A Linear Programming Algorithm for Computing the Stationary Distribution of Semimartingale Reflected Brownian Motion

Denis Saure*
Columbia University

Peter W. Glynn†
Stanford University

Assaf Zeevi‡
Columbia University

March 2008

Abstract

This paper proposes a linear programming algorithm for computing the stationary distribution of semimartingale reflected Brownian motion (SRBM), which arises as an approximation of certain queueing networks operating in heavy-traffic. Our algorithm is based on a Basic Adjoint Relationship (BAR) which characterizes the stationary distribution. Approximating the state space with a finite grid of points and using a finite set of "test" functions, the BAR reduces to a set of linear equations that can be solved using standard linear programming techniques. As the set of test functions increases in complexity and the grid becomes finer, the sequence of stationary distribution estimates which arise as solutions to the algorithm converges, in a suitable sense, to the stationary distribution of the SRBM. The algorithm is seen to produce good estimates of the stationary moments as well as the entire distribution. Extension to settings with parameter uncertainty, and to Ornstein-Uhlenbeck-type diffusions arising in the many-servers heavy traffic regime, are discussed as well.

Short Title: Computing the stationary distribution of SRBM

Keywords: queueing, diffusion approximations, reflected Brownian motion, Halfin-Whitt regime, numerical methods, linear programming.

2008 AMS Subject Classification (Primary): XXXX - XXXX - XXXX

*Columbia Graduate School of Business, e-mail: dsaure05@gsb.columbia.edu

†Management Science and Engineering, e-mail: glynn@stanford.edu

‡Columbia Graduate School of Business, e-mail: assaf@gsb.columbia.edu. Research partially supported by NSF grant DMI-0447562 and BSF grant #2006075

1 Introduction

Background and motivation. Multiclass queueing networks are widely used as mathematical models of real-world complex systems, e.g., communication networks, manufacturing operations, and service systems. At the same time, relevant queueing network models are mostly intractable for purposes of exact performance analysis. A possible remedy for this is to approximate the queueing network model with a more simplified structure that supports the derivation of the pertinent performance measures. Perhaps the most prevalent approach in this context has focused on a class of multidimensional diffusion processes, known as semimartingale reflected Brownian motions (SRBM) which serves as a proxy for the dynamics of the underlying queueing network. Loosely speaking, a d-dimensional SRBM living on the nonnegative orthant $S := \mathbb{R}_+^d$ behaves like an ordinary Brownian motion while in the interior of S, and is confined to the orthant by "instantaneous" reflection at the boundary faces; a formal definition will be given in §2.

The connection between the queuing network and its "Brownian" counterpart was first established in Reiman's seminal paper [25]. There, it is shown that for a certain family of open single-class queueing networks, the normalized vector of queue length processes converges in distribution to a SRBM on the positive orthant, as the traffic intensity approaches one. The scope of queueing network models for which such weak limit theory exists has been greatly expanded since, culminating in the influential papers of Branson [4] and Williams [30].

Regardless of whether supporting limit theory exists, one can always informally replace the original network with a suitable SRBM and pursue performance analysis of the latter as an approximation for the former. This is in essence the proposal advocated by Harrison and Nguyen [17], an approach that was dubbed QNET. In that paper, Harrison and Nguyen explain how to "map" the parameters of the underlying queueing network to those of the approximating Brownian model (the SRBM), and then use the latter to derive estimates of steady-state performance measures such as mean queue length, workload, etc. It should be noted that even if the SRBM can be arrived at as a rigorous weak limit, it is still unclear whether it is justified to approximate the stationary distribution of the underlying queueing network with that of the SRBM (for cases when this interchange-of-limits has been established, see e.g., Gamarnik and Zeevi [11] and Gurvich and Zeevi [14]). Assuming that the approximating SRBM indeed admits a steady-state, the main task, in order to carry out the program in Harrison and Nguyen [17], is to compute the stationary distribution of the SRBM. This is the main focal point of the present paper.

Consider an SRBM $Z = (Z(t) : t \geq 0)$ with state-space $S := \mathbb{R}_+^d$, whose drift vector $\mu = (\mu_1, \ldots, \mu_d)$ and covariance matrix $\sigma \in \mathbb{R}^{d \times d}$ dictate the dynamics of Z in the interior of the orthant, and with matrix $R \in \mathbb{R}^{d \times d}$ which dictates the directions of reflection at the boundary. Suppose Z admits a unique stationary distribution π. In most cases, it is impossible to compute π

in closed form, and a plausible alternative is to compute it numerically. To this end, define

$$Gf := \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\sigma_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^{d}\mu_i \frac{\partial f}{\partial x_i}, \tag{1}$$

$$D_i f(x) := R_i \cdot \nabla f(x) \text{ for } x \in F_i := \{x \in S\,;\, x_i = 0\}, \tag{2}$$

where R_i is the i-th column of the reflection matrix, and for two vectors $a, b \in \mathbb{R}^d$, $a \cdot b$ denotes the scalar product. Using Ito's lemma it has beeb shown that the stationary distribution π together with a *boundary measure* ν must satisfy

$$\int_S (Gf)\,(x)\,\pi(dx) + \sum_{i=1}^{d}\int_{F_i}(D_i f)\,(x)\,\nu_i(dx) = 0, \tag{3}$$

for all $f \in C_b^2$. The equation above is known as the *basic adjoint relationship* (BAR), and the aforementioned necessity was established in Harrison and Williams [18] (a more formal definition of the BAR will be advanced in §2). The first term on the left hand side of (3) is driven by the behavior of Z in the interior of S, while the second term is related to the behavior on the boundary faces $\{F_i\}$, and it is associated with the d-dimensional pushing process keeping Z in the orthant. Dai an Kurtz [8] established that this relationship is also sufficient to characterize the stationary distribution (π) and boundary measure (ν).

With this characterization at hand, Dai and Harrison [6] developed an algorithm for computing the stationary distributions of SRBM, by viewing the BAR as an orthogonality condition between an infinite dimensional functional space and the stationary distribution. By considering an increasing sequence of finite dimensional approximations of C_b^2, they obtain a sequence of approximating densities of π by means of orthogonal projections into a sequence of finite dimensional spaces. They prove that this sequence of densities converges to the stationary density of the SRBM in L^2, but their proof relies on a certain conjecture regarding densities that arise as solutions to the BAR. Later, Shen et al. [32] considered a variant of the Dai-Harrison algorithm for the case of an SRBM on a hypercube using a finite element method (or piecewise polynomials) to form the finite dimensional approximations of the functional space C^2 (Shen and Chen [31] later extend this algorithm to SRBM on the positive orthant).

Our approach. We view (3) as characterizing π via a solution to an infinite dimensional system of linear equations. Obviously it is impractical to try to solve this directly, but one can try approximating S and C_b^2 by suitable chosen sequences $\{S_n : n = 1, \ldots\}$ and $\{\mathcal{F}_m : m = 1, \ldots\}$ of finite cardinality subsets of S and finite-dimensional subspaces of C_b^2, respectively, such that $S_n \uparrow S$ and $\mathcal{F}_m \uparrow C_b^2$ in a suitable sense. For fixed values of n and m, (3) reduces to a finite-dimensional linear program (LP), that can be solved efficiently. The hope is then, that roughly speaking, π will emerge as a limit of a sequence of optimal solutions to the aforementioned LP's.

To establish the convergence of the sequence of approximations to the stationary distribution π, we need to ensure that this sequence is *tight*. This leads us to impose additional constraints in the LP formulation that enforce tightness. This, in turn, necessitates estimates of interior and boundary moments. One of the contributions of this paper is to provide a mean for enforcing such constraints, and in doing so we establish a certain relationship between the interior and boundary measure (Harrison and Williams [18] conjectured that the latter is a "projection" of the former). It is worth emphasizing that our proof of convergence does not require the conjecture advanced in Dai and Harrison [7] and follow up papers. The LP formulation is also quite flexible in terms of incorporating additional constraints, and we indicate how this can be used to get more stable numerical results, and obtain accurate approximations to the entire distribution as opposed to *first moments* as reported in Dai an Harrison [7] and some of the follow up papers. Finally, we also illustrate the flexibility of the LP approach in two ways: (i) by indicating how it can be adapted to account for parameter uncertainty in the primitives; and (ii) explaining how it can be used to compute approximations to the stationary distribution of diffusion processes arising in the many-server heavy-traffic regime.

Related literature. The linear programming approach pursued in the present paper originates with the work of Manne [24] in the context of discrete time and finite state-space Markov chains. Hernandez and Lasserre [20] extend this to analyze the convergence of linear-programming approximations for discrete-time general state-space controlled Markov chains. In [20], the authors approximate a discrete time analog of the BAR using a discrete probability distribution and a finite subspace of test functions. The main focus there is not on the steady-state distribution, but rather on minimizing a steady-state cost function (see also [21] for further discussion and references). Mendiondo and Stockbridge [22] extends this work to the continuous time setting in the context of long-run average and discounted control problems, when the state and control spaces are assumed to be compact. The key features that distinguish our work relative to these papers are: (i) feasibility in our sequence of approximating problems is guaranteed, while the work cited above assumes feasibility in the respective measure-control space; (ii) our focus is on directly computing the stationary distribution; and (iii) we deal with particulars of the SRBM problem which involves a non-compact state space, and involves reflection at the boundaries.

The remainder of the paper. Section 2 provides a formal definition of the SRBM, a summary of known existence, uniqueness, stability results, and the BAR condition. Section 3 formulates our algorithm, and establishes its convergence. Section 4 presents applications to several illustrative SRBM instances, with extensive computational results. Section 5 extends the algorithm to the setting where there is uncertainty with regard to certain parameters in the underlying queueing system. Finally, §6 provides an application to an Ornstein-Uhlenbeck-type diffusion that arises in the so-called many server heavy-traffic regime, illustrating that the method is not limited to SRBM.

2 Semi-martingale Reflected Brownian Motion in the Orthant

This section first reviews some basic proporties of SRBM in the orthant, and then derives a result which pertains to the boundary behavior of SRBM that is essential for the purposes of the algorithm we develop in §3.

2.1 Review and Basic Properties

We start with a formal definition.

Definition 1 *For $d \geq 1$, integer, let $S := \mathbb{R}_+^d$ (the positive d-dimensional orthant). Let μ be a constant vector in $\mathbb{R}^d$, σ a $d \times d$ non-degenerate covariance matrix (symmetric and strictly positive definite), and R a $d \times d$ matrix. For each $x \in S$, an SRBM associated with the data (S, μ, σ, R, x) is a $\mathcal{F}_t$-adapted, d-dimensional process $Z = (Z(t) : t \geq 0)$ defined on some filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ such that:*

(i) $Z = X + RY$, $\mathbb{P}_x$*-a.s.,*

(ii) $\mathbb{P}_x$*-a.s., Z has continuous paths and $Z(t) \in S$ for all $t \geq 0$,*

(iii) X is a d-dimensional Brownian motion with drift vector μ, covariance matrix σ and $X(0) = x$. In addition $X(t) - \mu t$ is a $\mathcal{F}_t$-martingale,

(iv) Y is an $\mathcal{F}_t$-adapted d-dimensional process such that under $\mathbb{P}$ it satisfies for each $j = 1, \ldots, d$:

a.) $Y(0) = 0$,

b.) $(Y_i(t) : t \geq 0)$ is continuous and non-decreasing,

c.) $Y_i(t)$ can increase only when Z hits the face $F_i = \{x \in S : x_i = 0\}$.

Loosely speaking, an SRBM Z behaves like the Brownian motion X in the interior of S, and is confined to the orthant by instantaneous "reflection" at the boundary faces: when Z hits F_i, the process Y_i increases, pushing Z in the direction R_i and keeping it in S. The most general condition currently known to ensure existence and uniqueness (in law) of SRBM in the orthant is that the matrix R be *completely* $\mathcal{S}$.

Definition 2 *A $d \times d$ matrix R is said to be $\mathcal{S}$ if there exits a d-dimensional vector $u \geq 0$ such that $Ru > 0$, and to be a completely $\mathcal{S}$ if each of its principal submatrices is an $\mathcal{S}$ matrix.*

This condition is in fact both necessary (Reiman and Williams [26]) and sufficient (Taylor and Williams [28]).

With regard to the stationary distribution of SRBM, which constitutes the main focus of the present paper, Dupuis and Williams [9] proved that a sufficient condition for existence is that all solutions of an associated deterministic Skorohod problem are attracted to the origin in finite time. In addition, any stationary distribution π of the SRBM Z is unique and equivalent to Lebesgue measure on S. To complete the picture, for each $i \in \{1, \ldots, d\}$ there exists a finite Borel measure ν_i on F_i such that ν_i is equivalent to the $(d-1)$-dimensional Lebesgue measure on F_i, and for each bounded Borel function f on F_i and $t \geq 0$,

$$\mathbb{E}_\pi \left[\int_0^t f(Z(s)) dY_i(s) \right] = t \int_{F_i} f(x) \nu_i(dx), \tag{4}$$

where $\mathbb{E}_\pi$ denotes expectation with respect to the stationary distribution π (i.e., when $Z(0) \sim \pi$). As alluded to earlier, the algorithm proposed in this paper is based on the BAR condition (restated here)

$$\int_S (Gf)(x)\, \pi(dx) + \sum_{i=1}^d \int_{F_i} (D_i f)(x)\, \nu_i(dx) = 0 \quad \text{for all } f \in C_b^2\,, \tag{5}$$

where the second order differential operator, G and the directional derivative, D_i are defined in (1) and (2) respectively. Necessity of the BAR was first shown by Harrison and Williams [18] when the matrix R is Minkowski ($I - R \geq 0$ and $I - R$ transient), and later by Day and Harrison [7] for the completely $\mathcal{S}$ case. Sufficiency was later proven by Dai and Kurtz [8] (see Dai [6]).

2.2 Relationship of the Interior and Boundary Measures

As mentioned in §1, our plan is to approximate the BAR (5) by a finite system of linear equations with a finite number of unknowns, and using these solutions to construct a sequence of probability measures converging to π. Since S is not compact, we will need to ensure the tightness of such a sequence, which can be done by bounding expectations for both interior and boundary measures.

Using an explicit Lyapunov function argument, Glynn and Zeevi [12] prove the existence of exponential moments of the interior measure for the case of R being symmetric and positive definite. Budhiraja and Lee [5] provide a more general result which hinges on the existence of a suitable Lyapunov function for the completely $\mathcal{S}$ case. Using this, we can show the following for the boundary measures.

