Math Meth Oper Res (2003) 57:21–39

Mathematical Methods
of Operations Research

Whittle's index policy for a multi-class queueing system with convex holding costs

P. S. Ansell*, K. D. Glazebrook†, J. Niño-Mora and M. O'Keeffe***

* Department of Statistics, University of Newcastle upon Tyne, NE1 7RU
† School of Management, University of Edinburgh, Edinburgh, EH8 9JY
** Department of Economics and Business, Universitat Pompeu Fabra, E-08005, Barcelona

Manuscript received: January 2002/Final version received: September 2002

Abstract. Multi-class service systems are of increasing importance in the practical modelling world but present a significant challenge for analysis. Most results to date concerning the optimal dynamic control of such systems have assumed holding cost rates to be linear in the number of customers present. In response to arguments that such an assumption is often inappropriate, we develop an index heuristic for a multi-class $M/M/1$ system with increasing convex holding cost rates. We use a prescription of Whittle to develop the required indices. A numerical study elucidates very strong performance of the index policy.

Key words: Achievable region approach, dynamic programming, indexability, index policy, restless bandit

1 Introduction

Much of the literature concerning the optimal service control of multi-class queueing systems has focused on the design of policies to minimise some measure of the system's holding cost rate. A frequent assumption that such holding costs are linear in the numbers of customers from each class present in the system has been in part motivated by the relative tractability of the resulting models. In particular, simple priority policies in which the server(s) chooses from among the customers waiting for service according to a fixed ordering of the classes have shown to be optimal for linear costs in a variety of contexts. See, for example, Cobham (1954), Cox and Smith (1961), Klimov (1974), Harrison (1975), Meilijson and Weiss (1977), Gittins (1979) and Weiss (1988).

However, van Meighem (1995) argues the inappropriateness of an assumption of linear holding costs in practice. In a related contribution, Ansell et al. (1999) point to unsatisfactory features of the resulting priority policies. In

response to such concerns, this paper considers the optimal service control of a multi-class $M/M/1$ queueing system with a system cost rate which is additive across classes and increasing convex in the numbers present within each class. The existing literature concerning such nonlinear costs is sparse. Contributions have usually incorporated holding costs via delay cost functions on job flowtimes. See, for example, Weber (1988), Righter and Xu (1991) and van Meighem (1995).

We introduce the multi-class queueing system of interest and the associated stochastic optimisation problems in Section 2. A discounted costs version of the server control problem is formulated in Section 3 as a restless bandit problem with dependent arms. Restless bandits constitute a famously intractable class of decision problems which were introduced by Whittle (1988). He proposed an index-based heuristic for those problems for which a requirement of indexability was met. See also Weber and Weiss (1990, 1991). We follow Whittle's prescription for the development of an index appropriate for our multi-class queueing system. Much of Section 3 is devoted to demonstrating that an appropriate notion of indexability holds here. The form of the index for the discounted costs version of our queueing model is suggested by a simple argument and the index appropriate for the undiscounted problem of primary interest is then inferred by taking an appropriate limit. The resulting index heuristic for our queueing model will always seek to serve that class whose holding cost rate may most rapidly be diminished thereby.

An alternative account of indexability for this system is available from the work of Niño-Mora (2001a,b) on systems satisfying partial conservation laws (PCL). This is sketched in Section 4. The results of a numerical investigation into the performance of the index policy in some simple cases involving two customer classes and quadratic costs are reported in Section 5. The index policy clearly outperforms the threshold policies proposed by Ansell et al. (1999) and is very close to optimal is all cases studied.

2 The model

We consider a multi-class $M/M/1$ queue in which customers from classes $\{1, 2, \ldots, K\}$ receive service provided by a single server. Class k arrival and service rates are λ_k, μ_k and we ensure finite queue lengths by requiring that the traffic intensity (load on the system),

$$\rho = \sum_{k=1}^{K} \frac{\lambda_k}{\mu_k}$$

is less than one. The stochastic optimisation problems of interest concern the allocation of the server to waiting customers to minimise some measure of expected holding cost. The class k holding cost rate $C_k : \mathbb{N} \to \mathbb{R}^+$ is assumed to be increasing, convex and bounded above by some polynomial of finite order. Server controls u must be non-anticipative and non-idling, and the server is allocated to customers in a preemptive fashion. Hence there are no penalties imposed when the server switches between customers and all such switches are deemed instantaneous. We write $\mathscr{U}$ for the class of admissible controls.

The stochastic optimisation problem of primary interest is expressed as the determination of minimum cost and the control achieving it via the equation

$$C^{OPT} = \inf_{u \in \mathscr{U}} \tilde{E}_u \left\{ \sum_{k=1}^{K} C_k(N_k) \right\}. \tag{1}$$

In (1), N_k denotes the number of class k customers present in the system and $\tilde{E}_u$ is the expectation taken with respect to the steady-state distribution of the system state under control u. In the linear case $C_k(n) = c_k n$, $n \in \mathbb{N}$, the optimisation problem in (1) is solved by the so-called $c\mu$-rule in which the server prioritises customer classes on the basis of the index $c_k \mu_k$.

As a step towards solving (1), we consider a related stochastic optimisation with discounted costs, expressed as

$$C(\mathbf{n}, \alpha) = \inf_{u \in \mathscr{U}} E_u \left[\int_0^{\infty} \sum_{k=1}^{K} C_k\{N_k(t)\} \alpha e^{-\alpha t} \, dt \mid N(0) = \mathbf{n} \right] \tag{2}$$

where $\alpha > 0$ is a discount rate, $N_k(t)$ is the number of class k customers present in the system at time t and E_u denotes an expectation taken over all realisations of the system under control u. Lemma 1 follows from standard results in dynamic programming (DP). See, for example, Puterman (1994).

Lemma 1. *For all initial states* $\mathbf{n}$,

$$\lim_{\alpha \to 0} C(\mathbf{n}, \alpha) = C^{OPT}.$$

In light of Lemma 1, we shall develop "good" policies for the average cost problem in (1) as limits (i.e. as $\alpha \to 0$) of "good" policies for the discounted costs problem in (2). This is the primary concern of Section 3.

