$ for all \mathbf{x} such that $x_1 = 0$ tends to 0 as $N \rightarrow \infty$. Therefore, the change of measure that is designed for $x_1 > 0$, has hardly any influence when $x_1 = 0$.

The second inequality, Equation (30), is satisfied for all $x_2 > (d_3 - d_1)\delta/\gamma_2$. Thus, $W_1^\delta(\mathbf{x})$ is smaller than $W_3^\delta(\mathbf{x})$ for all $x_2 > (d_3 - d_1)\delta/\gamma_2$. As a result, for all $x_2 \leq (d_3 - d_1)\delta/\gamma_2$ we have that $W_3^\delta(\mathbf{x})$ is the minimum function (which is not a limitation for the resulting change of measure to be asymptotically efficient since this function can be used for the whole state space). Recall that it is more important that $W_2^\delta(\mathbf{x})$ is *not* the minimum function for all $x_1 = 0$ and $x_2 > 0$, and this requirement is satisfied.

To derive a lower bound on $\alpha_{2,1}$ and an upper bound on $\alpha_{2,2}$, in contrast to the bounds in Equations (27) and (32), we use Condition 1 of Theorem 3.1. In this case, we only consider the term of the summation from this condition that considers $k = 2$, since this involves α_2 . That is, $\rho_2(\mathbf{x})\mathbb{H}(\mathbf{x}, \alpha_2)$ should be non-negative for large enough N .

Recall that we have derived an upper bound for $\rho_2(\mathbf{x})$ for all \mathbf{x} such that $x_1 = 0$, see Equation (33), that tends to 0 as $N \rightarrow \infty$. Thus, for all \mathbf{x} such that $x_1 > 0$ we see that, using Equation (18) for $\alpha = \alpha_2$, we can only satisfy Condition 1, if we have the following:

$$\lambda e^{-\alpha_{2,1}} + \mu_1 e^{\alpha_{2,1} - \alpha_{2,2}} + \mu_2 e^{\alpha_{2,2}} \leq 1, \quad (34)$$

$$\lambda e^{-\alpha_{2,1}} + \mu_1 e^{\alpha_{2,1} - \alpha_{2,2}} + \mu_2 \leq 1. \quad (35)$$

It is clear that, as we have $\alpha_{2,2} \geq 0$ from Equation (32), the second inequality is implied by the first inequality. Using Lemma 3.2, Figure 5, and Equations (27) and (32), we find that the first inequality can only be satisfied when:

$$\alpha_{2,1} \in [-\gamma_1, -\gamma_2] \quad \text{and} \quad \alpha_{2,2} \in [0, \gamma_1 - \gamma_2].$$

Clearly, $\alpha_{2,1}$ and $\alpha_{2,2}$ have a dependence, e.g., $\alpha_{2,1} = -\gamma_1$ implies $\alpha_{2,2} = 0$ and $\alpha_{2,1} = -\gamma_2$ implies $\alpha_{2,2} \in [0, \gamma_1 - \gamma_2]$, see Lemma 3.2 and Figure 5. The dependence can be found in Equation (34), however, this equation cannot be simplified. For future reference, we remark that as a result of these conditions on α_2 , we find from Equation (18), using $\alpha_{2,1} \geq -\gamma_1$ and $\alpha_{2,2} \leq \gamma_1 - \gamma_2$:

$$\rho_2(\mathbf{x}) \mathbb{H}(\mathbf{x}, \alpha_2) \geq \begin{cases} 0 & \text{if } x_1 > 0, x_2 > 0, \\ 0 & \text{if } x_1 > 0, x_2 = 0, \\ -\rho_2(\mathbf{x}) \log(3\mu_1) & \text{if } x_1 = 0, x_2 > 0, \end{cases} \quad (36)$$

since $\lambda + \mu_1 + \mu_2 = 1$.

4.1.3. Summary and proof that all conditions are satisfied. To summarize, we have found the following values for α_1 , α_2 , and α_3 that intuitively satisfy all conditions for an asymptotically efficient change of measure based on Theorem 3.1, see Table 2.

We show that these possibilities for α_k , $k = 1, 2, 3$, indeed give an asymptotically efficient change of measure, by considering all conditions in Theorem 3.1. Recall that $\lambda + \mu_1 + \mu_2 = 1$ implies $\mathbb{H}(\mathbf{x}, \alpha_3) = 0$. To start with Condition 1, using Equations (23), (28), (33), and (36), we find the following lower bound:

$$\begin{aligned} & \sum_{k=1}^3 \rho_k(\mathbf{x})(\mathbf{x}, \alpha_k) \\ & \geq \begin{cases} 0 & \text{if } x_1 > 0, x_2 > 0 \\ -e^{(d_1 - d_2)\delta/\varepsilon} \log(\mu_1 + 2\mu_2) & \text{if } x_1 > 0, x_2 = 0 \\ -e^{(d_2 - d_3)\delta/\varepsilon} \log(3\mu_1) & \text{if } x_1 = 0, x_2 > 0 \end{cases} \\ & \geq -e^{\max\{d_1 - d_2, d_2 - d_3\}\delta/\varepsilon} \log(3\mu_1), \end{aligned}$$

where the last step follows since queue 2 is the bottleneck queue, and thus, $\mu_2 \leq \mu_1$. It follows that Condition 1 is satisfied. For Condition 2, we note that for all $\mathbf{x} \in \delta_e$ we

Table 2. Possibilities for α_k when queue 2 is the bottleneck queue, provided that $d_1 < d_2 < d_3$.

