UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 07-43
Statistics and Econometrics Series 11
May 2007

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 6249849

Characterization and Computation of Restless Bandit Marginal Productivity Indices*

José Niño-Mora [1]

Abstract

The Whittle index [P. Whittle (1988). Restless bandits: Activity allocation in a changing world. J. Appl. Probab. 25A, 287-298] yields a practical scheduling rule for the versatile yet intractable multi-armed restless bandit problem, involving the optimal dynamic priority allocation to multiple stochastic projects, modeled as restless bandits, i.e., binary-action (active/passive) (semi-) Markov decision processes. A growing body of evidence shows that such a rule is nearly optimal in a wide variety of applications, which raises the need to efficiently compute the Whittle index and more general marginal productivity index (MPI) extensions in large-scale models. For such a purpose, this paper extends to restless bandits the parametric linear programming (LP) approach deployed in [J. Niño-Mora. A $(2/3)n^3$ fast-pivoting algorithm for the Gittins index and optimal stopping of a Markov chain, INFORMS J. Comp., in press], which yielded a fast Gittins-index algorithm. Yet the extension is not straightforward, as the MPI is only defined for the limited range of so-called indexable bandits, which motivates the quest for methods to establish indexability. This paper furnishes algorithmic and analytical tools to realize the potential of MPI policies in large-scale applications, presenting the following contributions: (i) a complete algorithmic characterization of indexability, for which two block implementations are given; and (ii) more importantly, new analytical conditions for indexability — termed LP-indexability — that leverage knowledge on the structure of optimal policies in particular models, under which the MPI is computed faster by the adaptive-greedy algorithm previously introduced by the author under the more stringent PCL-indexability conditions, for which a new fast-pivoting block implementation is given. The paper further reports on a computational study, measuring the runtime performance of the algorithms, and assessing by a simulation study the high prevalence of indexability and PCL-indexability.

Keywords: Dynamic programming, semi-Markov, finite state; stochastic scheduling; restless bandits; priority-index policy; indexability; Whittle index; marginal productivity index; parametric simplex; block algorithms; computational complexity

***JEL Classification**: C61, C63*

[1] Niño-Mora, Departamento de Estadística, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), e-mail: jose.nino@uc3m.es. Supported in part by the Spanish Ministry of Education & Science under grant MTM2004-02334 and a Ramón y Cajal Investigator Award, by the EU's Networks of Excellence Euro-NGI/FGI, and by the Autonomous Community of Madrid-UC3M's grants UC3M-MTM-05-075 and CCG06-UC3M/ESP-0767.

Characterization and Computation of Restless Bandit Marginal Productivity Indices

José Niño-Mora
Department of Statistics, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), Spain, jnimora@alum.mit.edu
http://alum.mit.edu/www/jnimora

Abstract

The Whittle index [P. Whittle (1988). Restless bandits: Activity allocation in a changing world. J. Appl. Probab. 25A, 287–298] yields a practical scheduling rule for the versatile yet intractable multi-armed restless bandit problem, involving the optimal dynamic priority allocation to multiple stochastic projects, modeled as restless bandits, i.e., binary-action (active/passive) (semi-) Markov decision processes. A growing body of evidence shows that such a rule is nearly optimal in a wide variety of applications, which raises the need to efficiently compute the Whittle index and more general marginal productivity index (MPI) extensions in large-scale models. For such a purpose, this paper extends to restless bandits the parametric linear programming (LP) approach deployed in [J. Niño-Mora. A $(2/3)n^3$ fast-pivoting algorithm for the Gittins index and optimal stopping of a Markov chain, INFORMS J. Comp., in press], which yielded a fast Gittins-index algorithm. Yet the extension is not straightforward, as the MPI is only defined for the limited range of so-called indexable bandits, which motivates the quest for methods to establish indexability. This paper furnishes algorithmic and analytical tools to realize the potential of MPI policies in large-scale applications, presenting the following contributions: (i) a complete algorithmic characterization of indexability, for which two block implementations are given; and (ii) more importantly, new analytical conditions for indexability — termed LP-indexability — that leverage knowledge on the structure of optimal policies in particular models, under which the MPI is computed faster by the adaptive-greedy algorithm previously introduced by the author under the more stringent PCL-indexability conditions, for which a new fast-pivoting block implementation is given. The paper further reports on a computational study, measuring the runtime performance of the algorithms, and assessing by a simulation study the high prevalence of indexability and PCL-indexability.

Key words: Dynamic programming, semi-Markov, finite state; stochastic scheduling; restless bandits; priority-index policy; indexability; Whittle index; marginal productivity index; parametric simplex; block algorithms; computational complexity
History: submitted May 8, 2007

---

1. Introduction

The *multi-armed restless bandit problem* (MARBP) furnishes a powerful modeling framework for a wide variety of problems where a decision-maker must dynamically prioritize the allocation of limited effort to multiple projects. The latter are modeled as *restless bandits*, i.e., binary-action (active/work; passive/rest) semi-Markov decision processes (SMDPs) that can change state even while rested. For a range of applications to problems of admission control, routing and scheduling see, e.g., Whittle (1988), Veatch and Wein (1996), Niño-Mora (2002, 2003, 2005, 2006b,c,d,e, 2007a,b,c,d), Raissi-Dehkordi and Baras (2002), Goyal et al. (2006), and La Scala and Moran (2006).

While the MARBP is generally intractable, Whittle (1988) introduced an index for restless bandits that extends the celebrated *Gittins index* rule, which is optimal in the classic case where passive bandits remain

frozen. See Gittins (1979). The *Whittle index* has been further extended in Niño-Mora (2002, 2006b,d) in the framework of the unifying and intuitive concept of *marginal productivity index* (MPI). A growing body of evidence (cf. the aforementioned papers) shows that the resulting *priority-index rule* that engages at each time a project of largest index is nearly optimal for a variety of applications. Further, the MPI characterizes optimal policies for problems modeling the optimal dynamic allocation of work to a project, which have intrinsic interest.

The prime goal of this paper is to furnish the required algorithmic and analytical tools that will allow researchers to fully realize the potential of such index policies in large-scale applications. We will accomplish such a goal by drawing on classic parametric linear programming (cf. Gass and Saaty (1955); Saaty and Gass (1954)), extending the approach that, first suggested in Kallenberg (1986), was developed in Niño-Mora (2006a) to obtain a Gittins-index algorithm of improved complexity, performing $(2/3)n^3 + O(n^2)$ arithmetic operations for a classic n-state bandit.

The required extension is, however, far from straightforward, as the MPI is only defined for the limited range of so-called *indexable* bandits, which motivates the quest for useful numerical and analytical methods to establish indexability. For such a purpose, we had introduced and developed in Niño-Mora (2001, 2002, 2006d) a set of sufficient conditions for indexability, termed *PCL-indexability* as they are based on satisfaction of *partial conservation laws* (PCL), under which a bandit's MPI is computed by an *adaptive-greedy algorithm*. Yet, though such work shows that several models of interest are PCL-indexable, our more recent work has revealed limitations to such an approach. Specifically: (i) one condition was that the index sequence produced by the aforementioned algorithm be nondecreasing, which we have found to be hard to verify analytically in models with a multi-dimensional state; and, (ii) more importantly, we have encountered in Niño-Mora (2007d) a relevant bandit model that is indexable, yet not PCL-indexable.

This paper overcomes such limitations, presenting the following contributions: (i) a complete algorithmic characterization of indexability, for which two block implementations are given, the *Complete-Pivoting Indexability* (CPI) algorithm and the *Reduced-Pivoting Indexability* (RPI) algorithm, which, after a common initialization stage involving the solution of a block linear equation system, perform $2n^3 + O(n^2)$ and $n^3 + O(n^2)$ arithmetic operations for an n-state bandit, respectively; and (ii) more importantly, new analytical sufficient conditions for indexability — termed *LP-indexability* — that leverage knowledge on the structure of optimal policies in particular models, under which the MPI is computed faster by the adaptive-greedy algorithm referred to above, for which a new fast-pivoting block implementation is given that performs — after the initialization stage — $(2/3)n^3 + O(n^2)$ operations; such conditions are also shown to be necessary, in that an indexable bandit is always *LP-indexable* relative to a certain family of policies; further, a more analytically tractable reformulation of the PCL-indexability conditions is presented. For examples where

such an approach is successfully deployed, we refer the reader to Niño-Mora (2006e, 2007a,d).

We must emphasize that the algorithms presented herein are described in a readily-implementable *block-partitioned* form, i.e., based on operations on submatrices (*blocks*) of a base matrix. Such implementations have been advocated in the scientific-computing literature to partly overcome the exponentially widening gap between processor speed and memory-access times in contemporary computers, which often render traditional complexity measures based on operation counts poor predictors of runtime performance. See Dongarra and Eijkhout (2000) and Baker et al. (2006).

The latter phenomenon is illustrated herein by a computational study comparing the runtime performance of the proposed algorithms, which reveals that the *fast-pivoting adaptive-greedy* (FPAG) algorithm is the fastest, consistently achieving a speedup factor of about 1.3 over the CPI algorithm, which in turn slightly outperforms the RPI algorithm. Such results reflect the influence of differing memory-access patterns in actual runtimes. Thus, the CPI algorithm manipulates whole matrices, which results in efficient handling of contiguous memory blocks, whereas the RPI and FPAG algorithms reduce operation counts at the expense of manipulating submatrices with complex patterns, which results in relatively inefficient noncontiguous data movement.

Another computational study was conducted to assess the prevalence of indexability and PCL-indexability among randomly generated restless bandits — with dense transition probability matrices — in a large-scale simulation study. The study reveals that such prevalences are extremely high, growing steeply as the number of states increases.

The remainder of the paper is organized as follows. Section 2 reviews the indexation theory for semi-Markov restless bandits. Section 3 elucidates the parametric simplex tableaux for the problem's LP formulation. Section 4 develops a simplex-based algorithmic characterization of indexability. Section 5 shows how to exploit special structure by introducing the new class of LP-indexable bandits, to which the adaptive-greedy index algorithm introduced in earlier work for PCL-indexable bandits is shown to extend, and revises the earlier definition of the PCL-indexability; further, a new fast-pivoting implementation is given of such an algorithm. While previous sections focus on the discounted criterion, Section 6 discusses the extension to the average criterion. Section 7 reports on the computational study's results.

2. Indexation for Semi-Markov Restless Bandits

This section reviews several key concepts from indexation theory to be used throughout the paper, as it applies to a finite-state semi-Markov restless bandit. The following discussion highlights the insightful relation of indexation with bicriteria optimization, which was implicit in Niño-Mora Niño-Mora (2002,

2006d), focusing on the discounted case. As in the previous section's model, we will find it useful to partition the state space N into the set $N^{\{0,1\}}$ of *controllable states*, where actions differ in some respect, and the set $N^{\{0\}} \triangleq N \setminus N^{\{0,1\}}$ of *uncontrollable states*. We will adopt the convention that the passive action is taken in the latter states, and denote the numbers of uncontrollable and controllable states by $m \triangleq |N^{\{0\}}| \geq 0$ and $n \triangleq |N^{\{0,1\}}| \geq 1$, respectively.

2.1 Semi-Markov Restless Bandits and Discrete-Stage Reformulation

Consider the problem of operating optimally a single dynamic and stochastic project, modelled as a binary-action (1/active/engage; 0/passive/rest) *semi-Markov decision process* (SMDP), whose *natural state* $X(t)$ evolves continually over time $t \geq 0$ through the finite state space N. The controller observes the *embedded state* $X_k \triangleq X(t_k)$ at an increasing sequence of *decision epochs* t_k, with $t_0 = 0$ and $\lim_k t_k \nearrow +\infty$, and takes an action $a_k \triangleq a(t_k) \in \{0,1\}$ that prevails during the ensuing *stage* $[t_k, t_{k+1})$. Processes $X(t)$ and $a(t)$ are thus piecewise constant, right-continuous with left limits. Actions are prescribed through a *policy* π, drawn from the class Π of *admissible* policies, which base decisions on the history of embedded states and actions up to the present decision epoch, and on the state observed at the latter. While the project occupies state i and action a prevails, *rewards* accrue and *work* is expended at rates R_i^a and $Q_i^a \geq 0$, respectively, with $Q_i^1 > 0$ and $Q_i^1 \geq Q_i^0 \geq 0$.

We complete next the model's description, by specifying its dynamics, and discuss its discrete-stage reformulation along the lines in (Puterman, 1994, Ch. 11)), which will be used in the subsequent analyses. If at decision epoch t_k the project occupies state $X_k = i$ and action $a_k = a$ is taken, the joint distribution of the duration $t_{k+1} - t_k$ of the ensuing (i,a)-*stage* and the next embedded state X_{k+1} is given by the transition distribution

$$F_{ij}^a(t) \triangleq \mathbb{P}\{t_{k+1} - t_k \leq t, X_{k+1} = j \mid X_k = i, a_k = a\},$$

having Laplace-Stieltjes transform (LST)

$$\phi_{ij}^a(\alpha) \triangleq \mathbb{E}\left[1_{\{X_{k+1}=j\}} e^{-\alpha(t_{k+1}-t_k)} \mid X_k = i, a_k = a\right] = \int_0^\infty e^{-\alpha t}\, dF_{ij}^a(t),$$

for $\alpha > 0$. The corresponding one-stage transition probabilities of the embedded process are

$$p_{ij}^a \triangleq \mathbb{P}\{X_{n+1} = j \mid X_k = i, a_k = a\} = \lim_{t\to\infty} F_{ij}^a(t) = \lim_{\alpha \searrow 0} \phi_{ij}^a(\alpha).$$

From $F_{ij}^a(t)$ we obtain the distribution of the duration of an (i,a)-stage,

$$F_i^a(t) \triangleq \mathbb{P}\{t_{k+1} - t_k \leq t \mid X_k = i, a_k = a\} = \sum_{j\in N} F_{ij}^a(t),$$

having LST

$$\phi_i^a(\alpha) \triangleq \mathbb{E}\left[e^{-\alpha(t_{k+1}-t_k)} \mid X_k = i, a_k = a\right] = \sum_{j\in N} \phi_{ij}^a(\alpha), \tag{1}$$

and mean

$$m_i^a \triangleq \mathbb{E}\left[t_{k+1} - t_k \mid X_k = i, a_k = a\right] = \int_0^\infty t\, dF_i^a(t).$$

In general, the natural-state process $X(t)$ might change state between decision epochs. Its evolution *within* an (i,a)-period is characterized by

$$\tilde{p}_{ij}^a(s) \triangleq \mathbb{P}\left\{X(t_k+s) = j \mid X_k = i, a_k = a, t_{k+1} - t_k > s\right\},$$

the probability that state j is occupied s time units after a decision epoch, given that the next epoch has not yet occurred. We can thus represent the expected total discounted work expended and the reward earned during an (i,a)-stage, respectively, as

$$q_i^a \triangleq \mathbb{E}\left[\int_{t_k}^{t_{k+1}} Q_{X(t)}^{a_n} e^{-\alpha(t-t_k)}\, dt \mid X_k = i, a_k = a\right] = \sum_{j\in N} Q_j^a \int_0^\infty \tilde{p}_{ij}^a(s)\{1 - F_i^a(s)\} e^{-\alpha s}\, ds \tag{2}$$

and

$$r_i^a \triangleq \mathbb{E}\left[\int_{t_k}^{t_{k+1}} R_{X(t)}^{a_k} e^{-\alpha(t-t_k)}\, dt \mid X_k = i, a_k = a\right] = \sum_{j\in N} R_j^a \int_0^\infty \tilde{p}_{ij}^a(s)\{1 - F_i^a(s)\} e^{-\alpha s}\, ds. \tag{3}$$

In our studies of several applications, we have found that it is often important to partition the state space N into the set of *uncontrollable states*

$$N^{\{0\}} \triangleq \left\{i \in N\colon q_i^0 = q_i^1, r_i^0 = r_i^1 \text{ and } F_{ij}^0(t) \equiv F_{ij}^1(t), j \in N\right\},$$

where both actions have identical consequences, and the remaining set $N^{\{0,1\}} \triangleq N \setminus N^{\{0\}}$ of *controllable states*. The notation $N^{\{0\}}$ reflects the convention we adopt whereby the passive action $a = 0$ is taken at uncontrollable states. We will denote by $n \triangleq |N^{\{0,1\}}|$ and $m \triangleq |N^{\{0\}}|$ the numbers of controllable and of uncontrollable states, respectively, and assume that $n \geq 1$. As we will see, the indices of concern in this paper, which are functions of the project's state, are only defined for controllable states.