3 Indexability and Whittle index policies

We formulate the discounted costs problem in (2) as a Markov Decision Problem (MDP) as follows:

(a) The state of the system at time t is $\mathbf{N}(t) = \{N_1(t), N_2(t), \ldots, N_K(t)\}$, the vector of queue lengths, $t \in \mathbb{N}$. The decision epochs are the set of arrival times together with all service completion epochs which do not result in an empty system. Let action a_k denote the allocation of service to a class k customer, $1 \leq k \leq K$. At each decision epoch, the controller chooses an action a_k from the set of k for which $N_k(t) \geq 1$;

(b) In describing how the process evolves as actions are applied to it, we shall write

$$\Lambda = \sum_{k=1}^{K} (\lambda_k + \mu_k)$$

and use standard uniformization in which successive decision epochs occur at the event times of a Poisson process of rate Λ, with some such events corresponding to virtual state transitions. For example, if action a_k is taken at decision epoch t when the system state is $\mathbf{N}(t) = \mathbf{n}$ with $n_k > 0$, then the next decision epoch occurs at time $t + X$, where $X \sim \exp(\Lambda)$. The system state following any state transition then is given by

$$\mathbf{N}\{(t+X)^+\} = \begin{cases} \mathbf{n} + \mathbf{1}^j, & \text{with probability } \lambda_j/\Lambda,\ 1 \le j \le K, \\ \mathbf{n} - \mathbf{1}^k, & \text{with probability } \mu_k/\Lambda, \\ \mathbf{n}, & \text{with probability } \sum_{j \ne k} \mu_k/\Lambda. \end{cases}$$

Between t and $t + X$ the system incurs discounted costs at rate

$$\alpha \sum_{l=1}^{K} C_l(n_l)$$

where the functions C_l satisfy the requirements outlined in Section 2;

(c) A policy is a rule for choosing actions in light of the history of the process to date. The general theory of stochastic dynamic programming (DP) indicates the existence of an optimal (cost minimising) policy which is stationary (i.e. makes decisions in light of the current state only) and whose value function satisfies the DP optimality equations. See Section 6.10 of Puterman (1994).

A pure DP approach to the problem is unlikely to be insightful and (especially for large K) may be computationally intractable. Following Whittle (1988), the search for natural control heuristics which perform well centres on *index policies*. Hence we would like to produce a class k index $W_k : \mathbb{Z}^+ \to \mathbb{R}^+$, $1 \le k \le K$, such that the policy which at all epochs t chooses the non-empty queue with maximal index $W_k\{N_k(t)\}$ is close to optimal. Whittle's (1988) contribution to the celebrated restless bandit problem contains the essential prescription for how suitable indices should be developed. Despite the fact that, from 3(b), our bandit model has stochastic dependencies between the arms, we shall see that this prescription works very effectively here. Note that we have used W in the notation in honour of Whittle's contribution and shall hereafter refer to *Whittle indices*. However, before we can develop indices, there is a prior consideration of *indexability*. We shall now discuss these issues in relation to an individual class and drop the class identifier k.

3.1 Indexability

Following Whittle, we consider a MDP, modelling the evolution of a single customer class, with state space $\mathbb{N}$ and with two actions {active, passive} in each state, save only the state 0 in which only passive is available. The active action corresponds to selection of the class for service in the original multi-class problem. Current state n enters one of $\{n+1, n-1\}$ with rates λ and μ respectively under this action, $n \in \mathbb{Z}^+$. Under the passive action current state n enters state $n + 1$ at rate λ, $n \in \mathbb{N}$. This MDP is a restless bandit.

We shall assume that costs are incurred at rate $\alpha C(n)$ under the active action and at rate $\alpha C(n) - W$ under the passive action, $n \in \mathbb{N}$. We shall refer to W as a *subsidy for passivity*. Plainly if $W = 0$, the active action is always preferable in all states. If $W \gg 0$, this may not be the case. We consider the stochastic optimisation problem whose aim is to design a policy for choosing actions to minimise the total expected discounted cost incurred over an infinite horizon. If it helps, the reader may think of the passive subsidy via an alternative source of income which earns revenues at rate W and which is available to the server whenever he chooses to stop serving the queue. Further, the problem may also be equivalently reformulated via a *charge for activity*.

We use $C(n, \alpha, W)$ for the optimal cost for the restless bandit from initial state $n \in \mathbb{N}$. Function $C(.\,, \alpha, W)$ satisfies the optimality equations

$$C(n, \alpha, W) = \min\{C_1(n, \alpha, W), C_2(n, \alpha, W)\}, \quad n \in \mathbb{Z}^+, \tag{3}$$

where

$$(\alpha + \lambda + \mu)C_1(n, \alpha, W) = \alpha C(n) + \mu C(n-1, \alpha, W) + \lambda C(n+1, \alpha, W) \tag{4}$$

and

$$(\alpha + \lambda + \mu)C_2(n, \alpha, W) = \alpha C(n) - W + \mu C(n, \alpha, W) + \lambda C(n+1, \alpha, W). \tag{5}$$

We note further that $C(0, \alpha, W) = C_2(0, \alpha, W)$. If $C_1(n, \alpha, W) \leq C_2(n, \alpha, W)$ then the active action is optimal in $n \in \mathbb{Z}^+$. If the direction of the inequality is reversed then the passive action is optimal. We write

$$\begin{aligned} \Pi_\alpha(W) &= \{0\} \cup \{n \in \mathbb{Z}^+; C_2(n, \alpha, W) \leq C_1(n, \alpha, W)\} \\ &= \{n \in \mathbb{N};\ \text{the passive action is optimal in } n \\ &\qquad \text{when the subsidy for passivity is } W\}, \quad W \in \mathbb{R}^+. \end{aligned}$$

The following expresses the notion of *indexability* for an individual class developed by Whittle (1988).

Definition 1. The class is *indexable* if $\Pi_\alpha : \mathbb{R} \to 2^{\mathbb{N}}$ is increasing, namely

$$W_1 > W_2 \Rightarrow \Pi_\alpha(W_1) \supseteq \Pi_\alpha(W_2).$$

Once we have indexability, the derived notion follows of a *state* **n** *index* as the minimum subsidy for passivity which renders the passive action optimal in state n.

Definition 2. When the class is indexable the *Whittle index* for state **n** is given by

$$W_\alpha(n) = \inf\{W; n \in \Pi_\alpha(W)\}, \quad n \in \mathbb{Z}^+,$$

where we take

$$W_\alpha(0) = 0.$$

The following is trivial to establish on the basis of the above.

Lemma 2. *For all states n of an indexable class*

$W \geq W_\alpha(n) \Rightarrow$ *the passive action is optimal*;

$W < W_\alpha(n) \Rightarrow$ *the active action is optimal*.

We shall now develop the form of the index by a simple heuristic argument which we trust the reader will find insightful. The later proof that the quantity we obtain is indeed the index is entirely rigorous.

