

Workload Interpretation for Brownian Models of Stochastic Processing Networks

J. M. Harrison

Graduate School of Business, Stanford University, Stanford, California 94305, USA,
harrison_michael@gsb.stanford.edu, <http://faculty-gsb.stanford.edu/harrison>

R. J. Williams

Department of Mathematics, University of California, San Diego, La Jolla, California 92093, USA,
williams@math.ucsd.edu, <http://www.math.ucsd.edu/~williams>

Brownian networks are a class of stochastic system models that can arise as heavy traffic approximations for stochastic processing networks. In earlier work we developed the “equivalent workload formulation” of a generalized Brownian network: denoting by $Z(t)$ the state vector of the generalized Brownian network at time t , one has a lower dimensional state descriptor $W(t) = MZ(t)$ in the equivalent workload formulation, where M is an arbitrary basis matrix for a linear space \mathcal{M} that is orthogonal to the space of so-called “reversible displacements.” Here we use the special structure of a stochastic processing network to develop a more extensive interpretation of the equivalent workload formulation associated with its Brownian network approximation. In particular, we (i) characterize and interpret the notion of a reversible displacement, and (ii) show how the basis matrix M can be constructed from the basic optimal solutions of a certain dual linear program. The latter provides a mechanism for reducing the choices for M from an infinite set to a finite one (when the workload dimension exceeds one). We illustrate our results for an example of a closed stochastic processing network.

Key words: Brownian network model; workload; state space collapse; reversible displacements; dual problem; stochastic control; singular control; stochastic processing network

MSC2000 subject classification: 60J60, 60K30, 90B15, 90C05, 93E20

OR/MS subject classification: queues: networks, diffusion models; networks: stochastic; probability: stochastic model applications

History: Received December 6, 2005; revised August 20, 2006.

1. Introduction. This paper continues a line of research that was initiated in Harrison [6] and further developed in Harrison and Williams [8]. The first of those antecedent works describes dynamic control problems (or to be more precise, dynamic resource allocation problems) associated with a broad class of “stochastic processing networks” (SPNs). It also describes a class of “generalized Brownian networks” that can be used to approximate SPNs of the identified type under heavy traffic conditions, offering purely formal arguments to support the proposed approximation. The SPNs identified in Harrison [6] include as special cases both open and closed multi-class queueing networks whose maximal content (in terms of jobs or customers) is large but finite, plus models where some processing activities involve simultaneous use of several resources, models where some activities require several different materials as inputs, and models where several different means exist for accomplishing a given task. Most importantly, the notion of “heavy traffic” advanced in Harrison [6] involves the system manager’s economic objective in a fundamental way, which allows a unified treatment of open and closed network control, and expands considerably the universe of network control problems that are amenable to heavy traffic diffusion approximations.

In Harrison and Williams [8] we considered the dynamic control problem, called a Brownian control problem, associated with a generalized Brownian network. We showed that this control problem is equivalent in a certain sense to a “reduced Brownian control problem” (RBCP) of lower dimension: denoting by $Z(t)$ the state vector of the generalized Brownian network at time t , one has a lower dimensional state descriptor $W(t) = MZ(t)$ in the RBCP, where M can be chosen as any basis matrix for a certain linear space \mathcal{M} . (Here the *rows* of M form a linearly independent set of basis vectors for \mathcal{M} ; this contrasts with the usual convention of taking the *columns* of a basis matrix to form a basis.) The space \mathcal{M} is characterized as the orthogonal complement of the space of so-called “reversible displacements.” Adopting language that is now standard in the literature, we called $W(\cdot) = \{W(t), t \geq 0\}$ a “workload process” in Harrison and Williams [8], and in a similar vein, one may describe the RBCP as an “equivalent workload formulation” of the original Brownian control problem. As Kelly and Laws [9] emphasized in their influential survey of heavy traffic approximations for dynamic routing problems, this model reduction is important not only for computation, but also for purposes of structural insight.

In this paper we consider the generalized Brownian network that corresponds to a given SPN, and we provide a more extensive interpretation of the workload process $W(\cdot)$ than was offered in Harrison and Williams [8]. In particular, we (i) characterize and interpret the notion of a reversible displacement, and (ii) show how the basis matrix M described above can be constructed from the basic optimal solutions of the dual to the linear program

used to define “heavy traffic” in Harrison [6]. The latter provides a mechanism for reducing the choices for M from an infinite set to a finite one, when the workload dimension exceeds one. The present paper is related to Harrison [6] in approximately the same way that Harrison [4] was related to Harrison and Van Mieghem [7]; broadly speaking, the goal here and in Harrison and Williams [8] is to cast the theory of “equivalent workload formulations” in a more general heavy traffic setting. It should also be said that the mathematical style of this paper differs in certain ways from what one sees in Harrison [4], following a pattern established in Bramson and Williams [2].

The remainder of the paper is organized as follows. Section 2 explains what is meant by the “first-order data” of a stochastic processing network, recapitulating relevant aspects of the general model formulation advanced in Harrison [6]. There we also present an example of a closed SPN that will be used to illustrate various results in the paper. Section 3 reviews the “static planning problem” (a linear program) that was used in Harrison [6] to define heavy traffic, and §4 describes the first-order data for the Brownian network approximation to an SPN satisfying our heavy traffic assumption. (The model descriptions provided in §2 and §4 of this paper are less than comprehensive, because the theory developed here involves only limited aspects of the original SPN and its Brownian approximation.) Section 5 defines the linear space \mathcal{M} referred to above, and §6 provides a characterization and dynamic control interpretation for the concept of reversible displacements that underlies our definition of workload. Theorem 7.1 of §7 describes the relationship between the linear space \mathcal{M} and the basic optimal solutions to the dual of the static planning problem. (The proof given in §7 depends on results developed in §6 and preliminary results developed in §7. This expository sequence has been chosen to maximize intuitive understanding, and more specifically, to interpret the notion of reversible displacements in a dynamic control context. It turns out that a much shorter algebraic proof of Theorem 7.1 can be given using the definitions and assumptions laid out in §2 through §5. For the interested reader, such a proof is provided in Appendix B.) In §8 we apply the general theory to our example.

1.1. Notation and terminology. For a positive integer k , \mathbb{R}^k will denote k -dimensional Euclidean space, \mathbb{R}_+^k will denote the nonnegative orthant in \mathbb{R}^k , and \mathbb{R}_{++}^k will denote the subset of vectors in \mathbb{R}_+^k having all components strictly positive. We consider each of these spaces to be endowed with its Borel σ -algebra. When $k = 1$, we shall often suppress the superscript in this notation. For convenience, we define \mathbb{R}^0 and \mathbb{R}_+^0 to be the real number zero. The Borel σ -algebras on these last two spaces consist of the empty set and the whole space. These conventions concerning \mathbb{R}^0 and \mathbb{R}_+^0 are used in treating the degenerate case of a zero-dimensional workload. For $x \in \mathbb{R}$, $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. All vectors will be assumed to be column vectors unless indicated otherwise. Inequalities between vectors are to be interpreted componentwise. The transpose of a vector or matrix will be denoted by using a superscript $'$. The dot product between two vectors $x, y \in \mathbb{R}^k$ will be denoted by $x'y$ or $x \cdot y$. The Euclidean norm of a vector $x \in \mathbb{R}^k$ will be denoted by $\|x\|$. We define the infimum of an empty set of real numbers to be $+\infty \equiv \infty$.