In the sequel we will focus on the discounted criterion based on measures (4)–(5), deferring to Section 6 discussion of the long-run average criterion.

2.2 Restless Bandit Indexation

We consider two measures to evaluate a policy π, relative to an initial state i and a discount rate $\alpha > 0$: the *reward measure*

$$f_i^\pi \triangleq \mathbb{E}_i^\pi\left[\int_0^\infty R_{X(t)}^{a(t)} e^{-\alpha t}\, dt\right] = \mathbb{E}_i^\pi\left[\sum_{k=0}^\infty r_{X_k}^{a_k} e^{-\alpha t_k}\right], \tag{4}$$

giving the expected total discounted value of rewards earned; and the *work measure*

$$g_i^\pi \triangleq \mathbb{E}_i^\pi \left[\int_0^\infty Q_{X(t)}^{a(t)} e^{-\alpha t}\, dt \right] = \mathbb{E}_i^\pi \left[\sum_{k=0}^{\infty} q_{X_k}^{a_k} e^{-\alpha t_k} \right], \tag{5}$$

giving the expected total discounted amount of work expended. Notice that the right-hand side's identities in (4)–(5) draw on the discrete-stage reformulation discussed above.

We will find it convenient to use the corresponding averaged measures obtained when the initial state i is drawn from an arbitrary distribution with positive probability mass $p_i > 0$ for $i \in N$:

$$f^\pi \triangleq \sum_{i \in N} p_i f_i^\pi \quad \text{and} \quad g^\pi \triangleq \sum_{i \in N} p_i g_i^\pi.$$

Introducing a wage rate ν at which work is paid for, we will address the *ν-wage problem*

$$\max_{\pi \in \Pi} f^\pi - \nu g^\pi, \tag{6}$$

which is to find an admissible project-operating policy maximizing the value of rewards earned minus labor costs incurred, and where ν will play the role of a parameter to be varied over $\mathbb{R}$.

The theory of finite-state and -action SMDPs ensures existence of an optimal policy for (6) that is: (i) deterministic stationary; and (ii) independent of the initial-state distribution. We represent each such a policy by its *active set* $S \subseteq N^{\{0,1\}}$, or subset of controllable states where the policy prescribes to engage the project at a decision epoch, and will refer to it as the *S-active policy*.

It appears reasonable to expect that, at least in some models, optimal active sets should expand monotonically from $\emptyset$ to $N^{\{0,1\}}$ as the wage ν decreases from $+\infty$ to $-\infty$., as a function of the state space's size Such an intuitive property was introduced by Whittle (1988), who termed it *indexability*, for Markovian restless bandits with state-independent work rates $q_i^a \equiv a$. His original definition readily extends to the present setting.

In *dynamic programming* (DP) terms, we may formulate the indexability property as follows. Letting $\vartheta_i^*(\nu)$ be the optimal value function starting at i for SMDP (6), the Bellman equations are

$$\vartheta_i^*(\nu) = \max_{a \in \{0,1\}} r_i^a - \nu q_i^a + \sum_{j \in N} \phi_{ij}^a \vartheta_j^*(\nu), \quad i \in N, \tag{7}$$

where we write $\phi_{ij}^a = \phi_{ij}^a(\alpha)$. In words, the project is indexable if, for each controllable state i, it is optimal to engage the project at i for ν small enough; namely, if there exists an *index* ν_i^*, for $i \in N^{\{0,1\}}$, such that it is optimal to engage the project in state i iff $\nu \leq \nu_i^*$; or, in formulas,

$$\vartheta_i^*(\nu) = r_i^1 - \nu q_i^1 + \sum_{j \in N} \phi_{ij}^1 \vartheta_j^*(\nu) \Longleftrightarrow \nu \leq \nu_i^* \tag{8}$$

Yet, in Niño-Mora (2006d) we have formulated the indexability property in an alternative — though equivalent — form yielding complementary insights, as reviewed next. Let $i_1,\ldots,i_n \in N^{\{0,1\}}$ be an ordering of the n controllable states, such that the *nested active-set family*

$$\mathscr{F}_0 \triangleq \{S_0, S_1, \ldots, S_n\}, \tag{9}$$

where $S_0 \triangleq \emptyset$ and $S_k \triangleq \{i_1,\ldots,i_k\}$ for $1 \leq k \leq n$, satisfies the work-regularity condition

$$g^{S_{k-1}} < g^{S_k}, \quad 1 \leq k \leq n. \tag{10}$$

Consider the *index* ν_i^*, for $i \in N^{\{0,1\}}$, defined by

$$\nu_{i_k}^* \triangleq \frac{f^{S_k} - f^{S_{k-1}}}{g^{S_k} - g^{S_{k-1}}}, \quad 1 \leq k \leq n. \tag{11}$$

Definition 2.1 (Indexability; MPI) We say that the bandit is *indexable* if there exists a nested active-set family $\mathscr{F}_0$ as above such that:

(i) index $\nu_{i_k}^*$ is nonincreasing in k, i.e., $\nu_{i_{k+1}}^* \geq \nu_{i_k}^*$ for $1 \leq k < n$; and

(ii) for ν-wage problem (6), the $\emptyset$-active policy is optimal iff $\nu \leq \nu_{i_1}^*$, the $N^{\{0,1\}}$-active policy is optimal iff $\nu \geq \nu_{i_n}^*$, and the S_k-active policy is optimal for ν-wage problem (6) iff $\nu \in [\nu_{i_{k+1}}^*, \nu_{i_k}^*]$, for $1 \leq k < n$.

We then say that the project is $\mathscr{F}_0$-*indexable*, and that ν_i^* is its *marginal productivity index* (MPI).

Note: as already noted in nm (give the reference), the optimal value function of an indexable bandit is given by

$$\vartheta_i^*(\nu) = \max_{S \in \mathscr{F}_0} f_i^S - \nu g_i^S = \max_{0 \leq k \leq n} f_i^{S_k} - \nu g_i^{S_k}, \quad i \in N, \nu \in \mathbb{R}$$

We introduced the term MPI in Niño-Mora (2006d), as it was shown there, and earlier in Niño-Mora (2002), that index ν_i^* measures the marginal value, or productivity, of work at each state i. The first paper gave a characterization of indexability in terms of the structure of the *achievable work-reward performance region*

$$\mathbb{H} \triangleq \left\{(g^\pi, f^\pi) \colon \pi \in \Pi\right\},$$

which is spanned by work-reward performance points under admissible policies. Such a region is the *convex polygon* given as the *convex hull* of the finite set of points (g^S, f^S), for all active sets $S \subseteq N$. Specifically, considering the *upper boundary* of $\mathbb{H}$, given by

$$\overline{\partial}\mathbb{H} \triangleq \{(g,f) \in \mathbb{H} \colon\ f^\pi \leq f \text{ for any } \pi \in \Pi \text{ with } g^\pi = g\},$$

it is shown in Niño-Mora (2006d, Th. 3.1) that the project is indexable iff there is a nested active-set family $\mathscr{F}_0$ as above that determines such an upper boundary.

Notice that the choice of $\mathscr{F}_0$ need not be unique, and that the MPI does not depend on such a choice. Consider, e.g., a discrete-time nonrestless (i.e., with $p^0_{ii} \equiv 1$) project with $q^a_i \equiv a$ and $r^a_i \equiv 0$, so that $f^\pi \equiv 0$ for any policy π. Then, *each* of the $n!$ orderings $i_1, \ldots, i_n$ of the n project states yields a nested family $\mathscr{F}_0$ relative to which the project is indexable — with MPI $\nu^*_i \equiv 0$.

2.3 Two Illustrative Examples

To help the reader unfamiliar with the above concepts to grasp them, we discuss next two illustrave examples, corresponding to discrete-time Markovian bandits with state space $N = \{1,2,3\}$ and one-period work expenditures $q^a_i = a$ — hence $N^{\{0,1\}} = N$. For each instance, a plot is displayed of the achievable work-reward performance region $\mathbb{H}$, where points (g^S, f^S) are labelled by their active sets S. We have taken the initial-state distribution to be uniform over N.

Figure 1 displays the achievable work-reward performance region for the instance with discrete-time discount factor $\beta = 0.9$, one-period active reward and one-period transition probabilities

$$\mathbf{r}^1 = \begin{bmatrix} 0.9016 \\ 0.10949 \\ 0.01055 \end{bmatrix}, \mathbf{P}^1 = \begin{bmatrix} 0.2841 & 0.4827 & 0.2332 \\ 0.5131 & 0.0212 & 0.4657 \\ 0.4612 & 0.0081 & 0.5307 \end{bmatrix}, \mathbf{P}^0 = \begin{bmatrix} 0.1810 & 0.4801 & 0.3389 \\ 0.2676 & 0.2646 & 0.4678 \\ 0.5304 & 0.2843 & 0.1853 \end{bmatrix},$$

and one-period passive reward $\mathbf{r}^0 = \mathbf{0}$. The plot shows that this is an indexable instance, relative to the nested active-set family $\mathscr{F}_0 = \big\{\emptyset, \{1\}, \{1,2\}, \{1,2,3\}\big\}$, which determines the region's upper boundary. The Whittle index/MPI values of states 1, 2 and 3 are given by the successive trade-off vs. work rates/slopes in such an upper boundary:

$$\nu^*_1 = \frac{f^{\{1\}} - f^{\emptyset}}{g^{\{1\}} - g^{\emptyset}} > \nu^*_2 = \frac{f^{\{1,2\}} - f^{\{1\}}}{g^{\{1,2\}} - g^{\{1\}}} > \nu^*_3 = \frac{f^{\{1,2,3\}} - f^{\{1,2\}}}{g^{\{1,2,3\}} - g^{\{1,2\}}}.$$

Figure 2 displays the achievable work-reward performance region for the instance with $\beta = 0.9$,

$$\mathbf{P}^1 = \begin{bmatrix} 0.7796 & 0.0903 & 0.1301 \\ 0.1903 & 0.1863 & 0.6234 \\ 0.2901 & 0.3901 & 0.3198 \end{bmatrix}, \mathbf{P}^0 = \begin{bmatrix} 0.1902 & 0.4156 & 0.3942 \\ 0.5676 & 0.4191 & 0.0133 \\ 0.0191 & 0.1097 & 0.8712 \end{bmatrix},$$

and

$$\mathbf{r}^1 = \begin{bmatrix} 0.9631 & 0.7963 & 0.1057 \end{bmatrix}^\top, \mathbf{r}^0 = \begin{bmatrix} 0.458 & 0.5308 & 0.6873 \end{bmatrix}^\top.$$

The plot shows that this is a nonindexable instance, since there is no nested active-set family that determines the region's upper boundary.

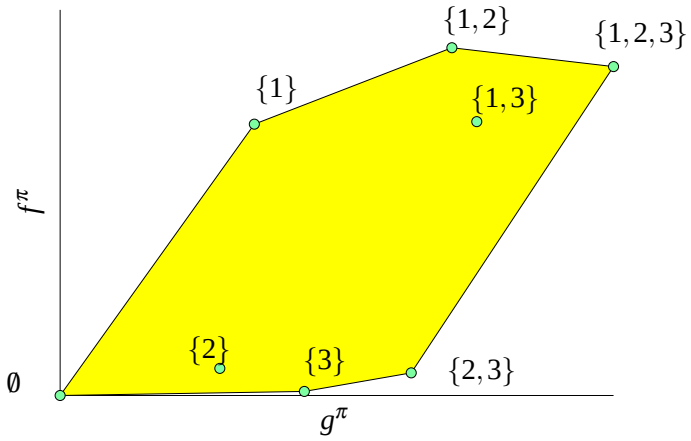

Figure 1: Indexable Instance: Achievable Work-Reward Performance Region.

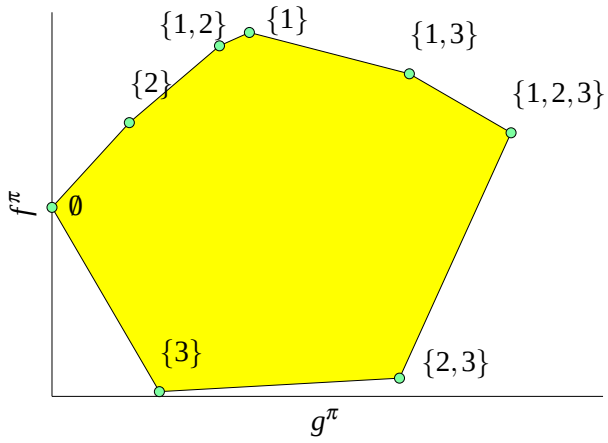

Figure 2: Nonindexable Instance: Achievable Work-Reward Performance Region.

2.4 Marginal Work, Reward and Productivity Measures

The analyses and algorithms below will use the *marginal measures* discussed next. For an action $a \in \{0,1\}$ and an active set $S \subseteq N^{\{0,1\}}$, denote by $\langle a, S\rangle$ the policy that takes action a in the initial stage, and adopts the *S-active policy* (having active set S) thereafter. Now, for a state i and an active set S, define the *marginal work measure* by

$$w_i^S \triangleq g_i^{\langle 1,S\rangle} - g_i^{\langle 0,S\rangle}, \tag{12}$$

i.e., as the marginal increase in work expended that results from taking initially the active instead of the passive action at state i, given that the S-active policy is adopted thereafter.

Further, define the *marginal reward measure* by

$$d_i^S \triangleq f_i^{\langle 1,S\rangle} - f_i^{\langle 0,S\rangle}, \tag{13}$$

i.e., as the corresponding marginal increase in value of rewards earned. Notice that marginal work and reward measures vanish at uncontrollable states:

$$w_i^S = d_i^S = 0, \quad i \in N^{\{0\}}. \tag{14}$$

Finally, for $w_i^S \neq 0$, define the *marginal productivity measure* by

$$v_i^S \triangleq \frac{d_i^S}{w_i^S}. \tag{15}$$

2.5 Reduction to the No Uncontrollable States Case

While we have found the distinction between controllable and uncontrollable states to be relevant in some applications of restless bandits, it would considerably complicate the notation in the analyses below. We thus show next that it suffices to restrict attention to bandits with no uncontrollable states, as these can be eliminated through suitable transformations.