	$\alpha_{k,1}$	$\alpha_{k,2}$	k	Condition
$x_2 > 0$	$-\gamma_2$	$-\gamma_2$	1	Equation (34)
$x_1 > 0$	$[-\gamma_1, -\gamma_2]$	$[0, \gamma_1 - \gamma_2]$	2	
$x_1 \geq 0, x_2 \geq 0$	0	0	3	

have $W_1^\delta(\mathbf{x}) = -d_1\delta$, and thus, Condition 2 is satisfied since $W^{\varepsilon,\delta}(\mathbf{x}) \leq W_1^\delta(\mathbf{x})$. To conclude with Condition 3, we find the following:

$$\begin{aligned} W^{\varepsilon,\delta}(\mathbf{0}) & \geq -\varepsilon \log(3e^{-W_3^\delta(\mathbf{0})/\varepsilon}) = -\varepsilon \log 3 + W_3^\delta(\mathbf{0}) \\ & = -\varepsilon \log 3 + \gamma_2 - d_3\delta, \end{aligned} \quad (37)$$

which goes to γ_2 as $N \rightarrow \infty$ and hence all conditions are satisfied. Therefore, the change of measure for the given possible values of $\alpha_1, \alpha_2, \alpha_3$ and d_k , $k = 1, 2, 3$, is asymptotically efficient. Note that also the conditions in Remark 3.1 are satisfied, based on the construction.

4.1.4. Discussion. It is clear that the choice of $\alpha_2 = (-\gamma_2, 0)$ results in the change of measure from Dupuis et al.,⁴ using $d_k = k$ for $k = 1, 2, 3$. For this case, $W^\delta(\mathbf{x})$ is illustrated in Figure 2. We show some other examples of the piecewise affine function $W^\delta(\mathbf{x})$ below, where we let $\alpha_2 = (-\gamma_1, 0)$ in Figure 7 and

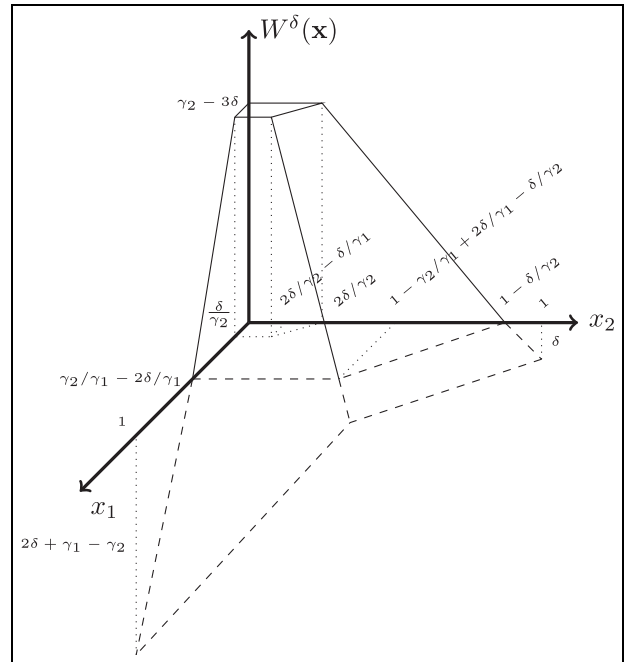


Figure 7. Display of $W^\delta(\mathbf{x})$ when queue 2 is the bottleneck queue, $\alpha_2 = (-\gamma_1, 0)$ and $d_k = k$, $k = 1, 2, 3$.

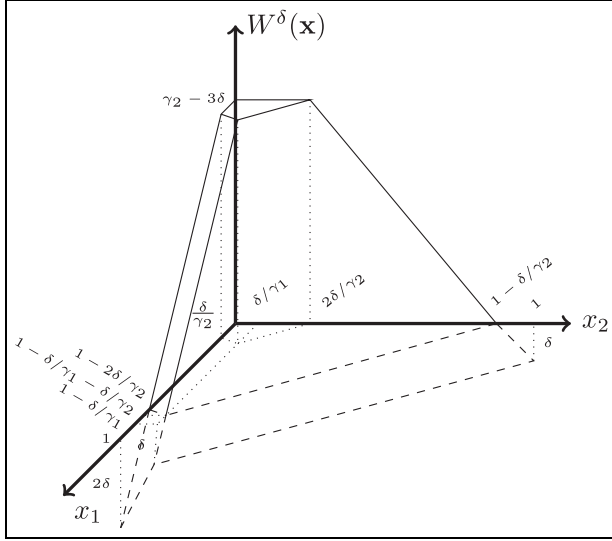


Figure 8. Display of $W^\delta(\mathbf{x})$ when queue 2 is the bottleneck queue, $\alpha_2 = (-\gamma_2, \gamma_1 - \gamma_2)$ and $d_k = k$, $k = 1, 2, 3$.