2. First-order data from a stochastic processing network. Assuming readers are familiar with the general formulation of an SPN that was developed in Harrison [6], we shall recapitulate in this section only the notation and definitions that are essential for current purposes. Further commentary on the model class, especially concerning its historical antecedents in economics and applied mathematics, can be found in Harrison [5].

Taken as primitive are positive integers ℓ , m , and n . We interpret ℓ as the number of processing *resources*, which may also be called *processors* or *servers* depending on the context, m as the number of distinct *materials* appearing anywhere in the network, and hence also the number of storage buffers, and n as the number of processing *activities*. The first-order data of concern in this section and the next are an m -vector λ^* , an $m \times n$ matrix R , an $\ell \times n$ nonnegative matrix A , and an n -vector v . These are interpreted as in §3 of Harrison [6]. By scaling appropriately we can take the ℓ -vector of resource capacities (denoted by q in Harrison [6]) to be $\mathbf{1}$ (the vector of ones).

An example of an open processing network was introduced and discussed in the earlier works (Harrison [6], Harrison and Williams [8]). One could use that example to illustrate the concepts treated here. However, to further illustrate the versatility of the SPN modeling framework and to provide an example where the Brownian workload is more than one-dimensional, here we use a *closed* processing network example. This system is portrayed in standard form in Figure 1(a). (A reduced form of this system will be described later, and Figure 1(b) is associated with that reduction.) For this example, we imagine that materials flow through the system as discrete jobs or customers, and we shall speak in terms of different “job classes” rather than different materials. The open-ended rectangles in Figure 1(a) represent buffers in which jobs of different classes are stored while

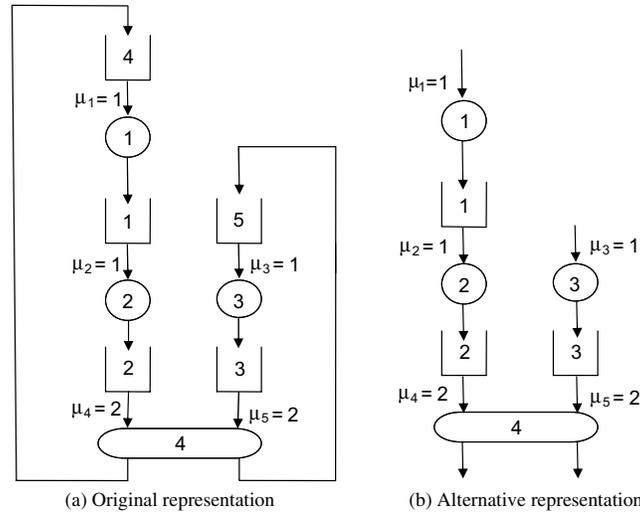


FIGURE 1. A closed network example.

awaiting service, and the circles and elongated ellipse represent servers. For concreteness in discussing this example, we shall express time in hours and economic value in dollars.

The system depicted in Figure 1(a) is closed, which means that jobs circulate perpetually, with no arrivals and no departures. There are five job classes and four servers. Each job class is processed by a unique server, so we have five processing activities. An activity is represented in Figure 1(a) by an arrow leading from the buffer involved to the server involved. Using notation that is standard in queueing theory, we denote by μ_j the reciprocal of the mean service time for activity j ($j = 1, \dots, 5$). Here we take the 5-vector $\mu = (1, 1, 1, 2, 2)'$. A complete specification of model data would include a service time distribution for each of the activities, but only the first moments (which we choose to express in the form of average service rates) are relevant for current purposes.

This example involves two closed populations: denoting by $Q_i(t)$ the number of class i jobs in the system at time t ($i = 1, \dots, 5$ and $t \geq 0$), and setting $b_1 = Q_1(0) + Q_2(0) + Q_4(0)$ and $b_2 = Q_3(0) + Q_5(0)$, one has that

$$Q_1(t) + Q_2(t) + Q_4(t) = b_1 \quad \text{and} \quad Q_3(t) + Q_5(t) = b_2 \quad \text{for all } t \geq 0. \quad (2.1)$$

Of course, these relationships imply that in order to track the evolution of job counts in the various classes over time, it suffices to observe the state vector

$$Q(t) = (Q_1(t), Q_2(t), Q_3(t))', \quad (2.2)$$

simultaneously adding the following state-space constraint to the system manager's dynamic control problem:

$$Q(t) \in \{q \in \mathbb{R}_+^3 : q_1 + q_2 \leq b_1 \text{ and } q_3 \leq b_2\}. \quad (2.3)$$

In the original model formulation, it is natural to constrain controls so that natural constraints on the controls such as activity 1 (respectively, 3) cannot be used when buffer 4 (respectively, 5) is empty. The equivalent constraint in the reduced formulation is that activity 1 (respectively, 3) cannot be used when $Q_1(t) + Q_2(t) = b_1$ (respectively, $Q_3(t) = b_2$). The reduced formulation corresponds to the system representation in Figure 1(b), where activities 1 and 3 are naturally described as “input activities,” and activities 4 and 5 are naturally described as “output activities.” The input-output matrix R and the capacity consumption matrix A for this alternative representation are

$$R = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 2 & 0 \\ 0 & 0 & -1 & 0 & 2 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \quad (2.4)$$

One interprets A_{kj} as the rate at which activity j consumes the capacity of server k , and a positive R_{ij} value as the average rate at which activity j removes jobs from buffer i . A negative R_{ij} value in (2.4) is interpreted as

INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s). Additional information, including rights and permission policies, is available at http://journals.informs.org/.

the average rate at which activity j creates class i jobs, or equivalently, the average rate at which it adds jobs to buffer i . Thus, for example, activity 2 removes jobs from buffer 1 at an average rate of 1 per hour, and it adds jobs to buffer 2 at that same average rate. The 3-dimensional vector of average arrival rates λ^* , which has one component for each of the buffers remaining in Figure 1(b), has all components zero because there are no exogenous arrivals for a closed network.

The remaining piece of first-order data for this closed network model is the 5-dimensional vector of average reward rates v . This has one component for each activity and we assume that

$$v = (0, 0, 0, v_4, v_5)' \quad \text{where } v_4 > v_5 > 0. \quad (2.5)$$

One interprets v_j as the average reward earned *per hour of service time* devoted to activity j . Thus the system manager earns positive rewards only for “output activities,” the average reward being $v_4/\mu_4 = v_4/2$ dollars per service completed for class 2 jobs, and $v_5/\mu_5 = v_5/2$ dollars per service completed for class 3 jobs.

The system portrayed in Figure 1(a) may be interpreted as a factory that makes two families of products and operates under a “closed-loop input control mechanism” (cf. Solberg [10]). For example, jobs of classes 1, 2, and 4 may represent orders for products in family A , with the class designation of each job indicating its stage of completion, and jobs of classes 3 and 5 similarly may represent outstanding orders (that is, orders released to the factory but not yet completed) for products in family B . To be perfectly concrete, one may suppose that class 4 jobs and class 5 jobs represent orders that have been released but not yet worked on, and that execution of a class 2 service or a class 3 service completes work on an order. The closed-loop control mechanism we have hypothesized is one that releases a new order for a product in family A immediately after completion of an outstanding order for such a product, thus keeping constant the number of outstanding orders for family A , and similarly it keeps constant the number of outstanding orders for products in family B . In formulating a dynamic control policy, the system manager has discretion as to how server 4 will divide its time between processing class 2 jobs and processing class 3 jobs.