Thus, consider a restless bandit as above, with controllable and uncontrollable state spaces $N^{\{0,1\}}$ and $N^{\{0\}}$, respectively. For a given active set $S \subseteq N^{\{0,1\}}$, we can evaluate the work measure g_i^S by solving the following linear equation system, which we decompose in blocks as

$$\begin{aligned}
\mathbf{g}_S^S &= \mathbf{q}_S^1 + \Phi_{SS}^1 \mathbf{g}_S^S + \Phi_{S,N^{\{0,1\}}\setminus S}^1 \mathbf{g}_{N^{\{0,1\}}\setminus S}^S + \Phi_{S,N^{\{0\}}}^1 \mathbf{g}_{N^{\{0\}}}^S \\
\mathbf{g}_{N^{\{0,1\}}\setminus S}^S &= \mathbf{q}_{N^{\{0,1\}}\setminus S}^0 + \Phi_{N^{\{0,1\}}\setminus S,S}^1 \mathbf{g}_S^S + \Phi_{N^{\{0,1\}}\setminus S,N^{\{0,1\}}\setminus S}^0 \mathbf{g}_{N^{\{0,1\}}\setminus S}^S + \Phi_{N^{\{0,1\}}\setminus S,N^{\{0\}}}^1 \mathbf{g}_{N^{\{0\}}}^S \\
\mathbf{g}_{N^{\{0\}}}^S &= \mathbf{q}_{N^{\{0\}}}^0 + \Phi_{N^{\{0\}},S}^0 \mathbf{g}_S^S + \Phi_{N^{\{0\}},N^{\{0,1\}}\setminus S}^0 \mathbf{g}_{N^{\{0,1\}}\setminus S}^S + \Phi_{N^{\{0\}}N^{\{0\}}}^0 \mathbf{g}_{N^{\{0\}}}^S,
\end{aligned} \tag{16}$$

writing, e.g., $\Phi^1_{S,N^{\{0,1\}}\setminus S} = (\phi_{ij})_{i\in S, j\in N^{\{0,1\}}\setminus S}$ and $\mathbf{g}^S_S = (g^S_i)_{i\in S}$. Now, solving in the last equation block above for $\mathbf{g}^S_{N^{\{0\}}}$, and defining

$$\begin{aligned}\widetilde{\mathbf{q}}^a_{N^{\{0,1\}}} &\triangleq \mathbf{q}^a_{N^{\{0,1\}}} + \Phi^a_{N^{\{0,1\}}N^{\{0\}}}\left(\mathbf{I}_{N^{\{0\}}N^{\{0,1\}}} - \Phi^0_{N^{\{0\}}N^{\{0\}}}\right)^{-1}\mathbf{q}^a_{N^{\{0\}}}\\ \widetilde{\Phi}^a_{N^{\{0,1\}}N^{\{0,1\}}} &\triangleq \Phi^a_{N^{\{0,1\}}N^{\{0,1\}}} + \Phi^a_{N^{\{0,1\}}N^{\{0\}}}\left(\mathbf{I}_{N^{\{0\}}N^{\{0,1\}}} - \Phi^0_{N^{\{0\}}N^{\{0\}}}\right)^{-1}\Phi^0_{N^{\{0\}}N^{\{0,1\}}},\end{aligned} \tag{17}$$

where $\mathbf{I}$ is the identity matrix, we can reformulate (16) as

$$\begin{aligned}\mathbf{g}^S_S &= \widetilde{\mathbf{q}}^1_S + \widetilde{\Phi}^1_{SS}\mathbf{g}^S_S + \widetilde{\Phi}^1_{S,N^{\{0,1\}}\setminus S}\mathbf{g}^S_{N^{\{0,1\}}\setminus S}\\ \mathbf{g}^S_{N^{\{0,1\}}\setminus S} &= \widetilde{\mathbf{q}}^0_{N^{\{0,1\}}\setminus S} + \widetilde{\Phi}^0_{N^{\{0,1\}}\setminus S,S}\mathbf{g}^S_S + \widetilde{\Phi}^1_{N^{\{0,1\}}\setminus S,N^{\{0,1\}}\setminus S}\mathbf{g}^S_{N^{\{0,1\}}\setminus S}.\end{aligned} \tag{18}$$

Similarly, we can evaluate the reward measure f^S_i by solving the linear equation system

$$\begin{aligned}\mathbf{f}^S_S &= \mathbf{r}^1_S + \Phi^1_{SS}\mathbf{f}^S_S + \Phi^1_{S,N^{\{0,1\}}\setminus S}\mathbf{f}^S_{N^{\{0,1\}}\setminus S} + \Phi^1_{S,N^{\{0\}}}\mathbf{f}^S_{N^{\{0\}}}\\ \mathbf{f}^S_{N^{\{0,1\}}\setminus S} &= \mathbf{r}^0_{N^{\{0,1\}}\setminus S} + \Phi^0_{N^{\{0,1\}}\setminus S,S}\mathbf{f}^S_S + \Phi^1_{N^{\{0,1\}}\setminus S,N^{\{0,1\}}\setminus S}\mathbf{f}^S_{N^{\{0,1\}}\setminus S} + \Phi^0_{N^{\{0,1\}}\setminus S,N^{\{0\}}}\mathbf{f}^S_{N^{\{0\}}}\\ \mathbf{f}^S_{N^{\{0\}}} &= \mathbf{r}^0_{N^{\{0\}}} + \Phi^0_{N^{\{0\}},S}\mathbf{f}^S_S + \Phi^0_{N^{\{0\}},N^{\{0,1\}}\setminus S}\mathbf{f}^S_{N^{\{0,1\}}\setminus S} + \Phi^0_{N^{\{0\}},N^{\{0\}}}\mathbf{f}^S_{N^{\{0\}}}.\end{aligned} \tag{19}$$

Proceeding as above, and defining

$$\widetilde{\mathbf{r}}^a_{N^{\{0,1\}}} \triangleq \mathbf{r}^a_{N^{\{0,1\}}} + \Phi^a_{N^{\{0,1\}}N^{\{0\}}}\left(\mathbf{I}_{N^{\{0\}}N^{\{0,1\}}} - \Phi^0_{N^{\{0\}}N^{\{0\}}}\right)^{-1}\mathbf{r}^a_{N^{\{0\}}}, \tag{20}$$

we can reformulate (19) as

$$\begin{aligned}\mathbf{f}^S_S &= \widetilde{\mathbf{r}}^1_S + \widetilde{\Phi}^1_{SS}\mathbf{f}^S_S + \widetilde{\Phi}^1_{S,N^{\{0,1\}}\setminus S}\mathbf{f}^S_{N^{\{0,1\}}\setminus S}\\ \mathbf{f}^S_{N^{\{0,1\}}\setminus S} &= \widetilde{\mathbf{r}}^0_{N^{\{0,1\}}\setminus S} + \widetilde{\Phi}^0_{N^{\{0,1\}}\setminus S,S}\mathbf{f}^S_S + \widetilde{\Phi}^1_{N^{\{0,1\}}\setminus S,N^{\{0,1\}}\setminus S}\mathbf{f}^S_{N^{\{0,1\}}\setminus S}.\end{aligned} \tag{21}$$

From the above discussion, it is readily seen how to eliminate uncontrollable states from the analyses: it suffices to consider a modified discrete-stage bandit having state space $N^{\{0,1\}}$ and work, reward and transition parameters defined by the transformations (17) and (20).

In the sequel we will assume that such transformations have been carried out, if required, focusing attention on the normalized case where all states are controllable.

3. Parametric LP Formulation and Simplex Tableau

We set out in this section to formulate the ν-wage problem (6) as a parametric LP problem, and to elucidate the structure of its simplex tableaux.

3.1 Bellman Equations and Parametric LP Formulation

While the LP formulation of concern is well-known in SMDP theory (see, e.g., Puterman (1994)), for ease of reference we outline next its derivation, starting from the Bellman equations for (6) in (7). The primal LP

formulation of such DP equations is

$$\begin{aligned}\vartheta^*(\nu) = &\min \sum_{j\in N} p_j \vartheta_j \\ &\text{subject to} \\ &x_i^a:\ \vartheta_i - \sum_{j\in N} \phi_{ij}^a \vartheta_j \geq r_i^a - \nu q_i^a, \quad (i,a) \in N\times\{0,1\},\end{aligned}$$

where $\mathbf{p} = (p_j)_{j\in N}$ represents the initial-state probability vector. It is well known that, if $\mathbf{p} > \mathbf{0}$ componentwise, such an LP has a unique solution that solves the DP equations.

Our analyses will be based instead on the dual standard-form LP,

$$\begin{aligned}\vartheta^*(\nu) = &\max \sum_{(j,a)\in N\times\{0,1\}} \left(r_j^a - \nu q_j^a\right) x_j^a \\ &\text{subject to} \\ &\vartheta_j:\ \sum_{a\in\{0,1\}} \left\{x_j^a - \sum_{j\in N} \phi_{ij}^a x_i^a\right\} = p_j, \quad j\in N \\ &x_j^a \geq 0, \quad (j,a)\in N\times\{0,1\}.\end{aligned}$$

We will work with the latter using matrix notation, writing

$$\begin{aligned}\vartheta^*(\nu) = &\max (\mathbf{r}^0 - \nu\mathbf{q}^0)\mathbf{x}^0 + (\mathbf{r}^1 - \nu\mathbf{q}^1)\mathbf{x}^1 \\ &\text{subject to} \\ &\left[\left(\mathbf{I} - \Phi^0\right)^{\mathsf{T}} \quad \left(\mathbf{I} - \Phi^1\right)^{\mathsf{T}}\right] \begin{bmatrix}\mathbf{x}^0 \\ \mathbf{x}^1\end{bmatrix} = \mathbf{p} \\ &\mathbf{x}^0, \mathbf{x}^1 \geq \mathbf{0},\end{aligned} \tag{22}$$

where $\mathbf{x}^a = (x_j^a)$ is a column vector, $\mathbf{r}^a = (r_j^a)$ and $\mathbf{q}^a = (q_j^a)$ are row vectors, and $^{\mathsf{T}}$ is the transposition operator .

Dual variables x_j^a correspond to the bandit's *discounted state-action occupancy measures*. For an admissible policy π, initial state i, action a and state j, let

$$x_{ij}^{a,\pi} \triangleq \mathbb{E}_i^\pi \left[\sum_{k=0}^{\infty} 1_{\{a(t_k)=a, X(t_k)=j\}} e^{-\alpha t_k}\right]$$

be the expected total discounted number of (j,a)-stages under policy π, starting at i. Thus, under initial state distribution $\mathbf{p}$, dual variable x_j^a corresponds to occupancy measure $x_j^{a,\pi} \triangleq \sum_i p_i x_{ij}^{a,\pi}$. Notice that reward and work measures are linear functions of occupancies: writing $\mathbf{x}^{a,\pi} = (x_j^{a,\pi})$,

$$\begin{aligned}f^\pi &= \sum_{(j,a)\in\{0,1\}\times N} r_j^a x_j^{a,\pi} = \mathbf{r}^0\mathbf{x}^{0,\pi} + \mathbf{r}^1\mathbf{x}^{1,\pi} \\ g^\pi &= \sum_{(j,a)\in\{0,1\}\times N} q_j^a x_j^{a,\pi} = \mathbf{q}^0\mathbf{x}^{0,\pi} + \mathbf{q}^1\mathbf{x}^{1,\pi}.\end{aligned} \tag{23}$$

3.2 Basic Feasible Solutions and Reduced Costs

We set out next to analyze parametric LP (22), starting with an elucidation of its *basic feasible solutions* (BFS). Clearly, these correspond to active sets $S \subseteq N^{\{0,1\}}$, and hence we will refer to the *S-active BFS*. For each such an S, we decompose the above vectors and matrices as

$$\mathbf{x}^a = \begin{bmatrix} \mathbf{x}^a_S \\ \mathbf{x}^a_{S^c} \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} \mathbf{p}_S \\ \mathbf{p}_{S^c} \end{bmatrix}, \quad \mathbf{\Phi}^a = \begin{bmatrix} \mathbf{\Phi}^a_{SS} & \mathbf{\Phi}^a_{SS^c} \\ \mathbf{\Phi}^a_{S^cS} & \mathbf{\Phi}^a_{S^cS^c} \end{bmatrix}, \quad \mathbf{I} = \begin{bmatrix} \mathbf{I}_{SS} & \mathbf{0}_{SS^c} \\ \mathbf{0}_{S^cS} & \mathbf{I}_{S^cS^c} \end{bmatrix},$$

where we write $S^c \triangleq N \setminus S$, and introduce the matrices

$$\begin{aligned} &\mathbf{\Phi}^S \triangleq \begin{bmatrix} \mathbf{\Phi}^1_{SS} & \mathbf{\Phi}^1_{S,N\setminus S} \\ \mathbf{\Phi}^0_{S^cS} & \mathbf{\Phi}^0_{S^cS^c} \end{bmatrix}, \quad \mathbf{\Phi}^{S^c} \triangleq \begin{bmatrix} \mathbf{\Phi}^0_{SS} & \mathbf{\Phi}^0_{SS^c} \\ \mathbf{\Phi}^1_{S^cS} & \mathbf{\Phi}^1_{S^cS^c} \end{bmatrix}, \\ &\mathbf{B}^S \triangleq \left(\mathbf{I} - \mathbf{\Phi}^S\right)^\top, \quad \mathbf{N}^S \triangleq \left(\mathbf{I} - \mathbf{\Phi}^{S^c}\right)^\top, \quad \mathbf{H}^S \triangleq \left(\mathbf{B}^S\right)^{-1}, \quad \mathbf{A}^S \triangleq \mathbf{H}^S\mathbf{N}^S. \end{aligned} \tag{24}$$

Notice that $\mathbf{\Phi}^S$ is the transition transform matrix under the S-active policy. Further, $\mathbf{B}^S$ is the *basis matrix* in LP (22) for the S-active BFS, whose *basic variables* are

$$\begin{bmatrix} \mathbf{x}^1_S \\ \mathbf{x}^0_{S^c} \end{bmatrix};$$

and $\mathbf{N}^S$ is the matrix of non-basic columns in LP (22), whose associated *non-basic variables* are

$$\begin{bmatrix} \mathbf{x}^0_S \\ \mathbf{x}^1_{S^c} \end{bmatrix}.$$

We thus rearrange the constraints in LP (22), decomposing them into basic and non-basic parts, as

$$\mathbf{B}^S \begin{bmatrix} \mathbf{x}^1_S \\ \mathbf{x}^0_{S^c} \end{bmatrix} + \mathbf{N}^S \begin{bmatrix} \mathbf{x}^0_S \\ \mathbf{x}^1_{S^c} \end{bmatrix} = \mathbf{p}.$$

We next draw on the above to evaluate performance measures under the S-active policy/BFS. The notation $x^{a,S}_j$ below refers to occupancy measure $x^{a,\pi}_j$ under the S-active policy, i.e., for $\pi = S$. Further, $\mathbf{g}^S = (g^S_j)_{j\in N}$, $\mathbf{f}^S = (f^S_j)_{j\in N}$, $\mathbf{w}^S = (w^S_j)_{j\in N}$ and $\mathbf{d}^S = (d^S_j)_{j\in N}$ are taken to be *row* vectors.

Lemma 3.1

(a) $\begin{bmatrix} \mathbf{x}^{0,S}_S \\ \mathbf{x}^{1,S}_{S^c} \end{bmatrix} = \mathbf{0}$ *and* $\begin{bmatrix} \mathbf{x}^{1,S}_S \\ \mathbf{x}^{0,S}_{S^c} \end{bmatrix} = \mathbf{H}^S\mathbf{p}$.

(b) $\mathbf{g}^S = \begin{bmatrix} \mathbf{q}^1_S & \mathbf{q}^0_{S^c} \end{bmatrix} \mathbf{H}^S$.

(c) $\mathbf{f}^S = \begin{bmatrix} \mathbf{r}^1_S & \mathbf{r}^0_{S^c} \end{bmatrix} \mathbf{H}^S$.

(d) $\begin{bmatrix} \mathbf{w}^S_S & -\mathbf{w}^S_{S^c} \end{bmatrix} = \begin{bmatrix} \mathbf{q}^1_S & \mathbf{q}^0_{S^c} \end{bmatrix} \mathbf{A}^S - \begin{bmatrix} \mathbf{q}^0_S & \mathbf{q}^1_{S^c} \end{bmatrix}$.