Theorem 1 *Assume R is completely-$\mathcal{S}$, and suppose π is the stationary distribution for Z associated with boundary measures ν_i , $i = 1, \ldots, d$. Then, there exits a d-dimensional vector u with all* *components being strictly positive such that*

$$\int_{F_i} e^{u \cdot x} \, \nu_i(dx) < \infty \quad , \text{ for } i = 1, \ldots, d.$$

Proof Consider a d-dimensional vector $v > 0$ such that $R^\top v > 0$ (we know such vector exists, since the completely-$\mathcal{S}$ class is closed under transposition; see Reiman and Williams [26] for a proof). Fix $a > 0$ and let $f(x) = \exp(ax \cdot v)$. Applying Itô's formula, we have that

$$f(Z(t)) - f(Z(0)) - \int_0^t (Gf)(Z(s))ds - \sum_{i=1}^{d} \int_0^t (D_i f)(Z(s))dY_i(s)$$

is a local P_x martingale with respect to a sequence of localizing stopping times $T_n := \inf\{t \geq 0, \|Z(t)\|_1 \geq n\}$ $n = 1, \ldots$. It follows that

$$\mathbb{E}_x[f(Z(t \wedge T_n))] - f(x) = \mathbb{E}_x[\int_0^{t \wedge T_n} (Gf)(Z(s))ds] + \mathbb{E}_x[\sum_{i=1}^{d} \int_0^{t \wedge T_n} (D_i f)(Z(s))dY_i(s)],$$

where $\mathbb{E}_x[\cdot] = \mathbb{E}[\cdot | Z(0) = x]$ and $a \wedge b := \min\{a, b\}$. For this choice of function f, $(Gf)(x) = f(x) \left[a(\mu \cdot v) + \frac{a^2}{2} \sum_{i=1}^d \sum_{j=1}^d \sigma_{ij} v_i v_j \right]$ and $(D_i f)(x) = af(x)[R_i \cdot v]$. Notice that

$$\mu \cdot v = \mu(R^\top)^{-1} R^\top v < 0, \tag{6}$$

since $R^{-1}\mu < 0$ must hold for π to exist. Define $\alpha := -a(\mu \cdot v) > 0$ and $\beta_i := aR_i \cdot v > 0$, $i = 1, \ldots, d$. The next inequality follows from the sign of α and the nonnegativity of f.

$$\mathbb{E}_x[f(Z(t \wedge T_n))] - f(x) + \alpha \mathbb{E}_x[\int_0^t f(Z(s))ds] \geq \sum_{i=1}^{d} \beta_i \mathbb{E}_x[\int_0^{t \wedge T_n} f(Z(s))dY_i(s)].$$

Now notice that

$$\begin{aligned}
\mathbb{E}_x[f(Z(t \wedge T_n))] &= \mathbb{E}_x[f(Z(t \wedge T_n)) \, ; \, T_n > t] + \mathbb{E}_x[f(Z(t \wedge T_n)) \, ; \, T_n \leq t] \\
&\leq \mathbb{E}_x[f(Z(t)) \, ; \, T_n > t] + \phi(n)\mathbb{P}[T_n \leq t] \\
&\leq \mathbb{E}_x[f(Z(t))] + e^{cn}\mathbb{P}[\max_{s \in [0\,,\,t]} \|Z(s)\|_1 \geq n],
\end{aligned}$$

where $\phi(n) = \max\{f(x) \, ; \, \|x\|_1 = n\} \leq e^{cn}$ for some $c > 0$. Using the El Kharroubi oscillation inequality (Lemma 1 in [3]) we have that

$$\begin{aligned}
\max_{s\in[0\,,\,t]} \|Z(s)\|_1 &\leq K \max_{s\in[0\,,\,t]} \|x+\mu s + AB(s)\|_1 \\
&\leq K(\|x\|_1 + c_1 t + c_2 \max_{s\in[0\,,\,t]} \|B(s)\|_1) \\
&\leq K(\|x\|_1 + c_1 t + (c_2 d) \max_{s\in[0\,,\,t]} \max_{i=1,\dots,d} |B_i(s)|) \\
&= K(\|x\|_1 + c_1 t + (c_2 d) \max_{i=1,\dots,d} \max_{s\in[0\,,\,t]} |B_i(s)|),
\end{aligned}$$

where $0 \leq K < \infty$, $B(s)$ is a d-dimensional standard Brownian Motion, A is the Cholesky decomposition of σ with $\sigma := AA^\top$, c_1 is a constant such that $\|\mu\|_1 \leq c_1$, and c_2 is such that $\sum_{ij} |A_{ij}| \leq c_2$. Using this, we have that

$$\begin{aligned}
\mathbb{P}[\Big\{\max_{s\in[0\,,\,t]} \|Z(s)\|_1 \geq n\Big\}] &\leq \mathbb{P}[\Big\{\max_{i=1,\dots,d} \max_{s\in[0\,,\,t]} |B_i(s)| \geq (n - K(\|x\|_1 + c_1 t))/c_2\Big\}] \\
&\leq d\mathbb{P}[\Big\{\max_{s\in[0\,,\,t]} |B_1(s)| \geq (n - K(\|x\|_1 + c_1 t))/Kc_2\Big\}] \\
&\leq 2d\mathbb{P}[\Big\{\max_{s\in[0\,,\,t]} B_1(s) \geq (n - K(\|x\|_1 + c_1 t))/Kc_2\Big\}].
\end{aligned}$$

A standard Gaussian tail bound gives $\mathbb{P}[\{\max_{s\in[0\,,\,t]} B_1(s) \geq x\}] \leq \sqrt{\frac{t}{2\pi}\frac{4}{x}} e^{-x^2/2t}$, therefore we can conclude that

$$\mathbb{P}[\Big\{\max_{s\in[0\,,\,t]} \|Z(s)\|_1 \geq n\Big\}] \leq k_1\sqrt{t}e^{-k_2 n^2/t}$$

for some $0 < k_1, k_2 < \infty$, for $n > K(\|x\|_1 + c_1 t)$. Combining these results we have that for n large enough,

$$\mathbb{E}_x\left[\sum_{i=1}^d \beta_i \int_0^{t\wedge T_n} f(Z(s))dY_i(s)\right] \leq \mathbb{E}_x[f(Z(t))] + k_1\sqrt{t}e^{cn-k_2n^2/t} - f(x) + \alpha\mathbb{E}_x\int_0^t f(Z(s))ds.$$

Letting $n \to \infty$, and invoking monotone convergence on the right-hand-side (since $f \geq 0$) we have;

$$\mathbb{E}_x\left[\sum_{i=1}^d \beta_i \int_0^t f(Z(s))dY_i(s)\right] \leq \mathbb{E}_x[f(Z(t)] - f(x) + \alpha\mathbb{E}_x\int_0^t f(Z(s))ds,$$

where $\mathbb{E}_x[f(Z(t))]$ is clearly finite valued for all $x \in \mathbb{R}_+^d$ and $t \geq 0$. Taking expectations with respect to π, and considering a small enough value of a for $\mathbb{E}_\pi[f(Z(0))]$ to exist, we have

$$\sum_{i=1}^d \beta_i \mathbb{E}_\pi\left[\int_0^t f(Z(s))dY_i(s)\right] \leq \mathbb{E}_\pi[f(Z(0))] - \mathbb{E}_\pi[f(x)] + t\alpha\mathbb{E}_\pi[f(x)].$$

Now, for $M > 0$ consider $f_M(x) = f(x)\mathbf{1}\{f(x) \le M\}$. Using (4) we have

$$\mathbb{E}_\pi \left[\int_0^t f_M(Z(s)) dY_i(s) \right] = t \int_{F_i} f_M(x) \nu_i(dx).$$

Leting $M \uparrow \infty$ and using monotone convergence, we conclude that

$$\mathbb{E}_\pi \left[\int_0^t f(Z(s)) dY_i(s) \right] = t \int_{F_i} f(x) \nu_i(dx), \quad i = 1, \ldots, d.$$

It follows that

$$\sum_{i=1}^d \beta_i \int_{F_i} f(x) \nu_i(dx) \le \alpha \mathbb{E}_\pi[f(x)], \tag{7}$$

and thus

$$\int_{F_i} e^{av \cdot x} \nu_i(dx) \le \frac{\alpha}{\beta_i} \mathbb{E}_\pi[e^{a\, v \cdot x}] < \infty, \; i = 1, \ldots, d.$$

This concludes the proof. ■

Remark 1 *A direct consequence of Theorem 1 is the finiteness of the moment generating function of the stationary boundary measures in a neighborhood of zero. The nature of the result displayed in (7) may be related to a conjecture made by Harrison and Williams [18] regarding the connection between the interior and boundary measures. In particular, based on the analysis on the skew-symmetric case, they conjecture that π should have a density function p on S such that the boundary measure ν_i has density $\sigma_{ii}/(2R_{ii})p$ on F_i.*

3 The Proposed Algorithm

3.1 LP Formulation

The BAR condition (5) which characterizes the stationary distribution π, can be visualized as an infinite dimensional system of linear equations, and as such it is impractical to solve it directly. The approach we adopt here it to develop a sequence of approximations which involve: (a) a finite system of linear equations; and (b) a suitable discretization of π. Let $\overline{S}$ be a dense countable subset of S, and consider an increasing sequence of finite cardinality sets $S_n \subseteq \overline{S}$ such that $S_n \uparrow \overline{S}$, and $0 \in S_n$ for all $n \ge 0$. Also, let $\{\mathcal{F}_m\}$ be an increasing sequence of finite dimensional subspaces of C_b^2, such that for all $f \in C_b^2$ and for any compact set $A \subseteq S$, there exists a sequence $\{f_m\}$ with $f_m \in \mathcal{F}_m$ for each $m = 1, \ldots$, such that

$$\lim_{m \to \infty} \sup_A |(G(f_m - f))(x)| = 0,$$

$$\lim_{m\to\infty} \sup_A |(D_j(f_m - f))(x)| = 0\,, \text{ for all } j = 1,\ldots,d,$$

where G is the second order operator defined in (1), and D_i are the differential operators defined in (2). Consider the following linear program.

$(\mathcal{P}_{nm})$ min u

$$\text{s.t.} \quad \left|\sum_{i:\, x_i\in S_n} (Gf)(x_i)\lambda_i + \sum_{j=1}^{d} \sum_{i:\, x_i\in F_j^n} (D_j f)(x_i)\gamma_{ji}\right| \le u\ , \text{ for all } f\in\mathcal{F}_m \ \ [\approx \text{BAR}] \tag{8}$$

$$\sum_{i:\, x_i\in S_n} \lambda_i = 1 \qquad \text{[Normalization]} \tag{9}$$

$$\sum_{i:\, x_i\in S_n} g(x_i)\lambda_i \le M \qquad \text{[Interior measure tightness]} \tag{10}$$

$$\sum_{i:\, x_i\in F_j^n} g_i(x_i)\gamma_{ji} \le M_j\ , \text{ for all } j\in\{1\ldots,d\} \qquad \text{[Boundary measure tightness]} \tag{11}$$

$$\lambda_i,\, \gamma_{ij} \ge 0\ , \ \text{ for all } \{i : x_i\in S_n\} \text{ and } j\in\{1,\ldots,d\}\,. \qquad \text{[Nonnegativity]} \tag{12}$$

In this formulation, $\{\lambda_i : i \text{ s.t. } x_i\in S_n\}$ is a discrete distribution, with S_n as its support, that approximates π, and $\left\{\gamma_{ji} : i \text{ s.t. } x_i\in F_j^n\right\}$ is a discrete approximation to ν_j , $j = 1,\ldots,d$. The constants M_j $j = 1,\ldots,d$ and M, and the functions $g_j(\cdot)$, $j = 1\ldots,d$ and $g(\cdot)$ in (10)) and (11) are specified as follows. First, the function $g : S\to\mathbb{R}_+$ is chosen such that $g(x)\le g(y)$ for $x\le y$ componentwise, and such that $g(x)\to\infty$ as $\|x\|_1\to\infty$, and with $M > 0$ such that $\mathbb{E}_\pi(g(Z))\le M$. The functions $g_i : F_i\to\mathbb{R}_+$ are chosen such that $g_j(x)\le g_j(y)$ for $x\le y$ componentwise, and such that $g_i(x)\to\infty$ as $\|x\|_1\to\infty$, with $M_i > 0$ such that $\int_{F_i} g_i(x)\nu_i(dx)\le M_i$. The constraint (8) is a discrete approximation for the BAR, while constraints (9) and (12) impose the required regularity on $\{\lambda_i\}$ and $\{\gamma_{ji}\}$, $j = 1,\ldots,d$. With a forward view towards the convergence result in §3.2, constraints (10) and (11) are put in place to impose tightness on the sequence of solutions to these LP's. The actual specification of the functions and constants in (10) and (11) will be discussed in §3.3.

Remark 2 *Our LP formulation proposes to approximate π by a discrete distribution with support on S_n. Notice that this is equivalent to approximate π by a convex combination of one point distribution functions, where λ_i is the weight given to the distribution function that assigns probability one to the point $x_i\in S_n$. In more generality, for a given n, one can try to approximate (π,ν) as a convex combination of generic distribution functions. For this approach to make sense, $\{S_n\}$, now viewed as a sequence of distribution function spaces, must be such that, for any probability measure μ and finite measures ν_i $i = 1,\ldots,d$, there exists a sequences $\{\mu_n\}$ and $\{\nu_{ni}\}$ $i = 1,\ldots,d$ with $(\mu_n,\nu_n)\in S_n$,* *such that* $\mu_n \Rightarrow \mu$ *and* $\nu_n \Rightarrow \nu$*. We express this idea in more detail on* §7.

3.2 Convergence Result

Let $(\lambda, \underline{\gamma})$ be a feasible solution to $(\mathcal{P}_{nm})$, where $\underline{\gamma} := (\gamma_1, \ldots, \gamma_d)$. Define

$$\pi^n_\lambda(x) := \sum_{i:\, x_i \in S_n} \lambda_i \delta_{x_i}(x),$$

$$\nu^n_{\underline{\gamma}j}(x) := \sum_{i:\, x_i \in F^n_j} \gamma_{ji} \delta_{x_i}(x) \quad\quad j = 1, \ldots, d,$$

where $\delta_y(x) = 1$ if $x = y$ and 0 otherwise. These discrete measures provide an approximation to the true underlying probability measure π and boundary measures $\nu := (\nu_1, \ldots, \nu_d)$. The following is the main result of this paper, and characterizes the properties of the sequence of approximating distributions generated by the LP $(\mathcal{P}_{nm})$. We use the notation "$\Rightarrow$" to mean weak convergence.

Theorem 2 *Consider an SRBM Z as given in definition 1. Let* π *be the stationary distribution for Z, with boundary measures* ν_j *,* $j = 1, \ldots, d$*. Then, there exist sequences* $\{n_k\}$ *and* $\{m_k\}$*,* $k = 1, 2, \ldots$*, of positive integers that diverge to infinity as* $k \to \infty$*, such that* $\pi^{n_k m_k}_{\lambda^k} \Rightarrow \pi$ *and* $\nu^{n_k m_k}_{\underline{\gamma}^k j} \Rightarrow \nu_j$ *as* $k \to \infty$*, for* $j = 1, \ldots, d$*, where* $(\lambda^k, \underline{\gamma}^k)$ *are optimal solutions to the LP* $(\mathcal{P}_{n_k m_k})$*.*

Proof First, consider sequences $\{\phi^n\}$, and $\{\varphi^n_1\}, \ldots, \{\varphi^n_d\}$, of probability and finite measures on S and $F_1, \ldots, F_d$, respectively, such that

(a) for every $n \geq 1$, ϕ^n (φ^n_j) has support on S_n (F^n_j), that is, ϕ^n (φ^n_j) is of the form $\phi^n(x) = \pi^{nm}_{\theta^n}(x)$ $(\varphi^n_j(x) = \nu^{nm}_{\vartheta^n j}(x))$ for some vector $\theta^n \geq 0$ $(\vartheta^n \geq 0)$ such that $\sum_i \theta^n_i = 1$ $(\sum_i \vartheta^n_i < \infty)$;

(b) the sequences $\{\phi^n\}$ and $\{\varphi^n_1\}, \ldots, \{\varphi^n_d\}$ converge weakly to π and $\nu_1, \ldots, \nu_d$, respectively;

(c) $\int_S g(x)\phi^n(dx) \leq \int_S g(x)\pi(dx)$, and $\int_{F_j} g(x)\varphi^n_j(dx) \leq \int_{F_j} g_j(x)\nu_j(dx)$, for $j = 1, \ldots, d$, for all $n \geq 1$.

