mathematics

MDPI

*Article*

# Fast Two-Stage Computation of an Index Policy for Multi-Armed Bandits with Setup Delays

**José Niño-Mora**

Department of Statistics, Carlos III University of Madrid, 28903 Getafe, Spain; jose.nino@uc3m.es

**Abstract:** We consider the multi-armed bandit problem with penalties for switching that include setup delays and costs, extending the former results of the author for the special case with no switching delays. A priority index for projects with setup delays that characterizes, in part, optimal policies was introduced by Asawa and Teneketzis in 1996, yet without giving a means of computing it. We present a fast two-stage index computing method, which computes the continuation index (which applies when the project has been set up) in a first stage and certain extra quantities with cubic (arithmetic-operation) complexity in the number of project states and then computes the switching index (which applies when the project is not set up), in a second stage, with quadratic complexity. The approach is based on new methodological advances on restless bandit indexation, which are introduced and deployed herein, being motivated by the limitations of previous results, exploiting the fact that the aforementioned index is the Whittle index of the project in its restless reformulation. A numerical study demonstrates substantial runtime speed-ups of the new two-stage index algorithm versus a general one-stage Whittle index algorithm. The study further gives evidence that, in a multi-project setting, the index policy is consistently nearly optimal.

**Keywords:** multi-armed bandits; setup delays; setup costs; index policies; semi-Markov decision processes; hysteresis



**Citation:** Niño-Mora, J. Fast Two-Stage Computation of an Index Policy for Multi-Armed Bandits with Setup Delays. *Mathematics* **2021**, *9*, 52. https://doi.org/10.3390/math9010052

Received: 7 December 2020
Accepted: 24 December 2020
Published: 29 December 2020



## 1. Introduction

### 1.1. Background

In a much-studied version of the *multi-armed bandit problem* (MABP), a decision-maker selects one project to engage from a finite set of dynamic and stochastic projects at each of an infinite sequence of discrete-time periods. Each project is modeled as a classic (non-restless) bandit, so the engaged (*active*) project gives rewards and its state changes in a Markovian fashion, while rested (*passive*) projects neither produce rewards nor change state. The goal is to find a policy that selects one project to be engaged at each time, for maximizing the expected total geometrically discounted reward. The MABP is widely applicable, being regarded as a modeling paradigm of the exploration versus exploitation trade-off, and it has generated a vast literature (see the monograph [1] and the cited references there). Although the *curse of dimensionality* hinders direct numerical solution of its *dynamic programming* (DP) optimality equations for realistic-size models, as the size of the multi-dimensional state space grows exponentially with the number of projects, the MABP is solved optimally by a remarkably simple type of policy, a so-called (priority-) *index policy*. Index policies are based on defining for each project $m$ an *index* $\lambda_m(i_m)$—a scalar mapping of the project state $i_m$ that depends only on the project parameters—and engage at each time a project of largest index. See, e.g., [2–6]. The index that is considered in [2], which is known in the literature as the *Gittins index*, extends to general Markovian bandits that which was introduced by Bellman in [7] for solving a Bernoulli bandit model.

However, appropriate modeling of potential applications often entails the incorporation of features that violate assumptions of the classic MABP. Regarding the assumption that passive projects do not give rewards, this is noncritical, since passive rewards can be

readily eliminated through a linear transformation, as shown in [8]. Yet, other assumptions turn out to be critical, as index policies are typically suboptimal when they are violated. Such is the case, as demonstrated in [9], with the requirement that switching from engaging one project to another be costless, which is hardly realistic in many, if not most, applications. As stated in (p. 1, [9]), "*it is difficult to imagine a relevant economic decision problem in which the decision-maker may costlessly move between alternatives*". This motivates the interest of investigating extensions of the MABP that incorporate costs and/or delays for switching projects, which we will refer to generically, as in [10], as the *multi-armed bandit problem with switching penalties* (MABPSP).

Despite its practical relevance, the MABPSP has received relatively scant research attention when compared to the standard MABP. We refer the reader to [11] for a review of research on the MABPSP until the early 2000s. Important references on such early work include [9,10,12–14]. Additionally, see the survey [15]. Yet, the last decade has witnessed growing interest on variants of the MABPSP, being motivated by the relevance of switching penalties in a variety of application areas, including hiring and retention of workers who learn over time [16], online marketing [17,18], experiential learning [19], opportunistic channel access in communication networks [20,21], and continuation and abandonment decisions for research projects [22]. For recent theoretical work on properties of the MABPSP, see [23].

While the aforementioned work concerns discrete-state projects, ref. [24,25] address Markovian continuous-state projects with constant setup penalties (costs or delays).

### *1.2. Index Policies, Histeresis, and the Asawa and Teneketzis Index for the MABPSP*

While switching penalties can generally be sequence-dependent, this paper will focus on the case that such penalties are separately defined for each project, while allowing them to depend on the project state. Specifically, we will assume that switching from engaging one project to another entails, similarly as in [26], a *setdown cost* to switch off the currently engaged project, and then a *setup cost* followed by a random *setup delay* to switch on the project about to be engaged. Note that setup delays can be used to model, e.g., time for preparing the ground or building infrastructure, as well as training or learning delays.

Although index policies are generally suboptimal for the MABPSP (see [9]), their ease of implementation motivates the interest of designing policies from such a class that perform well. An index policy in such a setting attaches to each project $m$ an index $\lambda_m(a_m^-, i_m)$, which now depends on both the previous action $a_m^- \in \{0, 1\}$ (passive: 0 or active: 1) and the current project state $i_m$. Thus, such an index decouples into a *continuation index* $\lambda_m(1, i_m)$, which applies when the project has already been set up, and a *switching index* $\lambda_m(0, i_m)$, to be used when the project has not yet been set up.

Drawing on intuition one would expect that switching penalties should discourage frequent switching and, hence, should cause a *histeresis effect* on the structure of optimal policies. Thus, it should be optimal to stick longer to the currently engaged project that would be the case in the absence of such penalties. As put in (p. 691 [9]), "*it is obvious that in comparing two otherwise identical arms, one of which was used in the previous period, the one which was in use must necessarily be more attractive than the one which was idle*". To be consistent with such a *hysteresis property*, the indices of a project $m$ must satisfy that

$$\lambda_m(1, i_m) \geqslant \lambda_m(0, i_m) \text{ for every project state } i_m. \tag{1}$$

Note that index policies can be optimal in special cases of the MABPSP, as shown in [13], in a model for scheduling a multi-class batch of stochastic jobs.

An intuitively appealing choice of index, extending that in [13], is that considered by Asawa and Teneketzis in [10]—which we will refer to in the sequel as the *AT index*—for a project having either a constant (not dependent on the project state) setup cost or a constant setup delay distribution, and no setdown costs. It is shown in [10] that the AT index provides a partial characterization of optimal policies for the version of the MABPSP considered there. The continuation AT index of a project is simply its Gittins index. As for

the switching AT index, it is the highest rate of discounted expected reward minus setup cost per unit of discounted expected *active time* (counting the setup delay as active time) that can be attained from an initially passive project by first setting it up and then engaging it for a random duration that is given by a stopping time.

### *1.3. Index Computation*

Efficient index computation is a key issue that must be addressed in practice for deploying an index policy for the MABPSP. For a project with $n$ states and constant setup cost, but without setup delays, (Section III.C [10]) shows that the $2n$ AT continuation and switching index values $\lambda^*(a^-, i)$ can be computed as the Gittins index of an appropriately defined $2n$-state project with augmented state $(a^-, i)$. Because computing the Gittins index has, in general, a cubic operation complexity in the number of states, such an approach results in an eightfold increase in complexity relative to that of computing the continuation index only.

A faster two-stage approach for a project with both setup and setdown costs—but no setup delays—that can be state-dependent was given by the author in [27]. The proposed algorithm computes, in the first stage, the continuation index and certain extra quantities by applying the $(4/3)n^3 + O(n^2)$ *fast-pivoting algorithm with extended output* presented in [28]. Subsequently, in the second stage, it computes the switching index in at most $O(n^2)$ operations. Hence, computing with that algorithm the $2n$ AT index values entails only a twofold complexity increase relative to the $(2/3)n^3 + O(n^2)$ operation count to compute the continuation (Gittins) index only through the *fast-pivoting algorithm* (without extended output) given in [28]. Further, ref. [27] reports on the results of a numerical study demonstrating that the resulting index policy for the version of the MABPSP considered there is close to optimal and outperforms the Gittins index policy by a wide margin, across a wide range of instances.

### *1.4. Approach via Restless Bandit Reformulation, Whittle Index, and Indexability*

The two-stage index algorithm shown in [27] exploits the reformulation of a project with switching costs and states $i$ as a *restless bandit*—i.e., a project that can change state while passive—without such costs, moving across *augmented states* $(a^-, i)$. In that way, the MABP with switching costs is cast as a *multi-armed restless bandit problem* (MARBP) without them, which allows for the deployment of theoretical and algorithmic results on restless bandit indexation, as introduced in [29] by Whittle. Such a theory has been developed in [30–33] by the author. Additionally, see the survey [34].

Thus, while the MARBP is generally intractable, as it is known to be PSPACE-hard (see [35]), Whittle introduced, in [29], a widely applied heuristic index policy. For a sample of recent applications, see, for example [36–48]. Yet, the *Whittle index* is only defined for a limited class of restless bandits, called *indexable*, and it is nontrivial to verify whether such an *indexability* property holds for a given model. The work of the author referred to above provides sufficient indexability conditions for general restless bandits, which are grounded on satisfaction by project performance metrics of *partial conservation laws* (PCLs), together with an *adaptive-greedy index algorithm* that computes the Whittle index (and extensions thereof) under such conditions.

Such a *PCL-indexability* approach is deployed in [27], using the result that the AT index of a non-restless bandit with switching costs (but no switching delays) is its Whittle index in the project's restless reformulation. The corresponding restless bandit model is shown to satisfy the PCL-indexability conditions, ensuring that its Whittle index can be computed by the adaptive-greedy algorithm. Special structure and the results in [49] are then used in [27] in order to decouple that algorithm into a faster two-stage method.

### *1.5. Motivation and Goals*

Yet, no method is given in Asawa and Teneketzis [10] in order to compute their proposed index under switching delays. The latter's relevance in applications, along with

the tractability and effectiveness of the AT index policy in the pure-switching-costs case, motivates the interest to extend the restless bandit indexation approach for developing an efficient index algorithm for bandits that incorporate both switching costs and delays, which is the first goal of this paper.

Carrying out such an extension turns out to raise methodological research challenges on restless bandit indexation. Thus, when a Markovian non-restless bandit with switching delays is reformulated as a semi-Markov restless bandit without them, it is found that the resultant model need not satisfy the PCL-indexability conditions that were the cornerstone to the analyses presented in Niño-Mora [27] for the pure-switching-costs case. This motivates us to significantly extend the scope of previous theory, obtaining more powerful sufficient indexability conditions, which are both easier to apply and applicable to a wider class of models, including that of concern herein. That is the second goal of this paper. The third goal entails assessing the runtime performance of the proposed index algorithm, and evaluating the performance of the resulting index policy, both in terms of its optimality gap and its improvement over alternative simpler index policies.

### *1.6. Contributions*

Concerning the second goal, on general restless bandit methodology, we introduce, for finite-state restless bandits, significantly simpler and less stringent sufficient conditions for indexability than the former PCL-based conditions, under which it is also assured that the adaptive-greedy algorithm computes the MPI. We further show such conditions to be necessary, in that any indexable finite-state restless bandit satisfies them. Thus, the new conditions furnish a complete characterization of indexability, which can be used in order to analytically establish a priori that a restless bandit model of concern is indexable—as opposed to numerically verifying a posteriori that a given instance is indexable.