(e) $\begin{bmatrix} \mathbf{d}^S_S & -\mathbf{d}^S_{S^c} \end{bmatrix} = \begin{bmatrix} \mathbf{r}^1_S & \mathbf{r}^0_{S^c} \end{bmatrix} \mathbf{A}^S - \begin{bmatrix} \mathbf{r}^0_S & \mathbf{r}^1_{S^c} \end{bmatrix}$.

Proof. (a) Set to zero non-basic variables: $\mathbf{x}_S^{0,S} = \mathbf{0}$ and $\mathbf{x}_{S^c}^{1,S} = \mathbf{0}$. Calculate basic variables from

$$\mathbf{B}^S \begin{bmatrix} \mathbf{x}_S^1 \\ \mathbf{x}_{S^c}^0 \end{bmatrix} = \mathbf{p} \Longrightarrow \begin{bmatrix} \mathbf{x}_S^{1,S} \\ \mathbf{x}_{S^c}^{0,S} \end{bmatrix} = \mathbf{H}^S \mathbf{p}.$$

(b) Use part (a) taking $\mathbf{p} = \mathbf{e}_j$ (the unit coordinate vector having the one in the position of state j) to represent work measures as

$$g_j^S = \begin{bmatrix} \mathbf{q}_S^1 & \mathbf{q}_{S^c}^0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_S^{1,S} \\ \mathbf{x}_{S^c}^{0,S} \end{bmatrix} = \begin{bmatrix} \mathbf{q}_S^1 & \mathbf{q}_{S^c}^0 \end{bmatrix} \mathbf{H}^S \mathbf{e}_j \Longrightarrow \mathbf{g}^S = \begin{bmatrix} \mathbf{q}_S^1 & \mathbf{q}_{S^c}^0 \end{bmatrix} \mathbf{H}^S.$$

(c) Proceed as in part (b) to represent reward measures as

$$f_j^S = \begin{bmatrix} \mathbf{r}_S^1 & \mathbf{r}_{S^c}^0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_S^{1,S} \\ \mathbf{x}_{S^c}^{0,S} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_S^1 & \mathbf{r}_{S^c}^0 \end{bmatrix} \mathbf{H}^S \mathbf{e}_j \Longrightarrow \mathbf{f}^S = \begin{bmatrix} \mathbf{r}_S^1 & \mathbf{r}_{S^c}^0 \end{bmatrix} \mathbf{H}^S.$$

(d) Represent marginal work measures (cf. (12)) as

$$\mathbf{w}_S^S = \mathbf{g}_S^S - \mathbf{q}_S^0 - \mathbf{g}^S \left(\Phi_{SN}^0\right)^\top \quad \text{and} \quad \mathbf{w}_{S^c}^S = \mathbf{q}_{S^c}^1 + \mathbf{g}^S \left(\Phi_{S^cN}^1\right)^\top - \mathbf{g}_{S^c}^S. \tag{25}$$

Reformulate now the identities in (25), using part (b), as

$$\begin{bmatrix} \mathbf{w}_S^S & -\mathbf{w}_{S^c}^S \end{bmatrix} = \mathbf{g}^S \mathbf{N}^S - \begin{bmatrix} \mathbf{q}_S^0 & \mathbf{q}_{S^c}^1 \end{bmatrix} = \begin{bmatrix} \mathbf{q}_S^1 & \mathbf{q}_{S^c}^0 \end{bmatrix} \mathbf{H}^S \mathbf{N}^S - \begin{bmatrix} \mathbf{q}_S^0 & \mathbf{q}_{S^c}^1 \end{bmatrix} = \begin{bmatrix} \mathbf{q}_S^1 & \mathbf{q}_{S^c}^0 \end{bmatrix} \mathbf{A}^S - \begin{bmatrix} \mathbf{q}_S^0 & \mathbf{q}_{S^c}^1 \end{bmatrix}.$$

(e) This part follows along the lines of part (d). □

The next result characterizes the marginal work and reward measures in (12)–(13) as *reduced costs* of LP problems. It further gives the reduced costs of parametric LP (22), and uses such results to obtain corresponding representations of the LPs objectives in terms of such reduced costs.

Lemma 3.2

(a) *The reduced costs for non-basic variables in the S-active BFS for LP*

$$\max \left\{ \mathbf{q}^0 \mathbf{x}^0 + \mathbf{q}^1 \mathbf{x}^1 : \begin{bmatrix} \left(\mathbf{I} - \Phi^0\right)^\top & \left(\mathbf{I} - \Phi^1\right)^\top \end{bmatrix} \begin{bmatrix} \mathbf{x}^0 \\ \mathbf{x}^1 \end{bmatrix} = \mathbf{p}, \quad \mathbf{x}^0, \mathbf{x}^1 \geq \mathbf{0} \right\}$$

are given in the left-hand side of Lemma 3.1(d)*. The LP's objective can thus be expressed as*

$$\sum_{(j,a) \in N \times \{0,1\}} q_j^a x_j^a = g^S - \sum_{j \in S} w_j^S x_j^0 + \sum_{j \in S^c} w_j^S x_j^1. \tag{26}$$

(b) *The reduced costs for non-basic variables in the S-active BFS for LP*

$$\max \left\{ \mathbf{r}^0 \mathbf{x}^0 + \mathbf{r}^1 \mathbf{x}^1 : \begin{bmatrix} \left(\mathbf{I} - \Phi^0\right)^\top & \left(\mathbf{I} - \Phi^1\right)^\top \end{bmatrix} \begin{bmatrix} \mathbf{x}^0 \\ \mathbf{x}^1 \end{bmatrix} = \mathbf{p}, \quad \mathbf{x}^0, \mathbf{x}^1 \geq \mathbf{0} \right\}$$

are given in the left-hand side of Lemma 3.1(e)*. The LP's objective can thus be expressed as*

$$\sum_{(j,a) \in N \times \{0,1\}} r_j^a x_j^a = f^S - \sum_{j \in S} d_j^S x_j^0 + \sum_{j \in S^c} d_j^S x_j^1. \tag{27}$$

(c) *The reduced costs for non-basic variables in the S-active BFS for LP* (22) *are given by*

$$\left[\mathbf{d}_S^S - \nu \mathbf{w}_S^S \quad -\mathbf{d}_{S^c}^S + \nu \mathbf{w}_{S^c}^S\right]. \tag{28}$$

Therefore, the LP's objective can be represented as

$$\sum_{(j,a)\in N\times\{0,1\}} (r_j^a - \nu q_j^a) x_j^a = f^S - \nu g^S - \sum_{j\in S} (d_j^S - \nu w_j^S) x_j^0 + \sum_{j\in S^c} (d_j^S - \nu w_j^S) x_j^1. \tag{29}$$

Proof. The results follow directly from the standard representation of reduced costs in LP theory, as given by Lemma 3.1(d,e), along with the standard representation of the LP's objective in terms of the current BFS' value and reduced costs. We have further used (14). □

The next result, which follows directly from Lemma 3.2, gives representations of measures g^π, f^π and objective $f^\pi - \nu g^\pi$ relative to the S-active policy. We first derived such *decomposition* identities in Niño-Mora (2001, 2002) through ad hoc algebraic arguments.

Lemma 3.3 *Under any policy* $\pi \in \Pi$:

(a) $g^\pi = g^S - \sum_{j\in S} w_j^S x_j^{0,\pi} + \sum_{j\in S^c} w_j^S x_j^{1,\pi}$.

(b) $f^\pi = f^S - \sum_{j\in S} d_j^S x_j^{0,\pi} + \sum_{j\in S^c} d_j^S x_j^{1,\pi}$.

(c) $f^\pi - \nu g^\pi = f^S - \nu g^S - \sum_{j\in S} (d_j^S - \nu w_j^S) x_j^{0,\pi} + \sum_{j\in S^c} (d_j^S - \nu w_j^S) x_j^{1,\pi}$.

The following result, first established in Niño-Mora (2002), clarifies the relation between work and reward measures and their marginal counterparts. We will use it later to prove Lemma 4.1.

Lemma 3.4

(a) *For* $j \in S^c$, $g^{S\cup\{j\}} = g^S + w_j^S x_j^{1,S\cup\{j\}}$ *and* $f^{S\cup\{j\}} = f^S + d_j^S x_j^{1,S\cup\{j\}}$.

(b) *For* $j \in S$, $g^{S\setminus\{j\}} = g^S - w_j^S x_j^{1,S\setminus\{j\}}$ *and* $f^{S\setminus\{j\}} = f^S - d_j^S x_j^{1,S\setminus\{j\}}$.

Proof. To obtain (a) (resp. (b)) use $\pi = S\cup\{j\}$ (resp. $\pi = S\setminus\{j\}$) in Lemma 3.3(a, b). □

Table 1: Parametric Simplex Tableau for S-Active BFS, Ready for Pivoting on a^S_{jj}.

	$(\mathbf{x}^0_S)^\top$	x^1_j	$(\mathbf{x}^1_{S^c\setminus\{j\}})^\top$
$\mathbf{x}^1_S$	$\mathbf{A}^S_{SS}$	$\mathbf{A}^S_{Sj}$	$\mathbf{A}^S_{S,S^c\setminus\{j\}}$
x^0_j	$\mathbf{A}^S_{jS}$	$\boxed{a^S_{jj}}$	$\mathbf{A}_{j,S^c\setminus\{j\}}$
$\mathbf{x}^0_{S^c\setminus\{j\}}$	$\mathbf{A}^S_{S^c\setminus\{j\},S}$	$\mathbf{A}^S_{S^c\setminus\{j\},j}$	$\mathbf{A}^S_{S^c\setminus\{j\},S^c\setminus\{j\}}$
	$\mathbf{w}^S_S$	$-w^S_j$	$-\mathbf{w}^S_{S^c\setminus\{j\}}$
	$\mathbf{d}^S_S$	$-d^S_j$	$-\mathbf{r}^S_{S^c\setminus\{j\}}$

Table 2: Tableau for $S\cup\{j\}$-Active BFS, Obtained by Pivoting on a^S_{jj}.

	$(\mathbf{x}^0_S)^\top$	x^0_j	$(\mathbf{x}^1_{S^c\setminus\{j\}})^\top$
$\mathbf{x}^1_S$	$\mathbf{A}^S_{SS}-\frac{\mathbf{A}^S_{Sj}\mathbf{A}^S_{jS}}{a^S_{jj}}$	$-\frac{\mathbf{A}^S_{Sj}}{a^S_{jj}}$	$\mathbf{A}^S_{S,S^c\setminus\{j\}}-\frac{\mathbf{A}^S_{Sj}\mathbf{A}^S_{j,S^c\setminus\{j\}}}{a^S_{jj}}$
x^1_j	$\frac{\mathbf{A}^S_{jS}}{a^S_{jj}}$	$\frac{1}{a^S_{jj}}$	$\frac{\mathbf{A}^S_{j,S^c\setminus\{j\}}}{a^S_{jj}}$
$\mathbf{x}^0_{S^c\setminus\{j\}}$	$\mathbf{A}^S_{S^c\setminus\{j\},S}-\frac{\mathbf{A}^S_{S^c\setminus\{j\},j}\mathbf{A}^S_{jS}}{a^S_{jj}}$	$-\frac{\mathbf{A}^S_{S^c\setminus\{j\},j}}{a^S_{jj}}$	$\mathbf{A}^S_{S^c\setminus\{j\},S^c\setminus\{j\}}-\frac{\mathbf{A}^S_{S^c\setminus\{j\},j}\mathbf{A}^S_{j,S^c\setminus\{j\}}}{a^S_{jj}}$
	$\mathbf{w}^S_S+\frac{w^S_j}{a^S_{jj}}\mathbf{A}^S_{jS}$	$\frac{w^S_j}{a^S_{jj}}$	$-\mathbf{w}^S_{S^c\setminus\{j\}}+\frac{w^S_j}{a^S_{jj}}\mathbf{A}^S_{j,S^c\setminus\{j\}}$
	$\mathbf{d}^S_S+\frac{d^S_j}{a^S_{jj}}\mathbf{A}^S_{jS}$	$\frac{d^S_j}{a^S_{jj}}$	$-\mathbf{d}^S_{S^c\setminus\{j\}}+\frac{d^S_j}{a^S_{jj}}\mathbf{A}^S_{j,S^c\setminus\{j\}}$

3.3 Parametric Simplex Tableau and Pivoting

We can now formulate the parametric simplex tableau under the S-active BFS, as shown in Table 1. The tableau is indexed by basic variables $\mathbf{x}^1_S$ and $\mathbf{x}^0_{S^c}$ in rows, and by nonbasic variables $\mathbf{x}^0_S$ and $\mathbf{x}^1_{S^c}$ in columns, and includes two rows of reduced costs for non-basic variables. It includes neither the conventional right-hand side nor the objective value, as they are not needed for our purposes.

Notice that the tableau is shown in a form that highlights its structure as it is ready for *pivoting* on element a^S_{jj}, with $j\in S^c$. Namely, for taking variable x^0_j *out* of the basis, and putting x^1_j *into* the basis, which corresponds to moving from the S-active to the $S\cup\{j\}$-active BFS. After such a pivot step is carried out, one obtains the updated tableau shown in Table 2.

3.4 Computing the Initial Tableau

We discuss next how to compute the initial tableau, corresponding to the $\emptyset$-active BFS, in a numerically-stable form that applies both to the discounted criterion of concern heretofore, and to the long-run average criterion to be addressed in Section 6 below. The time-average tableaux arise as limits of the discounted tableaux as the discount rate α vanishes.

Notice that (cf. (24))

$$\mathbf{B}^{\emptyset} = \big(\mathbf{I} - \Phi^0\big)^{\mathsf{T}}, \quad \mathbf{N}^{\emptyset} = \big(\mathbf{I} - \Phi^1\big)^{\mathsf{T}}, \quad \mathbf{H}^{\emptyset} = \big(\mathbf{B}^{\emptyset}\big)^{-1}, \quad \mathbf{A}^{\emptyset} = \mathbf{H}^{\emptyset}\mathbf{N}^{\emptyset}. \tag{30}$$

Hence, the direct approach to compute $\mathbf{A}^{\emptyset}$ would be to solve the linear equation system

$$\big(\mathbf{A}^{\emptyset}\big)^{\mathsf{T}}\big(\mathbf{I} - \Phi^0\big) = \big(\mathbf{I} - \Phi^1\big). \tag{31}$$

Yet, this has a major drawback: as the discount rate α vanishes, matrices $\mathbf{I} - \Phi^a$ become increasingly ill-conditioned, being singular for $\alpha = 0$ — as they converge to $\mathbf{I} - \mathbf{P}^a$, where $\mathbf{P}^a \triangleq (p^a_{ij})$.

To avoid such a difficulty, we will use the identity $(\mathbf{I} - \Phi^a)\mathbf{1} = \mathbf{1} - \phi^a$, which follows from (1). From this and (30) we obtain

$$\big(\mathbf{A}^{\emptyset}\big)^{\mathsf{T}}\big(\mathbf{1} - \phi^0\big) = \mathbf{1} - \phi^1.$$

The latter identity has the advantage that it yields a corresponding identity in the limit $\alpha \searrow 0$. Thus, denoting by ξ^a_i the duration of an (i,a)-stage (cf. Section 2.1), and using the McLaurin expansion

$$\phi^a_i = \mathbb{E}\left[e^{-\alpha\xi^a_i}\right] = 1 - \alpha m^a_i + O(\alpha^2), \quad \text{as } \alpha \searrow 0,$$

we obtain the limiting identity

$$\big(\mathbf{A}^{\emptyset}\big)^{\mathsf{T}}\mathbf{m}^0 = \mathbf{m}^1,$$

where m^a_i is the mean duration of an (i,a)-stage and $\mathbf{m} = (m^a_i)_{i\in N}$.