It is straightforward to ensure the first two properties above. We next explain how the third can be made to hold. Fix $n \geq 1$. Without lost of generality, assume the elements of $S_n = \left\{x_1, x_2, \ldots, x_{|S_n|}\right\}$ are such that $x_1 <_l x_2 <_l \ldots <_l x_{|S_n|}$, where $<_l$ represents lexicographic order, and $|\cdot|$ denotes the cardinality of a set. For $x_i \in S_n$ define $A_{x_i} := \{x \in S : x_i <_l x\} \cup \{x_i\}$. Put $\widehat{A}_i := \{x \in A_{x_i} : i = \max\{k : x \in A_{x_k}\}\}$. Notice that $\left\{\widehat{A}_i\right\}$ defines a partition of S. Set $\theta_i := \pi(\widehat{A}_i)$, i.e., $\int_{x \in \widehat{A}_i} \pi(dx)$. Put $\theta^n := (\theta_1, \ldots, \theta_{|S_n|})$, and $\phi^n := \pi^{nm}_{\theta^n}$. Consider a bounded continuous function $f : S \to \mathbb{R}$, and the simple function approximation $f_n(x) := f(x_i)$ for $x \in \widehat{A}_i$. Since $S_n \uparrow S$ we have that $f_n \to f$ by continuity of f. Hence,

$$\int_S f(x)\phi_n(x) = \int_S f_n(x)\pi(x) \to \int_S f(x)\pi(x) \quad \text{as } n \to \infty,$$

by bounded convergence. Since this holds for all bounded continuous f, we conclude that $\phi_n \Rightarrow \pi$ as $n \to \infty$. On the other hand, by construction $g(x_i) \leq g(x)$ for all $x \in \widehat{A}_i$, for all $i = 1, \ldots, |S_n|$. Therefore

$$\int_S g(x)\phi_n(dx) \leq \int_S g(x)\pi(dx).$$

The same argument applies to each of the boundary measure approximations.

Notice that (c) implies that $(\theta^n, \vartheta^n_j)$ is $(\mathcal{P}_{nm})$-feasible for all m and n. Let $\mathcal{M}$ denote the space of probability measures on S, and $\mathcal{M}_j$ the space of finite measures on F_j , $j = 1 \ldots, d$. For $\phi \in \mathcal{M}$, $\varphi_j \in \mathcal{M}_j$, $j = 1 \ldots, d$, and $f \in C^2_b$, let $B : \mathcal{M} \times \mathcal{M}_1 \times \ldots \times \mathcal{M}_d \times C^2_b \to \mathbb{R}$ denote the following operator:

$$B(\phi, \varphi, f) := \int_S (Gf)(x)\, \phi(dx) + \sum_{j=1}^{d} \int_{F_j} (D_j f)(x)\, \varphi_j(dx),$$

where $\varphi := (\varphi_1, \ldots, \varphi_d)$. Fix n, and consider the sequence $\left\{(\lambda^n_m, \underline{\gamma}^n_m)\right\}$ of optimal solutions to $(\mathcal{P}_{nm})$, $m = 1, 2 \ldots$. This sequence is tight, so it converges weakly, along a subsequence $\{m^n_k\}$, to a probability distribution $\hat{\pi}^n$ and finite measures $\hat{\nu}^n := (\hat{\nu}^n_1, \ldots, \hat{\nu}^n_d)$. Now, for any function $f \in C^2_b$ Consider the sequence $\{f_m\}$ with $f_m \in \mathcal{F}_m$ such that $\lim_{m\to\infty} \max_{x\in S_n} |(G(f - f_m))(x)| = 0$, and $\lim_{m\to\infty} \max_{x\in F^n_j} |(D_j(f - f_m))(x)| = 0$. Put $\pi^n_m := \pi^n_{\lambda^n_m}$, and $\nu^n_m := \nu^n_{\underline{\gamma}^n_m}$. We have that

$$\begin{aligned}
|B(\hat{\pi}^n, \hat{\nu}^n, f)| &= |\lim_{k\to\infty} B(\pi^n_{m^n_k}, \nu^n_{m^n_k}, f)| \\
&= |\lim_{k\to\infty} B(\pi^n_{m^n_k}, \nu^n_{m^n_k}, f - f_{m^n_k}) + \lim_{k\to\infty} B(\pi^n_{m^n_k}, \nu^n_m, f_{m^n_k})| \\
&\leq |\lim_{k\to\infty} B(\phi^n, \varphi^n, f_{m^n_k})| \\
&= |B(\phi^n, \varphi^n, f)|.
\end{aligned}$$

The sequence $\left\{(\hat{\pi}^n, \hat{\nu}^n)\right\}$, $n = 1, 2, \ldots$ is also tight, so it converges weakly, along a subsequence $\{n_k\}$, to a probability distribution $\hat{\pi}$ and finite measures $\hat{\nu} := (\hat{\nu}_1, \ldots, \hat{\nu}_d)$. From the above we have that

$$|B(\hat{\pi}, \hat{\nu}, f)| = |\lim_{k\to\infty} B(\hat{\pi}^{n_k}, \hat{\nu}^{n_k}, f)| \leq |\lim_{k\to\infty} B(\phi^{n_k}, \varphi^{n_k}, f)| = |B(\pi, \nu, f)| = 0.$$

Applying a diagonal argument, the result follows from the uniqueness of the stationary distribution, and hence $\hat{\pi}$ and $\hat{\nu}$ are solutions to the BAR. ■

3.3 Practical Considerations

In order to implement the algorithm described §3.1, we first need to specify the sequence of subsets of S, $\{S_n\}$, and the sequence of finite dimensional spaces $\{\mathcal{F}_m\}$ approximating C_b^2. Moreover, the specification of the algorithm requires a specification of the tightness conditions. That is, we need to specify feasible values of M and M_j for $j = 1 \dots, d$, as well as functions g and g_j , $j = 1, \dots, d$. In this section, we first discuss the specification of these constants and then we provide a specific choice $\{S_n\}$ and $\{\mathcal{F}_m\}$.

Tightness of the interior distribution. We use a Lyapunov-function type argument. For illustration purposes we will assume that R is symmetric and that R^{-1} is positive definite; this facilitates the construction of a simple Lyapunov function (a more general, but less implicit argument can be followed using the Lyapunov function identified in Dupuis and Williams [9]). Consider $f(x) = x \cdot R^{-1}x$. Following the steps in the proof of Theorem 1, applying Ito's formula and taking expectations with respect to $\mathbb{P}_x$, we have

$$\mathbb{E}_x[f(Z(t \wedge T_n))] - f(x) + \mathbb{E}_x[\int_0^{t \wedge T_n} Z(s) \cdot \varrho \, ds] = Ct + \mathbb{E}_x \left[\sum_{i=1}^d \int_0^t Z_i(s) dY_i(s) \right],$$

where $C = \sum_{ij} \sigma_{ij} (R^{-1})_{ij} > 0$, σ_{ij} is the (i,j)-element of σ, $\varrho := -R^{-1}\mu > 0$, and $T_n := \inf\{t \geq 0\,;\, |Z(t)| \geq n\}$. Notice that $\mathbb{E}_x[\sum_{i=1}^d \int_0^t Z_i(s) dY_i(s)] = 0$ by definition of the SRBM; see Definition 1. Considering the sign of ϱ, and that $f \geq 0$, we have that

$$-f(x) + \mathbb{E}_x[\int_0^{t \wedge T_n} Z(s) \cdot \varrho \, ds] \leq Ct.$$

After applying the Monotone Convergence to the left-hand-side, and taking expectations with respect to π, we conclude that

$$\int_S (e \cdot x) \, \pi(dx) \leq \frac{C}{\min_j \{\varrho_j\}},$$

with $e = (1, \dots, 1)^\top \in \mathbb{R}^d$. Here we can take $g(x) = e \cdot x$ and $M = C/\min_j\{\varrho_j\}$. For the case where R is Minkowski, a bound can be derived directly from lemma 8.4 in Harrison and Williams [18].

Tightness of the boundary measures. Take a d-dimensional vector $v > 0$ such that $R^\top v > 0$ (which follows from the completely-$\mathcal{S}$ condition) and construct a $d \times d$ symmetric matrix V by setting $V_{ij} = v_j v_i > 0$. It follows that $(RV)_{ij} > 0$ for $i, j = 1, \dots, d$. Let $f(x) = x \cdot Vx$. Note that $(Gf)(x) = [2\mu \cdot Vx + \sum_{ji} v_i v_j \sigma_{ij}]$. Following the steps in the proof of Theorem 1, for $n \geq 1$, we have

$$\mathbb{E}_x[f(Z(t))] + k_1\sqrt{t}n^2 e^{-k_2 n^2/t} + \mathbb{E}_x[\int_0^t Z(s)\cdot\zeta ds] - f(x) \quad \geq \quad \mathbb{E}_x\left[\sum_{i=1}^d \int_0^{t\wedge T_n} Z(s)\cdot(R_i V)\, dY_i(s)\right],$$

where $\zeta := V\mu > 0$, following (6). Letting $n \to \infty$ and taking expectations with respect to π, and using (4) as in proof of Theorem 1, we have that

$$\int_{F_i} (e\cdot x)\,\nu_i(dx) \leq \frac{\alpha}{\beta_i}\int_S e\cdot x\,\pi(dx),$$

$i = 1, \ldots, d$, where $\alpha := \max_j \{\zeta_j\} > 0$ and $\beta_i := \min_j \{(VR_i)_j\} > 0$, $i = 1, \ldots, d$. Combining this result with the previous one, we can choose $g_i(x) = e\cdot x$, and $M_i = (C\alpha)/(\beta_i \min_j \{\varrho_j\})$.

Support and function space approximation. Harrison and Williams [19] proved that the stationary distribution for a standard SRBM[1] has a separable density function (in the usual Cartesian coordinates) if and only if the covariance matrix σ satisfies a "skew symmetry" condition:

$$2\sigma_{ij} = R_{ij} + R_{ji} \quad \text{for} \quad i \neq j\, \in \{1, \ldots, d\}\,. \tag{13}$$

In this case, the marginal distribution for each coordinate i is exponential with rate $2\varrho_i := (-2R^{-1}\mu)_i$. Moreover, the boundary measures are the restriction of the joint distribution to the corresponding faces of S (up to a scaling factor). Recently, Budhiraja and Lee [5] established that the moment generating function of the stationary distribution is finite in a neighborhood of zero under the sufficient stability condition of Dupuis and Williams [9].

The above observation motivates taking $\{S_n\}$ to have an "exponential spacing" on the marginals, that is for each $n \geq 1$,

$$S_n = \{x \in S\,:\, x_i \in \{[\log(n) - \log(k)]/(2\varrho_i) \quad, \text{for } k = 1, \ldots, n\} \quad, \text{for } i = 1, \ldots, d\}\,. \tag{14}$$

For the boundary, we simply project this grid on the corresponding faces, $F_j^n = \{x \in S_n\,;\, x_j = 0\}$, $j = 1, \ldots, d$. This structure of S_n is related to the choice of a reference density in the Dai-Harrison algorithm [7]. Regarding the function spaces approximating C_b^2, for each $m \geq 1$ let $\mathcal{H}_m$ be the linear space spanned by $\left\{x_1^{p_1}\cdots x_d^{p_d}\,,\, p_1 + \ldots + p_d \leq m\,,\, p_1, \ldots, p_d \text{ are nonnegative integers}\right\}$. Take $\epsilon > 0$ and define $\mathcal{H}_m^l := \{f \in C_b^2\,:\, f(x) = g(x)$ for $\|x\|_1 \leq l$ for some $g \in \mathcal{H}_m$ such that $f(x) = c$ for $\|x\|_1 > l + \epsilon$ for some constant $c\}$. For each $m \geq 1$ we take $\mathcal{F}_m = \bigcup_{1\leq l\leq m} \mathcal{H}_m^l$.

Remark 3 *A practical choice of the exponential spacing can be had using a rate vector $C\varrho$, with $0.5 \leq C \leq 2$. Also, we use $\mathcal{F}_m := \mathcal{H}_m$ in the numerical experiences.*

[1] A SBRM is said to be standard if $\sigma_{ii} = 1\, i = 1, \ldots, d$, R is invertible and $\|\varrho\|_\infty = 1$.

m	3	4	5	6	7	8	9	10
$\mathbb{E}_\pi[Z_1]$	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
$\mathbb{E}_\pi[Z_2]$	0.793	0.821	0.764	0.768	0.762	0.758	0.760	0.750

Table 1: **Moment estimates for the SRBM in example 1**. The algorithm uses the grid S_n as in (14) with $n = 100$, and $m =$ degree of polynomials. True moments are $\mathbb{E}_\pi[Z_1] = 0.5$ and $\mathbb{E}_\pi[Z_2] = 0.75$.

4 Numerical Results

In this section we compare the performance of our algorithm to some known analytical results of particular instances of SRBM, as well as other numerical methods proposed in antecedent literature.

Example 1: A Two-Dimensional special case. Consider a two-dimensional SRBM $Z = (Z(t) : t \geq 0)$ associated to the following data:

$$R = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \quad , \quad \sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \mu = \begin{pmatrix} \mu_1 \\ 0 \end{pmatrix}.$$

For this SRBM the stationary condition reduces to $\mu_1 < 0$. Harrison [16] computed a closed form solution for the density of the stationary distribution in polar coordinates:

$$p(x_1, x_2) = (2|\mu_1|)^{3/2}/(\pi^{1/2})r^{-1/2} \exp(\mu_1 r(1 + \cos(\theta))) \cos(\theta/2),$$

where $(x_1, x_2) = (r\cos(\theta), r\sin(\theta))$. This SRBM is the only instance (excluding the skew symmetry condition) for which the stationary distribution can be computed in closed form.

Without loss of generality consider $\mu_1 = -1$. It can be shown (Greenberg [13]) that

$$\mathbb{E}_\pi[Z_1] = 0.5 \quad , \quad \mathbb{E}_\pi[Z_2] = 0.75.$$

Taking S_n as in (14), with $n = 100$, we use the proposed algorithm to get estimates for these moments. The results are displayed on Table 1 for various degrees of polynomials that compose $\mathcal{F}_m$. We see that the estimates are quite accurate even for low values of m. Each one of the instances shown in Table 1 ran in less than 1 minute on a regular Desktop PC with processor Intel Pentium D (3.2Ghz), using a MATLAB implementation of the algorithm. Figure 1 gives the marginal distribution estimates for $n = 100$ and $m = 6$ (solid line), compared with the true marginal distributions (doted line).

Example 2: A Symmetric case. A standard SRBM is said to be *symmetric* if its data has the following properties: $\sigma_{ij} = \rho$ for $1 \leq i < j \leq d$, $\mu_i = -1$ for $i = 1, \ldots, d$ and $R_{ij} = R_{ji} = -r \leq 0$ for $1 \leq i < j \leq d$. That σ is positive implies that $-1/(d-1) < \rho < 1$, and the completely $\mathcal{S}$

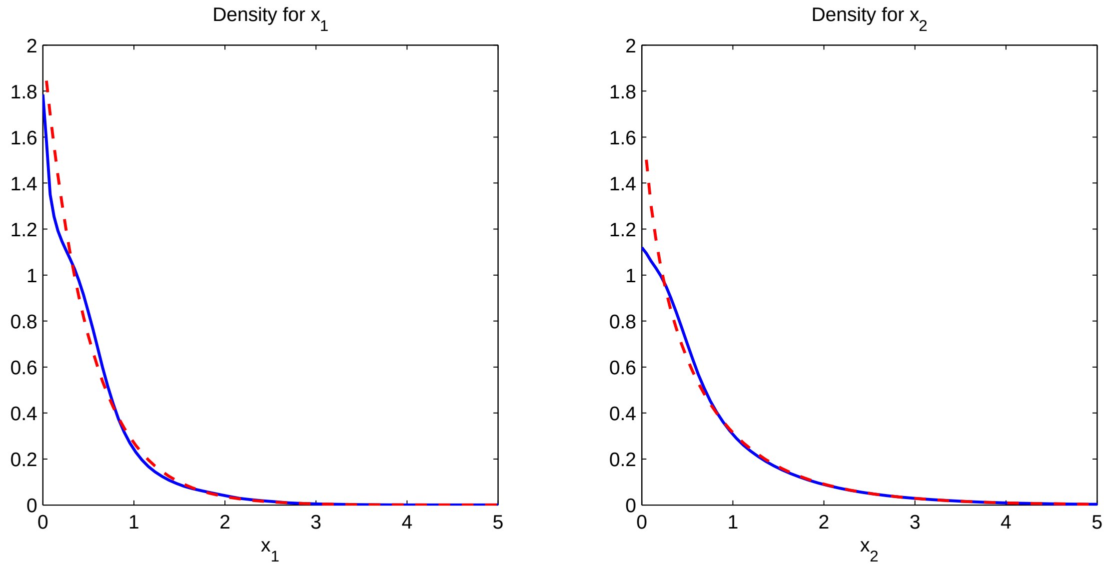

Figure 1: **Marginal distribution estimates for the SRBM in Example 1.** The algorithm uses the grid S_n as in (14) with $n = 100$, and $m = 6$ degree polynomials. The marginal distribution estimates (solid line) are superimposed on the true marginals (dotted line)

$r \backslash \rho$	-0.9	-0.5	0.0	0.5	0.9
0.2	1.98e-4	1.65e-4	1.26e-3	2.32e-3	6.66e-4
0.4	3.90e-4	4.88e-5	2.52e-5	2.92e-3	2.00e-4
0.6	1.76e-5	5.97e-4	2.05e-4	1.06e-6	5.70e-4
0.8	1.53e-5	4.27e-5	2.22e-5	1.17e-3	2.16e-3
0.9	1.65e-5	4.26e-5	5.43e-4	1.00e-3	3.24e-3
0.95	4.57e-6	3.42e-5	2.46e-4	5.98e-4	4.82e-3

Table 2: **Relative errors for the SRBM in Example 2.** Table depicts relative errors between moment estimates and true moments for various values of r and ρ. Here $n = 100$ and $m = 6$.

condition reduces to $r(d-1) < 1$. This type of SRBM arises as a heavy-traffic limit of a symmetric generalized Jackson network. Manipulating the BAR condition, Harrison and Dai [7] showed that

$$\mathbb{E}(Z_1) = \ldots = \mathbb{E}(Z_d) = \frac{1 - (d-2)r + (d-1)r\rho}{2(r+1)}.$$

We use our algorithm to compute estimates for these moments for the case $d = 2$. Imitating the work of Harrison and Dai [7], we let ρ take values in $\{-0.9, -0.5, 0.0, 0.5, 0.9\}$ and r in $\{0.2, 0.4, 0.6, 0.8, 0.9, 0.95\}$. Table 2 shows the relative errors between our estimates and the exact values for $m = 6$ and $n = 100$. The relative error is always lower that 1%.