As for the first goal, we deploy the new indexability conditions in the restless bandit reformulation of a non-restless bandit with switching delays and costs. Because the AT index emerges as the Whittle index in such a reformulation, we are thus assured that the adaptive-greedy algorithm will compute it. The complexity of such an algorithm is then reduced by exploiting special structure, which again yields a substantially faster two-stage method. In the first stage, the continuation index is computed in $(4/3)n^3 + O(n^2)$ arithmetic operations, and then the switching index is computed in the second stage in only—at most—$(5/2)n^2 + O(n)$ operations. Thus, we obtain a two-stage algorithm that computes both the continuation and switching index in roughly twice the time that is required to compute the continuation index alone (if the latter were computed using the fast-pivoting $(2/3)n^3 + O(n^2)$ algorithm in [34]).

Regarding the third goal, we report on a computational study demonstrating the substantial runtime speed-up that is achieved by the two-stage algorithm relative to direct application of the one-stage adaptive-greedy algorithm. This study further reports on experiments providing evidence that the index policy is close to optimal and it attains significant gains against a benchmark index policy across a wide range of randomly generated instances with two and three projects.

### *1.7. Structure of the Paper and Notation*

The rest of the paper proceeds as follows. Section 2 describes the MABPSP model of concern, reviews the AT index, and describes the restless bandit indexation approach to be applied. Section 3 lays the groundwork for such an approach in a general framework of finite-state restless bandits, introducing the new methological advances on restless bandit indexation. Section 4 deploys the new results in the special restless bandit model that arises from the reformulation of a non-restless bandit with switching penalties, which culminates in the development of the new two-stage index algorithm in Section 5. Section 6 presents some qualitative properties on how the index depends on setup and setdown penalties. Finally, Section 7 presents and discusses the numerical study.

Because the notation of the paper may be hard to follow, Table 1 summarizes it for the reader's convenience.

**Table 1.** Some notation employed in the paper.

| | |
|---|---|
| $\mathcal{M} \triangleq \{1, \ldots, M\}$ | set of projects |
| $t_k$ | decision periods |
| $X_m(t)$, $X(t)$ | project state in period $t$ |
| $\mathcal{X}_m$, $\mathcal{X}$ | project state space |
| $A_m(t)$, $A(t)$ | action chosen on a project in period $t$ |
| $A_m^-(t)$, $A^-(t)$ | previously chosen action |
| $R_m(i_m)$, $R(i)$ | rewards |
| $\beta$ | one-period discount factor |
| $p_m(i_m, j_m)$, $p(i, j)$ | state-transition probabilities |
| $c_m(i_m)$, $c(i)$ | setup costs |
| $d_m(i_m)$, $d(i)$ | setdown costs |
| $\xi_m(i_m)$, $\xi(i)$ | setup delays |
| $\phi_m(i_m)$, $\phi(i)$ | setup delay $z$-transforms, for $z = \beta$ |
| $\psi_m$, $\psi$ | setdown delay $z$-transform, for $z = \beta$ |
| $Y_m(t)$, $Y(t)$ | augmented state in period $t$ |
| $\mathcal{Y}_m$, $\mathcal{Y}$ | augmented state space |
| $F_i^\pi$, $F_y^\pi$, $F^\pi$ | reward metric |
| $G_i^\pi$, $G_y^\pi$, $G^\pi$ | resource consumption metric |
| $f_i^\pi$, $f_y^\pi$ | marginal reward metric |
| $g_i^\pi$, $g_y^\pi$ | marginal resource consumption metric |
| $\lambda_i^S$, $\lambda_y^S$ | marginal productivity metric |

## 2. MABPSP Model and Its Semi-Markov MARBP Reformulation

A decision-maker ponders how to prioritize the allocation of effort to $M$ dynamic and stochastic projects that are labelled by $m \in \mathcal{M} \triangleq \{1, \ldots, M\}$, one of which must be engaged (active) at each of a sequence of *decision periods* $t_k \in \mathbb{Z}_+ \triangleq \{0, 1, 2, \ldots\}$, with $t_0 = 0$ and $t_k \nearrow \infty$ as $k \nearrow \infty$, while others are rested (passive). Switching projects on and off entails setup and setdown delays and costs, respectively. A setup (resp. setdown) delay on a project is necessarily followed by a period in which the project is worked on (resp. rested), i.e., the times at which a setup or a setdown delay are completed are not decision periods. We will say that a project is "active" when it is either being engaged (worked upon) or undergoing a setup or a setdown delay. Let $X_m(t)$ and $A_m(t)$ denote the prevailing state, which belongs to the finite state space $\mathcal{X}_m$, and action for project $m$ at time $t$ ($A_m(t) = 1$: active; $A_m(t) = 0$: passive), and let $A_m^-(t) \triangleq A_m(t-1)$ denote the *previously chosen action*, with $A_m^-(0)$ indicating the initial setup status.

While project $m$ is passive, it neither accrues rewards nor changes state. Switching it on when it lies in state $i_m$ entails a lump setup cost $c_m(i_m)$, followed by a random setup delay of duration $\xi_m(i_m)$ periods, whose $z$-transform is $\phi_m(z; i_m) \triangleq \mathbb{E}\left[z^{\xi_m(i_m)}\right]$, over which no rewards are earned. After such a setup, the project must be engaged, yielding a reward $R_m(i_m)$, after which its state moves at the next period to $j_m$ with transition probability $p_m(i_m, j_m)$. After at least one period in which the project is engaged, it may be decided to switch it off. If this is done when the project lies in state $j_m$, then a lump setdown cost $d_m(j_m)$ is incurred, followed by a random setdown delay of duration $\eta_m$ with $z$-transform $\psi_m(z) \triangleq \mathbb{E}\left[z^{\eta_m}\right]$, over which no rewards accumulate. Subsequently, the project remains passive for one or more periods. Note that setup delay distributions are allowed to be state-dependent, whereas setdown delay's are not (cf. Section 2.1). Rewards and costs are geometrically time-discounted with factor $\beta < 1$. We write, in what follows, the above $z$-transforms evaluated at $z = \beta$ simply as $\phi_m(i_m)$ and $\psi_m$.

Actions are prescribed through a *scheduling policy* $\boldsymbol{\pi}$, which is chosen from the class $\mathbf{\Pi}$ of policies that are *admissible*, i.e., nonanticipative with respect to the history of states and actions, and engaging one project at a time. The MABPSP (cf. Section 1) is concerned

with finding an admissible scheduling policy that attains the maximum expected total discounted reward net of switching costs.

This problem can be cast into the framework of *semi-Markov decision problems* (SMDPs) by including into the state of each project $m$ the last action taken, i.e., by using the *augmented state* $Y_m(t) \triangleq (A_m^-(t), X_m(t))$, which belongs to the *augmented state space* $\mathcal{Y}_m \triangleq \{0,1\} \times \mathcal{X}_m$. Thus, one obtains a multidimensional SMDP having *joint state* $\mathbf{Y}(t) \triangleq (Y_m(t))_{m\in\mathcal{M}}$ and *joint action* $\mathbf{A}(t) \triangleq (A_m(t))_{m\in\mathcal{M}}$. This is a special type of semi-Markov MARBP (cf. Section 1), as the constituent projects become restless in such a reformulation.

Rewards and dynamics for the reformulated project $m$ are as follows, where $R_m^{a_m}(a_m^-, i)$ and $p_m^{a_m}((a_m^-, i_m), (b_m^-, j_m))$ denote the *one-stage* (i.e., from $t_k$ to $t_{k+1}$) expected reward and transition probability, which results from taking action $a_m$ in state $Y_m(t_k) = (a_m^-, i_m)$. On the one hand, if, in period $t_k$, the project lies in state $(1, i_m)$ and it is again engaged, it yields the reward $R_m^1(1, i_m) \triangleq R_m(i_m)$ and its state transitions at $t_{k+1} = t_k + 1$ to $(1, j_m)$ with probability $p_m^1((1, i_m), (1, j_m)) \triangleq p_m(i_m, j_m)$. If, instead, the project is switched off, it gives the reward $R_m^0(1, i_m) \equiv -d_m(i_m)$ and its state moves at $t_{k+1} = t_k + \eta_m + 1$ to $(0, i_m)$ with probability 1, i.e., $p_m^0((1, i_m), (0, i_m)) \equiv 1$. On the other hand, if the project occupies at time $t_k$ the state $(0, i_m)$ and is then switched on, it yields the expected reward

$$R_m^1(0, i_m) \triangleq \mathbb{E}[-c_m(i_m) + \beta^{\xi_m(i_m)} R_m(i_m)] = -c_m(i_m) + \phi_m(i_m) R_m(i_m) \tag{2}$$

until the following decision time $t_{k+1} = t_k + \xi_m(i_m) + 1$, in which the project state transitions to $(1, j_m)$ with probability $p_m^1((0, i_m), (1, j_m)) \triangleq p_m(i_m, j_m)$. If the project is kept idle, then it gives no reward, i.e., $R_m^0(0, i_m) \equiv 0$, and its state remains frozen up to $t_{k+1} = t_k + 1$, so $p_m^0((0, i_m), (0, i_m)) \equiv 1$.

Thus, the MABPSP of concern is formulated as the semi-Markov MARBP

$$\underset{\boldsymbol{\pi}\in\boldsymbol{\Pi}}{\text{maximize}}\ \mathbb{E}_{\mathbf{Y}(0)}^{\boldsymbol{\pi}}\left[\sum_{k=0}^{\infty}\sum_{m=1}^{M} R_m^{A_m(t_k)}(Y_m(t_k))\beta^{t_k}\right], \tag{3}$$

where $\mathbb{E}_{\mathbf{Y}(0)}^{\boldsymbol{\pi}}[\cdot]$ is expectation under policy $\boldsymbol{\pi}$ conditioned on starting from the joint state $\mathbf{Y}(0)$.

### *2.1. Reduction to the Case with No Setdown Penalties*

We next show that one can restrict attention with no loss of generality to the case that there are no setdown penalties, which will allow for us to simplify subsequent analyses. Imagine that, say, at time $t = 0$, a passive project is set up and is then worked on for a random number of periods determined by a stopping time $\tau \geqslant 1$, after which it is set down. Dropping the label $m$, denote, by $\mathbf{R} = (R_j)_{j\in\mathcal{X}}$, $\mathbf{c} = (c_j)_{j\in\mathcal{X}}$, and $\mathbf{d} = (d_j)_{j\in\mathcal{X}}$, the active reward vector, and the setup and setdown cost vectors. Denote, by $\boldsymbol{\phi} = (\phi_j)_{j\in\mathcal{X}}$, the setup delay $z$-transform vector and by $\psi$ the constant setdown delay transform, both evaluated at $z = \beta$. The total discounted expected net reward that is obtained from the project over such a time interval, starting from the augmented state $Y(0) = (0, i)$, is

$$F_{(0,i)}^{\tau}(\mathbf{R}, \mathbf{c}, \mathbf{d}, \boldsymbol{\phi}, \psi) \triangleq \mathbb{E}_{(0,i)}^{\tau}\left[-c_i + \beta^{\xi_i}\sum_{t=0}^{\tau-1} R_{X(t)}\beta^t - d_{X(\tau)}\beta^{\xi_i+\tau}\right], \tag{4}$$

where $\xi_i$ is the setup delay. The corresponding discounted *active time* expended on the project is

$$G_{(0,i)}^{\tau}(\boldsymbol{\phi}, \psi) \triangleq \mathbb{E}_{(0,i)}^{\tau}\left[\frac{1-\beta^{\xi_i}}{1-\beta} + \beta^{\xi_i}\sum_{t=0}^{\tau-1}\beta^t + \frac{1-\beta^{\eta}}{1-\beta}\beta^{\xi_i+\tau}\right], \tag{5}$$

where, as pointed out above, the setup and setdown delays $\xi_i$ and $\eta$ are both considered to be active time.