3. Defining heavy traffic by means of a linear program. As in §3 of Harrison [6] we consider the following *static planning problem*:

$$\text{maximize } v'x \quad \text{subject to } Rx = \lambda^*, \quad Ax \leq \mathbf{1} \quad \text{and } x \in \mathbb{R}_+^n. \quad (3.1)$$

We interpret an optimal solution x^* of (3.1) as a vector of desired average activity rates (cf. §3 of Harrison [6]). The following heavy traffic assumption will be assumed henceforth. This condition is the same as that introduced in §3 of Harrison [6].

ASSUMPTION 3.1. *The static planning problem (3.1) has a unique optimal solution x^* , and moreover, $Ax^* = \mathbf{1}$.*

Preparing the way for later developments, we denote by b the number of activities j such that $x_j^* > 0$, calling these *basic activities*, and we number the activities so that the basic ones are $1, \dots, b$. As in Harrison [6], activities $b + 1, \dots, n$ will be called *nonbasic*, and the matrices R and A will be partitioned as follows:

$$R = [H \quad J] \quad \text{and} \quad A = [B \quad N], \quad (3.2)$$

where H and B both have b columns. Thus H and B are the submatrices of R and A , respectively, that correspond to basic activities.

With an eye toward future developments, we note that the following linear program is dual to (3.1), or to be more precise, it is one way in which the dual of (3.1) can be written:

$$\text{minimize } y'\lambda^* + z'\mathbf{1} \quad \text{subject to } y'R + z'A \geq v' \quad \text{and } z \in \mathbb{R}_+^\ell. \quad (3.3)$$

We denote by \mathcal{D} the polyhedral set of feasible solutions for the dual linear program (3.3), meaning that

$$\mathcal{D} = \{(y, z) \in \mathbb{R}^{m+\ell}: y'R + z'A \geq v' \text{ and } z \in \mathbb{R}_+^\ell\}. \quad (3.4)$$

Note that \mathcal{D} does not depend on λ^* . By Assumption 3.1, the primal problem (3.1) has a finite optimal solution and hence so does the dual problem (3.3), which implies that the set \mathcal{D} is nonempty.

In the closed network example of Figure 1, for which $\lambda^* = 0$ and other first-order data are displayed in (2.4) and (2.5), it is easy to verify that Assumption 3.1 holds with

$$x^* = (1, 1, 1, \frac{1}{2}, \frac{1}{2})'. \quad (3.5)$$

That is, each of servers 1, 2, and 3 works full-time on the one activity of which it is capable, and server 4 divides its time evenly between its two activities. This processing plan uses all capacity of every server ($Ax^* = \mathbf{1}$),

maximizing the circulation rate for each of the two closed populations in Figure 1 and thereby generating rewards at the maximal rate of $v^*x^* = (v_4 + v_5)/2$ dollars per hour.

4. First-order data for the Brownian network approximation. Now set

$$p = \ell + n - b \quad (4.1)$$

and define a $p \times n$ matrix K via

$$K = \begin{bmatrix} B & N \\ 0 & -I \end{bmatrix}. \quad (4.2)$$

Comparing (4.2) with (3.2), one sees that the first ℓ rows of K form the capacity consumption matrix A , and the negative identity matrix $-I$ appearing in (4.2) is a square $(n - b) \times (n - b)$ matrix, where $n - b$ is the number of nonbasic activities in our static planning problem.

Sections 4 and 5 of Harrison [6] explain how to formulate a generalized Brownian network that approximates, in a certain formal sense, a stochastic processing network whose first-order data satisfy the heavy traffic Assumption 3.1 above. The first-order data of that Brownian network are the following: the $m \times n$ matrix R (see §2); the $p \times n$ matrix K defined via (4.2); a compact, convex state space $\mathcal{X} \subset \mathbb{R}^m$ that has nonempty interior; and the n -vector v (see §2).

For the closed queueing network example in §2, a generalized Brownian network approximation is appropriate when the number of jobs b_1, b_2 in each of the two closed populations is of the same large order of magnitude r and the behavior of the system is considered over suitably large intervals of time (of order r^2). Writing $b_1 = r\hat{b}_1$ and $b_2 = r\hat{b}_2$ where \hat{b}_1 and \hat{b}_2 are assumed to be of moderate size, and using Q to denote the three-dimensional queue length process described in (2.2) for the closed network example, it is natural to consider a renormalized state descriptor at time t given by

$$\hat{Q}^r(t) = Q(r^2t)/r.$$

Under suitable heavy traffic assumptions as described in §4 and §5 of Harrison [6], the Brownian network state process $Z(\cdot)$ is a formal approximation to the renormalized state process $\hat{Q}^r(\cdot)$, where the state space for $Z(\cdot)$ is $\mathcal{X} = [0, \hat{b}_1] \times [0, \hat{b}_2]$.

It was shown in Harrison and Williams [8] that the following condition is necessary and sufficient for well posedness of the Brownian control problem (that is, there exist admissible controls for the Brownian control problem if and only if the first-order data satisfy this condition).

ASSUMPTION 4.1. $\{Ry: Ky \geq 0, y \in \mathbb{R}^n\} = \mathbb{R}^m$.

Of course, Assumption 4.1 implies that the $m \times n$ matrix R has full row rank. To further explain the meaning of Assumption 4.1, it will be useful to define

$$\Lambda \equiv \{Rx: Ax \leq \mathbf{1}, x \in \mathbb{R}_+^n\}. \quad (4.3)$$

Using linear programming language, one may describe Λ as the set of all m -vectors λ such that problem (3.1) remains feasible when λ^* is replaced by λ . Given an m -vector $\lambda \in \Lambda$, we shall say that λ is *achieved* by an n -vector x , or that λ is *achievable using* x , if $Ax \leq \mathbf{1}$, $x \in \mathbb{R}_+^n$, and $Rx = \lambda$.

For the following lemma, recall that we are assuming that Assumption 3.1 holds. In particular, $\lambda^* \in \Lambda$. The proof of this lemma is given in Appendix A.

LEMMA 4.1. Assumption 4.1 holds if and only if λ^* lies in the interior of Λ , i.e.,

$$\text{there exists an } \varepsilon > 0 \text{ such that } \{\lambda \in \mathbb{R}^m: \|\lambda - \lambda^*\| < \varepsilon\} \subset \Lambda. \quad (4.4)$$

We assume that Assumption 4.1 holds henceforth. This assumption is satisfied by the closed network example described earlier in §2 (see §8 for elaboration).

5. Workload. Let

$$\mathcal{N} \equiv \{\beta \in \mathbb{R}^n: K\beta = 0\}, \quad (5.1)$$

$$\mathcal{R} \equiv \{R\beta: \beta \in \mathcal{N}\}. \quad (5.2)$$

Elements of the linear space \mathcal{R} are called *reversible displacements*. We shall denote by \mathcal{M} the orthogonal complement of \mathcal{R} in \mathbb{R}^m . It was shown in §4 of Harrison and Williams [8] that

$$\mathcal{M} = \{\gamma \in \mathbb{R}^m: \gamma'R = \eta'K \text{ for some } \eta \in \mathbb{R}^p\}. \quad (5.3)$$

We denote the dimension of \mathcal{M} by d , and refer to d as the *workload dimension*. If $d \geq 1$, we let M denote a linear mapping from \mathbb{R}^m onto \mathbb{R}^d represented by a $d \times m$ matrix whose rows are a maximal linearly independent set of vectors in \mathcal{M} . If $d = 0$, then M is the linear mapping from \mathbb{R}^m onto \mathbb{R}^0 (the real number zero); the degenerate case of $d = 0$ can occur in practice and in this case many manipulations simplify. In Harrison and Williams [8], the mapping M was used to define a workload process associated with a Brownian network. Consistent with the terminology used there, here we refer to M as a workload matrix.