Example 3: Skew-symmetric SRBM. As indicated earlier, Harrison and Williams [19] proved that a standard SRBM has a product form stationary distribution if and only if $\varrho < 0$ and condition (13) holds. In this, case we know that the i-th marginal distribution is exponential with

m	3	4	5	6	7	8	9	10
$\mathbb{E}_\pi[Z_1]$	0.4942	0.4943	0.4968	0.5014	0.5005	0.5011	0.4996	0.5002
$\mathbb{E}_\pi[Z_2]$	2.0095	2.0094	2.0053	1.9976	1.9990	1.9981	2.0005	1.9996

Table 3: **Moment estimates for the SRBM in Example 3.** The algorithm uses $n = 100$, and $m =$ degree of polynomials. True moments are $\mathbb{E}_\pi[Z_1] = 0.5$ and $\mathbb{E}_\pi[Z_2] = 2.0$.

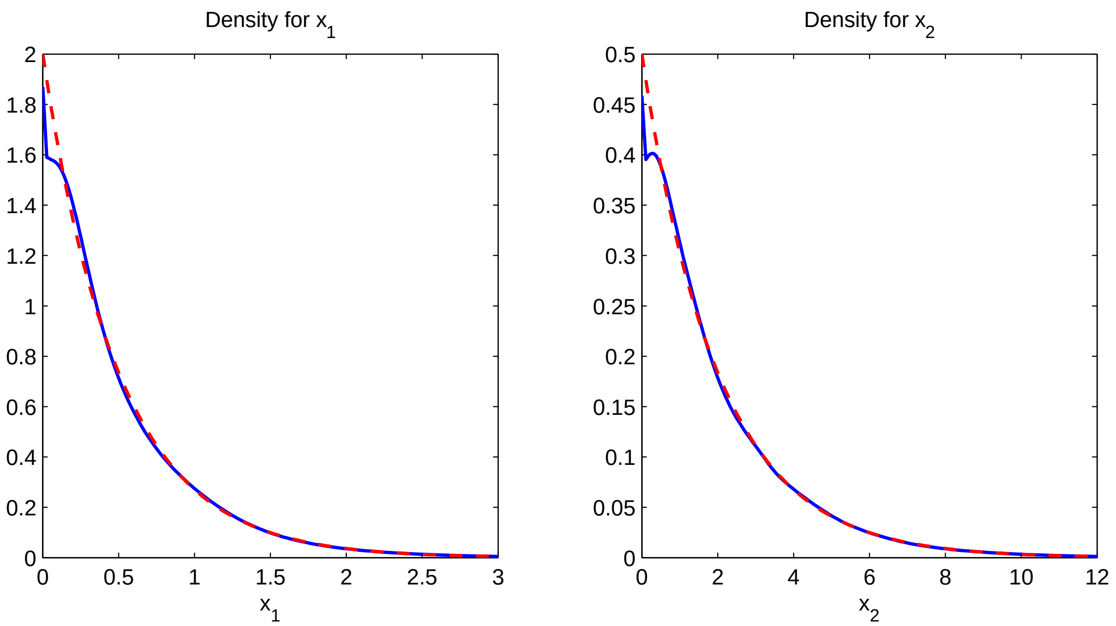

Figure 2: **Marginal distribution estimates for the SRBM in Example 3.** The algorithm uses the grid S_n as in (14) with $n = 150$, and $m = 10$ degree polynomials. The marginal distribution estimates (solid line) are superimposed on the true marginals (dotted line).

mean $1/(2\varrho_i)$, $i = 1, \ldots, d$. Consider a two-dimensional SRBM associated with the following data:

$$R = \begin{pmatrix} 1 & -0.6 \\ -0.25 & 1 \end{pmatrix} \quad , \quad \sigma = \begin{pmatrix} 1 & -0.425 \\ -0.425 & 1 \end{pmatrix} \quad , \quad \mu = \begin{pmatrix} -0.85 \\ 0 \end{pmatrix}.$$

One can check that condition (13) holds, and that $\varrho^\top = (-1, -0.25)$. This implies that $\mathbb{E}_\pi[Z_1] = 0.5$ and $\mathbb{E}_\pi[Z_2] = 2$. Taking $n = 100$ we use our algorithm to get estimates for these moments for various values of m, the mayor degree of the polynomials, and the results are summarized in Table 3. We see that our algorithm provides good estimates even for low degree polynomials. Each one of the instances shown in Table 3 ran in less than 1 minute on a regular desktop PC with processor Intel Pentium D (3.2Ghz), using a MATLAB implementation of the algorithm. Figure 2 depicts marginal distribution estimates for $n = 150$ and $m = 10$ (solid line), compared with the actual marginal distributions (doted line).

Example 4: Suresh and Whitt's Experiments. This section follows closely the analysis presented in Chapter 4 of Dai [6]. Consider a network of d queues in tandem. Let ρ_i denote the mean service time at station i, $C_{s_i}^2$ denote the squared coefficient of variation of the service time distribution at station i, and C_a^2 denote the squared coefficient of variation for the interarrival time distribution. Using the method proposed by Harrison and Nguyen [17], Dai [6] shows that the d-dimensional current workload process can be approximated by a d-dimensional SBRM as in Definition 1. When $d = 2$ the parameters (σ, R, μ) defining Z are given by

$$R = \begin{pmatrix} 1 & 0 \\ -\rho_2/\rho_1 & 1 \end{pmatrix} \quad , \quad \sigma = \begin{pmatrix} \rho_1^2(C_a^2 + C_{s_1}^2) & -\rho_1\rho_2 C_{s_1}^2 \\ -\rho_1\rho_2 C_{s_1}^2 & \rho_2^2(C_{s_1}^2 + C_{s_2}^2) \end{pmatrix} \quad , \quad \mu = \begin{pmatrix} \rho_1 - 1 \\ \rho_2/\rho_1 - 1 \end{pmatrix}.$$

The arrival rate is assumed to be 1, so that ρ_1 and ρ_2 represents the traffic intensities at station 1 and 2, respectively.

Suresh and Whitt [27] studied a system of two queues in tandem and considered various variability parameter triples $(C_a^2, C_{s_1}^2, C_{s_2}^2)$, for all combinations of the traffic intensities ρ_1 and ρ_2 in a representative range. For $C_{s_1}^2 \neq C_{s_2}^2$ they considered five variability combinations: (0.5,0.5,2.0), (1.0,0.5,8.0), (1.0,2.0,4.0), (4.5,0.5,1.0) and , (4.0,1.0,4.0). We will refer to these as *Case 1* to *Case 5*, respectively. For $C_{s_1}^2 = C_{s_2}^2$ they consider two variability combinations: (1.0,0.5,0.5) and (1.0,4.0,4.0). We will refer to these as *Case 6* and *Case 7*, respectively. When $C_{s_1}^2 \neq C_{s_2}^2$, they consider four values of ρ_i : 0.3, 0.6, 0.8, 0.9. When $C_{s_1}^2 \neq C_{s_2}^2$, for each queue they consider five values of ρ_i : 0.1, 0.2, 0.3, 0.6, 0.9, for each queue.

Extensive simulation experiments were conducted in order to obtain estimates for the expected steady-state waiting times. When $C_a^2 = 0.5$, the Erlang distribution was used with 2 degrees of freedom. When $C_a^2 = 1.0$, the exponential distribution was used. When $C_a^2 > 1.0$, the hyperexponential distribution with balanced means was used. They also compare their simulation results with the results obtained by using a software package described by Whitt [29], dubbed the QNA method. Dai [6] uses this set of experiments to test the algorithm proposed in Harrison and Dai [7], which they refer to as the QNET method. We use these results to benchmark the performance of our algorithm against the QNA and QNET methods. For the purposes of our algorithm, we used $n = 100$ and $m = 6$ for all experiments. Each run took less than 3 minutes on regular Desktop PC with processor Intel Pentium D (3.2Ghz), using a MATLAB implementation of the algorithm.

Tables 6 to 12 (see appendix A) give simulation estimates, QNET estimates, QNA estimates, and our LP-based estimates. Table 4 summarizes all balanced heavy traffic cases (equal traffic intensities in each queue, close to 1), and gives an overall comparison of QNA estimates, QNET estimates and the LP-based estimates. Order I refers to the setting where the arriving customers enters queue 1 first, then proceeds to queue 2, and then exits the system. In the same manner, Order II refers to the setting where the arriving customers enter queue 2 first, then proceed to

			Order I			Order II		
Case	ρ_1	ρ_2	QNET	QNA	LP-based	QNET	QNA	LP-based
1	0.9	0.9	0.01	0.01	0.00	0.02	0.13	0.01
	0.8	0.8	0.08	0.08	0.08	0.06	0.08	0.07
2	0.9	0.9	0.03	0.06	0.03	0.01	0.21	0.01
	0.8	0.8	0.04	0.01	0.02	0.08	0.07	0.07
3	0.9	0.9	0.04	0.02	0.04	0.04	0.19	0.05
	0.8	0.8	0.00	0.04	0.01	0.02	0.02	0.01
4	0.9	0.9	0.13	0.30	0.14	0.19	0.29	0.19
	0.8	0.8	0.14	0.08	0.11	0.09	0.00	0.07
5	0.9	0.9	0.01	0.12	0.01	0.01	0.01	0.00
	0.8	0.8	0.09	0.14	0.08	0.05	0.05	0.05
6	0.9	0.9	0.01	0.08	0.02	0.01	0.10	0.01
	0.8	0.8	0.01	0.04	0.01	0.01	0.04	0.01
7	0.9	0.9	0.04	0.18	0.05	0.06	0.07	0.04
	0.8	0.8	0.02	0.05	0.03	0.01	0.06	0.01
Average			0.05	0.09	0.04	0.05	0.09	0.04

Table 4: **Summary of comparisons between QNA, QNET and the LP-based method for Example 4.** The table depicts the minimum between the absolute relative error and absolute error with respect to the simulation estimates, for all three methods, for different orders of the queues, and for different values of $\rho_1 = \rho_2$. The last row depicts the average of the errors for the different combinations of methods/orders.

queue 1, and then exit the system. Each entry in Table 4 represents the minimum between two values: the absolute relative error of the respective method compared to the simulation estimates; and the absolute error of the respective method, again compared to the simulation estimates. As expected, the QNET and LP-based methods give more accurate estimates compared to the QNA method under balanced heavy traffic conditions. The LP-based method is seen to perform as least as well as the QNET method, and gives rise to more accurate estimates in most cases.

5 A Robust Formulation

Consider a network of 2 queues in tandem, i.e., the departure process from queue 1 forms the arrival process for queue 2. Jobs arrive to queue 1 according to an exogenous renewal process, and upon completion of processing at queue 2 they depart the system. We can characterize this simple queueing network by the 5-tuple

$$(C_a^2, C_{s_1}^2, C_{s_2}^2, \rho_1, \rho_2),$$

where C_a^2 is the squared coefficient of variation for the interarrival time distribution, ρ_i is the mean service time at station $i = 1, 2$, and $C_{s_i}^2$ is the squared coefficient of variation of the service time distribution at queue $i = 1, 2$. Assuming the arrival rate to queue 1 is one, the value of ρ_i

coincides with the traffic intensity at queue $i = 1, 2$. The 2-dimensional workload process can be approximated by a 2-dimensional SRBM Z in the orthant as given in Definition 1. Following the prescription in Harrison and Nguyen [17], it can be shown that

$$\mu = \begin{pmatrix} \rho_1 - 1 \\ \rho_2/\rho_1 - 1 \end{pmatrix}, \quad \sigma = \begin{pmatrix} \rho_1^2(C_a^2 + C_{s_1}^2) & -\rho_1\rho_2 C_{s_1}^2 \\ -\rho_1\rho_2 C_{s_1}^2 & \rho_2^2(C_{s_1}^2 + C_{s_2}^2) \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 \\ -\rho_2/\rho_1 & 1 \end{pmatrix}.$$

Using this, one can apply the LP-based algorithm to solve for the steady-state distribution of Z (see Example 4 in §4).