3.2 The form of the Whittle index for discounted costs

We shall suppose until further notice that the class is indeed indexable and that the Whittle index $W_\alpha : \mathbb{N} \to \mathbb{R}^+$ is increasing. The latter seems plausible from the assumption of increasing convex cost rates. Consider now the single class problem outlined in (3.1), with initial state n and with passive subsidy taken to be $W = \overline{W}_\alpha(n)$, the latter being the *assumed* value of the index. From the above, we can conclude the following concerning an optimal policy:

(i) the active action will be optimal for states $\{n+1, n+2, \ldots\}$;
(ii) the passive action will be optimal for states $\{0, \ldots, n-1\}$;
(iii) both active and passive actions will be optimal for state n.

Consider first the policy for the restless bandit determined by the choice in (iii) of the active action in state n along with (i) and (ii). Under this policy, the active action will continue to be applied from time zero until the process enters state $n-1$ for the first time at time T, say. Random variable T is stochastically identical to the busy period of an $M/M/1$ queue, starting with a single customer and having arrival rate λ and service rate μ. Having arrived in state $n-1$, from (ii) above the passive action will then be applied until the system returns to state n. This period of passivity has a duration which is exponentially distributed with rate λ. Once the system has returned to state n, the above cycle is repeated *ad infinitum*. The total expected discounted cost incurred over an infinite horizon by this policy may be readily computed as

$$[\overline{C}(n,\alpha) + E(e^{-\alpha T})\{\alpha C(n-1) - \overline{W}_\alpha(n)\}(\alpha+\lambda)^{-1}]\{1 - \lambda E(e^{-\alpha T})(\alpha+\lambda)^{-1}\}^{-1}, \tag{6}$$

where

$$\overline{C}(n,\alpha) = E\left[\int_0^T C\{N(t)\}\alpha e^{-\alpha t}\, dt \,|\, N(0) = n,\ \text{active}\right] \tag{7}$$

is the cost associated with the initial busy period. In future we shall drop "active" from the notation in the conditional expectation on the r.h.s. of (7).

Consider secondly the choice in (iii) of the passive action in n along with (i) and (ii). Under this action, the process will first enter state $n+1$ after a period of time during which the passive action was taken and which has an $\exp(\lambda)$ distribution. Having entered $n+1$, the active action is then applied according to (i) above until the process returns to n. This latter stage will have a duration which is stochastically identical to the busy period random variable T above. Once the process has returned to n, the above cycle is repeated *ad infinitum*. The total expected discounted cost earned over an infinite horizon may be readily computed as

$$[\{\alpha C(n) - \overline{W}_\alpha(n)\}(\alpha+\lambda)^{-1} + \lambda \bar{C}(n+1,\alpha)(\alpha+\lambda)^{-1}] \times \{1 - \lambda E(e^{-\alpha T})(\alpha+\lambda)^{-1}\}^{-1}. \tag{8}$$

However, the expressions in (8) and (9) should both represent the optimal cost associated with the restless bandit and hence should be equal. Equating (8) and (9) we obtain

$$\overline{W}_\alpha(n) = \{\lambda \bar{C}(n+1,\alpha) - (\alpha+\lambda)\bar{C}(n,\alpha) + \alpha C(n) - \alpha E(e^{-\alpha T})C(n-1)\}\{1 - E(e^{-\alpha T})\}^{-1}, \quad n \in \mathbb{Z}^+. \tag{9}$$

We can simplify the expression on the r.h.s. of (9) by use of two identities, both of which can be obtained from straightforward conditioning arguments. These are

$$\{\alpha + \lambda + \mu - \lambda E(e^{-\alpha T})\}\bar{C}(n,\alpha) = \alpha C(n) + \lambda \bar{C}(n+1,\alpha), \quad n \in \mathbb{Z}^+, \tag{10}$$

and

$$\lambda\{E(e^{-\alpha T})\}^2 - (\alpha + \lambda + \mu)E(e^{-\alpha T}) + \mu = 0. \tag{11}$$

Utilisation of (10) and (11) in (9) yields

$$\overline{W}_\alpha(n) = E(e^{-\alpha T})[\alpha \bar{C}(n,\alpha)\{1 - E(e^{-\alpha T})\}^{-1} - \alpha C(n-1)] \times \{1 - E(e^{-\alpha T})\}^{-1}, \quad n \in \mathbb{Z}^+. \tag{12}$$

The expression on the r.h.s of (12) has a natural interpretation as the (discounted) rate at which the holding cost rate is reduced by serving the class in state n. In Lemma 3, we take $\overline{W}_\alpha(0) = 0$.

Lemma 3. *$\overline{W}_\alpha(n)$ is increasing in n.*

Proof. From (7) we deduce that

$$\bar{C}(n,\alpha)\{1-E(e^{-\alpha T})\}^{-1} = \frac{E[\int_0^T C\{N(t)\}\alpha e^{-\alpha t}\,dt \mid N(0)=n]}{E[\int_0^T \alpha e^{-\alpha t}\,dt]}$$

$$= \sum_{m=0}^{\infty} C(n+m)x_m \tag{13}$$

where the set $\{x_m; m \geq 0\}$ form a probability mass function on $\mathbb{N}$. For an $M/M/1$ queue with arrival/service rates λ and μ respectively, having a single customer present at time 0 and with T as the duration of the first busy period, we have

$$x_m = \frac{E\{\int_0^T I_m(s)\alpha e^{-\alpha s}\,ds\}}{E\{\int_0^T \alpha e^{-\alpha t}\,dt\}}$$

where

$$I_m(s) = \begin{cases} 1, & \text{if } m \text{ customers are present at time } s, \\ 0, & \text{otherwise, } s \in \mathbb{R}^+,\ m \in \mathbb{N}. \end{cases}$$

From (13) it follows that

$$\{\bar{C}(n+1,\alpha) - \bar{C}(n,\alpha)\}\{1-E(e^{-\alpha T})\}^{-1}$$

$$= \sum_{m=0}^{\infty}\{C(n+1+m) - C(n+m)\}x_m$$

$$\geq C(n+1) - C(n) \geq C(n) - C(n-1), \quad n \in \mathbb{Z}^+, \tag{14}$$

by the increasing convex nature of C. It is an immediate consequence of (12) and (14) that

$$\overline{W}_\alpha(n+1) \geq \overline{W}_\alpha(n), \quad n \in \mathbb{Z}^+,$$

as required. The above analysis also yields $\overline{W}_\alpha(1) \geq 0 = \overline{W}_\alpha(0)$. This completes the proof.

We can now characterise optimal policies for the single class problem with passive subsidy W introduced in Section 3.1.