In the next result, which extends Lemma 3.4 of [27] to the present setting, $\mathbf{I}$ is the identity matrix indexed by $\mathcal{X}$, $\mathbf{P} = (p_{ij})_{i,j\in\mathcal{X}}$, $\mathbf{0}$ is a vector of zeros, and $\boldsymbol{\phi}\cdot\mathbf{d} \triangleq (\phi_j d_j)_{j\in\mathcal{X}}$.

**Lemma 1.**

(a) $F^{\tau}_{(0,i)}(\mathbf{R},\mathbf{c},\mathbf{d},\boldsymbol{\phi},\psi) = F^{\tau}_{(0,i)}(\psi^{-1}(\mathbf{R}+(\mathbf{I}-\beta\mathbf{P})\mathbf{d}),\mathbf{c}+\boldsymbol{\phi}\cdot\mathbf{d},\mathbf{0},\psi\boldsymbol{\phi},1)$.

(b) $G^{\tau}_{(0,i)}(\boldsymbol{\phi},\psi) = G^{\tau}_{(0,i)}(\psi\boldsymbol{\phi},1)$.

**Proof.** (a) Use the identity

$$d_{X(\tau)}\beta^{\tau} = d_i - \sum_{t=0}^{\tau-1}(d_{X(t)} - \beta d_{X(t+1)})\beta^t$$

to write

$$\begin{aligned}
F^{\tau}_{(0,i)}(\mathbf{R},\mathbf{c},\mathbf{d},\boldsymbol{\phi},\psi) &\triangleq \mathbb{E}^{\tau}_{(0,i)}\Big[-c_i+\beta^{\xi_i}\sum_{t=0}^{\tau-1}R_{X(t)}\beta^t - d_{Y(\tau)}\beta^{\xi_i+\tau}\Big]\\
&= -c_i+\phi_i\mathbb{E}^{\tau}_{(0,i)}\left[\sum_{t=0}^{\tau-1}R_{X(t)}\beta^t - d_{Y(\tau)}\beta^{\tau}\right]\\
&= -c_i+\phi_i\left(-d_i+\mathbb{E}^{\tau}_{(0,i)}\left[\sum_{t=0}^{\tau-1}(R_{X(t)}+d_{X(t)}-\beta d_{X(t+1)})\beta^t\right]\right)\\
&= -c_i-\phi_i d_i+\phi_i\mathbb{E}^{\tau}_{(0,i)}\left[\sum_{t=0}^{\tau-1}(R_{X(t)}+d_{X(t)}-\beta d_{X(t+1)})\beta^t\right]\\
&= -c_i-\phi_i d_i+\phi_i\psi\mathbb{E}^{\tau}_{(0,i)}\left[\psi^{-1}\sum_{t=0}^{\tau-1}(R_{X(t)}+d_{X(t)}-\beta d_{X(t+1)})\beta^t\right]\\
&= F^{\tau}_{(0,i)}(\psi^{-1}(\mathbf{R}+(\mathbf{I}-\beta\mathbf{P})\mathbf{d}),\mathbf{c}+\boldsymbol{\phi}\cdot\mathbf{d},\mathbf{0},\psi\boldsymbol{\phi},1).
\end{aligned}$$

(b) This part follows by writing

$$\begin{aligned}
G^{\tau}_{(0,i)} &\triangleq \mathbb{E}^{\tau}_{(0,i)}\left[\frac{1-\beta^{\xi_i}}{1-\beta}+\beta^{\xi_i}\sum_{t=0}^{\tau-1}\beta^t+\frac{1-\beta^{\eta}}{1-\beta}\beta^{\xi_i+\tau}\right] = \frac{1-\phi_i}{1-\beta}+\phi_i\mathbb{E}^{\tau}_{i}\left[\sum_{t=0}^{\tau-1}\beta^t+\frac{1-\psi}{1-\beta}\beta^{\tau}\right]\\
&= \frac{1-\phi_i}{1-\beta}+\phi_i\mathbb{E}^{\tau}_{(0,i)}\left[\sum_{t=0}^{\tau-1}\beta^t+\frac{1-\psi}{1-\beta}\left(1-(1-\beta)\sum_{t=0}^{\tau-1}\beta^t\right)\right]\\
&= \frac{1-\phi_i\psi}{1-\beta}+\phi_i\mathbb{E}^{\tau}_{(0,i)}\left[\sum_{t=0}^{\tau-1}(1-(1-\psi))\beta^t\right] = \frac{1-\phi_i\psi}{1-\beta}+\phi_i\psi\mathbb{E}^{\tau}_{(0,i)}\left[\sum_{t=0}^{\tau-1}\beta^t\right]\\
&= G^{\tau}_{(0,i)}(\psi\boldsymbol{\phi},1).
\end{aligned}$$

□

Lemma 1 can be used in order to eliminate setdown penalties: it suffices to incorporate them into new setup costs, setup delay transforms, and active rewards, while using the transformations

$$\widetilde{c}_j \triangleq c_j+\phi_j d_j, \quad \widetilde{\phi}_j \triangleq \psi\phi_j, \quad \text{and} \quad \widetilde{\mathbf{R}} \triangleq \psi^{-1}(\mathbf{R}+(\mathbf{I}-\beta\mathbf{P})\mathbf{d}). \tag{6}$$

Note that such a reduction would not have been accomplished had the setdown delay transform not been constant. In the case $c_j \equiv c$ and $d_j \equiv d$, we obtain $\widetilde{c}_j \equiv c+d\phi_j$ and $\widetilde{R}_j = (R_j+(1-\beta)d)/\psi$.

Accordingly, we will focus henceforth on the normalized case without setdown penalties $d_j \equiv 0$, $\psi \equiv 1$.

### *2.2. The AT Index*

We next consider the AT index of a project with setup penalties—dropping again the label $m$—extending the original definitions in [10]. The continuation AT index is

$$\lambda^{\mathrm{AT}}_{(1,i)} \triangleq \max_{\tau \geqslant 1} \frac{\mathbb{E}^{\tau}_{i}\left[\sum_{t=0}^{\tau-1} R_{X(t)}\beta^{t}\right]}{\mathbb{E}^{\tau}_{i}\left[\sum_{t=0}^{\tau-1} \beta^{t}\right]}, \tag{7}$$

where $\tau \geqslant 1$ is a stopping time for engaging the project starting in state $i$ when it is already set up; hence, $\lambda^{\mathrm{AT}}_{(1,i)}$ is just the project's Gittins index. As for the switching AT index, it is given by

$$\lambda^{\mathrm{AT}}_{(0,i)} \triangleq \max_{\tau \geqslant 1} \frac{-c_i + \mathbb{E}^{\tau}_{i}\left[\beta^{\xi_i}\sum_{t=0}^{\tau-1} R_{X(t)}\beta^{t}\right]}{\mathbb{E}^{\tau}_{i}\left[\sum_{t=0}^{\xi_i-1}\beta^{t} + \beta^{\xi_i}\sum_{t=0}^{\tau-1}\beta^{t}\right]} = \max_{\tau \geqslant 1} \frac{-c_i + \phi_i \mathbb{E}^{\tau}_{i}\left[\sum_{t=0}^{\tau-1} R_{X(t)}\beta^{t}\right]}{\frac{1-\phi_i}{1-\beta} + \phi_i \mathbb{E}^{\tau}_{i}\left[\sum_{t=0}^{\tau-1}\beta^{t}\right]}, \tag{8}$$

where now $\tau$ is a stopping-time rule that is followed after the project has been set up in state $i$.

The following requirements will be assumed henceforth on setup costs and setup delay transforms, which extend the corresponding conditions in [10].

**Assumption 1.** *The following holds:*

(i) *non-negative setup costs: $c_j \geqslant 0$ for $j \in \mathcal{X}$.*

(ii) *non-negative rewards: If some setup delay can be positive, i.e., $\boldsymbol{\phi} \neq \mathbf{1}$, then $R_j \geqslant 0$ for $j \in \mathcal{X}$.*

The next result shows that Assumption 1 ensures the satisfaction of the hysteresis property in (1).

**Lemma 2.** *Under Assumption 1, $\lambda^{\mathrm{AT}}_{(1,i)} \geqslant \lambda^{\mathrm{AT}}_{(0,i)}$ for $i \in \mathcal{X}$.*

**Proof.** For a given state $i \in \mathcal{X}$ and stopping-time rule $\tau$ as above, write $G^{\tau}_{i} \triangleq \mathbb{E}^{\tau}_{i}[\sum_{t=0}^{\tau-1}\beta^{t}]$ and $F^{\tau}_{i} \triangleq \mathbb{E}^{\tau}_{i}[\sum_{t=0}^{\tau-1} R_{X(t)}\beta^{t}]$. Now, Assumption 1 ensures that $c_i \geqslant 0$ and $F^{\tau}_{i} \geqslant 0$, and hence

$$\frac{F^{\tau}_{i}}{G^{\tau}_{i}} - \frac{-c_i + \phi_i F^{\tau}_{i}}{\frac{1-\phi_i}{1-\beta} + \phi_i G^{\tau}_{i}} = \frac{1}{G^{\tau}_{i}} \frac{(1-\beta)c_i G^{\tau}_{i} + (1-\phi_i)F^{\tau}_{i}}{1-\phi_i + (1-\beta)\phi_i G^{\tau}_{i}} \geqslant 0, \tag{9}$$

Further, (9), (7), and (8) immediately yield that $\lambda^{\mathrm{AT}}_{(1,i)} \geqslant \lambda^{\mathrm{AT}}_{(0,i)}$, which completes the proof. □

## 3. New Methodological Results on Restless Bandit Indexation

This section presents new results on restless bandit indexation, which, besides having an intrinsic interest, are required and form the basis for the approach to non-restless bandits with switching times that is deployed in later sections.

### *3.1. Indexable Restless Bandits and the Whittle Index*

Consider a semi-Markov restless bandit, representing a dynamic and stochastic project whose state $Y(t)$ transitions over time periods $t = 0, 1, 2, \ldots$ through the finite state space $\mathcal{Y}$. The project's evolution is governed by a policy $\pi$ that is taken from the class $\Pi$ of nonanticipative randomized policies, which, at each of an increasing sequence $t_k$ of

decision periods with $t_0 = 0$ and $t_k \nearrow \infty$ as $k \nearrow \infty$, prescribes an action $A(t_k) \in \{0,1\}$ that determines the status during the ensuing *stage* until the next decision period $t_{k+1}$ (1: active; 0: passive). Taking action $A(t_k) = a$ at time $t_k$ when the project occupies state $Y(t_k) = y$ has the following consequences over the following stage, relative to a given one-period discount factor $0 < \beta < 1$: an expected total discounted amount of reward $R_y^a$ and of a generic resource $Q_y^a \geqslant 0$ is earned and expended, respectively; further, the joint distribution of the stage's duration $t_{k+1} - t_k$ and its final state $Y(t_{k+1})$ is given through the discounted transition transform $\phi_{yy'}^a \triangleq \mathbb{E}\left[\beta^{t_{k+1}-t_k} 1_{\{Y(t_{k+1})=y'\}} \,|\, Y(t_k) = y, A(t_k) = a\right]$, where $1_{\{\cdot\}}$ denotes an event indicator.