One potential issue that arises in practice is that the value of certain inputs parameters may not be known exactly. In what follows, we will illustrate how such parameter uncertainty can be dealt with in the context of our algorithm, via a *robust* formulation. To be concrete, suppose we are interested in incorporating uncertainty with regard to the traffic intensities at each station. Assume that the true values of (ρ_1, ρ_2) belong to the following set

$$\mathcal{Q}_\delta := \left\{ \hat{\rho}_1 + \delta\rho_1, \hat{\rho}_2 + \delta\rho_2 \ : \ \delta\rho_1^2/a_1 + \delta\rho_2^2/a_2 \le r^2 \right\},$$

with $a_1\ a_2 > 0$, and $a_1 r$, $a_2 r \ll 1$,. That is, the values of (ρ_1, ρ_2) are contained in an ellipsoid with "small" radius r, centered on $(\hat{\rho}_1, \hat{\rho}_2)$. This ellipsoidal uncertainty set is motivated by the classical robust counterpart formulation in Nemirovski and Ben-Tal [1]. "Translating" this uncertainty to the SRBM data, we have

$$\mu = \begin{pmatrix} \hat{\rho}_1 + \delta\rho_1 - 1 \\ \hat{\rho}_2/\hat{\rho}_1 - (\hat{\rho}_2\delta\rho_1)/\hat{\rho}_1^2 + \delta\rho_2/\hat{\rho}_1 - 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0 \\ -\hat{\rho}_2/\hat{\rho}_1 + (\hat{\rho}_2\delta\rho_1)/\hat{\rho}_1^2 - \delta\rho_2/\hat{\rho}_1 & 1 \end{pmatrix},$$

$$\sigma = \begin{pmatrix} (\hat{\rho}_1^2 + 2\hat{\rho}_1\delta\rho_1)(C_a^2 + C_{s_1}^2) & -(\hat{\rho}_1\hat{\rho}_2 + \delta\rho_1\hat{\rho}_2 + \hat{\rho}_1\delta\rho_2)C_{s_1}^2 \\ -(\hat{\rho}_1\hat{\rho}_2 + \delta\rho_1\hat{\rho}_2 + \hat{\rho}_1\delta\rho_2)C_{s_1}^2 & (\hat{\rho}_2^2 + 2\hat{\rho}_2\delta\rho_2)(C_{s_1}^2 + C_{s_2}^2) \end{pmatrix}.$$

Notice that both the second order differential operator G, as well as the boundary differential operator D_i are linear in $\delta\rho_1$ and $\delta\rho_2$. Consider now the BAR constraints in $(\mathcal{P}_{nm})$:

$$\left| \sum_{i:\, x_i \in S_n} (Gf)(x_i)\lambda_i + \sum_{j=1}^{d} \sum_{i:\, x_i \in F_j^n} (D_i f)(x_i)\gamma_{ji} \right| \le u, \ \text{ for all } f \in \mathcal{F}_m.$$

Define $y = (\lambda, \gamma_1, \ldots, \gamma_d)$, where $\lambda \in \mathbb{R}^{|S_n|}$ and $\gamma_i \in \mathbb{R}^{|F_j^n|}$, $j = 1, \ldots, d$. The BAR constraints can then be represented as

$$|A^{nm} y| \le eu,$$

for some matrix $A^{nm} \in \mathbb{R}^{|\mathcal{F}_m| \times |S_n| + \sum |F_j^n|}$, where $|a|$ denotes componentwise absolute value for a vector a, e is a $|\mathcal{F}_m|$-dimensional vector with all its entries equal to one, and $|\mathcal{F}_m|$ refers to the cardinality of the set of basis functions spanning $\mathcal{F}_m$. It is easy to see that A^{nm} is also linear on $\delta\rho_1$ and $\delta\rho_2$, therefore it can be represented as $A^{nm} := \hat{A}^{nm} + \delta\rho_1 A^{nm}_{\delta\rho_1} + \delta\rho_2 A^{nm}_{\delta\rho_2}$ for some matrices $\hat{A}^{nm}$, $A^{nm}_{\delta\rho_1}$, $A^{nm}_{\delta\rho_2} \in \mathbb{R}^{|\mathcal{F}_m|\times|S_n|+\sum|F^n_j|}$ independent of $\delta\rho_1$ and $\delta\rho_2$. As a consequence, when uncertainty over traffic intensities is modeled as above, we have that A^{nm} belongs to the uncertainty set

$$\mathcal{U}^{nm} = \left\{ \hat{A}^{nm} + \delta\rho_1 A^{nm}_{\delta\rho_1} + \delta\rho_2 A^{nm}_{\delta\rho_2} \quad | \, \delta\rho_1^2/a_1 + \delta\rho_2^2/a_2 \le r^2 \right\}.$$

Following Ben-Tal and Nemirovski [1], we consider the *robust counterpart* of $(\mathcal{P}_{nm})$

$(\mathcal{R}_{nm})$ min u

$$\begin{aligned}
s.t. \quad & |A^{nm}y| \le eu \quad \text{for all } A^{nm} \in \mathcal{U}^{nm} && [\approx \text{BAR + uncertainty set}] \\
& \sum_{i:\, x_i \in S_n} \lambda_i = 1 && [\text{Normalization}] \\
& \sum_{i:\, x_i \in S_n} g(x_i)\lambda_i \le M && [\text{Interior measure tightness}] \\
& \sum_{i:\, x_i \in F^n_i} g_i(x_i)\gamma_{ji} \le M_j \quad \text{for all } j \in \{1\dots,d\} && [\text{Boundary measure tightness}] \\
& y \ge 0 && [\text{Nonnegativity}]
\end{aligned}$$

where $y = (\lambda, \gamma_1, \dots, \gamma_d)$. With this robust formulation we are looking for a probability distribution and finite measures that minimize the violation of the BAR condition, for all possible realizations of $(\rho_1, \rho_2) \in \mathcal{Q}_\delta$. Let us denote by A_i the i-th row of a matrix A. The point $y = (\lambda, \gamma)$ is $(\mathcal{R}_{nm})$-feasible if it is $(\mathcal{P}_{nm})$-feasible, and

$$|(\hat{A}^{nm})_i\, y + \delta\rho_1 (A^{nm}_{\delta\rho_1})_i\, y + \delta\rho_2 (A^{nm}_{\delta\rho_2})_i\, y| \le u \quad \text{for } \delta\rho_1^2/a_1 + \delta\rho_2^2/a_2 \le r^2\,,\ i = 1, \dots, |\mathcal{F}_m|.$$

Noting that

$$\begin{aligned}
\max_{\delta\rho_1^2/a_1 + \delta\rho_2^2/a_2 \le r^2} & |(\hat{A}^{nm})_i\, y + \delta\rho_1 (A^{nm}_{\delta\rho_1})_i\, y + \delta\rho_2 (A^{nm}_{\delta\rho_2})_i\, y| \\
&= |(\hat{A}^{nm})_i\, y| + \|\sqrt{a_1} r (A^{nm}_{\delta\rho_1})_i\, y, \sqrt{a_2} r (A^{nm}_{\delta\rho_2})_i\, y\|_2,
\end{aligned}$$

we can restate $(\mathcal{R}_{nm})$ as a *conic quadratic program*, with the BAR constraint above replaced by:

$$\left| (\hat{A}^{nm})_i\, y \right| \le e \cdot u - \|b_1 (A^{nm}_{\delta\rho_1})_i\, y, b_2 (A^{nm}_{\delta\rho_2})_i\, y\|_2 \ \text{ for all } i = 1, \dots, |\mathcal{F}_m|$$

where $b_i := \sqrt{a_i} r$, $i = 1, 2$. Solving such conic-quadratic problems can be done in polynomial time using interior point methods, with essentially the same computational complexity as LP problems of similar size (see [2]).

6 An Illustrative Application in the Many-server Heavy-traffic Regime

In this section we illustrate how the proposed algorithm can be used to approximate the stationary distributions of diffusions arising under the so called many-server (or Halfin-Whitt) heavy-traffic regime. For this purpose we will consider a two class queueing system operating under a first-come first-out discipline and with class 1 receiving static priority over class 2. Specifically, class 1 arrivals, say customers, occurs according to Poisson process with rate λ_1. Requests for class 1 service engage one *server* each for i.i.d. exponentially distributed amounts of time with rate μ_1, provided that the total number of class 1 customers in the system is less than the total number of servers N; otherwise, they wait in queue for an available server. Class 2 customers are assumed to arrive according to a Poisson process with rate λ_2, independent of the arrival process for class 1 customer. Requests for class 2 service are i.i.d. exponentially distributed amounts of time with rate μ_2. When there is at least one server free and there are not class 1 customers waiting on queue, that server is allocated to a class 2 service request, and when there are no free servers, the class 2 customers wait on queue. Despite its simple structure, exact analysis of this system is not straightforward.

6.1 Diffusion Approximation and BAR condition

Maglaras and Zeevi [23] studied this system and derived a diffusion approximation under the many-server regime developed by Halfin and Whitt [15]. This regime is defined by letting the number of servers grow, and concurrently letting the system utilization approach 1 at an appropriate rate. Specifically, consider a sequence of systems with n servers and arrival rates $\lambda_i^n := n\kappa_i\mu_i - \gamma_i\sqrt{n}\mu_i$ for some constants $\kappa_i > 0$ and $\gamma_i \in \mathbb{R}$ for $i = 1, 2$ such that $\kappa_1 + \kappa_2 = 1$.

Let $Q_i^n(t)$ denote the number of customers in the system, both in service and in queue, in each customer class $i = 1, 2$. Put $X_i^n(t) := (Q_i^n(t) - \kappa_i n)/\sqrt{n}$, $i = 1, 2$. Then, if $(X_1^n(0), X_2^n(0)) \to \xi \in \mathbb{R}^2$, then $X^n = (X_1^n(t), X_2^n(t) : t \geq 0)$ converges (as $n \to \infty$) to a diffusion process $X = (X(t) : t \geq 0)$ being the unique strong solution to

$$dX(t) = b(X(t))dt + \Sigma dB(t), \qquad X(0) = \xi. \tag{15}$$

Here $B = (B(t) : t \geq 0)$ is a standard Brownian motion in $\mathbb{R}^2$, and the infinitesimal drift function

$b_i(\cdot)$ for the i-th component is giving by

$$\begin{aligned} b_1(x) &:= -\mu_1\gamma_1 - \mu_1 x_1\,, \\ b_2(x) &:= \begin{cases} -\mu_2\gamma_2 - \mu_2 x_2 & x_1 + x_2 \le 0 \\ -\mu_2\gamma_2 + \mu_2 x_1 & x_1 + x_2 > 0 \end{cases}, \end{aligned}$$

and $\Sigma := \operatorname{diag}(\sigma_1, \sigma_2)$, with $\sigma_i^2 := 2\mu_i\kappa_i$, $i = 1, 2$. The first component of X is an Ornstein-Uhlenbeck process, and the second component has more complicated structure. The process X admits a unique stationary distribution π if and only if $\gamma := \sum_{i=1}^{k} \gamma_i > 0$ (see Maglaras and Zeevi [23] for further details). For $f \in C_b^2$ define the second order differential operator

$$Lf := \sum_{i=1}^{2} \left(b_i \frac{\partial f}{\partial x_i} + \frac{\sigma_i^2}{2} \frac{\partial^2 f}{\partial x_i^2} \right).$$

Under the above stability condition, $\gamma > 0$, a direct application of Ito's lemma shows that for all $f \in C_b^2$, the stationary distribution π must satisfy the following BAR:

$$\int_{\mathbb{R}^2} \left(\sum_{i=1}^{2} b_i(x) \frac{\partial f(x)}{\partial x_i} + \frac{\sigma_i^2}{2} \frac{\partial^2 f(x)}{\partial x_i^2} \right) \pi(dx) = 0. \tag{16}$$

The sufficiency of (16) is established in the next proposition.

Proposition 1 *Suppose π is a probability measure on $\mathbb{R}^2$. If it satisfies (16), then π is the stationary distribution for X.*

Proof Consider $f \in C_b^2$, and suppose that $x_0 \in \mathbb{R}^m$ is such that $\sup_{x \in \mathbb{R}^m} f(x) = f(x_0) \ge 0$. The application of Ito' formula gives

$$f(X(t)) - \int_0^t (Lf)(X(s))ds = f(x_0) + \int_0^t \nabla f(X(s)) \cdot \Sigma dW(s).$$

Taking expectations w.r.t. P_{x_0}, we have

$$\mathbb{E}_{x_0}[f(X(t))] - f(x_0) = \mathbb{E}_{x_0}\left[\int_0^t (Lf)(X(s))ds \right].$$

Divide by t and take $t \to 0$. By the continuity of L and the continuity of X, we get $(Lf)(x_0) \le 0$, since the left hand side above is non-positive. Therefore, the operator (L, C_b^2) satisfies the positive maximum principle. Consider a probability measure π on $\mathbb{R}^2$ satisfying (16). Since $f(X(t)) - \int_0^t (Lf)(X(s))ds$ is a P_x-martingale, the measure $P_\pi \equiv \int_{\mathbb{R}^m} P_x \pi(dx)$ is a solution for the martingale problem for (L, π). C_b^2 is an algebra, and is dense in $C(\mathbb{R}^m)$. The operator (L, C_b^2) satisfies the positive maximum principle, and therefore Echeverria's theorem (see [10], Theorem 9.14 of Chapter 4) applies to assert that π is a stationary distribution for a solution of the martingale problem for (L, π). We know that P_π is a solution of the martingale problem for (L, π), and this solution is unique, since (15) admits a unique strong solution; see [23]. This implies that π is a stationary distribution for X. The results follows from the uniqueness of the stationary distribution. ■

6.2 LP Formulation and Convergence

Adopting the approach used on §3, consider an increasing sequence of finite cardinality sets $S_n \subset \mathbb{R}^2$ such that $S_n \uparrow S$, and consider an increasing sequence $\{\mathcal{F}_m\}$ of finite subspaces of C_b^2 such that, for all $f \in C_b^2$ and for any compact set $A \subset \mathbb{R}^2$ there exists a $f_m \in \mathcal{F}_m$, such that $\lim_{m\to\infty} \sup_{x\in A} |(L(f_m - f))(x)| = 0$. For fixed values of n and m, the following LP mirrors the one in §3.1:

$$(\mathcal{P}_{nm}) \ \min \ u$$

$$\begin{aligned}
s.t. \quad & \left| \sum_{i:\, x_i \in S_n} (Lf)(x_i)\lambda_i \right| \le u, \ \text{ for all } f \in \mathcal{F}_m && [\approx \text{BAR}] \\
& \sum_{i:\, x_i \in S_n} \lambda_i = 1 && [\text{Normalization}] \\
& \sum_{i:\, x_i \in S_n} g(x_i)\lambda_i \le M && [\text{Interior measure tightness}] \\
& \lambda \ge 0, && [\text{Nonnegativity}]
\end{aligned}$$

for a suitable chosen $M \le \infty$. In this formulation, $\{\lambda_i : i \text{ s.t. } x_i \in S_n\}$ is a discrete distribution with S_n as its support, that approximates π. The constant M, and the function $g(\cdot)$ are specified as follows. The function $g : \mathbb{R}^2 \to \mathbb{R}$ is chosen such that $g(x) \le g(y)$ for $x \le y$ componentwise, and such that $g(x) \to \infty$ as $\|x\|_1 \to \infty$, and with $M > 0$ such that $\mathbb{E}_\pi(g(X)) \le M$. Convergence to π along a subsequence of solutions to these LP's can be proved by following the same steps as in proof of Theorem 2.

Remark 4 *As mentioned on Remark 2, our LP formulation proposes to approximate π by a convex combination of one point distribution functions. In more generality, one may try to approximate π as a convex combination of generic distribution functions. We express this idea in more detail on §7.*

6.3 Practical Considerations

In order to implement and use the algorithm described above, we need to specify the sequences $\{S_n\}$ and $\{\mathcal{F}_m\}$. Moreover, the specification of the algorithm requires a feasible specification of the tightness condition, via choice of of M and g.

Tightness. For illustration purposes we will assume $\gamma_2 > 0$ to facilitate the explicit construction of a simple Lyapunov function. Along the lines of §3.3, consider the function $f(x) = \|x\|_2^2$. For this function we have

$$\begin{aligned}(Lf)(x) &= -\mu_1\gamma_1 x_1 - \mu_1 x_1^2 - (\mu_2\gamma_2 x_2 + \mu_2 x_2^2)\mathbf{1}\{x_1 + x_2 \le 0\} \\ &\qquad -(\mu_2\gamma_2 x_2 - \mu_2 x_2 x_1)\mathbf{1}\{x_1 + x_2 > 0\} + \sigma_1^2 + \sigma_2^2,\end{aligned}$$

where $\mathbf{1}\{B\}$ is the indicator of the set B. It then follows that $L(x) \le -c_1\|x\|_1 + c_2$ for all x such that $\|x\|_1 \ge r$, for some $r > 0$, and for some constants $c_1, c_2 > 0$. Therefore, we conclude that $L(x) \le -c_1\|x\|_1 + c_3$ for all $x \in \mathbb{R}^2$, for a suitable constant $c_3 \ge c_2$. It follows from Corollary 2 on Glynn and Zeevi [12] that $\mathbb{E}_\pi(\|X\|_1) \le c_3/c_1$, therefore we can take $g(x) = \|x\|_1$, and $M = c_3/c_1$.

Support and function space approximation. As mentioned, $X_1(\infty) \sim N(-\gamma_1, \kappa_1)$. Based on this, we set the grid to have an "Normal-based" spacing on the first coordinate, and in absence of prior information on the distribution of x_2, we take the grid to be equispaced on the second coordinate:

$$\begin{aligned}S_n &= \{x \in S : \{x_1 = (\Phi^{-1}(1/2 + k/(2n)) + \gamma_1)/\sqrt{\kappa_1} \quad k = 0, \pm 1 \ldots, \pm n - 1\} \\ &\qquad \times\{x_2 = \alpha j/n \quad j = -n, \ldots, n\}\}, \end{aligned} \tag{17}$$

for some constant $\alpha > 0$. Regarding the approximation of the function space C_b^2, we take the same sequence specified in §3.3.