$\alpha_2 = (-\gamma_2, \gamma_1 - \gamma_2)$ in Figure 8. In both cases, we choose $d_k = k$, $k = 1, 2, 3$.

Choosing $\alpha_{2,2} = 0$, we find $\alpha_{2,1} \in [-\gamma_1, -\gamma_2]$, and thus, $W^\delta(\mathbf{x})$ can be “anything in between” Figures 2 and 7. That is, the narrow area next to the x_1 -axis in Figure 2 can be as steep as in Figure 7, and anything in between, while the resulting change of measure still gives an asymptotically efficient estimator. Choosing $\alpha_{2,1} = -\gamma_2$, we find $\alpha_{2,2} \in [0, \gamma_1 - \gamma_2]$, and thus, $W^\delta(\mathbf{x})$ can be “anything in between” Figures 2 and 8. That is, the narrow area next to the x_1 -axis in Figure 2 can be slightly tilted, while the resulting change of measure still gives an asymptotically efficient estimator.

We remark that the change of measure that is used for $x_2 > 0$ is the state-independent change of measure from Parekh and Walrand.¹ The changes of measure that are found here, can be interpreted as “protecting” the x_1 -axis in the sense that we have to apply a different change of measure in that part of the state space. In Figure 7, we protect the x_1 -axis quite a lot, rather than in Figure 8, where we only protect it slightly. Recall that the most likely path goes along the x_2 -axis, where we have to apply the change of measure from Parekh and Walrand,¹ which also follows from the subsolution approach. It turns out that the change of measure along the most likely path is very important in the construction of an asymptotically efficient change of measure based on subsolutions and that along the most likely path there is no variation possible for the change of measure. However, it turns out that for all other parts of the state space it is possible to apply a (slightly) different change of measure than the one from Dupuis et al.⁴ As we have seen in the work of Dupuis and Wang⁶ and as we will see in Section 4.3, it is also possible to apply a different change of measure in the interior of the state space, rather than the same change of measure as

along the most likely path or the change of measure that is applied along the x_1 -axis.

4.2. Three regions and queue 1 bottleneck

In this section, we again consider three functions $W_k^\delta(\mathbf{x})$, $k = 1, 2, 3$, but in contrast to Section 4.1, we now let queue 1 be the bottleneck queue, so $\gamma_1 \leq \gamma_2$. The regions that we consider in this section are the same as in the previous section (see Table 1). By changing the bottleneck queue, the most likely path also changes and, therefore, we start the construction of an asymptotically efficient change of measure based on Theorem 3.1 by finding α_2 .

4.2.1. Finding α_2 . To find α_2 , we start with Condition 2, as in Section 4.1.1, and we use Remark 3.1 to determine $\alpha_{2,1}$. When queue 1 is the bottleneck queue, the most likely path is covered by $W_2^\delta(\mathbf{x})$. Therefore, we find $\alpha_{2,1} = -\gamma_1$.

Next, we use Condition 1 to determine $\alpha_{2,2}$. Considering the term of the summation in this condition that involves $k = 2$, we need $\rho_2(\mathbf{x})\mathbb{H}(\mathbf{x}, \alpha_2)$ to be non-negative for large enough N . For those parts of the state space where $x_1 = 0$, we will find in the sequel that the weight factor $\rho_2(\mathbf{x})$ tends to 0 as $N \rightarrow \infty$. Since the function $W_2^\delta(\mathbf{x})$ is designed for $x_1 > 0$, it is expected that in that case the weight factor $\rho_2(\mathbf{x})$ does not tend to 0 as $N \rightarrow \infty$. Thus, using Equation (18) for $\alpha = \alpha_2$, for all \mathbf{x} such that $x_1 > 0$ we see that we can only satisfy Condition 1 from Theorem 3.1 if we have the following:

$$\begin{aligned} \mu_1 + \lambda e^{-\alpha_{2,2}} + \mu_2 e^{\alpha_{2,2}} &\leq 1, \\ \mu_1 + \lambda e^{-\alpha_{2,2}} + \mu_2 &\leq 1, \end{aligned}$$

where we used $\alpha_{2,1} = -\gamma_1$. Using the second bullet of Lemma 3.2 and Lemma 3.3, we find that both of the above conditions hold when $\alpha_{2,2} = 0$. Thus, to get an asymptotically efficient change of measure based on Theorem 3.1, we need the following:

$$\alpha_2 = (-\gamma_1, 0).$$

For future reference, we remark that as a result of this condition we find, using Equation (18) and $\lambda + \mu_1 + \mu_2 = 1$, the following:

$$\rho_2(\mathbf{x})\mathbb{H}(\mathbf{x}, \alpha_2) = \begin{cases} 0 & \text{if } x_1 > 0, x_2 > 0, \\ 0 & \text{if } x_1 > 0, x_2 = 0, \\ -\rho_2(\mathbf{x}) \log(2\mu_1 + \mu_2) & \text{if } x_1 = 0, x_2 > 0, \end{cases} \quad (38)$$

4.2.2. Finding α_1 . As in Section 4.1.2, we use the underlying idea of the construction of subsolutions to determine conditions on α_2 and d_k for $k = 1, 2, 3$. Since this approach has been fully explained in Section 4.1.2, we will skip most of the details and highlight the results.