Lemma 4. *If $\overline{W}_\alpha(m) \leq W < \overline{W}_\alpha(m+1)$ then the policy for the restless bandit which chooses the passive action in states $\{0,1,\ldots,m\}$ and the active action otherwise is optimal, $m \in \mathbb{N}$.*

Proof. Fix $W \in [\overline{W}_\alpha(m), \overline{W}_\alpha(m+1))$ and use $\hat{C}(.\,,\alpha,W)$ for the value function of the policy described in the statement of the Lemma. It is enough to show that $\hat{C}(.\,,\alpha,W)$ satisfies the optimality equations in (3). From (3)–(5), it follows simply that we need to show that

$$\mu\{\hat{C}(n,\alpha,W)-\hat{C}(n-1,\alpha,W)\}\geq W,\quad n\geq m+1, \tag{15}$$

and

$$\mu\{\hat{C}(n,\alpha,W)-\hat{C}(n-1,\alpha,W)\}\leq W,\quad n\leq m. \tag{16}$$

We prove (15) and (16) by considering four separate cases. Note that in the case $n=0$ only (15) is required.

Case 1: $\mu\{\hat{C}(m+1,\alpha,W)-\hat{C}(m,\alpha,W)\}\geq W$.

Following the calculation to (6), we have that

$$\hat{C}(m+1,\alpha,W)=[\bar{C}(m+1,\alpha)+E(e^{-\alpha T})\{\alpha C(m)-W\}(\alpha+\lambda)^{-1}]\times\{1-\lambda E(e^{-\alpha T})(\alpha+\lambda)^{-1}\}^{-1}, \tag{17}$$

and following that to (7), we deduce that

$$\hat{C}(m,\alpha,W)=[\{\alpha C(m)-W\}(\alpha+\lambda)^{-1}+\lambda\bar{C}(m+1,\alpha)(\alpha+\lambda)^{-1}]\times\{1-\lambda E(e^{-\alpha T})(\alpha+\lambda)^{-1}\}^{-1}. \tag{18}$$

From (17) and (18), we deduce that

$$\begin{aligned}
&\mu\{\hat{C}(m+1,\alpha,W)-\hat{C}(m,\alpha,W)\}\geq W\\
&\quad\Leftrightarrow \alpha\bar{C}(m+1,\alpha)-\alpha\{1-E(e^{-\alpha T})\}C(m)\\
&\quad\geq \frac{W}{\mu}\{\alpha+\lambda-\mu+(\mu+\lambda)E(e^{-\alpha T})\}\\
&\quad\Leftrightarrow \alpha\bar{C}(m+1,\alpha)-\alpha\{1-E(e^{-\alpha T})\}C(m)\\
&\quad\geq W\{1-E(e^{-\alpha T})\}^{2}\{E(e^{-\alpha T})\}^{-1}\\
&\quad\Leftrightarrow E(e^{-\alpha T})[\alpha\bar{C}(m+1,\alpha)\{1-E(e^{-\alpha T})\}^{-1}-\alpha C(m)]\\
&\qquad\times\{1-E(e^{-\alpha T})\}^{-1}\geq W \qquad (19)\\
&\quad\Leftrightarrow \overline{W}_{\alpha}(m+1)\geq W, \qquad (20)
\end{aligned}$$

which it is by supposition. Note that inequality (19) makes use of identity (11) and also that the l.h.s of (19) is $\overline{W}_{\alpha}(m+1)$ by (12). This concludes Case 1.

Case 2: $\mu\{\hat{C}(n,\alpha,W)-\hat{C}(n-1,\alpha,W)\}\geq W$, $n\geq m+1$

We prove Case 2 by induction, observing that $n=m+1$ corresponds to Case 1. We suppose that the required inequality holds for $m+1\leq n\leq k$ and

deduce it for $n = k + 1$. This will be enough. By the structure of the policy of interest, the active action is chosen for states $n \geq k$ and hence we have

$$\hat{C}(k+1, \alpha, W) = \bar{C}(k+1, \alpha) + E(e^{-\alpha T})\hat{C}(k, \alpha, W), \tag{21}$$

and

$$\hat{C}(k, \alpha, W) = \bar{C}(k, \alpha) + E(e^{-\alpha T})\hat{C}(k-1, \alpha, W).$$

Utilising the inductive hypothesis for $n = k$, we observe that it will be enough to show that

$$\mu\{\bar{C}(k+1, \alpha) - \bar{C}(k, \alpha)\} \geq W\{1 - E(e^{-\alpha T})\}.$$

However, from (10) and (11) we readily infer that

$$\begin{aligned}
&\mu\{\bar{C}(k+1, \alpha) - \bar{C}(k, \alpha)\} \\
&\quad = E(e^{-\alpha T})[\alpha \bar{C}(k+1, \alpha)\{1 - E(e^{-\alpha T})\}^{-1} - \alpha C(k)] \\
&\quad = \overline{W}_{\alpha}(k+1)\{1 - E(e^{-\alpha T})\} \qquad (22) \\
&\quad \geq W\{1 - E(e^{-\alpha T})\}, \qquad (23)
\end{aligned}$$

since by Lemma 3 and by the hypothesis of Lemma 4,

$$\overline{W}_{\alpha}(k+1) \geq \overline{W}_{\alpha}(m+1) > W.$$

As indicated above, inequality (23) is sufficient to demonstrate the inductive step and Case 2 is proved.

Case 3: $\mu\{\hat{C}(m, \alpha, W) - \hat{C}(m-1, \alpha, W)\} \leq W$

Since the passive action is deployed at states $n = m - 1$, m by the policy under study we have that

$$\hat{C}(m-1, \alpha, W) = \{\alpha C(m-1) - W\}(\alpha + \lambda)^{-1} + \lambda \hat{C}(m, \alpha, W)(\alpha + \lambda)^{-1} \tag{24}$$

with $\hat{C}(m, \alpha, W)$ given by (18). From (11), (18) and (24) and straightforward algebra we conclude that

$$\begin{aligned}
&\mu\{\hat{C}(m, \alpha, W) - \hat{C}(m-1, \alpha, W)\} \leq W \\
&\quad \Leftrightarrow \alpha\{\lambda \bar{C}(m+1, \alpha) + \alpha C(m)\}[\alpha + \lambda\{1 - E(e^{-\alpha T})\}]^{-1} - \alpha C(m-1) \\
&\quad \leq W\{1 - E(e^{-\alpha T})\}\{E(e^{-\alpha T})\}^{-1}. \qquad (25)
\end{aligned}$$

However, making use of (11) we note that

$$\overline{W}_\alpha(m) \le W \Leftrightarrow E(e^{-\alpha T})[\alpha\overline{C}(m,\alpha)\{1-E(e^{-\alpha T})\}^{-1} - \alpha C(m-1)]$$
$$\times \{1-E(e^{-\alpha T})\}^{-1} \le W$$
$$\Leftrightarrow \alpha[\alpha+\mu+\lambda\{1-E(e^{-\alpha T})\}]\overline{C}(m,\alpha)[\alpha+\lambda\{1-E(e^{-\alpha T})\}]^{-1}$$
$$-\alpha C(m-1) \le W\{1-E(e^{-\alpha T})\}\{E(e^{-\alpha T})\}^{-1}. \quad (26)$$

We can now utilise (10) for the case $n = m$ to infer that inequalities (25) and (26) are equivalent. This establishes Case 3.