6.4 Numerical Results

In the 2-class queue described in §6.1, take $\kappa_1 = \kappa_2 = 0.5$ and $\gamma_1 = \gamma_2 = 0.25$ (these parameters correspond to the example studied in Maglaras and Zeevi [23]). Figure 3 compares the approximation of the marginal distributions resulting from the algorithm by taking S_n as in (17) with $n = 50$ and $m = 4$ (solid line), against the true marginal distribution for the first coordinate, and against an approximation computed via Monte Carlo simulation (doted line), using an Euler scheme to approximate (15), for the second coordinate. Despite the pour fit, table 5 shows that the algorithm provide reasonable moment approximations for each coordinate. There, estimates from the

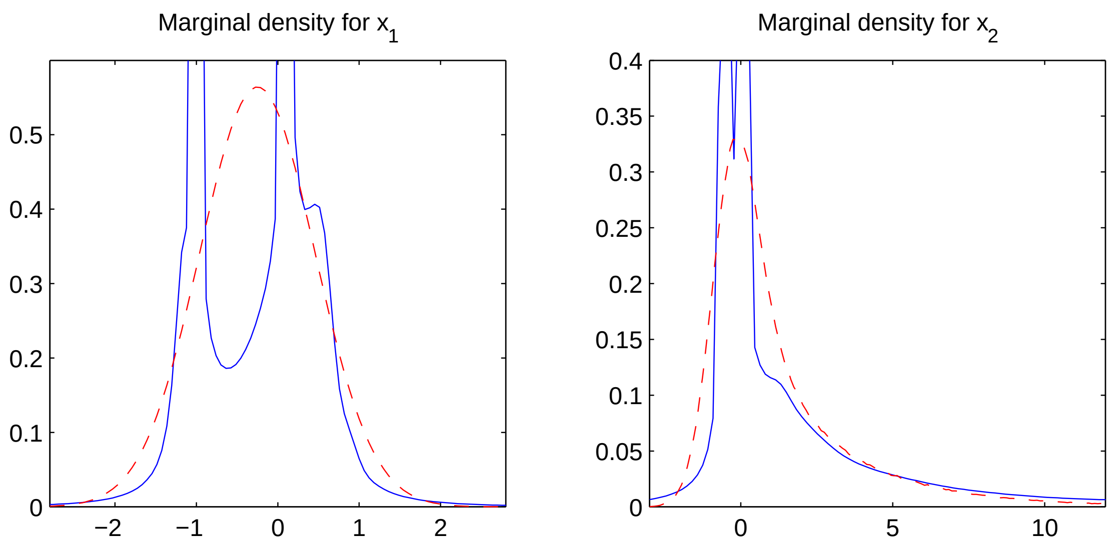

Figure 3: **Marginal distribution estimates for the two-dimensional diffusion**. The dotted line is computed via Monte Carlo simulation, and the solid line represents the algorithm estimates based on $n = 50$ and $m = 4$.

algorithm are compared against true moments, for the first coordinate, and against approximations computed via Monte Carlo simulation, labeled as "true values", for the second coordinate.

One observation that is evident from Figure 3 is that the LP approximation to the BAR condition is not imposing any reasonable smoothness on the resulting distribution (at least not for low values of m). Given the nature of the diffusion process under study here, one would anticipate much smoother stationary distribution for the second coordinate relative to what is computed by the algorithm (this is also seen to be the case in the dashed lines in Figure 3 which were obtained by simulation). To try to encode this into our LP formulation, we incorporate "smoothness" constrains as follows. Consider a *neighborhood parameter* $r > 0$, and a *smoothness parameter* δ. The following constraint can then be added to $(\mathcal{P}_{nm})$:

$$|\pi^n_m(x) - \pi^n_m(y)| \leq \delta \quad \text{for all } (x, y) \in S_n \quad \text{s.t. } \|x - y\|_2 < r. \tag{18}$$

This constraint is in essence imposing a type of a Lipschitz continuity on the distribution. In practice we introduce these constraints on each marginal distribution separately. We choose r so that the r-ball contains only one neighbor in each marginal, and δ sufficiently small to exclude "peaks" as those observed in figure 3. In particular, setting $\delta_1 = 0.0026$ (the first coordinate) and $\delta_2 = 0.0011$ (the second coordinate) we obtain the moment approximations reported in Table 5.

Remark 5 *Assuming that π is indeed Lipschitz continuous, one can implement constraint (18) by getting an upper bound on the Lipschitz continuity constant. This upper bound can be obtained as*

	$E[X_1^1]$	$E[X_1^2]$	$E[X_1^3]$	$E[X_1^4]$	$E[X_2^1]$	$E[X_2^2]$	$E[X_2^3]$	$E[X_2^4]$
"True value"	-0.25	0.56	-0.39	0.94	1.19	8.31	70.67	854
Estimate	-0.25	0.56	-0.39	0.94	1.26	7.79	69.68	812
Smoothed estimate	-0.25	0.56	-0.39	0.94	1.16	7.44	66.88	771
Relative error	0	0	0	0	0.058	0.063	0.014	0.049
Smoothed relative error	0	0	0	0	0.025	0.105	0.053	0.097

Table 5: **Moment approximations for first and second coordinates for the 2-dimensional diffusion**. The algorithm uses $n = 100$ and $m = 6$ degree polynomials. The table shows relative errors of algorithm estimates (incorporating smoothness constraints) with respect to true/simulated-values.

follows. For any set $A \subset \mathbb{R}^d$, we have that $\pi(A) = \int_{\mathbb{R}^d} K(y, A)\pi(dy)$, where $K(A, y)$ represents the kernel associated to the diffusion process. Assuming that π has a density ξ with respect to the Lebesgue measure on $\mathbb{R}^d$, and that this is differentiable, we have that

$$\xi(x) = \int_{\mathbb{R}^d} K(y,x)\xi(y)dy \Rightarrow \|\nabla\xi(x)\|_1 = \int_{\mathbb{R}^d} \|\nabla K(y,x)\|_1\xi(y)dy \Rightarrow \|\nabla\xi(x)\|_1 \leq \mathbb{E}_\pi|e\cdot\nabla K(y,x)|,$$

where $e \in \mathbb{R}^d$ has all components equal to one. Now, suppose we can find a non-negative function $g : \mathbb{R}^d \to \mathbb{R}_+$, twice continuous differentiable, and a constant c for which

$$Lg(x) \leq -|e\cdot\nabla K(y,x)| + c$$

uniformly on $y \in \mathbb{R}^d$, for all $x \in \mathbb{R}^d$, then we can use Corollary 2 on Glynn and Zeevi [12] to assert that

$$\|\nabla\xi(x)\|_1 \leq \mathbb{E}_\pi|e\cdot\nabla K(y,x)| \leq c.$$

Figure 4 depicts the approximation to the marginal distributions obtained by incorporating these smoothing constraints. The above considerations and implementation of the constraints is not meant to give definitive prescriptions, but rather to illustrate the flexibility of the LP approach, and the ability to incorporate such a priori information

7 Probability measure and function space approximation choice

Consider the LP formulation presented on §3.1. There, for fixed values of n and m, a feasible solution (λ, γ) to $(\mathcal{P}_{nm})$ is used to approximate π and $\nu := (\nu_1, \ldots, \nu_d)$ through a probability measure and finite measures assigning mass to a finite and discrete set of points on S and F_i $i = 1, \ldots, d$, respectively. More specifically, one can construct the following approximations to π

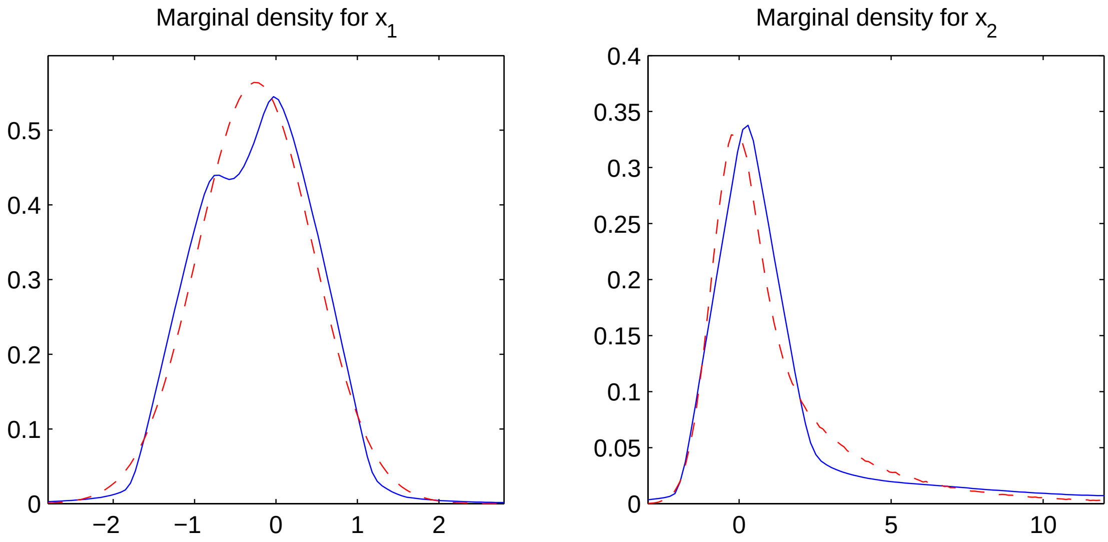

Figure 4: **Smoothed marginal distribution estimates for the two-dimensional diffusion.** The dotted line is computed via Monte Carlo simulation, and the solid line represents the algorithm estimates based on $n = 50$ and $m = 4$, incorporating smoothness constraints.

and ν:

$$\pi_\lambda^n(x) := \sum_{i:\, x_i \in S_n} \lambda_i \delta_{x_i}(x), \qquad \nu_{\gamma j}^n(x) := \sum_{i:\, x_i \in F_j^n} \gamma_{ji} \delta_{x_i}(x) \quad j = 1, \ldots, d,$$

where $\delta_y(x) = 1$ if $x = y$ and 0 otherwise. Notice that this is equivalent to approximate π by a convex combination of probability measures assigning all mass to single points in S and at the same time approximate ν_i by linear (positive) combination of probability measures assigning mass to single points in F_i, $i = 1, \ldots, d$. Using this interpretation, there is no need to restrict the approximating "architecture" to probability measures assigning all mass to single points.

Let $\mathcal{M}$ denote the space of probability measures on S, and $\mathcal{M}_j$ the space of finite measures on F_j, $j = 1 \ldots, d$. For arbitrary (finite) sets $\overline{M} \subset \mathcal{M}$ and $\overline{M}_j \subset \mathcal{M}_j$, $j = 1, \ldots, d$, and vectors $\lambda \in \mathbb{R}^{|\overline{\mathcal{M}}|}$ and $\gamma := (\gamma_1, \ldots, \gamma_d)$, $\gamma_j \in \mathbb{R}^{\overline{\mathcal{M}}_j}$, $j = 1, \ldots, d$, define:

$$\pi(\overline{\mathcal{M}}, \lambda) := \sum_{i:\, \phi_i \in \overline{\mathcal{M}}} \lambda_i \phi_i, \qquad \nu_j(\overline{\mathcal{M}}_j, \gamma_j) := \sum_{i:\, \varphi_i \in \overline{\mathcal{M}}_j} \gamma_{ji} \varphi_i \quad j = 1, \ldots, d. \tag{19}$$

Consider a sequence $\{\mathcal{M}_n\}$ such that for all $n > 0$, $\mathcal{M}_n \subset \mathcal{M}$, $|\mathcal{M}_n| < \infty$, and such that for all $\phi \in \mathcal{M}$ there exist a sequence of stochastic vectors $\{\lambda_n\}$ such that $\int_S g(x)\pi(\mathcal{M}_n, \lambda_n)(dx) \leq \int_S g(x)\phi(dx)$ for all $n > 0$, where $g : S \to \mathbb{R}_+$ specified as in §3.1, and $\pi(\mathcal{M}_n, \lambda_n) \Rightarrow \phi$. Analogously, consider sequences $\left\{\mathcal{M}_j^n\right\}$ $j = 1, \ldots, d$, such that for all $n > 0$, $\mathcal{M}_j^n \subset \mathcal{M}_j$, $|\mathcal{M}_j^n| < \infty$, $j = 1, \ldots, d$, and such that for all $\varphi \in \mathcal{M}_j$ there exist a sequence of positive (finite) vectors $\left\{\gamma_j^n\right\}$ such that

$\int_{F_j} g_j(x)\nu_j(\mathcal{M}_j^n, \gamma_j^n)(dx) \leq \int_{F_j} g_j(x)\varphi(dx)$ for all $n > 0$, where $g_j : F_j \to \mathbb{R}_+$ specified as in §3.1, and $\nu_j(\mathcal{M}_j^n, \gamma_j^n) \Rightarrow \varphi$, for $j = 1, \ldots, d$.

For $f \in C$ and for $\phi \in \mathcal{M}$ define $B(\phi, f) := \int_S f(x)\,\phi(dx)$, and for $\varphi \in \mathcal{M}_j$ define $B_j(\varphi, f) := \int_{F_j} f(x)\,\varphi_j(dx)$. Fixing $n > 0$ and $m > 0$, we can restate $\mathcal{P}_{nm}$ as follows:

$(\widetilde{\mathcal{P}}_{nm})$ min u

$$\text{s.t.} \quad \left| \lambda_i B(\pi(\mathcal{M}_n, \lambda), Gf) + \sum_{j=1}^{d} B_j(\nu_j(\mathcal{M}_j^n, \gamma_j^n), D_j f) \right| \leq u \ , \text{ for all } f \in \mathcal{F}_m \quad [\approx \text{BAR}] \quad (20)$$

$$\sum_{i:\, x_i \in S_n} \lambda_i = 1 \quad [\text{Normalization}] \quad (21)$$

$$B(\pi(\mathcal{M}_n, \lambda_n), g) \leq M \quad [\text{Interior measure tightness}] \quad (22)$$

$$B_j(\nu_j(\mathcal{M}_j^n, \gamma_j^n), g_j) \leq M_j \ , \text{ for all } j \in \{1 \ldots, d\} \quad [\text{Boundary measure tightness}] \quad (23)$$

$$\lambda_i,\, \gamma_{ij} \geq 0 \ , \text{ for all } \{i : x_i \in S_n\} \text{ and } j \in \{1, \ldots, d\}\,. \quad [\text{Nonnegativity}] \quad (24)$$

Where the constants M, M_j, $j = 1, \ldots, d$, are chosen as in §3.1. Notice that $\widetilde{\mathcal{P}}_{nm}$ remains a finite dimensional linear program. The following corollary generalizes Theorem 2.

Corollary 1 *Consider an SRBM Z as given in definition 1. Let π be the stationary distribution for Z, with boundary measures ν_j , $j = 1, \ldots, d$. Then, there exist sequences $\{n_k\}$ and $\{m_k\}$, $k = 1, 2, \ldots$, of positive integers that diverge to infinity as $k \to \infty$, such that $\pi_{\lambda^k}^{n_k m_k} \Rightarrow \pi$ and $\nu_{\underline{\gamma}^k j}^{n_k m_k} \Rightarrow \nu_j$ as $k \to \infty$, for $j = 1, \ldots, d$, where $(\lambda^k, \underline{\gamma}^k)$ are optimal solutions to the LP $(\widetilde{\mathcal{P}}_{n_k m_k})$.*

The proof of theorem 2 carries on under the new assumptions on $\{\mathcal{M}_n\}$ and $\left\{\mathcal{M}_j^n\right\}$, $j = 1, \ldots, d$. Also, extending this formulation for cases like the Many-server Heavy-traffic regime presented on §6 is straightforward.