By considering the origin of the state space, i.e., $x_1 = x_2 = 0$, we immediately find the same result as in Equation (24), since that result is independent of α_1 and α_2 .

Next, we consider the boundary $x_2 = 0$ (and so $x_1 > 0$) and recall that in that case, we want $W_2^\delta(\mathbf{x})$ to be the minimum function. The conditions that follow from this observation can be found in Equations (25) and (26). Since we have determined that $\alpha_2 = (-\gamma_1, 0)$, these conditions are equivalent to the following:

$$\begin{aligned} (\alpha_{1,1} + \gamma_1)x_1 &> (d_1 - d_2)\delta, \\ \gamma_1 x_1 &> (d_3 - d_2)\delta. \end{aligned} \quad (39)$$

The first condition is satisfied when:

$$\alpha_{1,1} \geq -\gamma_1 \quad \text{and} \quad d_1 < d_2, \quad (40)$$

since then the right-hand side of the inequality is negative. Similarly to Equation (28) we find the following:

$$\rho_1(\mathbf{x}) \leq e^{((- \gamma_1 - \alpha_{1,1})x_1 + (d_1 - d_2)\delta)/\varepsilon} \leq e^{(d_1 - d_2)\delta/\varepsilon}, \quad (41)$$

and so the weight factor $\rho_1(\mathbf{x})$ for \mathbf{x} such that $x_2 = 0$ tends to zero when $N \rightarrow \infty$.

For the second condition, see Equation (39), we find that this is satisfied for all $x_1 > \frac{(d_3 - d_2)\delta}{\gamma_1}$, and so for small values of x_1 the function $W_3^\delta(\mathbf{x})$ is the minimum function. It turns out that this is not a problem for the resulting change of measure to be asymptotically efficient, since the function $W_3^\delta(\mathbf{x})$ can be used throughout the whole state space and since $W_2^\delta(\mathbf{x}) < W_1^\delta(\mathbf{x})$ for all \mathbf{x} such that $x_2 = 0$. Therefore, it is ruled out that $W_1^\delta(\mathbf{x})$ is the minimum function for these \mathbf{x} , as desired.

At the boundary $x_1 = 0$ (and so $x_2 > 0$), we want $W_1^\delta(\mathbf{x})$ to be the minimum function. As a result, we want Equations (29) and (30) to hold. These conditions are equivalent to the following:

$$\begin{aligned} \alpha_{1,2}x_2 &< (d_1 - d_2)\delta, \\ \alpha_{1,2}x_2 &< (d_1 - d_3)\delta. \end{aligned}$$

Clearly, both conditions are satisfied when the second condition holds, since $d_2 < d_3$. In particular, we need the following:

$$\alpha_{1,2} < 0,$$

since $d_1 < d_3$, so that both conditions are satisfied for all $x_2 > \frac{(d_3 - d_1)\delta}{-\alpha_{1,2}}$. To prevent $W_2^\delta(\mathbf{x})$ from being the minimum function for \mathbf{x} such that $x_1 = 0$ and $x_2 \leq \frac{(d_3 - d_1)\delta}{-\alpha_{1,2}}$, we need $W_3^\delta(\mathbf{x}) < W_2^\delta(\mathbf{x})$ for those states \mathbf{x} . This condition is equivalent to Equation (31), and since $\alpha_{2,2} = 0$ and $d_2 < d_3$, this condition is always satisfied. As a result, we find for all \mathbf{x} such that $x_1 = 0$ that Equation (33) holds.

To get a tighter condition for $\alpha_{1,2}$ we use Condition 2 from Theorem 3.1. That is, for all \mathbf{x} on the exit boundary we need to have $W^{e,\delta}(\mathbf{x}) \leq g(N)$ such that $g(N) \rightarrow 0$ when $N \rightarrow \infty$. As we have $W^{e,\delta}(\mathbf{x}) \rightarrow W^\delta(\mathbf{x})$ when $N \rightarrow \infty$, we in particular need $W_1^\delta(\mathbf{x})$ to be non-positive for large enough N when $x_1 = 0$ and $x_2 = 1$, since $W_1^\delta(\mathbf{x})$ is designed to be the minimum function at the boundary $x_1 = 0$ (and so $x_2 > 0$). Therefore we find the following:

$$\alpha_{1,2} \leq -\gamma_1. \quad (42)$$

Using the same condition of Theorem 3.1, we also derive an upper bound on $\alpha_{1,1}$. Consider $W_2^\delta(\mathbf{x}) = -\gamma_1 x_1 + \gamma_1 - d_2\delta$, which is non-positive for all $x_1 \geq (\gamma_1 - d_2\delta)/\gamma_1$. Thus, for all $x_1 < (\gamma_1 - d_2\delta)/\gamma_1$, we need $W_1^\delta(\mathbf{x})$ to be non-positive for all $\mathbf{x} \in \delta_e$. On the exit boundary we have $x_1 + x_2 = 1$, so we find that for all $\mathbf{x} \in \delta_e$, using Equation (42),

$$\begin{aligned} W_1^\delta(\mathbf{x}) &= \alpha_{1,1}x_1 + \alpha_{1,2}(1 - x_1) + \gamma_1 - d_1\delta \\ &\leq (\alpha_{1,1} + \gamma_1)x_1 - d_1\delta, \end{aligned}$$

which goes to zero as $N \rightarrow \infty$ if and only if $\alpha_{1,1} \leq -\gamma_1$. Combining with Equation (40) we find the following:

$$\alpha_{1,1} = -\gamma_1.$$

To conclude, we derive a lower bound on $\alpha_{1,2}$ using Condition 1. We only consider the term of the summation in Condition 1 that involves $k = 1$. That is, $\rho_1(\mathbf{x})\mathbb{H}(\mathbf{x}, \alpha_1)$ should be non-negative for large enough N . We do not expect that $\rho_1(\mathbf{x}) \rightarrow 0$ when $N \rightarrow \infty$ for all \mathbf{x} such that $x_2 > 0$, as $W_1^\delta(\mathbf{x})$ is designed for $x_2 > 0$. Thus, using Equation (18) for $\alpha = \alpha_1$, for all \mathbf{x} such that $x_2 > 0$, we see that we can only satisfy Condition 1, if we have the following:

$$\mu_1 + \lambda e^{-\alpha_{1,2}} + \mu_2 e^{\alpha_{1,2}} \leq 1, \quad (43)$$

$$\mu_1 + \mu_1 + \mu_2 e^{\alpha_{1,2}} \leq 1, \quad (44)$$

where we have used that $\alpha_{1,1} = -\gamma_1$. Using Equation (42), it follows that the second inequality is implied by the first inequality. Using Lemma 3.2, the second bullet, we find that Equation (43) is satisfied when $\alpha_{1,2} \in [-\gamma_2, -\gamma_1]$, and so is Equation (44). For future reference, we remark that as a result of this condition we find, using Equation (18) and $\alpha_{1,2} \geq -\gamma_2$,

$$\rho_1(\mathbf{x})\mathbb{H}(\mathbf{x}, \alpha_1) \geq \begin{cases} 0 & \text{if } x_1 > 0, x_2 > 0, \\ -\rho_1(\mathbf{x}) \log(\mu_1 + 2\mu_2) & \text{if } x_1 > 0, x_2 = 0, \\ 0 & \text{if } x_1 = 0, x_2 > 0, \end{cases} \quad (45)$$

since $\lambda + \mu_1 + \mu_2 = 1$.

The derivation for using four regions is very similar as when using two regions and, therefore, it is omitted in this paper. A detailed derivation can be found in Buijsrogge.¹²

Table 4. Overview of proposed regions for the case $r = 4$.

Region	k
$x_1 > 0, x_2 > 0$	1
$x_1 = 0, x_2 > 0$	2
$x_1 > 0, x_2 = 0$	3
$x_1 \geq 0, x_2 \geq 0$	4

4.3.1. Summary and proof that all conditions are satisfied. We have found the following values for $\alpha_1, \alpha_2, \alpha_3$, and α_4 that intuitively satisfy all requirements for an asymptotically efficient change of measure based on Theorem 3.1, see Table 5.

We show that these possibilities for $\alpha_k, k = 1, 2, 3, 4$, indeed give an asymptotically efficient change of measure, by considering all conditions in Theorem 3.1. Recall that $\lambda + \mu_1 + \mu_2 = 1$ implies $\mathbb{H}(\mathbf{x}, \alpha_4) = 0$. To start with Condition 1 of Theorem 3.1, we find the following:

$$\begin{aligned} & \sum_{k=1}^3 \rho_k(\mathbf{x}) \mathbb{H}(\mathbf{x}, \alpha_k) \\ & \geq \begin{cases} 0 & \text{if } x_1 > 0, x_2 > 0 \\ -e^{(d_1-d_2)\delta/\varepsilon} \log(\mu_1 + 2\mu_2) & \text{if } x_1 > 0, x_2 = 0 \\ -e^{(d_2-d_3)\delta/\varepsilon} \log(3\mu_1) & \text{if } x_1 = 0, x_2 > 0 \end{cases} \\ & \geq -e^{\max\{d_1-d_2, d_2-d_3\}\delta/\varepsilon} \log(3\mu_1), \end{aligned}$$