Case 4: $\mu\{\hat{C}(n,\alpha,W) - \hat{C}(n-1,\alpha,W)\} \le W$, $n \le m$

We prove Case 4 by induction, observing that $n = m$ corresponds to Case 3. We suppose that the required inequality holds for $k+1 \le n \le m$ and deduce it for $n = k$. Since under the policy of interest, the passive action is chosen for states $n \le k$ we have that

$$\hat{C}(n,\alpha,W) = \{\alpha C(n) - W\}(\alpha+\lambda)^{-1} + \lambda\hat{C}(n+1,\alpha,W)(\alpha+\lambda)^{-1}, \quad n \le k. \quad (27)$$

Utilising the inductive hypothesis for $n = k+1$, in order to establish the required inequality for $n = k$ it is enough to show that

$$\mu\{C(k) - C(k-1)\} \le W. \quad (28)$$

Utilising the calculation to (14), we have that

$$\mu\{C(k) - C(k-1)\} \le \mu\{\overline{C}(k,\alpha) - \overline{C}(k-1,\alpha)\}\{1-E(e^{-\alpha T})\}^{-1}$$
$$= \overline{W}_\alpha(k) \quad (29)$$
$$\le \overline{W}_\alpha(m) \le W, \quad (30)$$

as required. Note that equality (29) utilises the calculation to (22), while (30) follows from Lemma 3. This establishes the inductive step and hence Case 4.

We have now established (15) and (16) and hence have proved Lemma 4.

Theorem 1 is the main result of this Section.

Theorem 1 (Indexability and the Whittle index for discounted costs). *The restless bandit is indexable with Whittle index* $W_\alpha(n) = \overline{W}_\alpha(n)$, $n \in \mathbb{N}$.

Proof. By Lemma 4 we have that

$$\Pi_\alpha(W) = \{0,1,\ldots,n\}, \quad \overline{W}_\alpha(n) \le W < \overline{W}_\alpha(n+1), \quad n \in \mathbb{N}. \quad (31)$$

Indexability follows from Lemma 4. That $\overline{W}_\alpha(n)$ is the Whittle index in state n follows from (31) and Definition 2.

3.3 The form of the Whittle index for the average cost problem

Motivated by Lemma 1, we recover a Whittle index $W : \mathbb{N} \to \mathbb{R}^+$ for the average cost problem from the limit

$$W(n) = \lim_{\alpha \to 0} W_\alpha(n)$$

$$= \lim_{\alpha \to 0} \overline{W}_\alpha(n), \quad n \in \mathbb{N}, \tag{32}$$

by Theorem 1. Utilising (12) within (32), we obtain the following result.

Theorem 2 (The Whittle index for average costs). *The Whittle index for the average cost problem is given by* $W(0) = 0$ *and*

$$W(n) = [\overline{C}(n)\{E(T)\}^{-1} - C(n-1)]\{E(T)\}^{-1}, \quad n \in \mathbb{Z}^+, \tag{33}$$

$$= \frac{\mu(\mu - \lambda)}{\lambda}[E\{C(n-1+N)\} - C(n-1)], \quad n \in \mathbb{Z}^+, \tag{34}$$

where in (33) we have

$$\overline{C}(n) = E\left[\int_0^T C\{N(t)\}\, dt \,|\, N(0) = n\right], \quad n \in \mathbb{Z}^+, \tag{35}$$

and in (34), N is random variable with probability mass function

$$P(N = n) = \rho^n(1 - \rho), \quad n \in \mathbb{N}, \tag{36}$$

where $\rho = \lambda/\mu$.

Proof. The expression in (33) is a straightforward consequence of (12) and (32). For (34), observe that from standard renewal theory arguments, the average cost incurred under the policy which chooses the passive action in states $\{0, 1, \ldots, n-1\}$ and active otherwise when $C(m)$ is the cost rate in state $m \in \mathbb{N}$ is given by

$$\{\overline{C}(n) + C(n-1)\lambda^{-1}\}\{E(T) + \lambda^{-1}\}^{-1} = E\{C(n-1+N)\}, \tag{37}$$

where N is a random variable with the steady-state distribution for the number of customers present in an $M/M/1$ system with arrival rate λ and service rate μ, as in (36). The expression in (34) follows from (33), (37) and the fact that $E(T) = (\mu - \lambda)^{-1}$. That $W(0) = 0$ is immediate from (32).

Comments

1. An interpretation of the index $W(n)$ as the rate at which the holding cost rate is reduced by serving the class in state n is clear from the expression on the r.h.s. of (33).

2. The expression in (34) allows for ready computation of $W(n)$. If we have $C(n) = a + bn + cn^2$ with a, b and c all non-negative then it is straightforward to show that

$$W(n) = b\mu + \frac{c(3\lambda - \mu)\mu}{(\mu - \lambda)} + 2c\mu n, \quad n \in \mathbb{Z}^+,$$

which becomes $b\mu$ when $c = 0$. Hence the Whittle index policy, which always allocates service to whichever class from amongst those having customers present in the system has highest Whittle index, is optimal when cost rates are linear. Note also that if cost rate $C(n)$ is polynomial in n of order p then, from (34), the corresponding index $W(n)$ is polynomial in n of order $p - 1$.
3. We shall explore the performance of the Whittle index policy in two class problems with quadratic costs in Section 5.

4 PCL – indexability

Niño-Mora (2001a) maps out an alternative route to the demonstration of (Whittle)-indexability for restless bandits and to index calculation which utilises the stronger notion of PCL (partial conservation laws) – indexability. This in turn is a development of ideas based on GCL (generalised conservation laws) which played a fundamental role in the account of Gittins indexation given by Bertsimas and Niño-Mora (1996). In brief, let us suppose that we wish to schedule a stochastic system which is servicing a countably infinite collection of job classes indexed by the natural numbers $\mathbb{N}$. Denote by $\mathscr{U}$ the collection of admissible scheduling policies. The stochastic optimisation problem of interest is the minimisation of some linear objective

$$\sum_{i \in \mathbb{N}} c_i x_i^u \tag{38}$$

where $c_i > 0$ is a cost rate for job class i and x_i^u a performance measure for class i under scheduling policy u. When the system satisfies a collection of so-called partial work conservation laws (PCL) then the stochastic optimisation problem in (38) is solved by an index policy for *some* choices of the cost rate vector $\mathbf{c}$. Whether a particular choice is in this admissible class or not may be determined by running an adaptive greedy algorithm. A system which satisfies PCL and whose cost-rate vector $\mathbf{c}$ is in the admissible class is called PCL-indexable.