The complexity of formulating $\widetilde{\mathcal{P}}_{n_k m_k}$ depends directly on the ability to efficiently compute $B(\phi, f)$ for all pairs $(\phi, f) \in \mathcal{M}_n \times \mathcal{F}_m$, and $B_j(\varphi, f)$ for all pairs $(\varphi, f) \in \mathcal{M}_j \times \mathcal{F}_m$, $j = 1, \ldots, d$. This indicates that the selection of $\{\mathcal{M}_n\}$, $\{\mathcal{M}_j\}$, $j = 1, \ldots, d$, and $\{\mathcal{F}_m\}$ could be made considering all possible synergies for formulating $\widetilde{\mathcal{P}}_{n_k m_k}$. For the implementations presented on §4 and §6, the selection of the probability measure sequences as single mass point measures "decouples" from the selection of $\mathcal{F}_m$ since, for computing B and B_j, $j = 1, \ldots, d$, one just need to be able to evaluate the functions on $\mathcal{F}_m$ on any point of S and F_j, $j = 1, \ldots, d$, respectively. Nevertheless, other choice for the probability measure-function space sequences, may result on an easier formulation of the corresponding linear programs.

Example 1. Consider the SRBM-implementation of the algorithm, shown on §4. For $\{\mathcal{M}_n\}$ let $\mathcal{M}_n$ to be a finite number of suitable truncated gaussian distributions, that differs on their first and second moments, for all $n > 0$. If $\{\mathcal{F}_m\}$ remains as specified ($\mathcal{F}_m$ form by polynomials of degree up to m), $B(\phi, Gf)$ unfolds as nothing but a linear combination of moments of a truncated normal random variable, so it can be computed efficiently. Considering truncated normal random variables to approximate the boundary measures results on the same type of efficiency.

Example 2. Consider any affine diffusion on $\mathbb{R}^d$, such that the BAR is necessary and sufficient to characterize a unique stationary distribution (this is the case of the diffusion that solves (15), presented on §6). Here the state space has no boundaries, so one only need to chose $\{\mathcal{M}_n\}$ and $\{\mathcal{F}_m\}$. One can take $\{\mathcal{M}_n\}$ such that $\mathcal{M}_n$ contains a finite number gaussian distributions, that differs on their first and second moments, for all $n > 0$. If $\{\mathcal{F}_m\}$ is such that $\mathcal{F}_m$ is form by polynomials of degree up to m, for all m, then $B(\phi, Gf)$ is, again, a linear combination of moments of a truncated normal random variable (here the truncation is given by the nature of the drift coefficients of the diffusion process), so it can be computed efficiently.

References

[1] A. Ben-Tal and A. Nemirovski. Robust solution of uncertain linear programs. *Operations Research Letters*, 25(1):1–13, 1999.

[2] A. Ben-Tal and M. Zibulevsky. Penalty/barrier multiplier methods for convex programming problems. *SIAM Journal of Optimization*, 7:347–366, 1997.

[3] A. Bernard and A. El Kharroubi. Régulations de processus dans le premier “orthant” de r^n. *Stochastics and stochastics reports*, 34:149–167, 1991.

[4] M. Bramson. State space collapse with applications to heavy traffic limits for multiclass queueing networks. *Queueing Systems: Theory and Applications*, 30(1-2):89–148, 1998.

[5] A. Budhiraja and C. Lee. Long time asymptotics for constrained diffusions in polyhedral domains. *Stochastic Processes and their Applications*, 117(8):1014–1036, 2007.

[6] J. G. Dai. *Steady-state analysis of reflected Brownian motions: Characterization, numerical methods and queueing applications.* PhD thesis, Dept. Math. Stanford University, 1990.

[7] J. G. Dai and J. M. Harrison. Reflected brownian motion in an orthant: Numerical methods for steady-state analysis. *The Annals of Applied probability*, 2(1):65–86, 1992.

[8] J.G. Dai and T. G. Kurtz. The sufficiency of the basic adjoint relationship.

[9] P. Dupuis and R. J. Williams. Lyapunov functions for semimartingale brownian motions. *Annals of Probability*, 22(2):680–702, 1994.

[10] S. N. Either and T. G. Kurtz. *Markov Processes, Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics, 1996.

[11] D. Gamarnik and A. Zeevi. Validity of heavy traffic steady-state approximations in generalized jackson networks. *Annals of Applied Probability*, 16(1):56–90, 2006.

[12] P. W. Glynn and A. Zeevi. Bounding stationary expectations of markov procesess. 2007.

[13] B. S. Greenberg. *Queueing systems with returning customers and the order of tandem queues*. PhD thesis, University of California, Berkeley, 1986.

[14] I. Gurvich and A. Zeevi. Validity of heavy-traffic steady-state approximations in open queueing networks: Sufficient conditions involving state-space collapse. 2007.

[15] S. Halfin and W. Whitt. Heavy-traffic limits for queues with many exponential servers. *Operations Research*, 29(3):567–588, 1981.

[16] J. M. Harrison. The diffusion approximation for tandem queues in heavy traffic. *Advances in Applied Probability*, 10:886–905, 1978.

[17] J. M. Harrison and V. Nguyen. The qnet method for two-moment analysis of open queueing networks. *Queueing Systems: Theory and Applications*, 6(1):1–32, 1990.

[18] J. M. Harrison and R. J. Williams. Brownian models of open queueing networks with homogeneous customer populations. *Stochastics*, 22(2):77–115, 1987.

[19] J. M. Harrison and R. J. Williams. Multidimensional reflected brownian motions having exponential stationary distributions. *Annals of Applied Probability*, 15(1):115–137, 1987.

[20] O. Hernandez-Lerma and J. B. Lasserre. Linear programming approximations for markov control processes in metric spaces. *Acta Applicandae Mathematicae*, 51:123–139, 1998.

[21] O. Hernandez-Lerma and J. B. Lasserre. *Markov Chains and Invariant Probabilities (Progress in Mathematics)*. Birkhäuser, Basel, Switzerland, 2003.

[22] R. Stockbridge M. Mendiondo. Approximation of infinite-dimensional linear programming problems which arise in stochastic control. *SIAM Journal of Control Optimization*, 36(4):1448–1472, 1998.

[23] C. Maglaras and A. Zeevi. Diffusion approximations for a multiclass markovian service system. *Mathematics of Operations Research*, 29(4):786–813, 2004.

[24] A. Manne. Linear programming and sequencial desicions. *Management Science*, 6(3):259–267, 1960.

[25] M. I. Reiman. Open queueing networks in heavy traffic. *Mathematics of Operations Research*, 9(3):441–458, 1984.

[26] M. I. Reiman and R. J. Williams. A boundary property of semimartingale reflecting brownian motions. *Probability Theory and Related Fields*, 77(1):87–97, 1988.

[27] S. Suresh and W. Whitt. Arranging queues in series: a simulation experiment. *Management Sciences*, 36(9):1080–1090, 1990.

[28] L.M. Taylor and R. J. Williams. Existence and uniqueness of semimartingale reflecting brownian motion in an orthant. *Probability Theory and Related Fields*, 96(3):283–317, 1993.

[29] W. Whitt. The queueing network analyzer. *Bell System Technical Journal*, 62(9):2779–2815, 1983.

[30] R. J. Williams. Diffusion approximations for open multiclass queueing networks: Sufficient conditions involving state space collapse. *Queueing Systems: Theory and Applications*, 30(1-2):27–88, 1998.

[31] H. Chen X. Shen. Computing the stationary distribution of an srbm in an orthant with applications to queueing networks. *Queueing Systems*, 45(1):27–45, 2003.

[32] J. G. Dai X. Shen, H. Chen and W. Dai. The finite element method for computing the stationary distribution of an srbm in a hypercube with applications to finite buffer queueing networds. *Queueing Systems: Theory and Applications*, 42:33–62, 2002.

A Numerical Results

		$C_a^2 = 0.5$, $C_{s_1}^2 = 0.5$, $C_{s_2}^2 = 2.0$							
		Order I				Order II			
ρ_1	ρ_2	QNET	QNA	LP	Sim	QNET	QNA	LP	Sim
0.9	0.9	10.12	10.13	10.01	10.010	7.75	8.97	7.84	7.946
		0.01	0.01	0.00	0.051	0.02	0.13	0.01	0.430
0.9	0.8	4.00	4.00	4.07	3.856	6.58	7.94	6.58	6.292
		0.04	0.04	0.06	0.219	0.05	0.26	0.05	0.368
0.9	0.6	1.12	1.12	1.14	1.039	5.07	6.24	5.33	4.976
		0.08	0.08	0.10	0.035	0.02	0.25	0.07	0.260
0.9	0.3	0.16	0.16	0.16	0.127	4.17	4.60	4.60	4.309
		0.03	0.03	0.03	0.003	0.03	0.07	0.07	0.154
0.8	0.9	10.12	10.13	10.25	10.153	3.48	3.54	3.49	3.444
		0.00	0.00	0.01	0.670	0.01	0.03	0.01	0.116
0.8	0.8	4.00	4.00	3.99	3.706	3.07	3.14	3.10	2.895
		0.08	0.08	0.08	0.158	0.06	0.08	0.07	0.085
0.8	0.6	1.12	1.12	1.15	1.068	2.46	2.46	2.50	2.374
		0.05	0.05	0.08	0.030	0.04	0.04	0.05	0.042
0.8	0.3	0.16	0.16	0.16	0.129	1.74	1.82	1.97	1.670
		0.03	0.03	0.03	0.005	0.04	0.09	0.18	0.044
0.6	0.9	10.12	10.13	10.50	9.687	1.08	1.00	1.08	0.966
		0.04	0.05	0.08	0.786	0.11	0.03	0.11	0.024
0.6	0.8	4.00	4.00	4.09	4.007	1.00	0.88	1.00	0.912
		0.00	0.00	0.02	0.125	0.09	0.03	0.09	0.011
0.6	0.6	1.12	1.12	1.12	1.072	0.86	0.69	0.87	0.736
		0.04	0.04	0.04	0.024	0.12	0.05	0.13	0.020
0.6	0.3	0.16	0.16	0.16	0.133	0.63	0.51	0.66	0.491
		0.03	0.03	0.03	0.002	0.14	0.02	0.17	0.006
0.3	0.9	10.12	10.13	10.40	9.562	0.16	0.14	0.16	0.136
		0.06	0.06	0.09	0.719	0.02	0.00	0.02	0.002
0.3	0.8	4.00	4.00	4.00	4.024	0.16	0.13	0.15	0.129
		0.01	0.01	0.01	0.272	0.03	0.00	0.02	0.002
0.3	0.6	1.12	1.12	1.16	1.090	0.15	0.10	0.15	0.112
		0.03	0.03	0.06	0.032	0.04	0.01	0.04	0.001
0.3	0.3	0.16	0.16	0.15	0.131	0.12	0.07	0.12	0.075
		0.03	0.03	0.02	0.003	0.05	0.01	0.05	0.001
Average		0.035	0.035	0.046		0.055	0.069	0.072	

Table 6: **Expected Waiting Times at the Second Queue in Example 4: Case 1**. Entries at the table depict expected waiting time estimates (above) and the minimum between the absolute relative errors and absolute errors with respect to the simulation estimates (below), for all three methods, for different orders of the queues, and for different values of ρ_1 and ρ_2. The last row depicts the average of the errors for the different combinations of methods/orders.

		$C_a^2 = 1.0$, $C_{s_1}^2 = 0.5$, $C_{s_2}^2 = 8.0$							
		Order I				Order II			
ρ_1	ρ_2	QNET	QNA	LP	Sim	QNET	QNA	LP	Sim
0.9	0.9	36.09	34.81	35.80	37.040	24.32	29.04	24.32	24.050
		0.03	0.06	0.03	4.410	0.01	0.21	0.01	1.760
0.9	0.8	14.05	13.75	14.05	13.350	18.36	24.22	19.00	18.120
		0.05	0.03	0.05	1.260	0.01	0.34	0.05	0.930
0.9	0.6	3.89	3.87	3.92	3.800	10.54	16.28	13.20	12.630
		0.02	0.02	0.03	0.240	0.17	0.29	0.05	0.800
0.9	0.3	0.55	0.55	0.55	0.520	6.60	8.63	8.94	7.550
		0.03	0.03	0.03	0.020	0.13	0.14	0.18	0.380
0.8	0.9	36.42	35.15	36.70	37.580	11.38	11.47	11.35	10.740
		0.03	0.06	0.02	5.260	0.06	0.07	0.06	0.450
0.8	0.8	14.28	13.89	14.08	13.780	9.65	9.57	9.60	8.950
		0.04	0.01	0.02	1.570	0.08	0.07	0.07	0.470
0.8	0.6	3.94	3.91	3.96	3.840	6.29	6.43	7.08	6.330
		0.03	0.02	0.03	0.250	0.01	0.02	0.12	0.270
0.8	0.3	0.55	0.56	0.56	0.490	2.94	3.41	4.48	3.360
		0.06	0.07	0.07	0.020	0.13	0.01	0.33	0.100
0.6	0.9	36.45	35.72	36.79	34.610	3.68	3.23	3.59	2.860
		0.05	0.03	0.06	3.370	0.29	0.13	0.26	0.140
0.6	0.8	14.40	14.11	14.56	13.160	3.31	2.69	3.29	2.570
		0.09	0.07	0.11	1.500	0.29	0.05	0.28	0.110
0.6	0.6	4.02	3.97	3.99	3.960	2.72	1.81	2.70	2.070
		0.02	0.00	0.01	0.240	0.31	0.13	0.30	0.040
0.6	0.3	0.56	0.57	0.57	0.520	1.38	0.96	1.80	1.110
		0.04	0.05	0.05	0.020	0.24	0.14	0.62	0.030
0.3	0.9	36.45	36.27	38.55	31.120	0.55	0.46	0.54	0.240
		0.17	0.17	0.24	4.440	0.31	0.22	0.30	0.010
0.3	0.8	14.40	14.34	14.06	13.330	0.54	0.38	0.52	0.250
		0.08	0.08	0.05	0.600	0.29	0.13	0.27	0.000
0.3	0.6	4.05	4.03	4.10	4.100	0.50	0.26	0.48	0.270
		0.01	0.02	0.00	0.250	0.23	0.01	0.21	0.010
0.3	0.3	0.57	0.58	0.57	0.550	0.39	0.14	0.39	0.200
		0.02	0.03	0.02	0.020	0.19	0.06	0.19	0.000
Average		0.048	0.047	0.052		0.171	0.125	0.206	

Table 7: **Expected Waiting Times at the Second Queue in Example 4: Case 2**. Entries at the table depict expected waiting time estimates (above) and the minimum between the absolute relative errors and absolute errors with respect to the simulation estimates (below), for all three methods, for different orders of the queues, and for different values of ρ_1 and ρ_2. The last row depicts the average of the errors for the different combinations of methods/orders.