We are thus led to the following numerically-stable approach to compute the initial tableau, for $\alpha \geq 0$ — where the case $\alpha = 0$ corresponds to the limiting tableau obtained as α vanishes. Letting

$$\widetilde{m}^a_i \triangleq \begin{cases} (1 - \phi^a_i)/\alpha & \text{if } \alpha > 0 \\ m^a_i & \text{if } \alpha = 0, \end{cases}$$

and $\widetilde{\mathbf{m}}^a = (\widetilde{m}^a_i)_{i\in N}$, choose an arbitrary state $j^* \in N$, and solve the *block linear system* (cf. Baker et al. (2006))

$$\Big[\mathbf{I}_{N,N\setminus\{j^*\}} - \Phi^0_{N,N\setminus\{j^*\}} \quad \widetilde{\mathbf{m}}^0\Big]^{\mathsf{T}}\mathbf{A}^{\emptyset} = \Big[\mathbf{I}_{N,N\setminus\{j^*\}} - \Phi^1_{N,N\setminus\{j^*\}} \quad \widetilde{\mathbf{m}}^1\Big]^{\mathsf{T}} \tag{32}$$

to obtain $\mathbf{A}^{\emptyset}$. Then, compute the initial reduced costs from (30) and Lemma 3.1(d, e):

$$\begin{aligned} \mathbf{w}^{\emptyset} &= \mathbf{q}^1 - \mathbf{q}^0\mathbf{A}^{\emptyset} \\ \mathbf{d}^{\emptyset} &= \mathbf{r}^1 - \mathbf{r}^0\mathbf{A}^{\emptyset}. \end{aligned} \tag{33}$$

4. Simplex-Based Characterization of Indexability

This section draws on the above results, and on the classic parametric-objective LP theory in Gass and Saaty (1955); Saaty and Gass (1954), adapted to the present setting, to develop a simplex-based characterization of indexability. In what follows, $S \subseteq N$ denotes an arbitrary active set.

4.1 Optimality Conditions for a BFS

We start by addressing the following question: For which range of values of the wage ν is the S-active BFS optimal for parametric LP (22)? Or, equivalently: For which range of values of the wage ν is the S-active policy optimal for ν-wage problem (6)? Though the answer is well-known in general from parametric LP theory to be given by the so-called *characteristic interval* of such a BFS, we next elucidate it in the present context.

Since the signs of marginal work and reward measures will play a key role in the answer, we next clarify the meaning of such signs in terms of work and reward measures.

Lemma 4.1

(a) *For* $j \in S^c$:

$$\begin{aligned} w_j^S > 0 &\Longleftrightarrow g_j^{S\cup\{j\}} > g_j^S \quad \textit{and} \quad d_j^S > 0 \Longleftrightarrow f_j^{S\cup\{j\}} > f_j^S \\ w_j^S < 0 &\Longleftrightarrow g_j^{S\cup\{j\}} < g_j^S \quad \textit{and} \quad d_j^S < 0 \Longleftrightarrow f_j^{S\cup\{j\}} < f_j^S \\ w_j^S = 0 &\Longleftrightarrow g_j^{S\cup\{j\}} = g_j^S \quad \textit{and} \quad d_j^S = 0 \Longleftrightarrow f_j^{S\cup\{j\}} = f_j^S. \end{aligned}$$

(b) *For* $j \in S$:

$$\begin{aligned} w_j^S > 0 &\Longleftrightarrow g_j^{S\setminus\{j\}} < g_j^S \quad \textit{and} \quad d_j^S > 0 \Longleftrightarrow f_j^{S\setminus\{j\}} < f_j^S \\ w_j^S < 0 &\Longleftrightarrow g_j^{S\setminus\{j\}} > g_j^S \quad \textit{and} \quad d_j^S < 0 \Longleftrightarrow f_j^{S\setminus\{j\}} > f_j^S \\ w_j^S = 0 &\Longleftrightarrow g_j^{S\setminus\{j\}} = g_j^S \quad \textit{and} \quad d_j^S = 0 \Longleftrightarrow f_j^{S\setminus\{j\}} = f_j^S. \end{aligned}$$

Proof. Part (a) follows from Lemma 3.4(a) taking $i = j$ and noting that $x_{jj}^{1,S\cup\{j\}} > 0$.

Part (b) follows from Lemma 3.4(b) taking $i = j$ and noting that $x_{jj}^{0,S\setminus\{j\}} > 0$. □

Thus, e.g., the condition $w_j^S > 0$ for some $j \in S^c$ means that expanding the active set from S to $S \cup \{j\}$ increases the work expended starting at j. Similarly, the condition $w_j^S > 0$ for some $j \in S$ means that shrinking the active set from S to $S \setminus \{j\}$ decreases the work expended.

We next use the characterization of reduced costs in Lemma 3.2(c) to give a necessary and sufficient optimality test for the S-active BFS in parametric LP (22), and hence for the S-active policy in ν-wage problem (6).

In the following result we assume that $\mathbf{p} > \mathbf{0}$ in LP (22). Note that its part (a) gives the characteristic or optimality interval for the S-active BFS, having lower and upper breakpoints

$$\underline{\nu}^S \triangleq \max_{j \in S^c, w_j^S > 0 \text{ or } j \in S, w_j^S < 0} \nu_j^S \quad \text{and} \quad \overline{\nu}^S \triangleq \min_{j \in S, w_j^S > 0 \text{ or } j \in S^c, w_j^S < 0} \nu_j^S, \tag{34}$$

respectively, while part (b) refers to concepts discussed at the end of Section 2.2. We further write

$$\underline{d}^S \triangleq \max_{j \in S^c, w_j^S = 0} d_j^S \quad \text{and} \quad \overline{d}^S \triangleq \min_{j \in S, w_j^S = 0} d_j^S. \tag{35}$$

Here and below we adopt the convention that the maximum (resp. minimum) over an empty set has the value $-\infty$ (resp. $+\infty$).

Lemma 4.2

(a) *The S-active BFS is optimal for LP* (22) *iff*

$$\underline{\nu}^S \leq \nu \leq \overline{\nu}^S, \tag{36}$$

and

$$\underline{d}^S \leq 0 \leq \overline{d}^S. \tag{37}$$

Further, it is the unique optimal solution iff the inequalities in (36)–(37) *hold strictly.*

(b) *The deterministic stationary policies determining the upper boundary $\overline{\partial}\mathbb{H}$ of the achievable work-reward performance region $\mathbb{H}$ are those with active sets $S \subseteq N$ satisfying* (37) *and*

$$\underline{\nu}^S \leq \overline{\nu}^S. \tag{38}$$

Proof. (a) The "if" part follows from the LP sufficient optimality condition given by nonnegativity of reduced costs for non-basic variables. The inequalities in (36) follow by reformulating such a condition, using Lemma 3.2(c) and (14), in terms of the marginal productivity measures ν_j^S in (15).

The "only if" part follows by considering LP (22). From the latter's MDP interpretation and the assumption $\mathbf{p} > \mathbf{0}$ it immediately follows that such an LP is *nondegenerate*, i.e., for any BFS, basic variables take positive values, and hence the LP optimality condition is also necessary.

The uniqueness result follows by invoking the result that, for a nondegenerate LP, an optimal BFS is the unique optimal solution iff the reduced costs of its non-basic variables are positive.

(b) The stationary deterministic policies determining the upper boundary $\overline{\partial}\mathbb{H}$ are those that are optimal for the ν-wage problem, hence for LP (22), for some wage value $\nu \in \mathbb{R}$. Therefore, by part (a), such sets are precisely those satisfying (37)–(38). □

4.2 Indexability Characterization and the CPI Algorithm

We next proceed to put together the above elements to give a complete characterization of indexability, both in combinatorial and algorithmic terms. We will refer to the *Complete-Pivoting Indexability* (CPI) algorithm described in Table 3. To avoid an unwieldy notation, we have used there a more algorithm-like notation, replacing superscript sets by numeric superscripts, e.g., writing $a_{ij}^{(k)}$ instead of $a_{ij}^{S_k}$. The algorithm seeks to construct a state ordering $i_1, \ldots, i_n$ relative to which the bandit is indexable (cf. Definition 2.1), with MPI values $\nu_{i_k}^*$ and active sets S_k as in Section 2.2, in which case the Boolean variable INDEXABLE returns the value `true`. It adapts to the present setting the *parametric-objective simplex algorithm* of Gass and Saaty (1955), letting the wage ν decrease from $+\infty$ to $-\infty$, and draws on Lemma 4.2 to test for the structure of successive optimal bases that ensures indexability. For moving from one basis to the next, the algorithm updates the tableau performing a complete simplex pivot step (cf. Table 2), hence its name.

The following result gives a complete, combinatorial characterization of indexability in terms of properties of active sets S.

Theorem 4.3 *The bandit is indexable iff* $\underline{d}^{\emptyset} \leq 0 \leq \overline{d}^{N}$ *and, for any active set* $S \subseteq N$ *satisfying* (37)–(38)*, it holds that*

$$\begin{aligned} \underline{\nu}^N &= -\infty, \quad \overline{\nu}^{\emptyset} = +\infty \\ \underline{\nu}^S &= \max_{j \in S^c:\, w_j^S > 0} \nu_j^S > -\infty, \quad \text{if } S \neq N \\ \overline{\nu}^S &= \min_{j \in S:\, w_j^S > 0} \nu_j^S < +\infty, \quad \text{if } S \neq \emptyset. \end{aligned} \tag{39}$$

Proof. Consider the "if" part. Under the corresponding assumptions, the reader can easily verify that, by construction, the CPI algorithm will terminate in n steps with Boolean variable INDEXABLE returning the value `true`. The algorithm hence constructs a state ordering $i_1, \ldots, i_n$ relative to which the bandit is indexable, as it satisfies the requirements in Definition 2.1.

Consider now the "only if" part. Suppose thus that the bandit is indexable, and let $\emptyset \subset S \subset N$ be an active set satisfying (37)–(38). Hence, by Lemma 4.2, the S-active policy is optimal for ν-wage problem (6) iff ν lies in the interval $[\underline{\nu}^S, \overline{\nu}^S]$. If we let the wage ν drop at or below the lower breakpoint $\underline{\nu}^S$, by indexability it must be possible to pivot to a new *expanded* active set of the form $S \cup \{j\}$, for some $j \in S^c$, which is optimal for an adjacent interval of ν values. Using Lemma 3.2(c), such a requirement is readily formulated as the second line in (39).

Table 3: The Complete-Pivoting Indexability (CPI) Algorithm.

solve $\left[\mathbf{I}_{N,N\setminus\{j^*\}} - \Phi^0_{N,N\setminus\{j^*\}} \quad \widetilde{\mathbf{m}}^0_N\right]^{\mathsf{T}} \mathbf{A}^{(0)} = \left[\mathbf{I}_{N,N\setminus\{j^*\}} - \Phi^1_{N,N\setminus\{j^*\}} \quad \widetilde{\mathbf{m}}^1\right]^{\mathsf{T}}$

$\begin{bmatrix}\mathbf{w}^{(0)}\\ \mathbf{d}^{(0)}\end{bmatrix} := \begin{bmatrix}\mathbf{q}^1\\ \mathbf{r}^1\end{bmatrix} - \begin{bmatrix}\mathbf{q}^0\\ \mathbf{r}^0\end{bmatrix}\mathbf{A}^{(0)}$; $S_0 := \emptyset$; $k := 1$; INDEXABLE := `true`

if $\max_{j\in N} w_j^{(0)} \le 0$ **or** $\min_{j\in N} w_j^{(0)} < 0$ **or** $\max_{j\in N:\, w_j^{(0)}=0} d_j^{(0)} > 0$, INDEXABLE := `false`

while INDEXABLE **and** $k \le n$ **do**

$v_j^{(k-1)} := d_j^{(k-1)}/w_j^{(k-1)}$, **for** $j \in S_{k-1}^c, w_j^{(k-1)} > 0$ **and** $j \in S_{k-1}, w_j^{(k-1)} < 0$

pick $i_k \in \underset{j\in S_{k-1}^c, w_j^{(k-1)}>0}{\arg\max} v_j^{(k-1)}$; $v^*_{i_k} := v_{i_k}^{(k-1)}$; $S_k := S_{k-1}\cup\{i_k\}$

if $\max_{j\in S_{k-1}, w_j^{(k-1)}<0} v_j^{(k-1)} > v^*_{i_k}$, INDEXABLE := `false`

else if $k < n$

$p^{(k-1)} = 1/a_{i_k i_k}^{(k-1)}$; $\mathbf{y}^{(k-1)} := p^{(k-1)}\mathbf{A}_{Ni_k}^{(k-1)}$; $\mathbf{z}^{(k-1)} := \mathbf{A}_{i_k N}^{(k-1)}$

$$\begin{bmatrix}\mathbf{w}_{S_k}^{(k)} & -\mathbf{w}_{S_k^c}^{(k)}\\ \mathbf{d}_{S_k}^{(k)} & -\mathbf{d}_{S_k^c}^{(k)}\end{bmatrix} := \begin{bmatrix}\mathbf{w}_{S_{k-1}}^{(k-1)} & -\mathbf{w}_{S_{k-1}^c}^{(k-1)}\\ \mathbf{d}_{S_{k-1}}^{(k-1)} & -\mathbf{d}_{S_{k-1}^c}^{(k-1)}\end{bmatrix} + p^{(k-1)}\begin{bmatrix}w_{i_k}^{(k-1)}\\ d_{i_k}^{(k-1)}\end{bmatrix}\left\{\mathbf{A}_{i_k N}^{(k-1)} + \mathbf{e}_{i_k}^{\mathsf{T}}\right\}$$

$\mathbf{A}^{(k)} := \mathbf{A}^{(k-1)} - \mathbf{y}^{(k-1)}\mathbf{z}^{(k-1)}$

$\mathbf{A}_{Ni_k}^{(k)} := -\mathbf{y}^{(k-1)}$; $\mathbf{A}_{i_k N}^{(k)} := p^{(k-1)}\mathbf{z}^{(k-1)}$; $a_{i_k i_k}^{(k)} := p^{(k-1)}$

$\mathbf{A}^{(k)} := \mathbf{A}^{(k-1)} - p^{(k-1)}\left\{\mathbf{A}_{Ni_k}^{(k-1)}\mathbf{A}_{i_k N}^{(k-1)} + \mathbf{A}_{Ni_k}^{(k-1)}\mathbf{e}_{i_k}^{\mathsf{T}} - \mathbf{e}_{i_k}\mathbf{A}_{i_k N}^{(k-1)} - \mathbf{e}_{i_k}\mathbf{e}_{i_k}^{\mathsf{T}}\right\}$

if $\max_{j\in S_k^c} w_j^{(k)} \le 0$, INDEXABLE := `false`

end { if }

$k := k+1$

end { while }

if $k = n+1$ **and** $\{\max_{j\in N} w_j^{(n)} \le 0$ **or** $\min_{j\in N} w_j^{(n)} < 0\}$, INDEXABLE := `false`

Similarly, if we let the wage ν rise at or above the upper breakpoint $\overline{\nu}^S$, by indexability it must be possible to pivot to a new *shrinked* active set of the form $S \setminus \{j\}$, for some $j \in S$, which would be optimal for an adjacent interval of ν values. Using Lemma 3.2(c), such a requirement is formulated as the third line in (39).