It will be convenient to partition $\mathcal{Y}$ into the (possibly empty) set of *uncontrollable states*

$$\mathcal{Y}^{\{0\}} \triangleq \left\{i \in \mathcal{Y} \colon Q_y^0 = Q_y^1 \text{ and } \phi_{yy'}^0 \equiv \phi_{yy'}^1,\ y \in \mathcal{Y}\right\},$$

where both actions entail identical resource consumptions and dynamics, and the remaining set $\mathcal{Y}^{\{0,1\}} \triangleq \mathcal{Y} \setminus \mathcal{Y}^{\{0\}}$ of *controllable states*, which is assumed to consist of $N = |\mathcal{Y}^{\{0,1\}}| \geqslant 1$ elements. The notation $\mathcal{Y}^{\{0\}}$ is meant to reflect the convention that the passive action $a = 0$ is chosen in uncontrollable states.

The value of the rewards earned and amount of resource expended by a policy $\pi$ starting from state $y$ is evaluated, respectively, by the discounted reward and resource consumption metrics

$$F_y^\pi \triangleq \mathbb{E}_y^\pi\left[\sum_{k=0}^{\infty} R_{Y(t_k)}^{A(t_k)} \beta^{t_k}\right] \quad \text{and} \quad G_y^\pi \triangleq \mathbb{E}_y^\pi\left[\sum_{k=0}^{\infty} Q_{Y(t_k)}^{A(t_k)} \beta^{t_k}\right].$$

Let us introduce a parameter $\lambda$ representing the resource unit price, and consider the *$\lambda$-price problem*

$$\underset{\pi \in \Pi}{\text{maximize}}\ F_y^\pi - \lambda G_y^\pi, \tag{10}$$

which concerns finding a policy that maximizes the value of rewards earned minus the cost of resources expended. Because (10) is an infinite-horizon finite-state and -action SMDP, by standard results it is solved by stationary deterministic policies that are characterized by the solutions to the following DP equations, where $V_y^*(\lambda)$ denotes the optimal value starting from $y$ under price $\lambda$:

$$V_y^*(\lambda) = \max_{a \in \{0,1\}} R_y^a - \lambda Q_y^a + \sum_{y' \in \mathcal{Y}} \phi_{yy'}^a V_{y'}^*(\lambda), \quad y \in \mathcal{Y}. \tag{11}$$

Such a project is said to be *indexable* (cf. [29]), if, for each controllable state $y \in \mathcal{Y}^{\{0,1\}}$, there exists a unique break-even price $\lambda_y^*$, such that: it is optimal to engage the project in state $y$ if and only if $\lambda \leqslant \lambda_y^*$, and it is optimal to rest it if and only if $\lambda \geqslant \lambda_y^*$. Or, in terms of the DP Equation (11),

$$R_y^1 - \lambda Q_y^1 + \sum_{y' \in \mathcal{Y}} \phi_{yy'}^1 V_{y'}^*(\lambda) \geqslant R_y^0 - \lambda Q_y^0 + \sum_{y' \in \mathcal{Y}} \phi_{yy'}^0 V_{y'}^*(\lambda) \iff \lambda_y^* \geqslant \lambda, \quad y \in \mathcal{Y}^{\{0,1\}}$$

and

$$R_y^1 - \lambda Q_y^1 + \sum_{y' \in \mathcal{Y}} \phi_{yy'}^1 V_{y'}^*(\lambda) \leqslant R_y^0 - \lambda Q_y^0 + \sum_{y' \in \mathcal{Y}} \phi_{yy'}^0 V_{y'}^*(\lambda) \iff \lambda^* \leqslant \lambda_y, \quad y \in \mathcal{Y}^{\{0,1\}}.$$

We will refer to the mapping $i \mapsto \lambda_y^*$ as the project's *Whittle index*. See [29].

### *3.2. Exploiting Special Structure: Indexability Relative to a Family of Policies*

While one can readily numerically test whether a given restless bandit instance is indexable, a researcher investigating a particular restless bandit model will instead be

concerned with analytically establishing its indexability under an appropriate range of model parameters. The key to achieving such a goal is—as in optimal-stopping problems—to exploit special structure by *guessing* a family of policies (stationary deterministic), among which there exists an optimal policy for (10) for every resource price $\lambda \in \mathbb{R}$.

We represent a stationary deterministic policy by its *active (state) set*, consisting of those controllable states where it prescribes engaging the project. Thus, a family of such policies is given as a family $\mathcal{F}$ of active sets $S \subseteq \mathcal{Y}^{\{0,1\}}$, and, hence, we will refer to the family of $\mathcal{F}$*-policies*. Relative to such a family, we will call the project $\mathcal{F}$*-indexable* if (i) it is indexable, and (ii) $\mathcal{F}$-policies are optimal for $\lambda$-price problem (10) for every resource price $\lambda \in \mathbb{R}$.

We will impose the following connectivity requirements on $\mathcal{F}$.

**Assumption 2.** *The active-set family $\mathcal{F}$ satisfies the following conditions*:

(i) $\emptyset, \mathcal{Y}^{\{0,1\}} \in \mathcal{F}$;
(ii) *for any $S, S' \in \mathcal{F}$, with $S \subset S'$, there exist $y, y' \in S' \setminus S$ such that $S \cup \{y\}, S' \setminus \{y'\} \in \mathcal{F}$;*
(iii) *for any $S, S' \in \mathcal{F}$, $S \cup S', S \cap S' \in \mathcal{F}$.*

Note that condition (iii) in Assumption 2 means that $\mathcal{F}$ is a *lattice* relative to set inclusion. As for condition (ii), it ensures that any two nested active sets $S, S' \in \mathcal{F}$ with $S \subset S'$ can be connected by an increasing *chain* $S = S_0 \subset \cdots \subset S_k = S'$ of *adjacent* (i.e., differing by one state) sets in $\mathcal{F}$. Further, condition (i) ensures that one can connect in such a fashion $\emptyset$ with $\mathcal{Y}^{\{0,1\}}$. We will call a set family $\mathcal{F}$ satisfying Assumption 2(ii, iii) a *monotonically connected lattice*.

### *3.3. New Sufficient Conditions for $\mathcal{F}$-Indexability and Adaptive-Greedy Index Algorithm*

Suppose that, for a particular restless bandit model, a suitable active-set family $\mathcal{F}$, as above, has been posited relative to which one aims to analytically establish $\mathcal{F}$-indexability. While, in the aforementioned earlier work of the author, sufficient conditions for $\mathcal{F}$-indexability are given, which further ensure that the project's Whittle index can be computed by using an adaptive-greedy index algorithm that was introduced in such work, we next introduce new sufficient conditions that are significantly less restrictive.The new conditions are motivated by the model of concern in this paper, as we will see that it need not satisfy the former conditions, as mentioned in Section 1.

In order to formulate the new conditions and the index algorithm we need to define certain *marginal metrics*, as follows. Given an action $a \in \{0,1\}$ and active set $S \subseteq \mathcal{Y}^{\{0,1\}}$, write, as $\langle a, S \rangle$, the policy that initially chooses action $a$, and then follows the *S-active policy*. For a given state $y$ and active set $S$, consider the *marginal work metric*

$$g_y^S \triangleq G_y^{\langle 1,S \rangle} - G_y^{\langle 0,S \rangle}, \tag{12}$$

which represents the marginal increase in the amount of resource expended resulting from taking first the active rather than the passive action and, then, following the $S$-active policy. Note that such a marginal work metric vanishes at uncontrollable states:

$$g_y^S = 0, \quad y \in \mathcal{Y}^{\{0\}}. \tag{13}$$

Further, define the *marginal reward metric*

$$f_y^S \triangleq F_y^{\langle 1,S \rangle} - F_y^{\langle 0,S \rangle}, \tag{14}$$

which represents the marginal increase in rewards earned. Finally, for $g_y^S \neq 0$, define the *marginal productivity metric*

$$\lambda_y^S \triangleq \frac{f_y^S}{g_y^S}. \tag{15}$$

We will consider the *adaptive-greedy index algorithm* that is given in Algorithm 1 in its *top-down* version, where index values are meant to be computed from highest to lowest; one could similarly consider the symmetric *bottom-up* version. Such an algorithm has a very simple structure, as it constructs in $n$ steps (recall that $N \triangleq |\mathcal{Y}^{\{0,1\}}|$), an increasing chain of successive active sets $S^0 = \emptyset \subset S^1 \subset \ldots \subset S^N = \mathcal{Y}^{\{0,1\}}$ in $\mathcal{F}$, proceeding at each step in a greedy fashion. Thus, once active set $S^{k-1} \in \mathcal{F}$ has been obtained, the next active set $S^k$ is constructed by augmenting $S^{k-1}$ with a controllable state $y \in \mathcal{Y}^{\{0,1\}} \setminus S^{k-1}$ that maximizes marginal productivity metric $\lambda_y^{S^{k-1}}$, restricting attention to states $y$ for which the following active set is in $\mathcal{F}$, so $S^k = S^{k-1} \cup \{y\} \in \mathcal{F}$. Ties are broken arbitrarily.

Note that Algorithm 1 only shows an algorithmic scheme, as it is not specified how to compute the metrics that are required for computations. A complete fast-pivoting implementation of such an algorithm is given by the author in [49].

Additionally, note that the algorithm's input consists of all the project's primitive parameters, namely states, rewards, transition probabilities, and discount factor.

The same considerations apply to Algorithm 2.

**Algorithm 1:** Top-down adaptive-greedy index algorithm $\mathrm{AG}_{\mathcal{F}}$.

**Output:** $\{y_k, \lambda^*_{y_k}\}_{k=1}^N$

$S^0 := \emptyset$
**for** $k := 1$ **to** $N$ **do**
  **choose** $y_k \in \arg\max \left\{\lambda_y^{S^{k-1}} : y \in \mathcal{Y}^{\{0,1\}} \setminus S^{k-1}, S^{k-1} \cup \{y\} \in \mathcal{F}\right\}$
  $\lambda^*_{y_k} := \lambda_{y_k}^{S^{k-1}}$; $S^k := S^{k-1} \cup \{y_k\}$
**end** { for }

The main result of this section, giving the new indexability conditions and ensuring the validity of the adaptive-greedy index algorithm for computing the Whittle index, is stated next.

**Algorithm 2:** Geometrically intuitive reformulation of adaptive-greedy index algorithm $\mathrm{AG}_{\mathcal{F}}$.

**Output:** $\{y_k, \lambda^*_{y_k}\}_{k=1}^N$

$S^0 := \emptyset$
**for** $k := 1$ **to** $N$ **do**
  **choose** $j_k \in \arg\max \left\{ \dfrac{F^{S^{k-1}\cup\{j\}} - F^{S^{k-1}}}{G^{S^{k-1}\cup\{y\}} - G^{S^{k-1}}} : y \in \mathcal{Y}^{\{0,1\}} \setminus S^{k-1}, S^{k-1} \cup \{y\} \in \mathcal{F} \right\}$
  $\lambda^*_{y_k} := \lambda_{y_k}^{S^{k-1}}$; $S^k := S^{k-1} \cup \{y_k\}$
**end** { for }

**Theorem 1.** *The following holds:*

(a) *Suppose that the project satisfies the following conditions*:

(i) *for every active set* $S \in \mathcal{F}$,

$$\begin{aligned} &g_y^S > 0, \quad y \in S, S \setminus \{y\} \in \mathcal{F}, \\ &g_y^S > 0, \quad y \in \mathcal{Y}^{\{0,1\}} \setminus S, S \cup \{y\} \in \mathcal{F}; \end{aligned} \tag{16}$$

*or, equivalently, for every nested active-set pair* $S \subset S'$ *with* $S, S' \in \mathcal{F}$,

$$\left(G_y^S\right)_{y\in\mathcal{Y}} \lesssim \left(G_y^{S'}\right)_{y\in\mathcal{Y}}. \tag{17}$$

(ii) *for every resource price* $\lambda \in \mathbb{R}$*, there exists an optimal* $\mathcal{F}$*-policy for* $\lambda$*-price problem* (10).