		$C_a^2 = 1.0$, $C_{s_1}^2 = 2.0$, $C_{s_2}^2 = 4.0$							
		Order I				Order II			
ρ_1	ρ_2	QNET	QNA	LP	Sim	QNET	QNA	LP	Sim
0.9	0.9	22.03	23.53	22.06	23.026	19.13	21.99	19.36	18.447
		0.04	0.02	0.04	2.776	0.04	0.19	0.05	1.520
0.9	0.8	9.07	9.30	9.24	8.887	16.23	9.93	16.68	16.562
		0.02	0.05	0.04	0.544	0.02	0.40	0.01	1.427
0.9	0.6	2.64	2.61	2.69	2.829	13.00	16.52	14.00	14.538
		0.07	0.08	0.05	0.116	0.11	0.14	0.04	1.192
0.9	0.3	0.39	0.37	0.38	0.402	12.22	13.24	13.28	12.967
		0.01	0.03	0.02	0.008	0.06	0.02	0.02	1.738
0.8	0.9	20.92	22.84	21.26	23.698	8.45	8.69	8.52	8.981
		0.12	0.04	0.10	2.050	0.06	0.03	0.05	0.426
0.8	0.8	8.70	9.02	8.64	8.686	7.57	7.87	7.66	7.747
		0.00	0.04	0.01	0.378	0.02	0.02	0.01	0.376
0.8	0.6	2.57	2.54	2.59	2.758	5.59	6.53	6.39	6.438
		0.07	0.08	0.06	0.079	0.13	0.01	0.01	0.343
0.8	0.3	0.38	0.36	0.38	0.399	4.88	5.23	5.55	4.896
		0.02	0.04	0.02	0.004	0.00	0.07	0.13	0.146
0.6	0.9	20.31	21.71	21.04	21.969	2.61	2.44	2.60	2.618
		0.08	0.01	0.04	2.683	0.00	0.07	0.01	0.067
0.6	0.8	8.13	8.58	8.37	8.131	2.43	2.21	2.43	2.352
		0.00	0.06	0.03	0.311	0.03	0.06	0.03	0.079
0.6	0.6	2.45	2.41	2.48	2.485	2.13	1.84	2.15	2.012
		0.01	0.03	0.00	0.086	0.06	0.09	0.07	0.063
0.6	0.3	0.37	0.34	0.37	0.368	1.52	1.47	1.72	1.556
		0.00	0.03	0.00	0.015	0.02	0.06	0.11	0.044
0.3	0.9	20.25	20.61	21.83	21.480	0.39	0.35	0.38	0.345
		0.06	0.04	0.02	3.439	0.05	0.01	0.04	0.009
0.3	0.8	8.01	8.14	8.50	8.408	0.38	0.32	0.37	0.345
		0.05	0.03	0.01	0.472	0.04	0.03	0.03	0.008
0.3	0.6	2.26	2.29	2.37	2.337	0.36	0.26	0.35	0.321
		0.03	0.02	0.01	0.081	0.04	0.06	0.03	0.009
0.3	0.3	0.35	0.33	0.35	0.340	0.30	0.21	0.31	0.248
		0.01	0.01	0.01	0.007	0.05	0.04	0.06	0.005
Average		0.037	0.037	0.029		0.045	0.080	0.043	

Table 8: **Expected Waiting Times at the Second Queue in Example 4: Case 3**. Entries at the table depict expected waiting time estimates (above) and the minimum between the absolute relative errors and absolute errors with respect to the simulation estimates (below), for all three methods, for different orders of the queues, and for different values of ρ_1 and ρ_2. The last row depicts the average of the errors for the different combinations of methods/orders.

		$C_a^2 = 4.0$, $C_{s_1}^2 = 0.5$, $C_{s_2}^2 = 1.0$							
		Order I				Order II			
ρ_1	ρ_2	QNET	QNA	LP	Sim	QNET	QNA	LP	Sim
0.9	0.9	10.73	8.77	10.79	12.541	9.56	8.38	9.60	11.788
		0.14	0.30	0.14	1.650	0.19	0.29	0.19	2.494
0.9	0.8	3.00	3.46	3.09	3.325	9.43	10.45	13.64	15.032
		0.10	0.04	0.07	0.102	0.37	0.30	0.09	1.716
0.9	0.6	0.69	0.97	0.73	0.725	4.12	13.85	17.65	16.603
		0.04	0.25	0.00	0.016	0.75	0.17	0.06	1.426
0.9	0.3	0.10	0.14	0.10	0.076	3.12	17.13	20.59	19.682
		0.02	0.06	0.02	0.001	0.84	0.13	0.05	2.000
0.8	0.9	7.70	11.18	15.63	16.269	2.87	3.31	2.95	3.130
		0.53	0.31	0.04	2.358	0.08	0.06	0.06	0.155
0.8	0.8	4.17	4.42	4.27	4.803	3.75	4.13	3.83	4.114
		0.13	0.08	0.11	0.398	0.09	0.00	0.07	0.240
0.8	0.6	0.80	1.24	0.85	0.907	3.21	5.47	5.74	6.147
		0.11	0.33	0.06	0.026	0.48	0.11	0.07	0.526
0.8	0.3	0.10	0.18	0.10	0.088	1.20	6.77	7.28	6.941
		0.01	0.09	0.01	0.001	0.83	0.02	0.05	0.387
0.6	0.9	3.41	15.15	19.31	18.706	0.69	0.93	0.72	0.788
		0.82	0.19	0.03	2.371	0.10	0.14	0.07	0.020
0.6	0.8	2.62	5.98	6.61	6.987	0.77	1.16	0.81	0.934
		0.63	0.14	0.05	0.717	0.16	0.23	0.12	0.018
0.6	0.6	1.17	1.68	1.21	1.471	1.05	1.54	1.07	1.302
		0.20	0.14	0.18	0.039	0.19	0.18	0.18	0.018
0.6	0.3	0.10	0.24	0.11	0.129	0.75	1.90	1.74	1.755
		0.03	0.11	0.02	0.002	0.57	0.08	0.01	0.049
0.3	0.9	2.11	18.97	22.10	20.014	0.10	0.13	0.10	0.105
		0.89	0.05	0.10	2.027	0.01	0.03	0.01	0.001
0.3	0.8	0.96	7.50	8.10	8.165	0.10	0.17	0.11	0.117
		0.88	0.08	0.01	0.492	0.02	0.05	0.01	0.002
0.3	0.6	0.62	2.11	1.96	2.031	0.10	0.22	0.11	0.142
		0.69	0.04	0.03	0.051	0.04	0.08	0.03	0.002
0.3	0.3	0.17	0.30	0.17	0.205	0.15	0.27	0.15	0.178
		0.04	0.10	0.04	0.005	0.03	0.09	0.02	0.002
Average		0.329	0.145	0.058		0.296	0.123	0.067	

Table 9: **Expected Waiting Times at the Second Queue in Example 4: Case 4**. Entries at the table depict expected waiting time estimates (above) and the minimum between the absolute relative errors and absolute errors with respect to the simulation estimates (below), for all three methods, for different orders of the queues, and for different values of ρ_1 and ρ_2. The last row depicts the average of the errors for the different combinations of methods/orders.

ρ_1	ρ_2	$C_a^2 = 4.0$, $C_{s_1}^2 = 1.0$, $C_{s_2}^2 = 4.0$							
		Order I				Order II			
		QNET	QNA	LP	Sim	QNET	QNA	LP	Sim
0.9	0.9	26.11	22.56	26.10	25.776	20.25	20.25	20.40	20.453
		0.01	0.12	0.01	2.604	0.01	0.01	0.00	1.104
0.9	0.8	9.14	8.91	9.28	10.105	20.25	20.25	20.21	25.409
		0.10	0.12	0.08	0.854	0.20	0.20	0.20	2.870
0.9	0.6	2.32	2.51	2.42	2.652	20.25	20.25	21.66	23.394
		0.13	0.05	0.09	0.130	0.13	0.13	0.07	2.348
0.9	0.3	0.32	0.36	0.32	0.357	20.25	20.25	22.72	21.580
		0.04	0.00	0.04	0.012	0.06	0.06	0.05	2.733
0.8	0.9	29.08	24.62	29.08	33.220	8.00	8.00	8.04	9.201
		0.12	0.26	0.12	5.395	0.13	0.13	0.13	0.534
0.8	0.8	10.30	9.73	10.46	11.360	8.00	8.00	8.04	8.455
		0.09	0.14	0.08	0.723	0.05	0.05	0.05	0.535
0.8	0.6	2.51	2.74	2.55	2.868	8.00	8.00	7.97	8.643
		0.12	0.04	0.11	0.188	0.07	0.07	0.08	0.441
0.8	0.3	0.32	0.39	0.34	0.399	8.00	8.00	8.03	7.991
		0.08	0.01	0.06	0.009	0.00	0.00	0.01	0.327
0.6	0.9	31.58	28.03	33.45	31.279	2.25	2.25	2.28	2.298
		0.01	0.10	0.07	3.484	0.02	0.02	0.01	0.063
0.6	0.8	11.89	11.07	11.99	12.949	2.25	2.25	2.26	2.443
		0.08	0.15	0.07	1.016	0.08	0.08	0.08	0.099
0.6	0.6	2.90	3.11	2.94	3.389	2.25	2.25	2.28	2.523
		0.14	0.08	0.13	0.150	0.11	0.11	0.10	0.045
0.6	0.3	0.34	0.44	0.36	0.444	2.25	2.25	2.26	2.316
		0.10	0.00	0.08	0.011	0.03	0.03	0.02	0.054
0.3	0.9	32.35	31.31	37.00	27.840	0.32	0.32	0.32	0.269
		0.16	0.12	0.33	1.955	0.05	0.05	0.05	0.005
0.3	0.8	12.73	12.37	13.17	13.667	0.32	0.32	0.32	0.286
		0.07	0.09	0.04	1.304	0.03	0.03	0.03	0.005
0.3	0.6	3.48	3.48	3.50	3.611	0.32	0.32	0.33	0.330
		0.04	0.04	0.03	0.226	0.01	0.01	0.00	0.009
0.3	0.3	0.41	0.50	0.42	0.528	0.32	0.32	0.32	0.328
		0.12	0.03	0.11	2.000	0.01	0.01	0.01	0.006
Average		0.089	0.086	0.091		0.063	0.063	0.056	

Table 10: **Expected Waiting Times at the Second Queue in Example 4: Case 5**. Entries at the table depict expected waiting time estimates (above) and the minimum between the absolute relative errors and absolute errors with respect to the simulation estimates (below), for all three methods, for different orders of the queues, and for different values of ρ_1 and ρ_2. The last row depicts the average of the errors for the different combinations of methods/orders.

		$C_a^2 = 1.0$, $C_{s_1}^2 = 0.5$, $C_{s_2}^2 = 0.5$							
		Order I				Order II			
ρ_1	ρ_2	QNET	QNA	LP	Sim	QNET	QNA	LP	Sim
0.9	0.9	4.84	4.43	4.88	4.793	4.84	4.43	4.86	4.910
		0.01	0.08	0.02	0.269	0.01	0.10	0.01	0.215
0.9	0.6	0.46	0.49	0.47	0.410	5.54	5.35	6.00	5.658
		0.05	0.08	0.06	0.006	0.02	0.05	0.06	0.410
0.9	0.3	0.06	0.07	0.07	0.043	6.01	5.89	6.25	5.635
		0.02	0.03	0.02	0.001	0.07	0.05	0.11	0.293
0.9	0.2	0.03	0.03	0.03	0.013	6.05	5.99	6.42	6.180
		0.02	0.02	0.01	0.001	0.02	0.03	0.04	0.442
0.9	0.1	0.01	0.01	0.01	0.002	6.07	6.05	6.60	6.047
		0.01	0.01	0.00	0.000	0.00	0.00	0.09	0.253
0.6	0.6	0.54	0.59	0.54	0.552	0.54	0.59	0.54	0.552
		0.01	0.04	0.01	0.008	0.01	0.04	0.01	0.009
0.6	0.3	0.07	0.08	0.07	0.057	0.61	0.65	0.65	0.644
		0.01	0.02	0.01	0.001	0.03	0.01	0.01	0.011
0.6	0.2	0.03	0.03	0.03	0.018	0.64	6.70	0.70	0.669
		0.01	0.01	0.01	0.000	0.03	6.03	0.03	0.011
0.6	0.1	0.01	0.01	0.01	0.003	0.67	0.67	0.74	0.668
		0.01	0.01	0.00	0.000	0.00	0.00	0.07	0.015
0.3	0.3	0.08	0.09	0.08	0.079	0.08	0.08	0.08	0.079
		0.00	0.01	0.00	0.001	0.00	0.00	0.00	0.001
0.3	0.2	0.03	0.04	0.03	0.026	0.08	0.10	0.08	0.087
		0.00	0.01	0.00	0.000	0.01	0.01	0.01	0.001
0.3	0.1	0.01	0.01	0.01	0.004	0.09	0.10	0.09	0.094
		0.01	0.01	0.00	0.000	0.00	0.01	0.00	0.001
0.2	0.2	0.03	0.04	0.03	0.031	0.03	0.04	0.03	0.031
		0.00	0.01	0.00	0.000	0.00	0.01	0.00	0.001
0.2	0.1	0.01	0.01	0.01	0.005	0.03	0.04	0.03	0.035
		0.01	0.01	0.00	0.000	0.01	0.01	0.01	0.001
0.1	0.1	0.01	0.01	0.01	0.004	0.01	0.01	0.01	0.006
		0.01	0.01	0.00	0.000	0.00	0.00	0.00	0.000
Average		0.011	0.023	0.010		0.015	0.423	0.030	

Table 11: **Expected Waiting Times at the Second Queue in Example 4: Case 6**. Entries at the table depict expected waiting time estimates (above) and the minimum between the absolute relative errors and absolute errors with respect to the simulation estimates (below), for all three methods, for different orders of the queues, and for different values of ρ_1 and ρ_2. The last row depicts the average of the errors for the different combinations of methods/orders.

		$C_a^2 = 1.0$, $C_{s_1}^2 = 4.0$, $C_{s_2}^2 = 4.0$							
		Order I				Order II			
ρ_1	ρ_2	QNET	QNA	LP	Sim	QNET	QNA	LP	Sim
0.9	0.9	26.43	30.09	26.80	25.505	26.43	30.09	26.86	28.072
		0.04	0.18	0.05	2.955	0.06	0.07	0.04	2.921
0.9	0.6	3.47	3.34	3.46	3.405	20.57	24.62	22.34	22.857
		0.02	0.02	0.02	0.144	0.10	0.08	0.02	4.542
0.9	0.3	0.51	0.48	0.51	0.530	20.27	21.34	23.80	20.511
		0.02	0.05	0.02	0.019	0.01	0.04	0.16	1.629
0.9	0.2	0.20	0.19	0.20	0.199	20.26	20.74	23.00	19.132
		0.00	0.01	0.00	0.005	0.06	0.08	0.20	2.606
0.9	0.1	0.04	0.04	0.45	0.046	20.25	20.37	29.07	20.067
		0.01	0.01	0.40	0.001	0.01	0.02	0.45	1.768
0.6	0.6	2.94	2.74	2.95	2.874	2.94	2.74	2.96	2.918
		0.02	0.05	0.03	0.001	0.01	0.06	0.01	0.116
0.6	0.3	0.48	0.39	0.48	0.457	2.32	2.37	2.55	2.494
		0.02	0.07	0.02	0.013	0.07	0.05	0.02	0.114
0.6	0.2	0.20	0.15	0.19	0.188	2.27	2.30	2.50	2.231
		0.01	0.04	0.00	0.004	0.02	0.03	0.12	0.088
0.6	0.1	0.04	0.03	0.04	0.045	2.25	2.26	2.44	2.271
		0.01	0.02	0.00	0.002	0.01	0.00	0.07	0.135
0.3	0.3	0.42	0.34	0.43	0.378	0.42	0.34	0.42	0.371
		0.04	0.04	0.05	0.011	0.05	0.03	0.05	0.009
0.3	0.2	0.17	0.13	0.18	0.153	0.39	0.33	0.39	0.356
		0.02	0.02	0.02	0.006	0.03	0.03	0.03	0.006
0.3	0.1	0.04	0.03	0.04	0.039	0.34	0.32	0.37	0.346
		0.00	0.01	0.00	0.002	0.01	0.03	0.02	0.012
0.2	0.2	0.16	0.13	0.17	0.141	0.16	0.13	0.17	0.145
		0.02	0.01	0.03	0.003	0.02	0.02	0.02	0.006
0.2	0.1	0.04	0.03	0.04	0.034	0.15	0.13	0.15	0.134
		0.01	0.00	0.01	0.002	0.02	0.00	0.02	0.003
0.1	0.1	0.04	0.03	0.04	0.034	0.04	0.03	0.36	0.033
		0.01	0.00	0.00	0.002	0.01	0.00	0.33	0.003
Average		0.016	0.035	0.044		0.031	0.036	0.105	

Table 12: **Expected Waiting Times at the Second Queue in Example 4: Case 7**. Entries at the table depict expected waiting time estimates (above) and the minimum between the absolute relative errors and absolute errors with respect to the simulation estimates (below), for all three methods, for different orders of the queues, and for different values of ρ_1 and ρ_2. The last row depicts the average of the errors for the different combinations of methods/orders.