The relations for the cases $S = \emptyset$ and $S = N$ follow along similar lines, as indexability implies that the $\emptyset$-active (resp. N-active) BFS must be optimal for ν large (resp. small) enough. □

Theorem 4.3 immediately yields the following algorithmic characterization of indexability.

Proposition 4.4 *The bandit is indexable iff algorithm CPI terminates in n steps with* INDEXABLE = `true`. *Then, the computed index ν_j^* is the bandit's MPI, and the following relations hold*:

$$\max_{j \in N:\, w_j^{S_0} > 0} \nu_j^{S_0} = \nu_{i_1}^{S_0} = \nu_{i_1}^*, \tag{40}$$

$$\nu_{i_n}^* = \nu_{i_n}^{S_n} = \min_{j \in N:\, w_j^{S_n} > 0} \nu_j^{S_n}, \tag{41}$$

and, for $2 \le k \le n$,

$$\max_{j \in S_{k-1}^c:\, w_j^{S_{k-1}} > 0} \nu_j^{S_{k-1}} = \nu_{i_k}^{S_{k-1}} = \nu_{i_k}^* \le \nu_{i_{k-1}}^* = \nu_{i_{k-1}}^{S_{k-1}} = \min_{j \in S_{k-1}:\, w_j^{S_{k-1}} > 0} \nu_j^{S_{k-1}}. \tag{42}$$

We next assess the computational complexity of the CPI algorithm's (while) loop, i.e., excluding the initialization stage. We use the term "arithmetic operations" to include both additions/subtractions and multiplications/divisions.

Proposition 4.5 *The CPI algorithm's loop performs at most $2n^3 + O(n^2)$ arithmetic operations.*

Proof. Observation of Table 3 shows that the more expensive operation at each step k is the matrix update $\mathbf{A}^{(k)} := \mathbf{A}^{(k-1)} - p^{(k-1)}\mathbf{y}^{(k-1)}\mathbf{z}^{(k-1)}$, which takes $2n^2$ arithmetic operations. Carrying out n steps yields the stated count. □

4.3 Reduced Tableaux and the RPI Algorithm

We seek next to eliminate unnecessary operations from the CPI algorithm. The key observation is that the tableau's rows corresponding to basic variables $\mathbf{x}_S^1$ are not used to update reduced costs in the CPI algorithm. Hence, it suffices to store and update only *reduced tableaux*, such as that shown in Table 4, which is set up for pivoting on element a_{jj}^S, for $j \in S^c$. Observation of Table 2 shows that a reduced tableau can be updated without using the deleted rows. Simplifying the CPI algorithm accordingly yields the *Reduced-Pivoting Indexability* (RPI) algorithm in Table 5.

As shown next, the RPI improves the operation count of the CPI algorithm by a factor of two.

Table 4: Reduced Tableau for S-Active BFS, Ready for Pivoting on a^S_{jj}.

	$\left(\mathbf{x}^0_S\right)^\mathsf{T}$	x^1_j	$\left(\mathbf{x}^1_{S^c\setminus\{j\}}\right)^\mathsf{T}$
x^0_j	$\mathbf{A}^S_{jS}$	a^S_{jj}	$\mathbf{A}^S_{j,S^c\setminus\{j\}}$
$\mathbf{x}^0_{S^c\setminus\{j\}}$	$\mathbf{A}^S_{S^c\setminus\{j\},S}$	$\mathbf{A}^S_{S^c\setminus\{j\},j}$	$\mathbf{A}^S_{S^c\setminus\{j\},S^c\setminus\{j\}}$
	$\mathbf{w}^S_S$	$-w^S_j$	$-\mathbf{w}^S_{S^c\setminus\{j\}}$
	$\mathbf{d}^S_S$	$-d^S_j$	$-\mathbf{d}^S_{S^c\setminus\{j\}}$

Table 5: The Reduced-Pivoting Indexability (RPI) Algorithm.

solve $\left[\mathbf{I}_{N,N\setminus\{j^*\}} - \mathbf{\Phi}^0_{N,N\setminus\{j^*\}} \quad \widetilde{\mathbf{m}}^0_N\right]^\mathsf{T} \mathbf{A}^{(0)} = \left[\mathbf{I}_{N,N\setminus\{j^*\}} - \mathbf{\Phi}^1_{N,N\setminus\{j^*\}} \quad \widetilde{\mathbf{m}}^1\right]^\mathsf{T}$

$\begin{bmatrix}\mathbf{w}^{(0)}\\ \mathbf{d}^{(0)}\end{bmatrix} := \begin{bmatrix}\mathbf{q}^1\\ \mathbf{r}^1\end{bmatrix} - \begin{bmatrix}\mathbf{q}^0\\ \mathbf{r}^0\end{bmatrix}\mathbf{A}^{(0)}$; $S_0 := \emptyset$; $k := 1$; INDEXABLE := `true`

if $\max_{j\in N} w^{(0)}_j \le 0$ **or** $\min_{j\in N} w^{(0)}_j < 0$ **or** $\max_{j\in N:\, w^{(0)}_j=0} d^{(0)}_j > 0$, INDEXABLE := `false`

while INDEXABLE **and** $k \le n$ **do**

$\nu^{(k-1)}_j := d^{(k-1)}_j / w^{(k-1)}_j$, **for** $j \in S^c_{k-1}, w^{(k-1)}_j > 0$ **and** $j \in S_{k-1}, w^{(k-1)}_j < 0$

pick $i_k \in \underset{j\in S^c_{k-1}, w^{(k-1)}_j>0}{\arg\max} \nu^{(k-1)}_j$; $\nu^*_{i_k} := \nu^{(k-1)}_{i_k}$; $S_k := S_{k-1}\cup\{i_k\}$

if $\max_{j\in S_{k-1}, w^{(k-1)}_j<0} \nu^{(k-1)}_j > \nu^*_{i_k}$, INDEXABLE := `false`

else if $k < n$

$p^{(k-1)} = 1/a^{(k-1)}_{i_k i_k}$; $\mathbf{y}^{(k-1)} := p^{(k-1)}\mathbf{A}^{(k-1)}_{S^c_k i_k}$; $\mathbf{z}^{(k-1)} := \mathbf{A}^{(k-1)}_{i_k N}$

$$\begin{bmatrix}\mathbf{w}^{(k)}_{S_k} & -\mathbf{w}^{(k)}_{S^c_k}\\ \mathbf{d}^{(k)}_{S_k} & -\mathbf{d}^{(k)}_{S^c_k}\end{bmatrix} := \begin{bmatrix}\mathbf{w}^{(k-1)}_{S_{k-1}} & -\mathbf{w}^{(k-1)}_{S^c_{k-1}}\\ \mathbf{d}^{(k-1)}_{S_{k-1}} & -\mathbf{d}^{(k-1)}_{S^c_{k-1}}\end{bmatrix} + p^{(k-1)}\begin{bmatrix}w^{(k-1)}_{i_k}\\ d^{(k-1)}_{i_k}\end{bmatrix}\left\{\mathbf{A}^{(k-1)}_{i_k N} + \mathbf{e}^\mathsf{T}_{i_k}\right\}$$

$\mathbf{A}^{(k)}_{S^c_k N} := \mathbf{A}^{(k-1)}_{S^c_k N} - \mathbf{y}^{(k-1)}\mathbf{z}^{(k-1)}$; $\mathbf{A}^{(k)}_{S^c_k i_k} := -\mathbf{y}^{(k-1)}$

if $\max_{j\in S^c_k} w^{(k)}_j \le 0$, INDEXABLE := `false`

end { if }

$k := k+1$

end { while }

if $k = n+1$ **and** $\{\max_{j\in N} w^{(n)}_j \le 0$ **or** $\min_{j\in N} w^{(n)}_j < 0\}$, INDEXABLE := `false`

Proposition 4.6 *The RPI algorithm's loop performs at most $n^3 + O(n^2)$ arithmetic operations.*

Proof. The loop's operation count is dominated by the matrix update $\mathbf{A}^{(k)}_{S_k^c N} := \mathbf{A}^{(k-1)}_{S_k^c N} - \mathbf{y}^{(k-1)}\mathbf{z}^{(k-1)}$ shown in Table 5, which takes $2(n-k)n$ arithmetic operations. Adding up such counts over $k = 1, \ldots, n$ yields the result. □

5. Exploiting Special Structure

We proceed to discuss how one can leverage structural knowledge on a particular bandit model to obtain substantially simpler indexability conditions and a faster index algorithm. While we had addressed such an issue in Niño-Mora (2001, 2002, 2006d), by introducing and deploying the class of PCL-indexable bandits, the approach and results herein are both new, as they draw on the above simplex-based analyses, and of wider applicability. In fact, we were motivated to develop them by the difficulties encountered when trying to deploy the PCL-indexability approach in the analysis of several complex bandit models. The new approach below was successful in such cases, yielding sound indexability analyses and new index algorithms in Niño-Mora (2006e, 2007a,d).

5.1 LP($\mathscr{F}$)-Indexable Bandits and the FPAG($\mathscr{F}$) Index Algorithm

When investigating a particular restless bandit model, one is concerned with identifying analytically a range of model parameters for which the model is indexable. Similarly as in the earlier work mentioned, our approach to establish a priori indexability of a bandit model is based on identifying the structure of optimal active sets for ν-wage problem (6), in the form of an *active-set family* $\mathscr{F} \subseteq 2^N$ that *contains* an optimal active set $S \in \mathscr{F}$ for every wage value $\nu \in \mathbb{R}$. Note that such an $\mathscr{F}$ need not be a nested family, but should contain the nested families $\mathscr{F}_0$ discussed in Section 2.2 that can arise as the model's parameters are varied over the range of concern.

Hence, $(N, \mathscr{F})$ is a *set system* on *ground set* N having $\mathscr{F}$ as its family of *feasible sets*. Algorithmic considerations lead us to impose strong structural properties on $(N, \mathscr{F})$, which refer to the *outer* and *inner boundaries* of an active set $S \in \mathscr{F}$, defined respectively by

$$\partial^{\text{out}}_{\mathscr{F}} S \triangleq \left\{ j \in S^c \colon S \cup \{j\} \in \mathscr{F} \right\} \quad \text{and} \quad \partial^{\text{in}}_{\mathscr{F}} S \triangleq \left\{ j \in S \colon S \setminus \{j\} \in \mathscr{F} \right\}. \tag{43}$$

We will further say that two active sets S and $S \cup \{j\}$, with $j \in S^c$, are *adjacent*.

Definition 5.1 We say that $(N, \mathscr{F})$ is a *monotonically connected set system* if:

(i) $\emptyset, N \in \mathscr{F}$;

Table 6: Minimal Tableau for S-Active BFS.

	$\left(\mathbf{x}^1_{S^c}\right)^\top$
$\mathbf{x}^0_{S^c}$	$\mathbf{A}^S_{S^cS^c}$
	$\mathbf{w}^S_{S^c}$
	$\mathbf{d}^S_{S^c}$

(ii) for every $S, S' \in \mathscr{F}$ with $S \subset S'$ there exist $j \in \partial^{\text{out}}_{\mathscr{F}} S$ and $j' \in \partial^{\text{in}}_{\mathscr{F}} S'$ such that $S \subset S \cup \{j\} \subseteq S'$ and $S \subseteq S' \setminus \{j'\} \subset S'$; and

(iii) for any $S, S' \in \mathscr{F}$ with $S \neq S'$, it holds that $S \cup S' \in \mathscr{F}$.

While various types of set system have been previously investigated, e.g. matroids or greedoids, to the best of our knowledge the concept of monotonically connected set system in Definition 5.1 is first introduced herein. The term "monotonically connected" is motivated by the fact that, in such a set system, one can always connect two feasible sets $S \subset S'$ by a monotone increasing sequence $S_1 \subset \cdots \subset S_m$ of adjacent sets in $\mathscr{F}$, with $S_1 = S$, $S_m = S'$. Further, one can also connect two distinct feasible sets $S \neq S'$ through two successive monotone sequences of adjacent sets in $\mathscr{F}$, the first of which is monotone increasing and connects S to $S \cup S'$, while the second is monotone decreasing and connects $S \cup S'$ to S'.

Assumption 5.2 $(N, \mathscr{F})$ is a monotonically connected set system.

We will further refer to the *Fast-Pivoting Adaptive-Greedy* index algorithm FPAG$(\mathscr{F})$ described in Table 8. This is a simplex-based implementation of the adaptive-greedy index algorithm for PCL-indexable bandits introduced in Niño-Mora (2001, 2002), whose scope we extend herein to the present broader setting. The FPAG$(\mathscr{F})$ algorithm is obtained by simplifying the CPI and RPI algorithms above by (i) storing and updating only *minimal tableaux* as shown in Table 6; and (ii) eliminating the indexability test at each step. Note that the minimal tableau for the $S \cup \{j\}$-active BFS is readily computed from that for the S-active BFS in Table 6, as shown in Table 7.

The results in Section 4 motivate us to introduce the following class of bandits, which we term $LP(\mathscr{F})$-*indexable* as their are based on LP analyses.

Definition 5.3 (LP$(\mathscr{F})$-indexability) We say that a bandit is $LP(\mathscr{F})$-*indexable* if:

(i) $w_i^{\emptyset}, w_i^N \geq 0$ for $i \in N$, and $\underline{d}^{\emptyset} \leq 0 \leq \overline{d}^N$;

(ii) for each active set $S \in \mathscr{F}$, $w_i^S > 0$ for $i \in \partial^{\text{in}}_{\mathscr{F}} S \cup \partial^{\text{out}}_{\mathscr{F}} S$; and

Table 7: Minimal Tableau for $S\cup\{j\}$-Active BFS, Obtained by Pivoting on a^S_{jj}.

	$\left(\mathbf{x}^1_{S^c\setminus\{j\}}\right)^\mathsf{T}$
$\mathbf{x}^0_{S^c\setminus\{j\}}$	$\mathbf{A}^S_{S^c\setminus\{j\},S^c\setminus\{j\}} - \dfrac{\mathbf{A}^S_{S^c\setminus\{j\},j}\mathbf{A}^S_{j,S^c\setminus\{j\}}}{a^S_{jj}}$
	$\mathbf{w}^S_{S^c\setminus\{j\}} - \dfrac{w^S_j}{a^S_{jj}}\mathbf{A}^S_{S^c\setminus\{j\},j}$
	$\mathbf{d}^S_{S^c\setminus\{j\}} - \dfrac{d^S_j}{a^S_{jj}}\mathbf{A}^S_{S^c\setminus\{j\},j}$

Table 8: The Fast-Pivoting Adaptive-Greedy Index Algorithm FPAG($\mathscr{F}$).

solve $\mathbf{A}^{(0)}\left[\mathbf{I}_{N,N\setminus\{j^*\}} - \mathbf{\Phi}^0_{N,N\setminus\{j^*\}} \quad \widetilde{\mathbf{m}}^0\right] = \left[\mathbf{I}_{N,N\setminus\{j^*\}} - \mathbf{\Phi}^1_{N,N\setminus\{j^*\}} \quad \widetilde{\mathbf{m}}^1\right]$
$\begin{bmatrix}\mathbf{w}^{(0)}\\ \mathbf{d}^{(0)}\end{bmatrix} := \begin{bmatrix}\mathbf{q}^1\\ \mathbf{r}^1\end{bmatrix} - \begin{bmatrix}\mathbf{q}^0\\ \mathbf{r}^0\end{bmatrix}\mathbf{A}^{(0)}$; $S_0 := \emptyset$

for $k := 1$ **to** n **do**
  $v_i^{(k-1)} := d_i^{(k-1)}/w_i^{(k-1)}, \quad i \in \partial^{\text{out}}_{\mathscr{F}} S_{k-1}$
  pick $i_k \in \arg\max\left\{v_i^{(k-1)} : i \in \partial^{\text{out}}_{\mathscr{F}} S_{k-1}\right\}$; $v^*_{i_k} := v^{(k-1)}_{i_k}$; $S_k := S_{k-1}\cup\{i_k\}$
  if $k < n$ **then**
    $\mathbf{A}^{(k)}_{S^c_k i_k} := \mathbf{A}^{(k-1)}_{S^c_k i_k}/a^{(k-1)}_{i_k i_k}$; $\mathbf{A}^{(k)}_{S^c_k S^c_k} := \mathbf{A}^{(k-1)}_{S^c_k S^c_k} - \mathbf{A}^{(k)}_{S^c_k i_k}\mathbf{A}^{(k-1)}_{i_k S^c_k}$
  end { if }
  $\mathbf{w}^{(k)}_{S^c_k} := \mathbf{w}^{(k-1)}_{S^c_k} - w^{(k-1)}_{i_k}\mathbf{A}^{(k)}_{S^c_k i_k}$; $\mathbf{d}^{(k)}_{S^c_k} := \mathbf{d}^{(k-1)}_{S^c_k} - d^{(k-1)}_{i_k}\mathbf{A}^{(k)}_{S^c_k i_k}$
end { for }

(iii) for every wage $\nu \in \mathbb{R}$ there exists an optimal active set $S \in \mathscr{F}$ for (6).