REMARK 5.1. At this point our review of essential definitions from Harrison and Williams [8] is complete. As noted earlier in the Introduction, the matrix M has the following significance: denoting by $Z(t)$ the state vector of our original Brownian control problem at time t , then $W(t) = MZ(t)$ will be the lower dimensional state descriptor in the reduced Brownian control problem. The description of W as a “workload process” originated in Harrison [3], where a restricted class of stochastic processing networks was considered, namely, multiclass queueing networks with sequencing control. In that context, the term “workload” was rather obviously appropriate. In Harrison and Van Mieghem [7] and Harrison [4], Brownian approximations for a more general class of stochastic processing networks were considered; here again W was called a “workload process,” and an argument was advanced in §4 of Harrison [4] to justify that terminology. Our use of the term “workload” here and in Harrison and Williams [8], is motivated by analogy with this prior usage. In this context, there is no apparent clear interpretation for the terminology related to quantities associated with the motivating stochastic processing networks (Harrison [6]).

In the next two sections, for the Brownian networks associated with the SPNs introduced in Harrison [6], we expand on the notion of reversible displacements, eventually showing (Theorem 7.2) how the workload matrix M can be constructed from basic optimal solutions of the dual linear program (3.3).

6. Reversible displacements.

DEFINITION 6.1. For each $\lambda \in \Lambda$, let $\nu(\lambda)$ denote the optimal objective value in the linear programming problem (3.1) when λ is substituted for λ^* . For $\lambda \notin \Lambda$, we set $\nu(\lambda) = -\infty$. Also, let

$$\nu^* \equiv \nu(\lambda^*) = v'x^*. \quad (6.1)$$

REMARK 6.1. The dual problem (3.3) remains feasible when an arbitrary vector $\lambda \in \mathbb{R}^m$ is substituted for λ^* , because λ appears only in the objective function of the dual problem. Thus, by the duality theory of linear programming, the objective in the primal problem (3.1) remains bounded above when an arbitrary λ is substituted for λ^* . That is, $\nu(\lambda)$ cannot take the value $+\infty$.

Recall that a displacement $\delta \in \mathbb{R}^m$ is said to be reversible if $\delta = R\beta$ for some $\beta \in \mathbb{R}^n$ such that $K\beta = 0$. From the definition (4.2) of K one sees that the condition $K\beta = 0$ simply means that

$$A\beta = 0 \quad \text{and} \quad \beta_j = 0 \quad \text{for } j = b + 1, \dots, n. \quad (6.2)$$

6.1. Characterization of reversible displacements.

LEMMA 6.1. A displacement $\delta \in \mathbb{R}^m$ is reversible if and only if, for each sufficiently large $t > 0$ there exists $x \in \mathbb{R}^n$ satisfying

$$Rx = \lambda^* + \frac{1}{t}\delta, \quad Ax = \mathbf{1}, \quad x \in \mathbb{R}_+^n \quad \text{and} \quad x_j = 0 \quad \text{for } j = b + 1, \dots, n. \quad (6.3)$$

PROOF. First assume that δ is reversible. Let $\beta \in \mathbb{R}^n$ be such that $\delta = R\beta$ and (6.2) holds. Then $x = x^* + (1/t)\beta$ satisfies (6.3) for all sufficiently large $t > 0$. Conversely, if $x \in \mathbb{R}^n$ satisfies (6.3) for some $t > 0$, we can take $\beta = t(x - x^*)$ to satisfy the definition of reversibility. \square

LEMMA 6.2. For each $\delta \in \mathbb{R}^m$ and $t > 0$,

$$\nu\left(\lambda^* + \frac{1}{t}\delta\right) + \nu\left(\lambda^* - \frac{1}{t}\delta\right) \leq 2\nu^*. \quad (6.4)$$

PROOF. Let (y^*, z^*) be an arbitrary optimal solution for the dual problem (3.3). (Assumption 3.1 guarantees that the dual problem has an optimal solution.) If we consider the primal and dual problems (3.1) and (3.3) with some other $\lambda \in \mathbb{R}^m$ in place of λ^* , by the duality theory of linear programming, the pair (y^*, z^*) is still a

feasible solution for the dual problem, and one has

$$v(\lambda) \leq y^* \cdot \lambda + z^* \cdot \mathbf{1}. \tag{6.5}$$

Applying this inequality to both $\lambda = \lambda^* + (1/t)\delta$ and $\lambda = \lambda^* - (1/t)\delta$ gives (6.4), because $y^* \cdot \lambda^* + z^* \cdot \mathbf{1} = v^*$. \square

THEOREM 6.1. A vector $\delta \in \mathbb{R}^m$ is a reversible displacement if and only if (6.4) holds with equality for all sufficiently large $t > 0$.

REMARK 6.2. Examination of the proof below reveals that the above theorem remains true if “for all sufficiently large $t > 0$ ” is replaced with “for some $t > 0$.”

PROOF. First suppose that δ is reversible and let β be as in the definition of a reversible displacement. Just as in the proof of Lemma 6.1, $\lambda^* + (1/t)\delta$ and $\lambda^* - (1/t)\delta$ are achievable using $x^* + (1/t)\beta$ and $x^* - (1/t)\beta$, respectively, for all sufficiently large $t > 0$. Thus,

$$\begin{aligned} v\left(\lambda^* + \frac{1}{t}\delta\right) + v\left(\lambda^* - \frac{1}{t}\delta\right) &\geq v\left(x^* + \frac{1}{t}\beta\right) + v\left(x^* - \frac{1}{t}\beta\right) \\ &= 2v'x^* = 2v^*. \end{aligned} \tag{6.6}$$

Then, by Lemma 6.2, (6.4) must hold with equality.

Now suppose on the other hand that (6.4) holds with equality for some $t > 0$. Then $\lambda^* + (1/t)\delta$ and $\lambda^* - (1/t)\delta$ are achievable using vectors $x^{(1)}$ and $x^{(2)}$, respectively, where $v'x^{(1)} + v'x^{(2)} = 2v^*$. This implies that λ^* is achievable using $x \equiv \frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)}$, and that $v'x = v^*$. That is, this vector x is an optimal solution for the primal problem (3.1), so the uniqueness part of Assumption 3.1 implies that $x = x^*$. Then it must be that components $b+1, \dots, n$ of both $x^{(1)}$ and $x^{(2)}$ are all zero, because $x^{(1)} \geq 0$ and $x^{(2)} \geq 0$ and $x_{b+1}^* = \dots = x_n^* = 0$. Furthermore, since $Ax^{(1)} \leq \mathbf{1}$, $Ax^{(2)} \leq \mathbf{1}$, and $\frac{1}{2}(Ax^{(1)} + Ax^{(2)}) = \mathbf{1}$, it must be that $Ax^{(1)} = \mathbf{1}$ and $Ax^{(2)} = \mathbf{1}$. Thus, on setting $\beta = t(x^* - x^{(2)})$, we have that $K\beta = 0$, $R\beta = \delta$, implying that δ is reversible. \square

6.2. Fluid model interpretation of reversible displacements. Here we consider a fluid model associated with a stochastic processing network. (One may think of this fluid model as a formal functional law of large numbers approximation to the original SPN.) The data for this fluid model consist of the first-order data (λ^*, R, A, v) introduced in §2, plus the same compact, convex state space \mathcal{X} (having nonempty interior) that is associated with the approximating Brownian network.