*Then, the project is* $\mathcal{F}$*-indexable and algorithm* $\mathrm{AG}_{\mathcal{F}}$ *computes its Whittle index* $\lambda^*_{y_k}$ *in non-increasing order.*

(b) *If the project is indexable, then it satisfies conditions* (i) *and* (ii) *in part* (a) *for some nested family of adjacent active sets of the form* $\mathcal{F} = \{S^0, S^1, \ldots, S^N\}$ *with* $S^0 = \emptyset \subset S^1 \subset \ldots \subset S^N = \mathcal{Y}^{\{0,1\}}$.

In order to prove Theorem 1, we need to establish a number of preliminary results. Before doing so, let us clarify the improvement that the new sufficient $\mathcal{F}$-indexability conditions (i) and (ii) in Theorem 1(a) represent over those that were introduced in Niño-Mora [30,31] based on PCLs, which are:

(i) for every $S \in \mathcal{F}$, $g^S_y > 0$ for $y \in \mathcal{Y}^{\{0,1\}}$;

(ii) algorithm $\mathrm{AG}_{\mathcal{F}}$ computes index $\lambda^*_{y_k}$ in non-increasing order: $\lambda^*_{y_1} \geqslant \lambda^*_{y_2} \geqslant \cdots \geqslant \lambda^*_{y_N}$.

Thus, the new condition (i) in Theorem 1(a), as formulated in (16), is significantly less stringent than the old condition (i). Further, the reformulation in (17) clarifies its intuitive meaning: it means that resource consumption metric $G^S_y$ is monotone non-decreasing in the active set $S$ within the domain $\mathcal{F}$, and that two nested active sets $S \subset S'$ in $\mathcal{F}$ give different resource consumption vectors $(G^S_y)_{y\in\mathcal{Y}}$ and $(G^{S'}_y)_{y\in\mathcal{Y}}$.

As for the old condition (ii), the author has found that, in complex models with a multidimensional state, it can be elusive to establish it analytically. In contrast, the new condition (ii) in Theorem 1(a) allows one either to draw on the rich literature available on optimality of structured policies for special models, or to deploy ad hoc DP arguments to prove the optimality of $\mathcal{F}$-policies for the model at hand.

Note that [50] has proposed sufficient $\mathcal{F}$-indexability conditions, which are, however, significantly more restrictive than those herein. Thus, the conditions in [50] require, among further assumptions, including (i) and (ii) in Theorem 1(a), that the resource metric be submodular and reward metric be supermodular in the active set. Theorem 1(a) shows that such extra assumptions are unnecessary.

Theorem 1(b) further assures that the new conditions are also necessary for indexability, in the sense that any indexable restless bandit satisfies them relative to some nested active-set family $\mathcal{F}$, as stated.

We start by establishing the equivalence between the formulations in (16) and (17) of condition (i) in Theorem 1(a), by drawing on the results in Niño-Mora (Sect. 6 of [31]) (for Markovian restless bandits) and in Niño-Mora (Sect. 4 of [32]) for semi-Markov restless bandits. These refer to relations between resource and reward metrics and their marginal counterparts, via *state-action occupancy measures*

$$x^{a,\pi}_{yy'} \triangleq \mathbb{E}^{\pi}_y \left[ \sum_{k=0}^{\infty} 1_{\{Y(t_k)=y', A(t_k)=a\}} \beta^{t_k} \right]. \tag{18}$$

Note that $x^{a,\pi}_{yy'}$ measures the expected total discounted number of decision periods, in which action $a$ is chosen in state $y'$ while using policy $\pi$, starting from state $y$. In the present notation, the relevant relations are

$$\begin{aligned} G^{S\setminus\{y'\}}_y &= G^S_y - g^S_{y'} x^{0,S\setminus\{y'\}}_{yy'}, \quad y' \in S \\ G^{S\cup\{y'\}}_y &= G^S_y + g^S_{y'} x^{1,S\cup\{y'\}}_{yy'}, \quad y' \in \mathcal{Y}^{\{0,1\}} \setminus S, \end{aligned} \tag{19}$$

and

$$\begin{aligned} F^{S\setminus\{y'\}}_y &= F^S_y - f^S_{y'} x^{0,S\setminus\{y'\}}_{yy'}, \quad y' \in S \\ F^{S\cup\{y'\}}_y &= F^S_y + f^S_{y'} x^{1,S\cup\{y'\}}_{yy'}, \quad y' \in \mathcal{Y}^{\{0,1\}} \setminus S. \end{aligned} \tag{20}$$

**Lemma 3.** *Conditions* (16) *and* (17) *in Theorem* 1(a) *are equivalent.*

**Proof.** Suppose that (16) holds for a certain $S \in \mathcal{F}$. We then have, on the one hand, that $g^S_{y'} > 0$ for $y' \in S$ such that $S \setminus \{y'\} \in \mathcal{F}$, along with $x^{0,S\setminus\{j\}}_{yy'} \geqslant 0$ for any $y$, implies, via the first identity in (19), that $G^{S\setminus\{y'\}}_y \leqslant G^S_y$; further, by taking $y = y'$, we obtain $G^{S\setminus\{y'\}}_{y'} < G^S_{y'}$, since $x^{0,S\setminus\{y'\}}_{y'y'} > 0$. Hence, we have $\left(G^{S\setminus\{y'\}}_y\right)_{i\in\mathcal{Y}} \lneq \left(G^S_y\right)_{y\in\mathcal{Y}}$, for such $y'$. On the other hand, we have that $g^S_{y'} > 0$ for $y' \in \mathcal{Y}^{\{0,1\}} \setminus S$ such that $S \cup \{y'\} \in \mathcal{F}$, along with $x^{1,S\cup\{y'\}}_{yy'} \geqslant 0$ for any $y$, implies, via the second identity in (19), that $G^S_y \leqslant G^{S\cup\{j\}}_y$; further, by taking $y = y'$, we obtain $G^S_{y'} < Gy'^{S\cup\{y'j\}}$, since $x^{1,S\cup\{y'\}}_{y'y'} > 0$. Hence, we have $\left(G^S_y\right)_{y\in\mathcal{Y}} \lneq \left(G^{S\cup\{y'\}}_y\right)_{y\in\mathcal{Y}}$ for such $y'$. Now, the proven relations imply (17) via Assumption 2(ii).

Conversely, suppose that (17) holds for a certain $S \in \mathcal{F}$. Then, on the one hand, we have $\left(G^{S\setminus\{y'\}}_y\right)_{y\in\mathcal{Y}} \lneq \left(G^S_y\right)_{y\in\mathcal{Y}}$ for $y' \in S$ such that $S \setminus \{y'\} \in \mathcal{F}$. This, along with $x^{0,S\setminus\{j'\}}_{yy'} \geqslant 0$ for every $y$ implies, via the first identity in (19), that $g^S_{y'} > 0$ for such $y'$. On the other hand, we have $\left(G^S_y\right)_{y\in\mathcal{Y}} \lneq \left(G^{S\cup\{y'\}}_y\right)_{y\in\mathcal{Y}}$ for $y' \in \mathcal{Y}^{\{0,1\}} \setminus S$ such that $S \cup \{y'\} \in \mathcal{F}$. This, along with $x^{1,S\cup\{y'\}}_{yy'} \geqslant 0$ for every $y$ implies, via the second identity in (19), that $g^S_{y'} > 0$ for such $y'$. Therefore, (16) holds, which completes the proof. □

### *3.4. Proving Theorem 1: Achievable Resource-Reward Performance Region Approach*

We next deploy an approach in order to prove Theorem 1, which draws on first principles via an intuitive geometric and economic viewpoint introduced in [31,32]. We will find it convenient to consider, instead of (10), the $\lambda$-price problem that is obtained by using the averaged resource and reward metrics where the initial project state $Y(0)$ is drawn from a distribution $p$ with positive probability mass $p_y > 0$ at every state $y \in \mathcal{Y}$,

$$G^\pi \triangleq \sum_{y\in\mathcal{Y}} p_y G^\pi_y \quad \text{and} \quad F^\pi \triangleq \sum_{y\in\mathcal{Y}} p_y F^\pi_y, \tag{21}$$

i.e.,

$$\underset{\pi\in\Pi}{\text{maximize}}\ F^\pi - \lambda G^\pi. \tag{22}$$

Relative to such metrics, consider the project's *achievable resource-reward performance region*

$$\mathcal{H} \triangleq \left\{(G^\pi, F^\pi)\colon \pi \in \Pi\right\}, \tag{23}$$

which is defined as the region in the resource-reward plane that consists of all the performance points $(G^\pi, F^\pi)$ that can be achieved under admissible project operating policies $\pi \in \Pi$. The optimality of stationary deterministic policies for infinite-horizon finite-state and -action SMDPs ensures that $\mathcal{H}$ is the *closed convex polygon* spanned as the *convex hull* of points $(G^S, F^S)$ for active sets $S \subseteq \mathcal{Y}^{\{0,1\}}$. Thus, we can reformulate $\lambda$-price problem (22) as the *linear programming* (LP) problem

$$\underset{(G,F)\in\mathcal{H}}{\text{maximize}}\ F - \lambda G. \tag{24}$$

In order to illustrate and clarify such an approach, consider the concrete example of a certain restless bandit having state space $\mathcal{Y} = \mathcal{Y}^{\{0,1\}} = \{1, 2, 3\}$ that is discussed in (Sec. 2.2 of [34]) For such a project, Figure 1, in that paper, plots the achievable resource-reward performance region $\mathcal{H}$, with points $(G^S, F^S)$ being labeled by their active sets $S$.