We note that conditions (i, ii) are meant to be established through an ad hoc *work-reward analysis* for the model at hand, while condition (iii) will be typically established by DP arguments. See Niño-Mora (2006e, 2007a,d) for specific examples.

We are now ready to present what we consider the main result of this paper. While its part (a) says that LP($\mathscr{F}$)-indexability is a sufficient condition for indexability, with the MPI being computed by algorithm FPAG($\mathscr{F}$), its part (b) says that such a condition is also necessary, in that an indexable bandit is always LP-indexable, relative to some nested active-set family.

Theorem 5.4 *The following holds*:

(a) *An LP($\mathscr{F}$)-indexable bandit is indexable, and its MPI is computed in nondecreasing order by algorithm FPAG($\mathscr{F}$).*

(b) *An indexable bandit is LP($\mathscr{F}$)-indexable relative to some nested active-set family $\mathscr{F}$.*

Proof. (a) Since the core of the following proof is geometric, to help the reader visualize and grasp the following arguments we will refer to Figure 3 for illustration, which represents the achievable work-reward performance region $\mathbb{H}$ of a bandit (cf. Section 2.2).

Suppose the bandit is LP($\mathscr{F}$)-indexable. We first note that conditions (i, ii) in Definition 5.3 imply, by Lemma 4.2(a), that the $\emptyset$-active (resp. N-active) BFS is optimal for parametric LP problem (22) iff $\nu \geq \underline{\nu}^{\emptyset} > -\infty$ (resp. iff $\nu \leq \overline{\nu}^{N} < +\infty$). Imagine now that the parametric-objective simplex algorithm of Gass and Saaty (1955) is run on such an LP, by decreasing the wage parameter ν from $+\infty$ to $-\infty$. Since the LP is bounded, this will yield a finite decreasing sequence of distinct breakpoints in the ν axis, which is nonempty since it contains the finite values $\underline{\nu}^{\emptyset}$ and $\overline{\nu}^{N}$. Note also that multiple successive iterations of the algorithm might correspond to the same breakpoint. The sequence of adjacent closed intervals determined by such breakpoints have the property that there is a unique optimal BFS for values of ν lying strictly within each interval.

We may visualize the progress of the Gass-Saaty algorithm in Figure 3. The key observation is that, geometrically, *as the wage ν is decreased from $+\infty$ to $-\infty$ the algorithm traverses the upper boundary $\overline{\partial}\mathbb{H}$ of region $\mathbb{H}$ from left to right, pivoting through a sequence of BFS of LP* (22) *whose successive values in the g (work) axis are increasing*. Such a sequence of BFS will yield work-reward points that contain all vertices of $\mathbb{H}$ liying in its upper boundary, which are marked by black circles in Figure 3; yet, other BFS produced in the algorithm might yield points that are not vertices of $\mathbb{H}$, such as those marked by small black squares.

Notice that, in the figure, the interval of ν values for which a BFS yielding a point in the upper boundary is optimal is visualized as the interval between the left and right slopes in the upper boundary meeting at such a point.

Consider the case that there is just one breakpoint, say λ_1, so that $\lambda_1 = \underline{\nu}^{\emptyset} = \overline{\nu}^{N}$. For $\nu = \lambda_1$, the interpretation of LP (22) in terms of (6) ensures that the DP equations (7) satisfy

$$\vartheta_i^*(\lambda_1) = r_i^1 - \nu q_i^1 + \sum_{j \in N} \phi_{ij}^1 \vartheta_j^*(\lambda_1) = r_i^0 - \nu q_i^0 + \sum_{j \in N} \phi_{ij}^0 \vartheta_j^*(\lambda_1), \quad i \in N,$$

and hence every active set $S \subseteq N$ yields an optimal basis. Therefore, Definition 5.1(i, ii) ensures that there exists a monotone increasing sequence $S_0 \subset \cdots \subset S_n$ of adjacent active sets in $\mathscr{F}$, with $S_0 = \emptyset$ and $S_n = N$, which, by Definition 5.3(ii) and Lemma 4.1, satisfies the requirements of Definition 2.1, ensuring that the bandit is indexable. Further, such an active-set sequence can be constructed by running algorithm FPAG($\mathscr{F}$), which corresponds to taking n pivot steps in the Gass and Saaty algorithm at the only breakpoint λ_1.

Consider now the case that there are $L \geq 2$ distinct breakpoints, which we denote by $\lambda_1 > \cdots > \lambda_L$. Then, the $\emptyset$-active BFS and the N-active BFS will be the only optimal solutions for $\nu > \lambda_1$ and for $\nu < \lambda_L$, respectively. Further, for $2 \leq l \leq L$, the LP will have a unique optimal BFS in the interval $\nu \in (\lambda_{l-1}, \lambda_l)$, whose active set we denote by T_l. Such active sets satisfy $g^{T_l} < g^{T_{l+1}}$ and, by Definition 5.3(iii), $T_l \in \mathscr{F}$. Further, for $\nu = \lambda_l$, the interpretation of LP (22) in terms of ν-wage problem (6) ensures that the latter's DP equations (7) must satisfy

$$\vartheta_i^*(\lambda_l) = r_i^1 - \nu q_i^1 + \sum_{j \in N} \phi_{ij}^1 \vartheta_j^*(\lambda_l) = r_i^0 - \nu q_i^0 + \sum_{j \in N} \phi_{ij}^0 \vartheta_j^*(\lambda_l), \quad i \in (T_{l+1} \setminus T_l) \cup (T_l \setminus T_{l+1}),$$

and therefore every active set S with $T_l \subseteq S \subseteq T_l \cup T_{l+1}$ or $T_{l+1} \subseteq S \subseteq T_l \cup T_{l+1}$ yields an optimal solution for the λ_l-wage problem, and hence an optimal BFS for the LP.

We now argue by contradiction that such an active-set sequence must be monotone increasing, i.e., $T_l \subset T_{l+1}$ for all l. For suppose such is not the case, so that $T_l \cup T_{l+1} \supset T_{l+1}$ for some l. Then, Definition 5.1(ii, iii) ensures both that $T_l \cup T_{l+1} \in \mathscr{F}$, and that there exists a monotone decreasing sequence $S_1 \supset \cdots \supset S_m$ of adjacent sets in $\mathscr{F}$ connecting $S_1 = T_l \cup T_{l+1}$ to $S_m = T_{l+1}$. By the argument at the end of the previous paragraph, it follows that each such active set S_k must be optimal for $\nu = \lambda_l$, and hence satisfy $g^{T_l} \leq g^{S_k} \leq g^{T_{l+1}}$, as illustrated in Figure 3. Yet, construction of the S_k's, Definition 5.3(ii) and Lemma 4.1 imply that $g^{S_1} > \cdots > g^{S_m}$, and hence $g^{T_l \cup T_{l+1}} > g^{T_{l+1}}$, which contradicts the inequality $g^{T_l \cup T_{l+1}} \leq g^{T_{l+1}}$ argued before.

Therefore, set sequence T_l is monotone increasing and hence, by Definition 5.1(ii), there exists a monotone increasing sequence $S_1 \subset \cdots \subset S_m$ of adjacent active sets in $\mathscr{F}$ connecting $S_1 = T_l$ to $S_m = T_{l+1}$. By the above DP argument, each of the S_k's yields an optimal BFS for $\nu = \lambda_l$ and, further, Definition 5.3(ii)

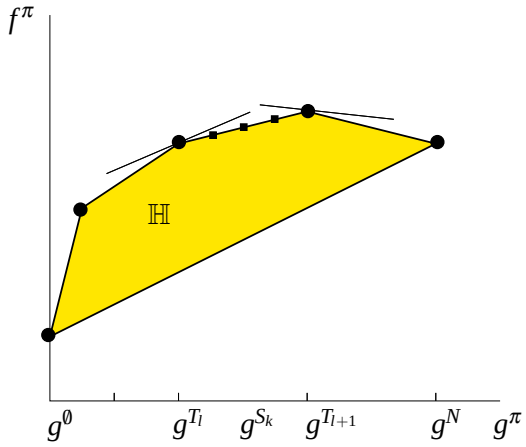

Figure 3: Geometry of the Gass-Saaty / FPAG($\mathscr{F}$) Algorithm for an LP($\mathscr{F}$)-Indexable Bandit.

and Lemma 4.1 imply that $g^{S_1} < \cdots < g^{S_m}$. Further, such a sequence of S_k's can be actually contructed using algorithm FPAG($\mathscr{F}$), since this is just a form of the Gass and Saaty algorithm that only considers BFS having active sets in $\mathscr{F}$.

The above shows that algorithm FPAG($\mathscr{F}$) will construct an increasing sequence of adjacent active sets in $\mathscr{F}$ connecting $\emptyset$ to N, which satisfies the requirements of Definition 2.1, implying that the bandit is indexable.

(b) This part follows by noticing that a bandit that has been shown to be indexable via Proposition 4.4, is LP($\mathscr{F}$)-indexable relative to the nested active-set family $\mathscr{F}$ constructed by algorithm CPI. This completes the proof. □

The following result assesses the computational complexity of algorithm FPAG($\mathscr{F}$), showing that it improves significantly upon that of algorithm RPI. In particular, the complexity of its "for" loop matches that of solving an $n \times n$ linear equation system by Gaussian elimination.

Proposition 5.5 *The FPAG($\mathscr{F}$) algorithm's loop performs* $(2/3)n^3 + O(n^2)$ *operations.*

Proof. The loop's operation count is dominated by the update of matrix $\mathbf{A}^{(k)}_{S_k^c S_k^c}$ at each step k, which takes $2(n-k)^2$ arithmetic operations, yielding the stated total arithmetic operation count. □

In the special case of nonrestless semi-Markov bandits, using algorithm FPAG($\mathscr{F}$) with $\mathscr{F} = 2^N$ yields a $(2/3)n^3 + O(n^2)$ method to compute the Gittins index, as the initialization step becomes trivial, thus matching the complexity result in Niño-Mora (2006a) for classic Markov bandits.

5.2 PCL($\mathscr{F}$)-Indexable Bandits Revisited

We next revisit the concept of PCL($\mathscr{F}$)-indexability, introduced and developed in Niño-Mora (2001, 2002, 2006d), in light of the above developments.

Definition 5.6 (PCL($\mathscr{F}$)-indexability) We say that a bandit is *PCL($\mathscr{F}$)-indexable* if:

(i) for each active set $S \in \mathscr{F}$, $w_i^S > 0$ for $i \in N$; and

(ii) for every wage $\nu \in \mathbb{R}$ there exists an optimal active set $S \in \mathscr{F}$ for (6); or

(ii') algorithm FPAG($\mathscr{F}$) produces a nonincreasing index sequence: $\nu_{i_1}^* \geq \nu_{i_2}^* \geq \cdots \geq \nu_{i_n}^*$.

Thus, a PCL($\mathscr{F}$)-indexable bandit is an LP($\mathscr{F}$)-indexable bandit having positive marginal work for active sets $S \in \mathscr{F}$. Note that Definition 5.6 differs slightly from those given in the earlier work mentioned, which only required satisfaction of conditions (i, ii'), and imposed less stringent requirements on set system $(N, \mathscr{F})$. Our motivation for introducing the above alternate form is applied: we have found that, in the analysis of bandit models with complex state spaces, condition (ii') can be much more difficult to establish than condition (ii). See, e.g., Niño-Mora (2007a).

Proposition 5.7 *In Definition* 5.6, *conditions* (i, ii) *and* (i, ii') *are equivalent.*

Proof. Suppose that conditions (i, ii) hold. Then, the bandit is LP($\mathscr{F}$)-indexable and, by Theorem 5.4(a), it is indexable, with algorithm FPAG($\mathscr{F}$) computing its MPI ν_i^* in nondecreasing order. Hence, condition (ii') holds.

Suppose now that conditions (i, ii') hold. Then, it is shown in Niño-Mora (2001, Cor. 2) and in Niño-Mora (2002, Th. 6.3) (in increasingly general settings) that, for a finite-state Markovian bandit, such conditions imply its indexability, from which (ii) follows. The extension of such a result to the present semi-Markov setting is straightforward. □

6. Extension to the Average Criterion

In applications of restless bandit indexation to problems under the (long-run) average criterion, one must address the version of ν-wage problem (6) based on reward and work measures

$$f_i^\pi \triangleq \liminf_{T\nearrow\infty} \frac{1}{T}\mathbb{E}_i^\pi\left[\int_0^T R_{X(t)}^{a(t)}\,dt\right] = \liminf_{K\nearrow\infty} \frac{1}{K}\mathbb{E}_i^\pi\left[\sum_{k=0}^K r_{X_k}^{a_k}\right], \tag{44}$$

and

$$g_i^\pi \triangleq \limsup_{T\nearrow\infty} \frac{1}{T}\mathbb{E}_i^\pi\left[\int_0^T Q_{X(t)}^{a(t)} e^{-\alpha t}\,dt\right] = \limsup_{K\nearrow\infty} \frac{1}{K}\mathbb{E}_i^\pi\left[\sum_{k=0}^K q_{X_k}^{a_k} e^{-\alpha t_k}\right]. \tag{45}$$

As in Niño-Mora (2002, Sec. 6.5), we must now assume that the embedded process X_n is *communicating*, i.e., every state can be reached from every other state under some stationary policy. This ensures that the above measures do not depend on the initial state i under a stationary deterministic policy, and hence one can write f^S and g^S for active sets $S \subseteq N$. Hence, the corresponding ν-wage problem (6) can be solved by a stationary deterministic policy independent of i, which allows one to readily extend the indexability theory above to the average criterion.

Regarding the above algorithms, they apply without modification to the average criterion, as the results in Section 3.4 show that the required tableaux emerge as limits of their discounted counterparts as the discount rate vanishes, and also shows how to compute the initial tableau. To extend the results in Section 5 one must further assume that the active-set family $\mathscr{F}$ of concern has the property that, for every $S \in \mathscr{F}$, the S-active policy is *unichain*, i.e., it induces on the embedded process X_n a single recurrent class plus a (possibly empty) set of transient states.

7. Computational Experiments

This section reports the results of several computational experiments, based on the author's MATLAB implementations of the algorithms discussed in this paper.

7.1 Assessing the Prevalence of Indexability and PCL-Indexability

We start by assessing experimentally the prevalence of the indexability and PCL-indexability properties, in two different classes of randomly generated restless bandit instances.