DEFINITION 6.2. An admissible control for the fluid model with initial state $\zeta^o \in \mathcal{X}$ is a nondecreasing function $\tau: [0, \infty) \rightarrow \mathbb{R}_+^n$ such that for ζ, u defined by

$$\zeta(t) \equiv \zeta^o + \lambda^*t - R\tau(t), \quad t \geq 0, \tag{6.7}$$

$$u(t) \equiv \mathbf{1}t - A\tau(t), \quad t \geq 0, \tag{6.8}$$

we have

- (i) $\zeta(t) \in \mathcal{X}$ for all $t \geq 0$,
- (ii) $u(\cdot)$ is nondecreasing with $u(0) \geq 0$.

For each $t \geq 0$, the cumulative value generated by such an admissible control τ over the time interval $[0, t]$ is

$$V_\tau(t) \equiv v'\tau(t). \tag{6.9}$$

LEMMA 6.3. Suppose that, for some $\zeta^o \in \mathcal{X}$ and $t > 0$, there is an admissible control τ for the fluid model such that $\zeta(t) = \zeta(0) = \zeta^o$. Then,

$$\frac{1}{t}V_\tau(t) \leq v^*, \tag{6.10}$$

where equality holds if and only if

$$\tau(t) = tx^*. \tag{6.11}$$

PROOF. Let $x = (1/t)\tau(t)$. Now, $x \in \mathbb{R}_+^n$ and

$$Ax = \mathbf{1} - \frac{1}{t}u(t) \leq \mathbf{1},$$

since $u(t) \geq u(0) \geq 0$. Furthermore, $Rx = \lambda^*$ because $\zeta(t) = \zeta(0) = \zeta^o$. Thus, x is a feasible solution for the linear program (3.1) and so

$$\frac{1}{t}V_\tau(t) = v'x \leq v^*, \tag{6.12}$$

where the last inequality is an equality if and only if $x = x^*$, by Assumption 3.1. \square

THEOREM 6.2. *Suppose that $\zeta^o, \zeta^\dagger \in \mathcal{X}$ and define $\delta = \zeta^\dagger - \zeta^o$. Then δ is reversible if and only if the following holds:*

$$\begin{aligned} & \text{There exists an admissible control } \tau \text{ for the fluid model with initial state } \zeta^o \text{ and} \\ & \text{a } T > 0 \text{ such that } \zeta(T) = \zeta^o, \zeta(t) = \zeta^\dagger \text{ for some } t \in (0, T), \text{ and } V_\tau(T) = T\nu^*. \end{aligned} \quad (6.13)$$

REMARK 6.3. In words, (6.13) says that starting from state ζ^o , one can travel to state ζ^\dagger and then back to ζ^o in finite time, achieving an average value rate of ν^* (the maximum achievable average value rate, cf. Lemma 6.3) for the period of the round trip.

PROOF. Suppose that δ is reversible and let β be as in the definition of reversible displacements. Then, as in the proof of Theorem 6.1, there exists $t > 0$ large enough that $\lambda^* - (1/t)\delta$ is achievable using $x^* - (1/t)\beta$, and $\lambda^* + (1/t)\delta$ is achievable using $x^* + (1/t)\beta$. Set $T = 2t$, $\tau(s) = (x^* - (1/t)\beta)s$, for $0 \leq s \leq t$, $\tau(s) = \tau(t) + (x^* + (1/t)\beta)(s - t)$ for $t < s \leq 2t \equiv T$, and $\tau(s) = \tau(T) + x^*(s - T)$ for all $s > T$. Then it is readily verified that $\zeta(s) = \zeta^o + \lambda^*s - R\tau(s)$, $0 \leq s < \infty$, satisfies $\zeta(0) = \zeta^o$, $\zeta(t) = \zeta^\dagger$, $\zeta(2t) = \zeta^o$, and $\zeta(s) = \zeta^o$ for all $s > T = 2t$. For $u(s) = \mathbf{1}s - A\tau(s)$, $0 \leq s < \infty$, since $Ax \leq \mathbf{1}$ for $x = x^* - (1/t)\beta$, $x = x^* + (1/t)\beta$ and $x = x^*$, we have that $u(\cdot)$ is nondecreasing and also $u(0) = 0$. Finally, using the convexity of \mathcal{X} , it is readily shown that $\zeta(s) \in \mathcal{X}$ for all $s \in [0, T]$ and hence for all $s \in [0, \infty)$. By direct computation we see that

$$V_\tau(T) = v' \left(t \left(x^* - \frac{1}{t}\beta \right) + t \left(x^* + \frac{1}{t}\beta \right) \right) = 2tv'x^* = 2t\nu^* = T\nu^*. \quad (6.14)$$

Thus (6.13) holds.

Now suppose on the other hand that (6.13) holds. Without loss of generality we can assume that $T = 2t$, because either the period of travel from ζ^o to ζ^\dagger or the period of return travel from ζ^\dagger to ζ^o can be extended by inserting an interval (s_1, s_2) of indefinite length and augmenting τ so that $\tau(s) = \tau(s_1) + x^*(s - s_1)$ for $s \in (s_1, s_2)$ (the state does not change during this interval and value is generated precisely at rate ν^*). Given that $T = 2t$, set

$$x^{(1)} = \frac{1}{t}\tau(t) \quad \text{and} \quad x^{(2)} = \frac{1}{t}(\tau(2t) - \tau(t)). \quad (6.15)$$

Then $\lambda^* - (1/t)\delta$ is achievable using $x^{(1)}$ and $\lambda^* + (1/t)\delta$ is achievable using $x^{(2)}$ (note that $x^{(1)}$ and $x^{(2)}$ have nonnegative components because τ is assumed to be nonnegative and nondecreasing). It follows from the definition of ν that

$$\nu \left(\lambda^* - \frac{1}{t}\delta \right) + \nu \left(\lambda^* + \frac{1}{t}\delta \right) \geq v'x^{(1)} + v'x^{(2)} = \frac{1}{t}V_\tau(2t). \quad (6.16)$$

From (6.13) we have that $V_\tau(2t) = 2t\nu^*$, so it follows from Lemma 6.2 that (6.16) must hold with equality. Then the remark following Theorem 6.1 implies that δ is reversible. \square

7. Optimal dual solutions and the workload matrix. Recall from (3.4) that \mathcal{D} denotes the set of feasible solutions (y, z) for our dual linear program (3.3). From Assumption 3.1 we know that the primal linear program (3.1) has an optimal solution and hence so does its dual (3.3). We denote by \mathcal{D}^* the set of *all* optimal solutions for the dual problem. It is a general result in linear programming theory that the primal and dual problems have the same optimal objective value, so one can equivalently define

$$\mathcal{D}^* = \{(y, z) \in \mathcal{D}: y'\lambda^* + z'\mathbf{1} = \nu^*\}. \quad (7.1)$$

LEMMA 7.1. *\mathcal{D}^* is bounded.*

REMARK 7.1. We note in passing that the boundedness of the nonempty set \mathcal{D}^* ensures that the kind of difficulty (e.g., no extreme points of \mathcal{D}) discussed in §7.1 of Bramson and Williams [2] in relation to the class of stochastic processing networks considered in Harrison [4] will not occur in the context of the SPN models introduced in Harrison [6] that we consider here.