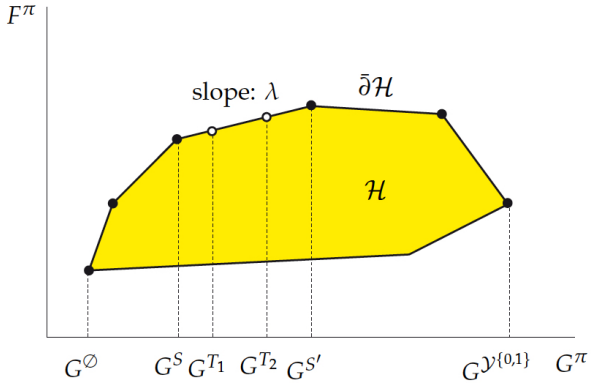


**Figure 1.** Illustration for the proof of Theorem 1.

The fact that such a project is indexable is apparent from the structure of the *upper boundary* of $\mathcal{H}$,

$$\bar{\partial}\mathcal{H} \triangleq \big\{(G,F) \in \mathcal{H} \colon \tilde{F} \leqslant F \text{ for every } (\tilde{G},\tilde{F}) \in \mathcal{H} \text{ having } \tilde{G} = G\big\}, \tag{25}$$

as this is determined from left to right by an *increasing nested family of adjacent active sets connecting* $\emptyset$ *to* $\mathcal{Y}^{\{0,1\}}$: $\mathcal{F} = \big\{\emptyset, \{1\}, \{1,2\}, \{1,2,3\}\big\}$. Thus, the Whittle indices of the states are given by the successive *slopes* measuring the *marginal reward versus resource trade-off rates*:

$$\lambda_1^* = \frac{F^{\{1\}} - F^{\emptyset}}{G^{\{1\}} - G^{\emptyset}} \geqslant \lambda_2^* = \frac{F^{\{1,2\}} - F^{\{1\}}}{G^{\{1,2\}} - G^{\{1\}}} \geqslant \lambda_3^* = \frac{F^{\{1,2,3\}} - F^{\{1,2\}}}{G^{\{1,2,3\}} - G^{\{1,2\}}}. \tag{26}$$

In this example, the geometry of the top-down adaptive-greedy algorithm $\mathrm{AG}_{\mathcal{F}}$ corresponds to traversing the upper boundary $\bar{\partial}\mathcal{H}$ from left to right, proceeding, at each step, by augmenting the current active set by a new state in a locally greedy fashion, as the slopes in (26) are equivalently formulated as

$$\lambda_1^* = \frac{f_1^{\emptyset}}{g_1^{\emptyset}} \geqslant \lambda_2^* = \frac{f_2^{\{1\}}}{g_2^{\{1\}}} \geqslant \lambda_3^* = \frac{f_3^{\{1,2\}}}{g_3^{\{1,2\}}}. \tag{27}$$

The insights that are conveyed by such an example extend to the general setting of concern herein, as elucidated in Niño-Mora [31,32,34]. Thus, the indexability of a project is recast as a property of the upper boundary $\bar{\partial}\mathcal{H}$ of region $\mathcal{H}$, whereby it is determined by a nested active-set family as in the example. Note that the equivalence between the geometric slopes in (27) and the marginal productivity rates (26) in follow from (19) and (20) or, more precisely, from the corresponding relations for the averaged metrics,

$$\begin{aligned} G^{S\setminus\{y'\}} &= G^S - g_{y'}^S x_{y'}^{0,S\setminus\{y'\}}, \quad y' \in S \\ G^{S\cup\{y'\}} &= G^S + g_{y'}^S x_{y'}^{1,S\cup\{y'\}}, \quad y' \in \mathcal{Y}^{\{0,1\}} \setminus S, \end{aligned} \tag{28}$$

and

$$\begin{aligned} F^{S\setminus\{y'\}} &= F^S - f_{y'}^S x_{y'}^{0,S\setminus\{y'\}}, \quad y' \in S \\ F^{S\cup\{y'\}} &= F^S + f_{y'}^S x_{y'}^{1,S\cup\{y'\}}, \quad y' \in \mathcal{Y}^{\{0,1\}} \setminus S, \end{aligned} \tag{29}$$

where $x_{y'}^{a,\pi}$ is the state-action occupancy measure that is obtained by drawing the initial state according to the probabilities $p_y$. Thus, assuming condition (i) in Theorem 1(a), we have, for $S \in \mathcal{F}$,

$$\frac{f_{y'}^{S}}{g_{y'}^{S}} = \begin{cases} \dfrac{F^S - F^{S\setminus\{y'\}}}{G^S - G^{S\setminus\{y'\}}}, & y' \in S, S \setminus \{y'\} \in \mathcal{F} \\ \dfrac{F^{S\cup\{y'\}} - F^S}{G^{S\cup\{y'j\}} - G^S}, & y' \in \mathcal{Y}^{\{0,1\}} \setminus S, S \cup \{y'\} \in \mathcal{F}. \end{cases} \tag{30}$$

Such relations allow for us to reformulate the adaptive-greedy algorithm $\mathrm{AG}_{\mathcal{F}}$ in Algorithm 1 into the geometrically intuitive form that is shown in Algorithm 2. Such a reformulation clarifies that this algorithm seeks to traverse, from left to right, the upper boundary $\bar{\partial}\mathcal{H}$, proceeding at each step by augmenting the current active set by a new state in a locally greedy fashion, while only using active sets in $\mathcal{F}$.

We next proceed to establish a number of preliminary results, on which the proof of Theorem 1 will draw. The first shows that the family of optimal active sets for the $\lambda$-price problem is a lattice that contains its intervals.

**Lemma 4.** *If $S$ and $S'$ are optimal active sets for (22), then so is any $S''$ satisfying $S \cap S' \subseteq S'' \subseteq S \cap S''$.*

**Proof.** The result is an immediate property of the DP Equations (11) characterizing the optimal stationary deterministic policies (i.e., the optimal active sets) for the $\lambda$-price problem. □

The following result shows that, under condition (i) in Theorem 1(a), resource consumption metric $G^S$ is strictly increasing relative to active-set inclusion in the domain $S \in \mathcal{F}$.

**Lemma 5.** *Suppose that condition* (i) *in Theorem* 1(a) *holds. Then, $G^S < G^{S'}$ for $S \subset S'$, $S, S' \in \mathcal{F}$.*

**Proof.** The result follows immediately from the formulation of such a condition (i) in (17), along with the assumption of positive initial state probabilities $p_y > 0$ for $y \in \mathcal{Y}$. □

The next result establishes, under conditions (i) and (ii) in Theorem 1(a), the *non-degeneracy* of the extreme points of $\mathcal{H}$ in upper boundary $\bar{\partial}\mathcal{H}$, showing that each is achieved by a unique active set in $\mathcal{F}$.

**Lemma 6.** *Suppose that conditions* (i) *and* (ii) *in Theorem* 1(a) *hold. Then, for every $(G^*, F^*) \in \bar{\partial}\mathcal{H}$ that is an extreme point of $\mathcal{H}$, there exists a unique active set $S^* \in \mathcal{F}$ achieving it, i.e., with $(G^*, F^*) = (G^{S'}, F^{S'})$.*

**Proof.** Because $(G^*, F^*)$ is an extreme point of $\mathcal{H}$ in $\bar{\partial}\mathcal{H}$, there exists a resource price $\lambda^*$, such that $(G^*, F^*)$ is the unique solution to the LP problem (24) for $\lambda = \lambda^*$. Now, condition (ii) in Theorem 1 ensures that there exists an active set $S^* \in \mathcal{F}$ that is optimal for $\lambda^*$-price problem (22), i.e., such that $(G^*, F^*) = (G^{S^*}, F^{S^*})$. Let us argue, by contradiction, that such an active set is unique, assuming that there exists a different active set $S^{**} \in \mathcal{F}$, for which $(G^*, F^*) = (G^{S^{**}}, F^{S^{**}})$. Then, by Assumption 2(iii) and Lemma 4, both $S^* \cap S^{**}$ and $S^* \cup S^{**}$ would belong in $\mathcal{F}$ and be optimal for the $\lambda^*$-price problem. Therefore,

$$(G^*, F^*) = (G^{S^*}, F^{S^*}) = (G^{S^* \cap S^{**}}, F^{S^* \cap S^{**}}) = (G^{S^* \cup S^{**}}, F^{S^* \cup S^{**}}). \tag{31}$$

Now, since it is assumed that $S^* \neq S^{**}$, there are two cases to consider: in the first case, if it were $S^* \not\subset S^{**}$, then it would be $S^* \cap S^{**} \subset S^* \subset S^* \cup S^{**}$ and, hence, by Lemma 5,

$G^{S^* \cap S^{**}} < G^{S^*} < G^{S^* \cap S^{**}}$, which contradicts (31). In the second case, if it were $S^{**} \not\subset S^*$, then it would be $S^* \cap S^{**} \subset S^{**} \subset S^* \cup S^{**}$ and, hence, by Lemma 5, $G^{S^* \cap S^{**}} < G^{S^{**}} < G^{S^* \cup S^{**}}$, which again contradicts (31). Therefore, there cannot exist such an $S^{**}$, which completes the proof. □

We can now prove Theorem 1.

**Proof of Theorem 1.** (a) We will show that the project is $\mathcal{F}$-indexable by using the geometric characterization of indexability that is reviewed in the present section. Namely, by showing that the upper boundary $\bar{\partial}\mathcal{H}$ is determined by an increasing nested family of adjacent active sets in $\mathcal{F}$ connecting $\emptyset$ to $\mathcal{Y}^{\{0,1\}}$. We refer the reader to the plot shown in Figure 1 for a geometric illustration of the following arguments.

Let us start by showing that the extreme points of $\mathcal{H}$, which determine $\bar{\partial}\mathcal{H}$, are attained, from left to right, by a unique increasing chain of active sets in $\mathcal{F}$—not necessarily adjacent. Thus, consider two adjacent extreme points of $\mathcal{H}$ in $\bar{\partial}\mathcal{H}$, i.e., joined by a line segment in $\bar{\partial}\mathcal{H}$. By Lemma 6, there exist two unique and distinct active sets $S, S' \in \mathcal{F}$, whose performance points $(G^S, F^S)$ and $(G^{S'}, F^{S'})$ achieve such extreme points, where we assume, without loss of generality, that $G^S < G^{S'}$. We will show that it must be $S \subset S'$. Letting $\lambda = (F^{S'} - F^S)/(G^{S'} - G^S)$ be the slope of the line segment joining such extreme points we have that both $S$ and $S'$ solve the $\lambda$-price problem and, hence, by Lemma 4, so do $S \cap S'$ and $S \cup S'$. Now, from the stated properties of $S$ and $S'$, it follows that points $(G^{S \cap S'}, F^{S \cap S'})$ and $(G^{S \cup S'}, F^{S \cup S'})$ must lie in the line segment joining $(G^S, F^S)$ and $(G^{S'}, F^{S'})$ and, hence, $G^{S \cap S'}, G^{S \cup S'} \in [G^S, G^{S'}]$. Further, since, by Assumption 2(iii) $S \cap S', S \cup S' \in \mathcal{F}$, Lemma 5 gives that $G^{S \cap S'} \leqslant G^S$ and $G^{S'} \leqslant G^{S \cup S'}$. Therefore,

$$G^S = G^{S \cap S'} = G^{S \cup S'} = G^{S'}. \tag{32}$$

We next argue, by contradiction, that $S \subset S'$: if such were not the case, i.e., $S \not\subset S'$, then it would follow that $S \cap S' \subset S \subset S \cup S'$ and, hence, by Lemma 5, $G^{S \cap S'} < G^S < G^{S \cup S'}$, contradicting (32).

Let us next show that, if any two adjacent extreme points $(G^S, F^S)$ and $(G^{S'}, F^{S'})$ in $\bar{\partial}\mathcal{H}$, with $G^S < G^{S'}$, are determined by active sets $S \subset S'$ in such a chain that are not adjacent, they can be connected from left to right by points in $\bar{\partial}\mathcal{H}$ that are attained by an increasing chain of adjacent active sets in $\mathcal{F}$. On the one hand, Assumption 2(ii) ensures the existence of an increasing chain of active sets in $\mathcal{F}$ that are adjacent and connect $S$ to $S'$: $S = T_0 \subset T_1 \subset \cdots \subset T_{k-1} \subset T_k = S'$. On the other hand, if $\lambda = (F^{S'} - F^S)/(G^{S'} - G^S)$ is the slope of the line segment joining such extreme points, then we have that both $S$ and $S'$ solve the $\lambda$-price problem and, hence, by Lemma 4, so does every intermediate active set $T_1, \ldots, T_{k-1}$ in such a chain. Hence, Lemma 5 ensures that $G^S < G^{T_1} < \cdots < G^{T_{k-1}} < G^{S'}$, as required.

In order to establish $\mathcal{F}$-indexability, it only remains to show that the leftmost (resp. rightmost) extreme point of $\mathcal{H}$ in $\bar{\partial}\mathcal{H}$ is that attained by active set $S = \emptyset$ (resp. $S = \mathcal{Y}^{\{0,1\}}$). This follows from Assumption 2(i), condition (ii) in Theorem 1(a), and Lemma 5 (ensuring that $G^{\emptyset} < G^S < G^{\mathcal{Y}^{\{0,1\}}}$ for $S \in \mathcal{F}, \emptyset \subset S \subset \mathcal{Y}^{\{0,1\}}$).