In the first class, we considered discrete-time bandits. We conducted a simulation study based on generating a random i.i.d. sample of 10^7 bandit instances with $q_i^a = a$ and dense transition probability matrices — obtained by appropriately scaling a matrix with Uniform$[0,1]$ entries — for each of the state-space sizes $n = 3, \ldots, 7$. For each instance, we used the above algorithms to test for indexability and PCL-indexability (relative to any $\mathscr{F}$), as the discount factor β varies. Note that the value $\beta = 1$ refers to the average criterion discussed in Section 6.

Table 9 reports the results. They show that the prevalence of nonindexable bandits fastly decreases as the discount factor gets smaller, and as the state space gets larger. The highest prevalence of nonindexable projects (1 out of 12225) was found for projects with 3 states under the average criterion. Indexability thus appears to be a highly prevalent property over this class of instances, and the more so the larger the state space and the smaller the discount factor. The table further shows the same pattern with the number of instances found to be indexable yet not PCL-indexable. The highest prevalence of such bandits was found

Table 9: Counts on Random i.i.d. Samples of 10^7 Bandit Instances.

	Nonindexable					Indexable non-PCL				
	number of states					number of states				
β	3	4	5	6	7	3	4	5	6	7
0.1	0	0	0	0	0	0	0	0	0	0
0.2	0	0	0	0	0	0	0	0	0	0
0.3	0	0	0	0	0	0	0	0	0	0
0.4	0	0	0	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0	0	0	0
0.6	0	0	0	0	0	0	0	0	0	0
0.7	0	0	0	0	0	30	0	0	0	0
0.8	16	1	0	0	0	574	32	1	0	0
0.9	135	7	0	0	0	4460	509	36	5	0
1.0	818	66	4	0	0	18631	3640	425	50	3

in the case of 3 states under the average criterion, being then of only about 1 non-PCL instance out of 537 indexable instances.

In the second class of instances, we considered continuous-time bandits with exponential transition rates λ_{ij}^a for states $i \neq j$, having the following structure:

$$\lambda_{ij}^1 = \lambda_{ij}^0 + \mu_{ij}, \quad i \neq j, \tag{46}$$

for some nonnegative μ_{ij}'s. The relations in (46) model a situation where the bandit is subject to two different types of events: "regular events" and "extra events." Regular events are driven by transition probabilities λ_{ij}^0 and are not subject to control. Extra events, which coexist with regular event, can be turned on and off. When activated, they are driven by transition rates μ_{ij}.

For such a system, two definitions for the Q_i^a's spring to mind. One is the conventional definition $Q_i^a \triangleq a$. The other is to set

$$Q_i^1 \triangleq \sum_{j \in N \setminus \{i\}} \mu_{ij}, \quad Q_i^0 \equiv 0, \tag{47}$$

so that Q_i^1 is the rate at which extra events occur in state i when they are turned on. In the first definition of the Q_i^a's, the wage parameter ν in (6) is the charge incurred per unit time that the extra-events stream is turned on. In the second, it is the cost incurred per extra event generated.

Table 10 reports the results of the corresponding simulation study for such a class of instances — reformulated into discrete-time via uniformization. The pairs shown give the counts under both definitions of the Q_i^a's, starting with (47). Thus, e.g., the pair $(19, 45)$ for $\beta = 0$ means that, out of 10^7 instances with 3 states, 19 of them were nondexable using the Q_i^a definition in (47), and 45 were nonindexable using the conventional definition $Q_i^a \equiv a$.

Table 10: Counts on Random Samples of 10^7 Bandit Instances for Two Definitions of Q_i^a.

	Nonindexable					Indexable non-PCL				
	number of states					number of states				
β	3	4	5	6	7	3	4	5	6	7
0.1	0	0	0	0	0	0	0	0	0	0
0.2	0	0	0	0	0	0	0	0	0	0
0.3	0	0	0	0	0	0	0	0	0	0
0.4	0	0	0	0	0	0	0	0	0	0
0.5	0	0	0	0	0	0	0	0	0	0
0.6	0	0	0	0	0	0	0	0	0	0
0.7	0	0	0	0	0	0	0	0	0	0
0.8	0	1	0	0	0	0	0	0	0	0
0.9	0	7	0	0	0	$(0,7)$	0	0	0	0
1.0	$(19,45)$	$(0,3)$	$(1,0)$	0	0	$(317,924)$	$(62,58)$	$(5,2)$	$(1,0)$	0

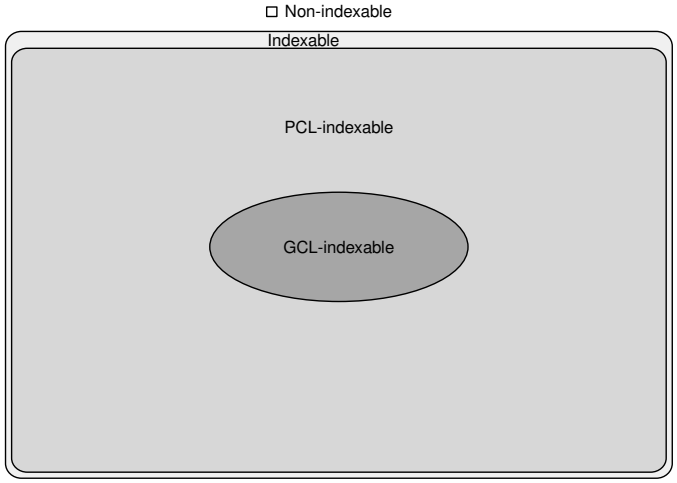

Figure 4: Classification of Restless Bandits.

The table shows that, in this class of instances, both indexability and PCL-indexability are even more highly prevalent properties than in the previous class. It further shows that, for $n = 3$ states, both the prevalences of instances that are indexable and of instances that are PCL-indexable are significantly higher under definition (47).

Such experimental evidence supports the claim that, at least for bandits with dense transition probability matrices, both indexability and PCL-indexability are highly prevalent properties. Figure 4 shows a modified version of the classification of restless bandits introduced in Niño-Mora (2001), updated to better reflect relative class sizes. Note that the figure refers to the class of *GCL-indexable* bandits, named after their satisfaction of *generalized conservation laws* (GCL), which are PCL-indexable relative to $\mathscr{F} = 2^N$.

7.2 Runtime Comparison of Index Algorithms

In contemporary computers, the actual runtime performance of an algorithm depends both on its arithmetic operation count and on its memory-access patterns, with the latter being often the dominant factor. To compare the performance of the algorithms discussed in this paper, we have thus conducted a computational study, using MATLAB implementations developed by the author. The experiments were performed on an HP xw9300 254 (2.8 GHz) Opteron workstation running MATLAB 2006b under Windows XP x64. For each of the state space sizes $n = 1000, 1500, \ldots, 6000$, a random discrete-time bandit instance with dense transition probability matrices was generated. Transition matrices were obtained by scaling matrices with Uniform$[0,1]$ entries, dividing each row by its sum. Active rewards were also generated with Uniform$[0,1]$ entries, while passive rewards were set to zero. The discrete-time discount factor used was $\beta = 0.8$.

For each instance, the CPI algorithm was used to test both for indexability and for PCL-indexability (by checking the signs of marginal work measures for the generated nested active-active set family). Since such tests turned out positive in each case, the MPI values were computed using the CPI, RPI and FPAG($\mathscr{F}$) algorithms, which was run taking $\mathscr{F} = 2^N$.

Figure 5 displays the recorded runtimes for each algorithm, where where the lines shown are obtained by cubic least-squares fits. The results show that the FPAG algorithm, having an operation count of $(2/3)n^3$, is indeed the fastest of the three, consistently achieving speedup factors of about 1.3 over the CPI and RPI algorithms, which exhibit similar runtimes, though the RPI algorithm was the slowest. Recall that the operation counts are $2n^3$ and n^3 for the CPI and the RPI algorithms, respectively. Such discrepancies between theoretical and actual speedup factors are accounted for by noticing the algorithms memory-access patterns. Thus, algorithm CPI, being based on complete pivoting steps, has efficient memory-access patterns, as the coefficient matrix $\mathbf{A}$ is always updated as a contiguous memory block. In contrast, both the RPI and the FPAG algorithms reduce the operation count at the expense of using and updating submatrices of $\mathbf{A}$, which results in costly noncontiguous memory-access patterns. Yet, in the case of the FPAG algorithm, the large reduction in arithmetic operations compensates such inefficiencies, rendering it the fastest algorithm.

Acknowledgments

This research has been supported in part by the Spanish Ministry of Education & Science under grant MTM2004-02334 and a Ramón y Cajal Investigator Award, by the EU's Networks of Excellence Euro-NGI and Euro-FGI, and by the Autonomous Community of Madrid-UC3M's grants UC3M-MTM-05-075 and CCG06-UC3M/ESP-0767.

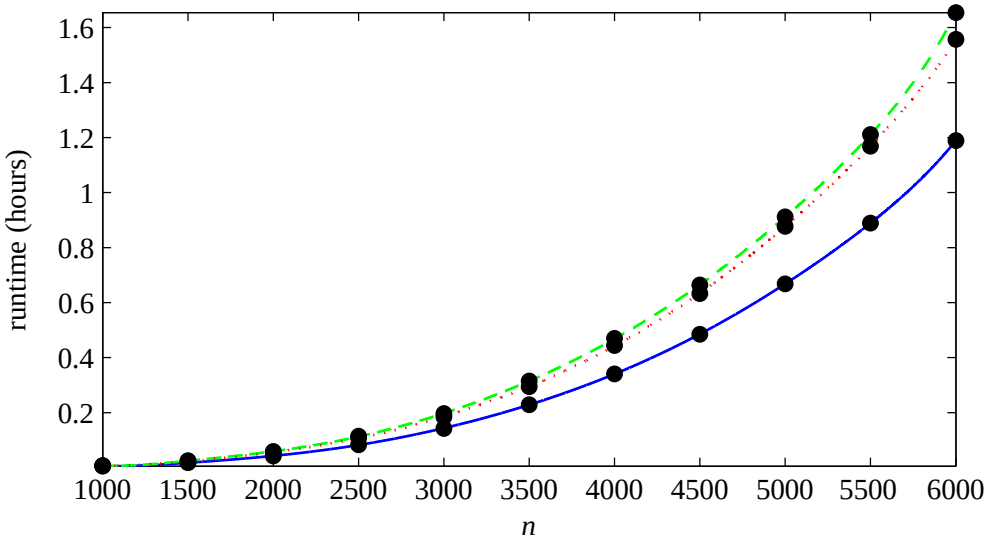

Figure 5: Runtimes with Cubic Least-Squares Fit: FPAG (solid), CPI (dotted) and RPI (dashed).

References

Baker, A. H, J. M. Dennis, E. R. Jessup. 2006. On improving linear solver performance: A block variant of GMRES. *SIAM J. Sci. Comput.* **27** 1608–1626.

Dongarra, J. J., V. Eijkhout. 2000. Numerical linear algebra algorithms and software. *J. Comput. Appl. Math.* **123** 489–514.

Gass, S., T. Saaty. 1955. The computational algorithm for the parametric objective function. *Naval Res. Log. Quart.* **2** 39–46.

Gittins, J. C. 1979. Bandit processes and dynamic allocation indices. *J. Roy. Statist. Soc. Ser. B* **41** 148–177. With discussion.

Goyal, M., A. Kumar, V. Sharma. 2006. A stochastic control approach for scheduling multimedia transmissions over a polled multiaccess fading channel. *Wireless Netw.* **12** 605–621.

Kallenberg, L. C. M. 1986. A note on M. N. Katehakis' and Y.-R. Chen's computation of the Gittins index. *Math. Oper. Res.* **11** 184–186.

La Scala, B. F., B. Moran. 2006. Optimal target tracking with restless bandits. *Digital Signal Processing* **16** 479–487.

Niño-Mora, J. 2001. Restless bandits, partial conservation laws and indexability. *Adv. in Appl. Probab.* **33** 76–98.

Niño-Mora, J. 2002. Dynamic allocation indices for restless projects and queueing admission control: a polyhedral approach. *Math. Program.* **93** 361–413.

Niño-Mora, J. 2003. Restless bandit marginal productivity indices, diminishing returns, and scheduling a multiclass make-to-order/-stock queue. *Proceedings of the 41st Annual Allerton Conference on Communication, Control and Computing*. 100–109.

Niño-Mora, J. 2005. A marginal productivity index policy for the finite-horizon multiarmed bandit problem. *Proceedings of the 44th IEEE Conference on Decision and Control and European Control Conference ECC 2005 (CDC-ECC'05)*. IEEE, 1718–1722.

Niño-Mora, J. 2006a. A $(2/3)n^3$ fast-pivoting algorithm for the Gittins index and optimal stopping of a Markov chain. *INFORMS J. Comp.* In press, `http://halweb.uc3m.es/jnino/eng/public.html`.

Niño-Mora, J. 2006b. Marginal productivity index policies for scheduling a multiclass delay-/loss-sensitive queue. *Queueing Syst.* **54** 281–312.

Niño-Mora, J. 2006c. Marginal productivity index policies for scheduling multiclass wireless transmissions. *Proceedings of the 2nd Euro-NGI Conference on Next Generation Internet Networks (NGI 2006)*. IEEE, 342–349.

Niño-Mora, J. 2006d. Restless bandit marginal productivity indices, diminishing returns and optimal control of make-to-order/make-to-stock $M/G/1$ queues. *Math. Oper. Res.* **31** 50–84.

Niño-Mora, J. 2006e. Two-stage index computation for bandits with switching penalties I: switching costs. Working Paper 07-41, Statistics and Econometrics Series 09, `http://halweb.uc3m.es/jnino/eng/public2.html`, Univ. Carlos III de Madrid, Spain. Submitted.

Niño-Mora, J. 2007a. An index policy and algorithm for bandit and optimal stopping problems with deadlines. Working Paper 07-44, Statistics and Econometrics Series 12, `http://halweb.uc3m.es/jnino/eng/public2.html`, Dept. of Statistics, Univ. Carlos III de Madrid, Spain. Submitted.

Niño-Mora, J. 2007b. Marginal productivity index policies for admission control and routing to parallel multi-server loss queues with reneging. *Proceedings of the First Euro-FGI Conference on Network Control and Optimization (NET-COOP 2007)*. LNCS, Springer, pp. to be assigned.

Niño-Mora, J. 2007c. Marginal productivity index policies for scheduling multiclass delay-/loss-sensitive traffic with delayed state observation. *Proceedings of the third Euro-NGI Conference on Next Generation Internet Networks (NGI 2007)*. IEEE, pp. to be assigned.

Niño-Mora, J. 2007d. Two-stage index computation for bandits with switching penalties II: switching delays. Working Paper 07-42, Statistics and Econometrics Series 10, `http://halweb.uc3m.es/jnino/eng/public2.html`, Dept. of Statistics, Univ. Carlos III de Madrid, Spain. Submitted.

Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, New York.

Raissi-Dehkordi, M., J. S. Baras. 2002. Broadcast scheduling in information delivery systems. *Proceedings of Global Telecommunications Conference (GLOBECOM '02)*. IEEE, 2935–2939.

Saaty, T., S. Gass. 1954. Parametric objective function (Part 1). *J. Operations Res. Soc. Amer.* **2** 316–319.

Veatch, M. H., L. M. Wein. 1996. Scheduling a multiclass make-to-stock queue: Index policies and hedging points. *Oper. Res.* **44** 634–647.

Whittle, P. 1988. Restless bandits: Activity allocation in a changing world. J. Gani, ed., *A Celebration of Applied Probability, J. Appl. Probab.*, vol. 25A. Applied Probability Trust, Sheffield, UK, 287–298.