PROOF. From (3.4) and (7.1) it follows that \mathcal{D}^* is a (nonempty) convex polyhedron, and so \mathcal{D}^* is unbounded if and only if it contains a ray, cf. p. 175 and Exercise 4.47 of Bertsimas and Tsitsiklis [1]. That is, \mathcal{D}^* is

unbounded if and only if there exists a pair $(\gamma, \eta) \in \mathbb{R}^{m+\ell} \setminus \{(0, 0)\}$ that satisfies

$$\gamma'R + \eta'A \geq 0, \quad \eta \geq 0, \quad \text{and} \quad \gamma'\lambda^* + \eta'\mathbf{1} = 0. \quad (7.2)$$

A ray in \mathcal{D}^* consists of pairs $(y, z) \in \mathbb{R}^{m+\ell}$ having the form

$$(y, z) = (y^*, z^*) + a(\gamma, \eta), \quad a \geq 0, \quad (7.3)$$

where $(y^*, z^*) \in \mathcal{D}^*$ and (γ, η) satisfies (7.2). The two inequalities in (7.2) are needed to ensure that all such (y, z) pairs are *feasible* solutions for the dual linear program, and then the equality in (7.2) is needed to ensure that they are all *optimal* solutions.

Arguing by contradiction, let us suppose that there exists a pair $(\gamma, \eta) \neq (0, 0)$ satisfying (7.2). Note that one cannot have $\gamma = 0$ because the second and third conditions in (7.2) would then imply $\eta = 0$ as well. Thus there exists $\delta \in \mathbb{R}^m$ such that $\gamma'\delta < 0$. Now consider the primal and dual linear programs (3.1) and (3.3) with $\lambda = \lambda^* + (1/t)\delta$ in place of λ^* , where $t > 0$ is arbitrary. Any (y, z) of the form (7.3) remains a feasible solution for the dual linear program, because λ enters the dual problem only through its objective function. Furthermore, when λ replaces λ^* in the dual linear program, we see from (7.1) and (7.2) that a feasible pair (y, z) having the form (7.3) has an associated objective value

$$y \cdot \lambda + z \cdot \mathbf{1} = \nu^* + \frac{1}{t}(y^* \cdot \delta) + \frac{a}{t}(\gamma \cdot \delta). \quad (7.4)$$

By taking a arbitrarily large, we can drive the right-hand side of (7.4) arbitrarily far in the negative direction. Thus our dual linear program (a minimization problem) is unbounded below, implying that the corresponding primal linear program is infeasible. That is equivalent to saying that $\lambda \notin \Lambda$, regardless of how large t may be. According to Lemma 4.1, that contradicts Assumption 4.1, so we conclude that \mathcal{D}^* is bounded. \square

Combining (3.4), (7.1), and Lemma 7.1, we know that \mathcal{D}^* is a nonempty, bounded, convex polyhedron. Thus it is generated by its nonempty, finite set of extreme points. Each of its extreme points is also an extreme point of \mathcal{D} (this is an easy exercise), and \mathcal{D} has only finitely many extreme points. Consequently, the following definition makes sense.

DEFINITION 7.1. Let $\{(y^l, z^l): l = 1, \dots, L\}$ be the extreme points of \mathcal{D} , and let those numbered $1, \dots, L^*$ be the extreme points of \mathcal{D}^* ($1 \leq L^* \leq L < \infty$).

REMARK 7.2. The extreme points of \mathcal{D} are precisely the basic feasible solutions for the system of equality and inequality constraints that define \mathcal{D} , see p. 50 of Bertsimas and Tsitsiklis [1]. Extreme points of \mathcal{D}^* are called basic optimal solutions of the dual linear program (3.3).

LEMMA 7.2. For each $\delta \in \mathbb{R}^m$ one has the following:

$$\nu\left(\lambda^* + \frac{1}{t}\delta\right) = \nu^* + \frac{1}{t} \min_{1 \leq l \leq L^*} (y^l \cdot \delta) \quad (7.5)$$

for all sufficiently large $t > 0$.

PROOF. Fix $\delta \in \mathbb{R}^m$. For large $t > 0$ we write $\lambda = \lambda^* + (1/t)\delta$ and then focus on the linear programs (3.1) and (3.3) with λ in place of λ^* . From Assumptions 3.1 and 4.1, and Lemma 4.1, we know that both problems are feasible for sufficiently large $t > 0$. So for all sufficiently large $t > 0$, each problem has at least one optimal solution. Because the feasible region \mathcal{D} for the dual problem, which does not depend on λ , has at least one extreme point (see above), it follows that the dual problem has an extreme point optimal solution for each sufficiently large $t > 0$; cf. Theorem 2.7 in Bertsimas and Tsitsiklis [1]. Because $\nu(\lambda)$ equals the optimal objective value for the dual problem, we thus have the following for all sufficiently large $t > 0$:

$$\nu\left(\lambda^* + \frac{1}{t}\delta\right) = \min_{1 \leq l \leq L^*} \left\{ y^l \cdot \left(\lambda^* + \frac{1}{t}\delta\right) + z^l \cdot \mathbf{1} \right\}. \quad (7.6)$$

For all $t > 0$ sufficiently large, the minimum in (7.6) is attained at one of the extreme points labeled by $l = 1, \dots, L^*$ that is optimal at $t = \infty$, and for each of those, one has $y^l \cdot \lambda^* + z^l \cdot \mathbf{1} = \nu^*$. \square

THEOREM 7.1. A vector $\delta \in \mathbb{R}^m$ is a reversible displacement if and only if there exists $c \in \mathbb{R}$ such that

$$y^l \cdot \delta = c \quad \text{for } l = 1, \dots, L^*. \quad (7.7)$$

REMARK 7.3. Because \mathcal{D}^* is a bounded polyhedron (see Lemma 7.1), each point $(y, z) \in \mathcal{D}^*$ can be written as a convex combination of the extreme points (y^l, z^l) , $l = 1, \dots, L^*$. Thus, defining

$$\mathcal{Y}^* = \{y \in \mathbb{R}^m: (y, z) \in \mathcal{D}^* \text{ for some } z \in \mathbb{R}_+^{\ell}\}, \quad (7.8)$$

one sees that (7.7) is equivalent to

$$y \cdot \delta = c \quad \text{for all } y \in \mathcal{Y}^*. \quad (7.9)$$

REMARK 7.4. The short proof below depends on results developed in §6 and earlier in this section. This expositional sequence has been chosen to maximize intuitive understanding, and more specifically, to interpret the notion of reversible displacements in a dynamic control context. Appendix B provides a much shorter proof of Theorem 7.1 that proceeds directly from the definitions and assumptions laid out in §2 through §5.

PROOF. It follows from Theorem 6.1 and Lemma 7.2 that $\delta \in \mathbb{R}^m$ is reversible if and only if

$$\min_{1 \leq l \leq L^*} y^l \cdot \delta = \max_{1 \leq l \leq L^*} y^l \cdot \delta. \quad (7.10)$$

The above occurs if and only if (7.7) holds for some constant $c \in \mathbb{R}$. \square

THEOREM 7.2. *The space \mathcal{M} is spanned by $\{y^1 - y^1, y^2 - y^1, \dots, y^{L^*} - y^1\}$.*

PROOF. By Theorem 7.1, $\delta \in \mathcal{R}$ if and only if (7.7) holds, which is equivalent to

$$(y^l - y^1) \cdot \delta = 0 \quad \text{for } l = 1, 2, \dots, L^*. \quad (7.11)$$

It follows that $\mathcal{M} = \mathcal{R}^\perp$ is spanned by $y^l - y^1$, $l = 1, 2, \dots, L^*$. We include the zero vector here to cover the case where $L^* = 1$ and $\mathcal{M} = \{0\}$. \square

Thus, for $d \geq 1$, one can form a “canonical” basis for the space $\mathcal{M} = \mathcal{R}^\perp$ by taking a maximal linearly independent subset from $\{y^l - y^1: l = 1, \dots, L^*\}$. This subset can be used to form the rows of the matrix M . For $d = 0$, M is always the (trivial) linear mapping onto \mathbb{R}^0 .