Niño-Mora (2001b) utilises the above ideas to develop sufficient conditions for the (Whittle)-indexability of countable state restless bandits in terms of model parameters. He further demonstrates that the restless bandit model associated with our multi-class $M/M/1$ system described in Section 3 does indeed satisfy these sufficient conditions and hence meets the requirements for PCL-indexability. A closed form expression for the discounted index in (12) in terms of model parameters emerges from the PCL approach via a (suitably modified) version of the adaptive greedy algorithm above. The average cost index in (32) is again obtained by considering the limit $\alpha \to 0$. This analysis is complex and the details may be found in Niño-Mora (2001b). PCL-

indexability is an important analytical tool which is sometimes available when the simple direct arguments of Section 3 are not.

5 Assessing the Whittle index policy in a simple case

We shall assess the performance of the Whittle index policy for a two class system with quadratic costs. We take $C_k(n) = b_k n + c_k n^2$, $k = 1, 2$ and, following (1), seek an admissible control to minimise C^u, i.e.,

$$C^{OPT} = \inf_{u \in \mathscr{U}} C^u,$$

where

$$C^u = \tilde{E}_u\{b_1 N_1 + b_2 N_2 + c_1 N_1^2 + c_2 N_2^2\}. \tag{39}$$

We shall consider a range of policies for this problem, including the index policy developed in Section 3. These will be compared with each other, and their associated costs assessed against exact values of and lower bounds on the minimised achievable cost C^{OPT}.

5.1 Classes of heuristic policies

Threshold policies

This class of policies was studied by Ansell et al. (1999) in the context of a queueing control model which sought to minimise a linear cost objective over admissible controls which satisfied imposed constraints on the second moments of queue lengths. For a two-class system, threshold policies are specified by a pair (k, T) where $k \in \{1, 2\}$ and $T \in \mathbb{Z}^+$. The policy (k, T) gives priority to class k, unless class not-k has a queue length which exceeds threshold T. For example, $(1, T)$ works as follows:

(1) If $N_1(t) > 0$ and $N_2(T) < T$ then a class 1 customer is served;
(2) If $N_2(t) \geq T$, or if $0 < N_2(t) < T$ and $N_1(t) = 0$ then a class 2 customer is served.

Ansell, Glazebrook and Mitrani (2001) developed approaches to performance analysis for this class of policies. In particular, they demonstrate how to calculate the joint steady-state distribution of the resulting queue lengths (N_1, N_2). The policies demonstrate strong performance when compared with heuristics which randomise between the pure priority policies which favour class 1 over class 2 and class 2 over class 1 respectively.

Linear switching policies

The linear form of the Whittle index for the problem in (38) identified in Comment 2 above suggests a development of the above threshold policies to those which choose between classes 1 and 2 on the basis of a *linear switching curve* $n_2 = \alpha n_1 + \beta$ as follows:

(3) If $N_1(t) > 0$ and $0 \le N_2(t) \le \max\{0, \alpha N_1(t) + \beta\}$ then a class 1 job is served;
(4) If $N_2(t) > \alpha N_1(t) + \beta$, or if $N_2(t) > 0$ and $N_1(t) = 0$ then a class 2 customer is served.

The joint steady state distribution

$$p_{i,j} = \lim_{t\to\infty} P\{N_1(t) = i, N_2(t) = j\} = P(N_1 = i, N_2 = j), \quad (i, j) \in \mathbb{N}^2,$$

under the policy described in (3) and (4) satisfies the balance equations

$$\begin{aligned}
&[\lambda_1 + \lambda_2 + \mu_1 I\{(i > 0, 0 < j \le \alpha i + \beta) \cup (i > 0, j = 0)\} \\
&\qquad + \mu_2 I\{(i > 0, j > 0, j > \alpha i + \beta) \cup (i = 0, j > 0)\}]\, p_{i,j} \\
&\quad = \lambda_1 p_{i-1,j} + \lambda_2 p_{i,j-1} + \mu_1 I\{0 \le j \le \alpha(i+1) + \beta\} p_{i+1,j} \\
&\qquad + \mu_2 I\{(i = 0) \cup (j + 1 > \alpha i + \beta)\} p_{i,j+1}, \quad (i, j) \in \mathbb{N}^2, \qquad (40)
\end{aligned}$$

where $p_{-1,j} = p_{i,-1} = 0$ and I is the indicator function.

In the numerical study which follows, solutions of (40) are obtained by application of the Power Series Algorithm; see Blanc (1993). In outline, this method takes the recursively unsolvable balance equations (40) and transforms them into a recursively solvable set of equations via the introduction of an additional parameter. Once the joint steady-state distribution has been computed then the cost in (39) corresponding to any linear switching policy may be inferred. In the numerical study reported in Section 5.2 this cost is then minimised over choices of α and β to obtain an optimal cost in the linear switching class.

Whittle index policies

From Comment 2 at the conclusion of Section 3, a Whittle index policy will choose between the two customer classes when both have members present in the system on the basis of the class indices

$$W_k(n) = b_k \mu_k + \frac{c_k(3\lambda_k - \mu_k)}{(\mu_k - \lambda_k)} + 2c_k\mu_k n, \quad n \in \mathbb{Z}^+, k = 1, 2. \qquad (41)$$

If $W_1\{N_1(t)\} \ge W_2\{N_2(t)\}$, then class 1 is served at t provided that $N_1(t) > 0$. Otherwise class 2 is served if $N_2(t) > 0$. Plainly, such a policy is a member of the class based on linear switching curves above. The corresponding switching curve is easily shown to be $n_2 = \alpha n_1 + \beta$ where

$$\alpha = \frac{c_1\mu_1}{c_2\mu_2}$$

and

$$\beta = \frac{(\beta_1 - \beta_2)}{2c_2\mu_2},$$

with

$$\beta_k = (b_k\mu_k - 2c_k\mu_k) + \frac{c_k\mu_k(\lambda_k + \mu_k)}{(\mu_k - \lambda_k)}, \quad k = 1, 2.$$

5.2 Numerical study

In the numerical investigation based on the problem in (39), we sought to compare (i) the minimum cost achievable by any threshold policy with (ii) that achievable by any linear switching policy with (iii) that achieved by the Whittle index policy. Note that romans (i)–(v) refer to columns in Tables 1 and 2 below. Since both the index policy and all threshold policies are members of the linear switching class, the best achievable from the latter must bound the others below. That fact notwithstanding, we shall see that the cost performance of the index policy is remarkably strong. The above costs from designated policy classes are further compared with (iv) C^{OPT}, the minimum cost achievable by any policy from the admissible class and (v) a lower bound on this minimum cost obtained by deployment of the achievable region approach to stochastic optimisation. See Dacre, Glazebrook and Niño-Mora (1999).