where the first step follows by conveniently substituting the upper and lower bounds on α_k from Table 5 and the last step follows since queue 2 is the bottleneck queue, and hence $\lambda < \mu_2 \leq \mu_1$, and since $d_1 < d_2 < d_3 < d_4$. It follows that Condition 1 is satisfied. For Condition 2, we note that for all $\mathbf{x} \in \delta_e$ we have $W_2^\delta(\mathbf{x}) = -d_2\delta$, and thus, Condition 2 is satisfied since $W^{e,\delta}(\mathbf{x}) \leq W_2^\delta(\mathbf{x})$. Condition 3 follows similarly as in Equation (37). Therefore, the change of measure for the given possible values of $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, and $\alpha_k, k = 1, 2, 3, 4$, is asymptotically efficient.

4.3.2. Discussion. From Table 5, we again see that the change of measure from Parekh and Walrand¹ is used for $x_1 = 0$ and $x_2 > 0$, as expected, since it is the region in

which the most likely path lies. Clearly, the result in the work of Dupuis and Wang⁶ satisfies the conditions in Table 5. Recall that in that paper also different overflow probabilities are considered, which imposes additional constraints for an asymptotically efficient change of measure based on subsolutions. The change of measure from Dupuis and Wang,⁶ where in particular $\alpha_1 = (-\gamma_1, -\gamma_2)$ and $\alpha_3 = (-\gamma_1, 0)$, can be found in Figure 3.

From the results in Table 5, it is also clear that we could adapt Figure 3 such that along the x_1 -axis we have a similar affine function as in Figure 7 or Figure 8. However, when choosing $\alpha_3 = (-\gamma_2, \gamma_1 - \gamma_2)$, as we did for α_2 in Figure 8, we need $\alpha_{1,1} = -\gamma_2$ and so the functions $W_1^\delta(\mathbf{x})$ and $W_2^\delta(\mathbf{x})$ are almost the same. That is, only their constants d_1 and d_2 would differ.

Summarizing, at the x_1 -axis, we can make similar adaptations as when we considered three regions in Section 4.1. The difference is that now also in the interior we can “push” the function a bit more toward the origin when we compare with Figure 2.

4.4. Four regions and queue 1 bottleneck

In this section, we again consider four functions $W_k^\delta(\mathbf{x})$, $k = 1, 2, 3, 4$, but in contrast to Section 4.3, we let queue 1 be the bottleneck queue, i.e., $\gamma_1 \leq \gamma_2$. The regions that we consider in this section are the same as in the previous section (see Table 4). Again, the derivation for using four regions is very similar as when using two regions and therefore it is omitted in this paper. Recall that a detailed derivation can be found in the work of Buijsrogge.¹²

4.4.1. Summary and proof that all conditions are satisfied. We have found the following values for $\alpha_1, \alpha_2, \alpha_3$, and α_4 that intuitively satisfy all conditions for an asymptotically efficient change of measure based on Theorem 3.1 see Table 6.

Again, we show by considering all conditions in that theorem, indeed these possibilities for $\alpha_k, k = 1, 2, 3, 4$, give an asymptotically efficient change of measure. We start with Condition 1, and recall that $\mathbb{H}(\mathbf{x}, \alpha_4) = 0$. Then, we find the following:

Table 5. Possibilities for α_k when queue 2 is the bottleneck queue, provided that $d_1 < d_2 < d_3 < d_4$.

	$\alpha_{k,1}$	$\alpha_{k,2}$	k	Conditions
$x_1 > 0, x_2 > 0$	$[-\gamma_1, -\gamma_2]$	$-\gamma_2$	1	
$x_1 = 0, x_2 > 0$	$-\gamma_2$	$-\gamma_2$	2	
$x_1 > 0, x_2 = 0$	$[-\gamma_1, \alpha_{1,1}]$	$[0, \gamma_1 - \gamma_2]$	3	$\lambda e^{-\alpha_{3,1}} + \mu_1 e^{\alpha_{3,1} - \alpha_{3,2}} + \mu_2 e^{\alpha_{3,2}} \leq 1$
$x_1 \geq 0, x_2 \geq 0$	0	0	4	$\alpha_{3,1} + \alpha_{3,2} \geq \alpha_{1,1} - \gamma_2$

Table 6. Possibilities for α_k when queue 1 is the bottleneck queue, provided that $d_1 < d_2 < d_3 < d_4$. There is no simple expression for the maximum value that $\alpha_{2,1}$ could possibly attain and it is therefore denoted with *. It should satisfy the conditions stated (see also Figure 6).