REMARK 7.5. The method used here for constructing the workload matrix M from basic optimal solutions of the dual linear program (3.3) is different from the analogous procedure developed in Harrison [4]. In that earlier analysis, the workload space \mathcal{M} was generated by the y -components of the basic optimal solutions for the dual linear program. Here one has to first “center” the analogous y -components. Of course, the static planning problem underlying the theory developed in Harrison [4] is different in character from the one underlying our current theory (see §3), so this distinction is not surprising.

8. Analysis of example. For the closed network example pictured in Figure 1, whose first-order data were displayed in (2.4) and (2.5), we recorded the unique optimal solution x^* of the static planning problem (3.1) in (3.5). In this case all five activities are basic, so (4.2) reduces to $K = A$, and using that fact, readers can easily verify that Assumption 4.1 is satisfied. For the dual linear program (3.3), using complementary slackness and the fact that the optimal value for the dual problem is equal to the optimal value for the primal problem, the set of dual optimal solutions is seen to be the set of $(y, z) \in \mathbb{R}^3 \times \mathbb{R}^4$ satisfying

$$y'R + z'A = v', \quad (8.1)$$

$$z \in \mathbb{R}_+^4, \quad (8.2)$$

$$z'1 = \frac{v_4 + v_5}{2}. \quad (8.3)$$

Using simple linear algebra, the set of all solutions of (8.1) is easily seen to be the set of $(y, z) \in \mathbb{R}^3 \times \mathbb{R}^4$ such that

$$(y', z') = (0, 0, \frac{1}{2}(v_5 - v_4), 0, 0, \frac{1}{2}(v_5 - v_4), v_4) + a_1(1, 0, 0, 1, -1, 0, 0) + a_2(0, 1, 1, 0, 1, 1, -2), \quad (8.4)$$

where $a_1, a_2 \in \mathbb{R}$. The constraint (8.2) reduces this to the set of $(y, z) \in \mathbb{R}^3 \times \mathbb{R}^4$ satisfying (8.4) where $a_1, a_2 \in \mathbb{R}$ satisfy $0 \leq a_1 \leq a_2$, $\frac{1}{2}(v_4 - v_5) \leq a_2 \leq \frac{1}{2}v_4$. We note that for each such (y, z) , $z'1 \equiv (v_4 + v_5)/2$. Thus, the set specified above is precisely the set of optimal solutions of the dual problem and the extreme points of this set

are the basic optimal solutions of the dual problem. The four extreme points of this set are readily seen to be given by

$$\begin{aligned} y^1 &= (0, \frac{1}{2}(v_4 - v_5), 0)' & \text{and} & & z^1 &= (0, \frac{1}{2}(v_4 - v_5), 0, v_5)'; \\ y^2 &= (\frac{1}{2}(v_4 - v_5), \frac{1}{2}(v_4 - v_5), 0)' & \text{and} & & z^2 &= (\frac{1}{2}(v_4 - v_5), 0, 0, v_5)'; \\ y^3 &= (0, \frac{1}{2}v_4, \frac{1}{2}v_5)' & \text{and} & & z^3 &= (0, \frac{1}{2}v_4, \frac{1}{2}v_5, 0)'; \\ y^4 &= (\frac{1}{2}v_4, \frac{1}{2}v_4, \frac{1}{2}v_5)' & \text{and} & & z^4 &= (\frac{1}{2}v_4, 0, \frac{1}{2}v_5, 0)'. \end{aligned}$$

In particular,

$$y^2 - y^1 = \frac{1}{2}(v_4 - v_5)(1, 0, 0)', \quad y^3 - y^1 = \frac{1}{2}v_5(0, 1, 1)', \quad y^4 - y^1 = \frac{1}{2}(v_4, v_5, v_5)'.$$

According to Theorem 7.2, the first two of these three vectors constitute a basis for the linear space \mathcal{M} (they are linearly independent, and the third is redundant). Rescaling for simplicity, we then have the following potential choice of basis matrix:

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}. \quad (8.5)$$

Appendix A. Proof of Lemma 4.1. Suppose that Assumption 4.1 holds. Let $e^{(1)}, \dots, e^{(m)}$ be the unit basis vectors in \mathbb{R}^m . For $i = 1, \dots, m$, let $f^{(i)} = e^{(i)}$ and for $i = m + 1, \dots, 2m$, let $f^{(i)} = -e^{(i-m)}$. Since R and K are linear, we see that Assumption 4.1 is equivalent to the following:

$$\{Ry: Ky \leq 0, y \in \mathbb{R}^n\} = \mathbb{R}^m. \quad (A1)$$

Thus, for each $i = 1, \dots, 2m$, there is $y^{(i)} \in \mathbb{R}^n$ such that

$$Ry^{(i)} = f^{(i)}, \quad Ky^{(i)} \leq 0. \quad (A2)$$

Note that $Ky^{(i)} \leq 0$ is equivalent to $Ay^{(i)} \leq 0$ and $y_j^{(i)} \geq 0$ for $j = b + 1, \dots, n$. Let

$$\gamma = \sum_{i=1}^{2m} \|y^{(i)}\| \quad \text{and} \quad \varepsilon = \min_{j=1}^b x_j^* / \gamma.$$

Note that $\varepsilon > 0$.

Fix $u \in \mathbb{R}^m$ satisfying $\|u\| < 1$. This can be written as

$$u = \sum_{i=1}^{2m} w_i f^{(i)} \quad (A3)$$

where $w_i = u_i^+$ for $i = 1, \dots, m$, and $w_i = u_{i-m}^-$ for $i = m + 1, \dots, 2m$. Consider $x = x^* + \varepsilon \sum_{i=1}^{2m} w_i y^{(i)}$. Then,

$$Rx = \lambda^* + \varepsilon \sum_{i=1}^{2m} w_i f^{(i)} = \lambda^* + \varepsilon u, \quad (A4)$$

$$Ax = \mathbf{1} + \varepsilon \sum_{i=1}^{2m} w_i Ay^{(i)} \leq \mathbf{1}, \quad (A5)$$

since $w_i \geq 0$, $Ay^{(i)} \leq 0$ for all i . Furthermore, for $j = 1, \dots, b$,

$$\begin{aligned} x_j &= x_j^* + \varepsilon \sum_{i=1}^{2m} w_i y_j^{(i)} \\ &\geq x_j^* - \varepsilon \sum_{i=1}^{2m} \|y^{(i)}\|, \quad \text{since } |w_i| \leq 1, \\ &\geq x_j^* - \varepsilon \gamma \geq 0 \end{aligned}$$

and for $j = b + 1, \dots, n$,

$$x_j = x_j^* + \varepsilon \sum_{i=1}^{2m} w_i y_j^{(i)} = 0 + \varepsilon \sum_{i=1}^{2m} w_i y_j^{(i)} \geq 0,$$

since $w_i \geq 0$ for all i and $y_j^{(i)} \geq 0$ for all i and $j = b + 1, \dots, n$.