Having established $\mathcal{F}$-indexability, the result that algorithm $\mathrm{AG}_{\mathcal{F}}$ computes the project's Whittle index follows immediately from the algorithm's geometric interpretation, as revealed by its reformulation in Algorithm 2.

(b) Suppose now that the project is indexable. Then, $\bar{\partial}\mathcal{H}$ is determined by some increasing chain of adjacent active sets connecting $\emptyset$ to $\mathcal{Y}^{\{0,1\}}$: $S^0 = \emptyset \subset S^1 \subset \cdots \subset S^N = \mathcal{Y}^{\{0,1\}}$. Letting $\mathcal{F} \triangleq \{S^0, S^1, \ldots, S^N\}$, it is readily seen that such an active-set family satisfies conditions (i) and (ii) in part (a). This completes the proof. □

## 4. Application to Projects with Setup Delays and Costs

This section deploys the framework and results above on restless bandit indexation in our motivating model: the restless bandit reformulation of a non-restless bandit with setup costs and delays (and no setdown penalties: cf. Section 2.1), as discussed in Section 2. The project label $m$ is dropped thereafter from the notation.

In this reformulation, all of the augmented states are controllable, i.e., $\mathcal{Y} = \mathcal{Y}^{\{0,1\}}$, and an active-state subset of the augmented state space $\mathcal{Y}$ representing a stationary deterministic policy is given by specifying the original-state subsets $S_0, S_1 \subseteq \mathcal{X}$, such that the project is engaged when it was rested (resp. engaged) previously if the state $X(t)$ belongs to $S_0$ (resp. in $S_1$). We will denote such an active set/policy, as in [27], by

$$S_0 \oplus S_1 \triangleq \{0\} \times S_0 \cup \{1\} \times S_1 \subseteq \mathcal{Y}.$$

We next address the issue of guessing an appropriate family $\mathcal{F}$ of active sets $S_0 \oplus S_1$, which contains optimal active sets for the $\lambda$-price problem of concern (cf. (10)), which is now formulated as

$$\underset{\pi \in \Pi}{\text{maximize}}\ F^{\pi}_{(a^-,i)} - G^{\pi}_{(a^-,i)}, \tag{33}$$

where $F^{\pi}_{(a^-,i)}$ and $G^{\pi}_{(a^-,i)}$ are the reward and resource (work) metrics that are given by

$$F^{\pi}_{(a^-,i)} \triangleq \mathbb{E}^{\pi}_{(a^-,i)}\left[\sum_{k=0}^{\infty} R^{a(t)}_{Y(t)}\beta^{t_k}\right] \quad \text{and} \quad G^{\pi}_{(a^-,i)} \triangleq \mathbb{E}^{\pi}_{(a^-,i)}\left[\sum_{k=0}^{\infty} Q^{a(t)}_{Y(t)}\beta^{t_k}\right]. \tag{34}$$

The intuition that, under Assumption 1, if engaging the project is optimal when it was not set up, then engaging it should also be optimal when it was set up, leads us to posit the following choice of $\mathcal{F}$:

$$\mathcal{F} \triangleq \left\{S_0 \oplus S_1 \colon S_0 \subseteq S_1 \subseteq \mathcal{X}\right\}. \tag{35}$$

Such an $\mathcal{F}$ represents a family of policies that satisfies Assumption 2. If $S_0 \subset S_1$, policy $S_0 \oplus S_1 \in \mathcal{F}$ has the *hysteresis region* $S_1 \setminus S_0$, i.e., when the original state $X(t)$ lies in $S_1 \setminus S_0$ the policy sticks to the previously chosen action. We will seek to prove indexability with respect to such a family of policies, i.e., $\mathcal{F}$-indexability.

Note that the marginal work, reward, and productivity metrics defined in general by (12)–(15) now take the form

$$g^{S_0\oplus S_1}_{(a^-,i)} \triangleq G^{\langle 1,S_0\oplus S_1\rangle}_{(a^-,i)} - G^{\langle 0,S_0\oplus S_1\rangle}_{(a^-,i)}, \tag{36}$$

$$f^{S_0\oplus S_1}_{(a^-,i)} \triangleq F^{\langle 1,S_0\oplus S_1\rangle}_{(a^-,i)} - F^{\langle 0,S_0\oplus S_1\rangle}_{(a^-,i)}, \tag{37}$$

and, for $g^{S_0\oplus S_1}_{(a^-,i)} \neq 0$,

$$\lambda^{S_0\oplus S_1}_{(a^-,i)} \triangleq \frac{f^{S_0\oplus S_1}_{(a^-,i)}}{g^{S_0\oplus S_1}_{(a^-,i)}}. \tag{38}$$

We next adapt to the present setting the general top-down adaptive-greedy algorithm $\mathrm{AG}_{\mathcal{F}}$ in Algorithm 1, which yields the algorithm in Algorithm 3, where $n \triangleq |\mathcal{X}|$ is now the number of project states in the non-restless formulation. The output of the algorithm has been decoupled, noting that, at every step, the algorithm expands the current active set $S_0^{k_0-1} \oplus S_1^{k_1-1}$ by adding a state that can be either of the form $(0, i_0^{k_0})$ or $(1, i_1^{k_1})$. Thus, instead of using a single counter $k$, ranging from 0 to $2n$, two counters $1 \leqslant k_0 \leqslant k_1 \leqslant n$ are used, with such counters being related by $k = k_0 + k_1 - 1$. Henceforth, we use a more algorithm-like notation, writing, e.g., $\lambda^{S_0^{k_0-1}\oplus S_1^{k_1-1}}_{(0,j)}$ as $\lambda^{(k_0-1,k_1-1)}_{(0,j)}$. Note that the active sets $S_0^{k_0}$ and $S_1^{k_1}$ that are generated in the algorithm are given by $S_0^{k_0} = \{i_0^1, \ldots, i_0^{k_0}\}$ and $S_1^{k_1} = \{i_1^1, \ldots, i_1^{k_1}\}$, and satisfy $S_0^{k_0} \subseteq S_1^{k_1}$, for $1 \leqslant k_0 \leqslant k_1 \leqslant n$, consistently with (35). Thus,

the algorithm produces a decoupled output consisting of two augmented-state strings strings $(0, i_0^{k_0})$ and $(1, i_1^{k_1})$, which jointly span $\mathcal{Y}$, along with corresponding switching and continuation index values $\lambda^*_{(0,i_0^{k_0})}$ and $\lambda^*_{(1,i_1^{k_1})}$.

**Algorithm 3:** Adaptation of index algorithm $\mathrm{AG}_{\mathcal{F}}$ to the present model.

**Output:** $\{(0, i_0^{k_0}), \lambda^*_{(0,i_0^{k_0})}\}_{k_0=1}^{n}$, $\{(1, i_1^{k_1}), \lambda^*_{(1,i_1^{k_1})}\}_{k_1=1}^{n}$

$S_0^0 := \emptyset$; $S_1^0 := \emptyset$; $k_0 := 1$; $k_1 := 1$
**while** $k_0 + k_1 \leqslant 2n + 1$ **do**
  **if** $k_1 \leqslant n$ **choose** $j_1^{\max} \in \arg\max \{\lambda^{(k_0-1,k_1-1)}_{(1,j)} : j \in \mathcal{X} \setminus S_1^{k_1-1}\}$
  **if** $k_0 < k_1$ **choose** $j_0^{\max} \in \arg\max \{\lambda^{(k_0-1,k_1-1)}_{(0,j)} : j \in S_1^{k_1-1} \setminus S_0^{k_0-1}\}$
  **if** $k_1 = n+1$ **or** $\{k_0 < k_1 \leqslant n$ **and** $\lambda^{(k_0-1,k_1-1)}_{(1,j_1^{\max})} < \lambda^{(k_0-1,k_1-1)}_{(0,j_0^{\max})}\}$
    $i_0^{k_0} := j_0^{\max}$; $\lambda^*_{(0,i_1^{k_0})} := \lambda^{(k_0-1,k_1-1)}_{(0,i_1^{k_0})}$; $S_0^{k_0} := S_0^{k_0-1} \cup \{i_0^{k_0}\}$; $k_0 := k_0 + 1$
  **else**
    $i_1^{k_1} := j_1^{\max}$; $\lambda^*_{(1,i_1^{k_1})} := \lambda^{(k_0-1,k_1-1)}_{(1,i_1^{k_1})}$; $S_1^{k_1} := S_1^{k_1-1} \cup \{i_1^{k_1}\}$; $k_1 := k_1 + 1$
  **end** { if }
**end** { while }

### *4.1. Proving That $\mathcal{F}$-Policies Are Optimal*

We next aim to establish that condition (ii) in Theorem 1(a) is satisfied by the present model, i.e., that $\mathcal{F}$-policies, i.e., those with active sets $S_0 \oplus S_1 \in \mathcal{F}$ that are defined by (35), suffice to solve the $\lambda$-price problem (33) for any price $\lambda \in \mathbb{R}$. We will use the DP optimality equations that characterize the optimal value function $V^*_{(a^-,i)}(\lambda)$ for problem (33), starting from each augmented state $(a^-, i) \in \mathcal{Y}$: thus, for each original state $i \in \mathcal{X}$,

$$\begin{aligned} V^*_{(1,i)}(\lambda) &= \max\{\beta V^*_{(0,i)}(\lambda), R_i - \lambda + \beta \sum_{j\in\mathcal{X}} p_{ij} V^*_{(1,j)}(\lambda)\} \\ V^*_{(0,i)}(\lambda) &= \max\{\beta V^*_{(0,i)}(\lambda), -c_i - \frac{1-\phi_i}{1-\beta}\lambda + \phi_i\big(R_i - \lambda + \beta \sum_{j\in\mathcal{X}} p_{ij} V^*_{(1,j)}(\lambda)\big)\}. \end{aligned} \tag{39}$$

We start by showing that the optimal value function is non-negative.

**Lemma 7.** $V^*_{(a^-,i)}(\lambda) \geqslant 0$.

**Proof.** Because no setdown penalties are assumed (cf. Section 2.1), a possible course of action incurring zero net reward is to set down the project and keep it that way, which yields the result. □

We can now prove the optimality of $\mathcal{F}$-policies.

**Lemma 8.** *For every $\lambda \in \mathbb{R}$, there exists an optimal active set $S_0 \oplus S_1 \in \mathcal{F}$ for $\lambda$-price problem* (33).

**Proof.** Fix $\lambda \in \mathbb{R}$ and $i \in \mathcal{X}$. It suffices to show that, if resting the project is optimal in state $(1, i)$, then it is also optimal doing so in state $(0, i)$. Let us formulate that hypothesis, as

$$\beta V^*_{(0,i)}(\lambda) \geqslant R_i - \lambda + \beta \sum_{j\in\mathcal{X}} p_{ij} V^*_{(1,j)}(\lambda). \tag{40}$$

We aim to show that, then, it is optimal resting the project in state $(0,i)$, so

$$\beta V^*_{(0,i)}(\lambda) \geqslant -c_i - \frac{1-\phi_i}{1-\beta}\lambda + \phi_i\big(R_i - \lambda + \beta \sum_{j\in\mathcal{X}} p_{ij} V^*_{(1,j)}(\lambda)\big).$$

Consider first the case $\lambda < 0$. We will argue, by contradiction, that hypothesis (40) then cannot hold, i.e., it cannot be optimal to rest the project once it is active. Drawing on non-restless bandit theory, note that, when the project is active, it is optimal to rest it only if it ever reaches an original state $j \in \mathcal{X}$ at which $\lambda \leqslant \lambda^*_j$, where $\lambda^*_j$ is the original (non-restless) bandit's Gittins index. Assumption 1(ii) now assures us that $\lambda^*_j \geqslant 0$ for each $j \in \mathcal{X}$, and, therefore, it is optimal to keep the project active forever.