	$\alpha_{k,1}$	$\alpha_{k,2}$	k	Conditions
$x_1 > 0, x_2 > 0$	$-\gamma_1$	$[-\gamma_2, -\gamma_1]$	1	
$x_1 = 0, x_2 > 0$	$[-\gamma_1, *]$	$[-\gamma_2, \alpha_{1,2}]$	2	$\lambda e^{-\alpha_{2,1}} + \mu_1 e^{\alpha_{2,1} - \alpha_{2,2}} + \mu_2 e^{\alpha_{2,2}} \leq 1$
$x_1 > 0, x_2 = 0$	$-\gamma_1$	0	3	$\alpha_{2,1} + \alpha_{2,2} \geq -\gamma_1 + \alpha_{1,2}$
$x_1 \geq 0, x_2 \geq 0$	0	0	4	

$$\begin{aligned}
& \sum_{k=1}^4 \rho_k(\mathbf{x}) \mathbb{H}(\mathbf{x}, \alpha_k) \\
& \geq \begin{cases} 0 & \text{if } x_1 > 0, x_2 > 0 \\ -e^{(d_1-d_3)\delta/\varepsilon} \log(2\mu_1 + \mu_2) & \text{if } x_1 > 0, x_2 = 0 \\ -e^{(d_2-d_3)\delta/\varepsilon} \log(\mu_1 + 2\mu_2) & \text{if } x_1 > 0, x_2 = 0 \\ -e^{(d_1-d_2)\delta/\varepsilon} \log(2\mu_1 + \lambda) & \text{if } x_1 = 0, x_2 > 0 \\ -e^{(d_3-d_4)\delta/\varepsilon} \log(3\mu_1) & \text{if } x_1 = 0, x_2 > 0 \end{cases} \\
& \geq -2e^{\max\{(d_2-d_3), (d_1-d_2), (d_3-d_4)\}\delta/\varepsilon} \log(3\mu_1),
\end{aligned}$$

where the first step follows by conveniently substituting the upper and lower bounds on α_k from Table 6 and the final inequality follows since queue 1 is the bottleneck queue, and thus, $\mu_1 \leq \mu_2$. It follows that Condition 1 is satisfied. For Condition 2, we note that for all $\mathbf{x} \in \delta_e$ we have $W_1^\delta(\mathbf{x}) \leq -d_1\delta$, and thus, Condition 2 is satisfied since $W^{\varepsilon,\delta}(\mathbf{x}) \leq W_1^\delta(\mathbf{x})$. Condition 3 follows similarly as in Equation (37), with γ_2 replaced by γ_1 , since queue 1 is the bottleneck queue. Therefore, the change of measure for the given possible values of $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, and d_k , $k = 1, 2, 3, 4$, is asymptotically efficient.

4.4.2. Discussion. From Table 6, we see that for $x_1 > 0$ and $x_2 = 0$ the change of measure from Parekh and Walrand¹ is used, as expected. The difference compared with using three regions in Section 4.2 is that there is a possibility to apply a slightly different change of measure for $x_1 = 0$ and $x_2 > 0$. In Figure 10, we show an example when choosing $\alpha_1 = (-\gamma_1, -\gamma_1)$, $\alpha_2 = (0, -\gamma_1)$ and $\alpha_3 = (-\gamma_1, 0)$. These values for α_2 can only occur when $-\gamma_1 \geq \gamma_1 - \gamma_2$.

Comparing with Figure 9, we see that in the area close to the x_2 -axis, we can apply a slightly different change of measure. Of course, there are several other possibilities.

5. Conclusion

In this paper, we determined sufficient conditions for subsolution-based changes of measure to give asymptotically efficient estimators. As a result, for the two-node $M|M|1$ tandem queue, we explicitly gave a whole family

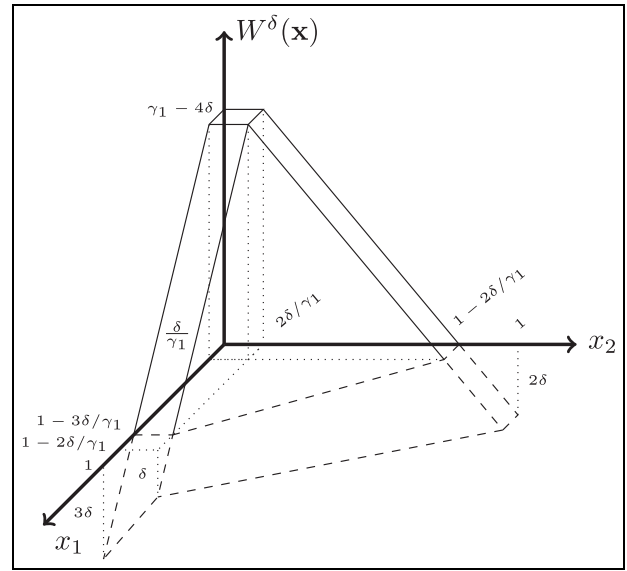


Figure 10. Display of $W^\delta(\mathbf{x})$ when queue 1 is the bottleneck queue, $\alpha_1 = (-\gamma_1, -\gamma_1)$, $\alpha_2 = (0, -\gamma_1)$ and $d_k = k$, $k = 1, 2, 3, 4$.