Since $u \in \mathbb{R}^m$ satisfying $\|u\| < 1$ was arbitrary, it follows that for each $\lambda \in \mathbb{R}^m$ satisfying $\|\lambda - \lambda^*\| < \varepsilon$, there is $x \in \mathbb{R}^n$ such that $Rx = \lambda$, $Ax \leq \mathbf{1}$, and $x \in \mathbb{R}_+^n$. Hence, the desired property (4.4) holds.

Conversely, suppose that (4.4) holds. Since $\Gamma = \{Ry: Ky \geq 0, y \in \mathbb{R}^n\}$ is a cone, to prove that Assumption 4.1 holds, it suffices to show that Γ contains a nonempty open neighborhood of the origin. By (4.4), for each $u \in \mathbb{R}^m$ satisfying $\|u\| < \varepsilon$, there is $y \in \mathbb{R}^n$ such that $R(x^* + y) = \lambda^* - u$, $A(x^* + y) \leq \mathbf{1}$, and $x^* + y \in \mathbb{R}_+^n$. Then, using the properties of x^* we see that $Ry = -u$ and $Ky \leq 0$ or equivalently, $R(-y) = u$ and $K(-y) \geq 0$. It follows that $u \in \Gamma$ as desired. \square

Appendix B. A Minimalist Proof of Theorem 7.1. Given all the definitions and assumptions enunciated in §2 through §5, we define the sets \mathcal{D}^* and \mathcal{Y}^* as in (7.1) and (7.8) respectively. From Assumption 3.1 and the complementary slackness theorem of linear programming (Theorem 4.5 in Bertsimas and Tsitsiklis [1]), we have the following: a feasible solution (y, z) for the dual problem (3.3) is an optimal solution if and only if

$$(y'R + z'A)_j = v_j \quad \text{for } j = 1, \dots, b. \quad (\text{B1})$$

That is, associating nonnegative slack variables s_1, \dots, s_n with the inequality constraints of the dual problem (3.3), a feasible dual solution (y, z, s) is one that satisfies

$$y'R + z'A - s' = v', \quad z \geq 0 \text{ and } s \geq 0, \quad (\text{B2})$$

and a feasible dual solution is optimal if and only if $s_1 = \dots = s_b = 0$. Thus, associating with each feasible dual solution (y, z, s) the nonnegative p -vector

$$\xi = (z_1, \dots, z_\ell, s_{b+1}, \dots, s_n)',$$

we see from the definition (4.2) of K that

$$\mathcal{Y}^* = \{y \in \mathbb{R}^m: y'R + \xi'K = v', \xi \in \mathbb{R}_+^p\}. \quad (\text{B3})$$

Furthermore, from the results concerning *strict* complementary slackness that are developed in Exercise 4.20 of Bertsimas and Tsitsiklis [1], we have the following: there exists an optimal solution $(\hat{y}, \hat{\xi})$ for the dual problem (3.3) such that

$$\hat{y}'R + \hat{\xi}'K = v' \quad \text{and} \quad \hat{\xi} \in \mathbb{R}_{++}^p. \quad (\text{B4})$$

THEOREM B.1. *A vector $\delta \in \mathbb{R}^m$ is a reversible displacement if and only if $y \cdot \delta = \hat{y} \cdot \delta$ for all $y \in \mathcal{Y}^*$.*

REMARK B.1. It follows from Remark 7.3 that Theorem B.1 and Theorem 7.1 are equivalent. (Because $\hat{y} \in \mathcal{Y}^*$, the constant appearing in the statement of Theorem 7.1 must be $c = \hat{y} \cdot \delta$.)

REMARK B.2. Recalling that $\mathcal{M} = \mathcal{R}^\perp$ by definition, one may restate Theorem B.1 as follows:

$$\mathcal{M} = \text{span}(\mathcal{Y}^* - \hat{y}). \quad (\text{B5})$$

PROOF. We shall prove the alternative form (B5) of the theorem. First let $y \in \mathcal{Y}^*$ be arbitrary, and let $\xi \in \mathbb{R}_+^p$ be such that y and ξ jointly satisfy the equality in (B3). From (B3) and (B4) we then have that

$$(y - \hat{y})'R = (\hat{\xi} - \xi)'K,$$

implying that $(y - \hat{y}) \in \mathcal{M}$ by (5.3). Thus $(\mathcal{Y}^* - \hat{y}) \subset \mathcal{M}$, since $y \in \mathcal{Y}^*$ was arbitrary, and hence $\text{span}(\mathcal{Y}^* - \hat{y}) \subset \mathcal{M}$ as well.

To prove the reverse inclusion, let $\gamma \in \mathcal{M}$ be arbitrary, and let $\eta \in \mathbb{R}^p$ be such that $\gamma'R = \eta'K$, in accordance with (5.3). Because every component of $\hat{\xi}$ is strictly positive, we can choose $\varepsilon > 0$ small enough to ensure that $\xi := \hat{\xi} + \varepsilon\eta \geq 0$. Defining $y = \hat{y} - \varepsilon\gamma$, we then have

$$y'R + \xi'K = (\hat{y}'R + \hat{\xi}'K) - \varepsilon(\gamma'R - \eta'K). \quad (\text{B6})$$

The first term of the right side of (B6) equals v' by (B4), and the second term is zero. Thus y and ξ jointly satisfy the two conditions in (B3), implying that $y \in \mathcal{Y}^*$. Because $-\varepsilon\gamma = y - \hat{y}$, this means that $-\varepsilon\gamma \in (\mathcal{Y}^* - \hat{y})$ and hence $\gamma \in \text{span}(\mathcal{Y}^* - \hat{y})$. \square

Acknowledgment. The research of R. J. Williams was supported in part by NSF Grants DMS-0305272 and DMS-0604537.

References

- [1] Bertsimas, D., J. N. Tsitsiklis. 1997. *Introduction to Linear Optimization*. Athena Scientific, Belmont, MA.
- [2] Bramson, M., R. J. Williams. 2003. Two workload properties for Brownian networks. *Queueing Systems* **45** 191–221.
- [3] Harrison, J. M. 1988. Brownian models of queueing networks with heterogeneous customer populations. W. Fleming, P. L. Lions, eds. *Stochastic Differential Systems, Stochastic Control Theory and Their Applications, IMA*, Vol. 10. Springer Verlag, New York, 147–186.
- [4] Harrison, J. M. 2000. Brownian models of open processing networks: Canonical representation of workload. *Ann. Appl. Probab.* **10** 75–103. Correction: **13** 2003 390–393.
- [5] Harrison, J. M. 2002. Stochastic networks and activity analysis. Yu. Suhov, ed. *Analytic Methods in Applied Probability, in Memory of Fridrih Karpelevich*. American Mathematical Society, Providence, RI, 53–76.
- [6] Harrison, J. M. 2003. A broader view of Brownian networks. *Ann. Appl. Probab.* **13** 1119–1150.
- [7] Harrison, J. M., J. A. Van Mieghem. 1997. Dynamic control of Brownian networks: State space collapse and equivalent workload formulations. *Ann. Appl. Probab.* **7** 747–771.
- [8] Harrison, J. M., R. J. Williams. 2005. Workload reduction of a generalized Brownian network. *Ann. Appl. Probab.* **15** 2255–2295.
- [9] Kelly, F. P., N. Laws. 1993. Dynamic routing in open queueing networks: Brownian models, cut constraints and resource pooling. *Queueing Systems* **13** 47–86.
- [10] Solberg, J. J. 1981. Capacity planning with a stochastic workflow model. *AIEE Trans.* **13** 116–122.