The costs in (i)–(iii) for designated policies were obtained by application of the approach to performance analysis outlined in (40) and following. The minimal cost C^{OPT} in (iv) was obtained by the DP method of value iteration and is based on the recursion

$$\begin{aligned} C_{t+1}(n_1, n_2) = \min\Bigg\{ & \frac{b_1 n_1 + b_2 n_2 + c_1 n_1^2 + c_2 n_2^2}{\lambda_1 + \lambda_2 + \mu_1} \\ & + \frac{\lambda_1 C_t(n_1 + 1, n_2) + \lambda_2 C_t(n_1, n_2 + 1) + \mu_1 C_t(n_1 - 1, n_2)}{\lambda_1 + \lambda_2 + \mu_1}; \\ & \frac{b_1 n_1 + b_2 n_2 + c_1 n_1^2 + c_2 n_2^2}{\lambda_1 + \lambda_2 + \mu_2} \\ & + \frac{\lambda_1 C_t(n_1 + 1, n_1) + \lambda_2 C_t(n_1, n_2 + 1) + \mu_2 C_t(n_1, n_2 - 1)}{\lambda_1 + \lambda_2 + \mu_2} \Bigg\}, \\ & (n_1, n_2) \in (\mathbb{Z}^+)^2, \end{aligned}$$

$$\begin{aligned} (\lambda_1 + \lambda_2 + \mu_1) C_{t+1}(n_1, 0) = {} & b_1 n_1 + c_1 n_1^2 + \lambda_1 C_t(n_1 + 1, 0) + \lambda_2 C_t(n_1, 1) \\ & + \mu_1 C_t(n_1 - 1, 0), \quad n_1 \in \mathbb{Z}^+, \end{aligned}$$

$$\begin{aligned} (\lambda_1 + \lambda_2 + \mu_2) C_{t+1}(0, n_2) = {} & b_2 n_2 + c_2 n_2^2 + \lambda_1 C_t(1, n_2) + \lambda_2 C_t(0, n_2 + 1) \\ & + \mu_2 C_t(0, n_2 - 1), \quad n_2 \in \mathbb{Z}^+, \end{aligned}$$

Table 1. Values of the costs associated with (i) the best threshold policy, (ii) the best linear switching policy, and (iii) the index policy, together with (iv) C^{OPT} and (v) a semidefinite lower bound on C^{OPT} when $b_1 = 5$, $b_2 = 1$

c_1	c_2	(i)	(ii)	(iii)	(iv)	(v)
0.1	0.1	9.344	9.334	9.335	9.333	9.305
0.1	0.2	9.581	9.575	9.575	9.574	9.566
0.1	0.5	10.101	10.101	10.101	10.101	10.095
0.1	1.0	10.969	10.969	10.969	10.969	10.964
0.1	2.0	12.703	12.703	12.703	12.703	12.700
0.2	0.1	9.926	9.882	9.885	9.878	9.858
0.2	0.2	10.244	10.199	10.199	10.199	10.184
0.2	0.5	10.764	10.763	10.763	10.763	10.740
0.2	1.0	11.631	11.631	11.631	11.631	11.619
0.2	2.0	13.366	13.366	13.366	13.366	13.358
0.5	0.1	11.476	11.275	11.276	11.272	11.242
0.5	0.2	12.053	11.906	11.917	11.902	11.866
0.5	0.5	12.752	12.700	12.701	12.699	12.604
0.5	1.0	13.620	13.615	13.615	13.615	13.544
0.5	2.0	15.354	15.354	15.354	15.354	15.316
1.0	0.1	13.513	13.016	13.026	13.004	12.898
1.0	0.2	14.742	14.304	14.307	14.258	14.205
1.0	0.5	15.962	15.713	15.725	15.706	15.494
1.0	1.0	16.933	16.848	16.848	16.848	16.638
1.0	2.0	18.668	18.660	18.660	18.660	18.517
2.0	0.1	16.473	15.404	15.427	15.348	15.089
2.0	0.2	19.074	17.990	17.990	17.969	17.778
2.0	0.5	21.799	20.992	21.096	20.962	20.660
2.0	1.0	23.482	22.917	22.917	22.915	22.418
2.0	2.0	25.294	25.146	25.146	25.146	24.703

$$(\lambda_1 + \lambda_2) C_{t+1}(0,0) = \lambda_1 C_t(1,0) + \lambda_2 C_t(0,1), \quad t \in \mathbb{N}.$$

Full details of this numerical scheme may be found in Tijms (1994).

A minimised cost based on the above DP iterative scheme is obtainable for the current two-class case but is not a realistic possibility for significantly larger problems. This fact remains an important part of the motivation for the development and analysis of heuristic policies. Hence, and with a view in part to larger problems than the present one, we have also produced in (v), lower bounds on C^{OPT} by a computationally efficient procedure based on the achievable region approach. The idea is to develop a range of constraints satisfied by the first and second moments of the queue lengths. Let algebraic variables (x_1, x_2, y_1, y_2) stand for the moments $\{E(N_1), E(N_2), E(N_1^2), E(N_2^2)\}$ respectively and let $P \subseteq (\mathbb{R}^+)^4$ be the region defined by the constraints generated on these variables. Plainly the value of the optimisation problem

$$\begin{aligned} &\min\ b_1 x_1 + b_2 x_2 + c_1 y_1 + c_2 y_2 \\ &\text{subject to} \quad (x_1, x_2, y_1, y_2) \in P \end{aligned} \tag{42}$$