(continuum) of changes of measure that all lead to asymptotically efficient estimators, and the previously known changes of measure are just three members of this family. For d -node tandem queues, it seems likely that we can use a similar analysis to find a family of changes of measure that are asymptotically efficient.

For the case $d = 2$, we like to highlight one particular change of measure based on the subsolution $W^{\varepsilon,\delta}(\mathbf{x})$ in Equation (6) (via either Equation (3) or Equation (8)), that uses the following three functions:

$$\begin{aligned}
W_1^\delta(\mathbf{x}) &= -\gamma x_1 - \gamma x_2 + \gamma - \delta, \\
W_2^\delta(\mathbf{x}) &= -\gamma x_1 + \gamma - 2\delta, \\
W_3^\delta(\mathbf{x}) &= \gamma - 3\delta,
\end{aligned}$$

where we recall that $\gamma = \min\{\gamma_1, \gamma_2\}$. We note that this is the same subsolution as in the work of Dupuis et al.⁴ if queue 2 is the bottleneck queue, while it also works when

queue 1 is the bottleneck queue (with γ_2 replaced by γ_1). This matches nicely with the known fact that interchanging the queues leaves our probability of interest unchanged.


From an implementation point of view with respect to subsolutions, in general, it makes sense to use as few regions as possible, i.e., three regions (or $d + 1$ in the d -node case). However, when the event of interest is not total buffer overflow, but, e.g., individual buffer overflow or simultaneous buffer overflows, it may be more useful to implement the change of measure from Dupuis and Wang⁶ that is based on four regions.

Finally, we mention that future work could aim at investigating whether or not the method generalizes to more general models, which we expect to be the case. In the work of Buijsrogge et al.,⁷ something similar has already been done for non-Markovian tandem queues, but one could also think about more general (non-Markovian) networks, for which we expect similar results to hold.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by the Netherlands Organisation for Scientific Research (NWO), project number 613.001.105.

ORCID iD

Anne Buijsrogge  <https://orcid.org/0000-0002-7434-7950>

References

1. Parekh S and Walrand J. A quick simulation method for excessive backlogs in networks of queues. *IEEE T Automat Contr* 1989; 34: 54–66.
2. Glasserman P and Kou SG. Analysis of an importance sampling estimator for tandem queues. *ACM T Model Comput S* 1995; 5: 22–42.
3. de Boer PT. Analysis of state-independent importance-sampling measures for the two-node tandem queue. *ACM T Model Comput S* 2006; 16: 225–250.
4. Dupuis P, Sezer AD and Wang H. Dynamic importance sampling for queueing networks. *Ann Appl Probab* 2007; 17: 1306–1346.
5. Dupuis P and Wang H. Subsolutions of an Isaacs equation and efficient schemes for importance sampling. *Math Oper Res* 2007; 32: 723–757.
6. Dupuis P and Wang H. Importance sampling for Jackson networks. *Queueing Syst* 2009; 62: 113–157.
7. Buijsrogge A, de Boer PT and Scheinhardt WRW. Importance sampling for non-Markovian tandem queues using subsolutions. *Queueing Syst* 2019; 93: 31–65.
8. Budhiraja A and Dupuis P. *Analysis and approximation of rare events: representations and weak convergence methods*, vol. 94. New York: Springer, 2019.
9. Blanchet J. Optimal sampling of overflow paths in Jackson networks. *Math Oper Res* 2013; 38: 698–719.
10. Sezer AD. Approximation of excessive backlog probabilities of two tandem queues. *J Appl Probab* 2018; 55: 968–997.
11. de Boer PT and Scheinhardt WRW. Alternative proof and interpretations for a recent state-dependent importance sampling scheme. *Queueing Syst* 2007; 57: 61–69.
12. Buijsrogge A. *Rare event simulation for non-Markovian tandem queues*. PhD Thesis, University of Twente, Enschede, 2019.
13. Weber RR. The interchangeability of queues in series. *J Appl Probab* 1979; 16: 690–695.
14. Buijsrogge A, de Boer PT, Rosen K, Scheinhardt W, et al. Large deviations for the total queue size in non-Markovian tandem queues. *Queueing Syst* 2017; 85: 305–312.

Author biographies

Anne Buijsrogge received both her M.Sc. and Ph.D. degree in applied mathematics from the University of Twente, in 2014 and 2019 respectively. Currently, she is a Postdoctoral researcher at Delft University of Technology in the Applied Probability group. Her research interest include queueing theory and rare event simulation.

Pieter-Tjerk de Boer received the M.Sc. degree in applied physics and the Ph.D. degree in computer science from the University of Twente in 1996 and 2000, respectively, where he currently is an Associate Professor in the Design and Analysis of Communication Systems group. His research interests include efficient simulation techniques, communication networks and their mathematical performance modeling, and software-defined radio.

Werner Scheinhardt received both his M.Sc. and Ph.D. degree in applied mathematics from the University of Twente, in 1994 and 1998 respectively, where he currently is an Assistant Professor in the Stochastic Operations Research group. His research interests are in applied probability, in particular queueing theory.