Table 2. Values of the costs associated with (i) the best threshold policy, (ii) the best linear switching policy, and (iii) the index policy, together with (iv) C^{OPT} and (v) a semidefinite lower bound on C^{OPT} when $b_1 = 4$, $b_2 = 2$

c_1	c_2	(i)	(ii)	(iii)	(iv)	(v)
0.1	0.1	8.550	8.549	8.550	8.549	8.520
0.1	0.2	8.724	8.723	8.724	8.723	8.709
0.1	0.5	9.244	9.244	9.244	9.244	9.238
0.1	1.0	10.112	10.111	10.112	10.111	10.109
0.1	2.0	11.846	11.846	11.846	11.846	11.845
0.2	0.1	9.213	9.209	9.213	9.209	9.100
0.2	0.2	9.387	9.385	9.386	9.385	9.327
0.2	0.5	9.907	9.906	9.907	9.906	9.883
0.2	1.0	10.774	10.772	10.774	10.772	10.762
0.2	2.0	12.509	12.508	12.509	12.508	12.503
0.5	0.1	11.201	11.131	11.133	11.113	10.688
0.5	0.2	11.375	11.346	11.346	11.344	11.020
0.5	0.5	11.895	11.889	11.890	11.889	11.747
0.5	1.0	12.762	12.756	12.762	12.756	12.687
0.5	2.0	14.497	14.491	14.497	14.491	14.459
1.0	0.1	14.515	13.813	13.813	13.687	12.945
1.0	0.2	14.688	14.321	14.329	14.291	13.662
1.0	0.5	15.208	15.100	15.100	15.099	14.637
1.0	1.0	16.076	16.052	16.052	16.051	15.780
1.0	2.0	17.810	17.796	17.808	17.796	17.660
2.0	0.1	19.314	17.525	17.525	17.108	16.619
2.0	0.2	20.592	19.042	19.042	18.814	18.175
2.0	0.5	21.776	20.896	20.896	20.872	19.974
2.0	1.0	22.703	22.351	22.351	22.350	21.560
2.0	2.0	24.437	24.359	24.359	24.356	23.846

is a lower bound on C^{OPT}. See Ansell et al. (1999) for details of how appropriate constraints determining P are developed. The resulting optimisation problem in (42) is a semidefinite program and is solved by using an interior point algorithm and software package (SDPA) developed by Kojima (1994).

The numerical results are presented in Tables 1 and 2 for problems with $\lambda_1 = 1, \mu_1 = 3, \lambda_2 = 5$ and $\mu_2 = 12$ and a range of values of the cost coefficients b_1, b_2, c_1 and c_2. All of the problems analysed in Table 1 have $b_1 = 5$, $b_2 = 1$ while those in Table 2 have $b_1 = 4$, $b_2 = 2$. Upon inspection of columns (i) and (ii) in both Tables, firstly observe that extension from the class of threshold policies to the larger linear switching class can effect a significant improvement in performance. Note from (ii) and (iii) that the performance of the Whittle index policy is very close to that of an optimal policy in the linear switching class in all cases studied and more significantly, a glance at (ii)–(iv) indicates that the index policy is also close to optimal in the class of all policies. While not relevant to the policy assessment which is our primary focus in the current section, the values in column (v) which bound C^{OPT} below, do so sufficiently tightly to believe that the achievable region approach which yielded them should provide an effective tool for analysis for larger problems in which C^{OPT} is not available.

6 Acknowledgements

We express our appreciation to the Engineering and Physical Sciences Research Council for supporting the work of the first and second authors by means of grant GR/M09308 and that of the fourth author by a research studentship. The work of the second and third authors were also supported by NATO under Collaborative Linkage Grant PST.CLG 976568 while the third author received support from the Spanish Ministry of Science and Technology under grant BEC 2000-1027.

References

Ansell PS, Glazebrook KD, Mitrani I (2001) Threshold policies for a single-server queueing network. Prob. Eng. Inf. Sci. 15:15–33

Ansell PS, Glazebrook KD, Mitrani I, Niño-Mora J (1999) A semidefinite programming approach to the optimal control of a single server queueing system with imposed second moment constraints. J. Oper. Res. Soc. 50:765–773

Bertsimas D, Niño-Mora J (1996) Conservation laws, extended polymatroids and multi-armed bandit problems: a polyhedral approach to indexable systems. Math. Oper. Res. 21:257–306

Blanc JPC (1993) Performance analysis and optimisation with the Power Series Algorithm. Lecture Notes in Computer Science 729:55–80, Springer-Verlag

Cobham A (1954) Priority assignment in waiting line problems. Oper. Res. 2:70–76

Cox DR, Smith WL (1961) Queues. Methuen, London

Dacre M, Glazebrook KD, Niño-Mora J (1999) The achievable region approach to the optimal control of stochastic systems (with discussion). J. Roy. Statist. Soc. B61:747–791

Gittins JC (1979) Bandit processes and dynamic allocation indices (with discussion). J. R. Statist. Soc. B41:148–177

Harrison JM (1975) Dynamic scheduling of a multiclass queue: discount optimality. Oper. Res. 23:270–282

Klimov GP (1974) Time sharing service systems I. Theory Prob. Appl. 19:532–551

Kojima M (1994) A primitive interior point algorithm for semidefinite programs. In Mathematica, Research Report #293, Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology, Oh-Okayama, Meguro, Tokyo 152, Japan

Meilijson I, Weiss G (1977) Multiple feedback at a single-server station. Stoch. Proc. Appl. 5:195–205

Niño-Mora J (2001a) Restless bandits, partial conservation laws and indexability. Adv. Appl. Prob. 33:76–98

Niño-Mora J (2001b) Countable partial conservation laws, Whittle's restless bandit index and a dynamic $c\mu$ rule for scheduling a multiclass $M/M/1$ queue with convex holding costs. Technical Report, Universitat Pompeu Fabra

Puterman ML (1994) Markov decision processes: Discrete stochastic dynamic programming. Wiley, New York

Righter R, Xu SH (1991) Scheduling jobs on nonidentical IFR processors to minimize general cost functions. Adv. Appl. Prob. 23:909–924

Tijms H (1994) Stochastic models, an algorithmic approach. Wiley, New York

van Meighem JA (1995) Dynamic scheduling with convex delay costs: the generalized $c\mu$-rule. Ann. Appl. Prob. 5:809–833

Weber RR (1988) Stochastic scheduling on parallel processors and minimization of concave functions of completion times. In "Stochastic differential systems, stochastic control theory and applications." IMA Vol. Math. Appl. 10:601–609, Springer-Verlag

Weber RR, Weiss G (1990) On an index policy for restless bandits. J. Appl. Prob. 27:637–648

Weber RR, Weiss G (1991) Addendum to "On an index policy for restless bandits". Adv. Appl. Prob. 23:429–430

Weiss G (1988) Branching bandit processes. Prob. Eng. Inf. Sci. 2:269–278

Whittle P (1988) Restless bandits: activity allocation in a changing world. J. Appl. Prob. A25:287–298