Next, consider the case $\lambda \geqslant 0$. Then, the following chain of inequalities holds:

$$\beta V^*_{(0,i)}(\lambda) \geqslant R_i - \lambda + \beta \sum_{j\in\mathcal{X}} p_{ij} V^*_{(1,j)}(\lambda) \geqslant -c_i - \frac{1-\phi_i}{1-\beta}\lambda + \phi_i\big(R_i - \lambda + \beta \sum_{j\in\mathcal{X}} p_{ij} V^*_{(1,j)}(\lambda)\big),$$

where the fact that the second inequality holds is apparent by reformulating it as

$$(1-\phi_i)(R_i + \beta \sum_{j\in\mathcal{X}} p_{ij} V^*_{(1,j)}(\lambda)) \geqslant -c_i - \beta\frac{1-\phi}{1-\beta}\lambda,$$

and noting that Assumption 1(ii) and Lemma 7, ensure that the latter inequality left-hand side is non-negative, and, further, Assumption 1(i) and $\lambda \geqslant 0$ ensure non-positivity of its right-hand side. This completes the proof. □

### *4.2. Work Metric Analysis and $\mathcal{F}$-Indexability Proof*

We now consider how to calculate work and marginal work metrics $G^{S_0\oplus S_1}_{(a^-,i)}$ and $g^{S_0\oplus S_1}_{(a^-,i)}$, by relating them to the corresponding metrics $G^S_i$ and $g^S_i$ for the underlying non-restless project. We will further use such analyses to establish that condition (i) in Theorem 1(a) holds for the model of concern, thus allowing for us to apply such a theorem.

For each $S \subseteq \mathcal{X}$, the $G^S_i$ are characterized by the unique solution to the evaluation equations

$$G^S_i = \begin{cases} 1 + \beta \sum_{j\in S} p_{ij} G^S_j & \text{if } i \in S \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

Further, the marginal work metric $g^S_i$ is evaluated by

$$g^S_i \triangleq G^{\langle 1,S\rangle}_i - G^{\langle 0,S\rangle}_i = 1 + \beta \sum_{j\in\mathcal{X}} p_{ij} G^S_j - \beta G^S_i = \begin{cases} (1-\beta) G^S_i & \text{if } i \in S \\ 1 + \beta \sum_{j\in S} p_{ij} G^S_j & \text{otherwise.} \end{cases} \tag{42}$$

Note that (41) and (42) imply that

$$g^S_i > 0, \quad i \in N. \tag{43}$$

We now go back to the project's restless bandit reformulation. The next result, whose proof is omitted, as it is immediate, gives the evaluation equations for work metric $G^{S_0\oplus S_1}_{(a^-,i)}$ under a given active set.

**Lemma 9.** *For $S_0 \oplus S_1 \in \mathcal{F}$,*

$$G^{S_0\oplus S_1}_{(0,i)} = \begin{cases} \frac{1-\phi_i}{1-\beta} + \phi_i G^{S_0\oplus S_1}_{(1,i)} & \text{if } i \in S_0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad G^{S_0\oplus S_1}_{(1,i)} = \begin{cases} 1 + \beta \sum_{j\in\mathcal{X}} p_{ij} G^{S_0\oplus S_1}_{(1,j)} & \text{if } i \in S_1 \\ 0 & \text{otherwise.} \end{cases}$$

The following result represents work metric $G^{S_0\oplus S_1}_{(a^-,i)}$ in terms of the $G^S_j$.

**Lemma 10.** *For $S_0 \oplus S_1 \in \mathcal{F}$:*

(a) $G^{S_0\oplus S_1}_{(a^-,i)} = G^{S_1}_i = 0$, *for* $a^- \in \{0,1\}, i \in \mathcal{X} \setminus S_1$.

(b) $G^{S_0\oplus S_1}_{(1,i)} = G^{S_1}_i$, *for* $i \in S_1$.

(c) $G^{S_0\oplus S_1}_{(0,i)} = (1-\phi_i)/(1-\beta) + \phi_i G^{S_1}_i$, *for* $i \in S_0$.

(d) $G^{S_0\oplus S_1}_{(0,i)} = 0$, *for* $i \in S_1 \setminus S_0$.

**Proof.** (a) The result follows readily from the definition of $S_0 \oplus S_1$.

(b) For $i \in S_1$, we have

$$G^{S_0\oplus S_1}_{(1,i)} = 1 + \beta \sum_{j\in S_1} p_{ij} G^{S_0\oplus S_1}_{(1,j)} + \beta \sum_{j\in \mathcal{X}\setminus S_1} p_{ij} G^{S_0\oplus S_1}_{(1,j)} = 1 + \beta \sum_{j\in S_1} p_{ij} G^{S_0\oplus S_1}_{(1,j)},$$

while using Lemma 9 and part (a). Thus, the $G^{S_0\oplus S_1}_{(1,i)}$ satisfy the equations in (41) characterizing the $G^{S_1}_i$ for $i \in S_1$, which gives the result.

(c) We have, for $i \in S_0$,

$$G^{S_0\oplus S_1}_{(0,i)} = \frac{1-\phi_i}{1-\beta} + \phi_i G^{S_0\oplus S_1}_{(1,i)} = \frac{1-\phi_i}{1-\beta} + \phi_i G^{S_1}_i,$$

using Lemma 9, the inclusion $S_0 \subseteq S_1$, and (a, b).

(d) The result follows readily from the definition of $S_0 \oplus S_1$. □

Concerning the marginal work metric $g^{S_0\oplus S_1}_{(a^-,i)}$, (36) and Lemma 9, they readily give that

$$\begin{aligned} g^{S_0\oplus S_1}_{(1,i)} &= 1 + \beta \sum_{j\in\mathcal{X}} p_{ij} G^{S_0\oplus S_1}_{(1,j)} - \beta G^{S_0\oplus S_1}_{(0,i)} \\ g^{S_0\oplus S_1}_{(0,i)} &= \frac{1-\phi_i}{1-\beta} + \phi_i\big(1 + \beta \sum_{j\in\mathcal{X}} p_{ij} G^{S_0\oplus S_1}_{(1,j)}\big) - \beta G^{S_0\oplus S_1}_{(0,i)}. \end{aligned} \tag{44}$$

The following result represents marginal work metric $g^{S_0\oplus S_1}_{(a^-,i)}$ in terms of the $g^S_j$.

**Lemma 11.** *For every $a^- \in \{0,1\}, S_0 \oplus S_1 \in \mathcal{F}$:*

(a) $g^{S_0\oplus S_1}_{(1,i)} = g^{S_1}_i$, *for* $i \in \mathcal{X} \setminus S_1$.

(b) $g^{S_0\oplus S_1}_{(0,i)} = \dfrac{1-\phi_i}{1-\beta} + g^{S_1}_i$, *for* $i \in \mathcal{X} \setminus S_1$.

(c) $g^{S_0\oplus S_1}_{(1,i)} = \dfrac{1-\beta\phi_i}{1-\beta}\Big(g^{S_1}_i - \beta\dfrac{1-\phi_i}{1-\beta\phi_i}\Big)$, *for* $i \in S_0$.

(d) $g^{S_0\oplus S_1}_{(0,i)} = 1 - \phi_i + \phi_i g^{S_1}_i$, *for* $i \in S_0$.

(e) $g^{S_0\oplus S_1}_{(1,i)} = \dfrac{g^{S_1}_i}{1-\beta}$, *for* $i \in S_1 \setminus S_0$.

(f) $g^{S_0\oplus S_1}_{(0,i)} = \dfrac{1-\phi_i}{1-\beta} + \dfrac{\phi_i}{1-\beta} g^{S_1}_i$, *for* $i \in S_1 \setminus S_0$.

**Proof.** (a) We have, for $i \in \mathcal{X} \setminus S_1$,

$$g^{S_0\oplus S_1}_{(1,i)} = 1 + \beta \sum_{j\in\mathcal{X}} p_{ij} G^{S_0\oplus S_1}_{(1,j)} - \beta G^{S_0\oplus S_1}_{(0,i)} = 1 + \beta \sum_{j\in S_1} p_{ij} G^{S_1}_j = g^{S_1}_i,$$

using (44), Lemma 10(a,b), and (42).

(b) We can write, for $i \in \mathcal{X} \setminus S_1$,

$$g_{(0,i)}^{S_0 \oplus S_1} = \frac{1-\phi_i}{1-\beta} + \phi_i \big(1 + \beta \sum_{j \in \mathcal{X}} p_{ij} G_{(1,j)}^{S_0 \oplus S_1}\big) - \beta G_{(0,j)}^{S_0 \oplus S_1}$$
$$= \frac{1-\phi_i}{1-\beta} + \phi_i \big(1 + \beta \sum_{j \in S_1} p_{ij} G_j^{S_1}\big) = \frac{1-\phi_i}{1-\beta} + \phi_i g_i^{S_1},$$

while using (44), Lemma 10(a,b), and (42).

(c) We have, for $i \in S_0$,

$$g_{(1,i)}^{S_0 \oplus S_1} = G_{(1,i)}^{S_0 \oplus S_1} - \beta G_{(0,i)}^{S_0 \oplus S_1} = G_i^{S_1} - \beta \Big(\frac{1-\phi_i}{1-\beta} + \phi_i G_i^{S_1}\Big)$$
$$= (1-\beta\phi_i) G_i^{S_1} - \beta \frac{1-\phi_i}{1-\beta} = \frac{1-\beta\phi_i}{1-\beta} \Big(g_i^{S_1} - \beta \frac{1-\phi_i}{1-\beta\phi_i}\Big),$$

using (44), $S_0 \subseteq S_1$, Lemma 9, Lemma 10(b,c), and (42).

(d) We obtain, for $i \in S_0$,

$$g_{(0,i)}^{S_0 \oplus S_1} = \frac{1-\phi_i}{1-\beta} + \phi_i G_{(1,i)}^{S_0 \oplus S_1} - \beta G_{(0,i)}^{S_0 \oplus S_1} = \frac{1-\phi_i}{1-\beta} + \phi_i G_i^{S_1} - \beta \Big(\frac{1-\phi_i}{1-\beta} + \phi_i G_i^{S_1}\Big)$$
$$= 1 - \phi_i + \phi_i (1-\beta) G_i^{S_1} = 1 - \phi_i + \phi_i g_i^{S_1},$$

while using Lemma 9, $S_0 \subseteq S_1$, Lemma 10(b,c), and (42).

(e) We have, for $i \in S_1 \setminus S_0$,

$$g_{(1,i)}^{S_0 \oplus S_1} = G_{(1,i)}^{S_0 \oplus S_1} - \beta G_{(0,i)}^{S_0 \oplus S_1} = G_i^{S_1} = \frac{g_i^{S_1}}{1-\beta},$$

using (44), Lemma 9, Lemma 10(d), and (42).

(f) We have, for $i \in S_1 \setminus S_0$,

$$g_{(0,i)}^{S_0 \oplus S_1} = \frac{1-\phi_i}{1-\beta} + \phi_i G_{(1,i)}^{S_0 \oplus S_1} = \frac{1-\phi_i}{1-\beta} + \phi_i G_i^{S_1} = \frac{1-\phi_i}{1-\beta} + \frac{\phi_i}{1-\beta} g_i^{S_1},$$

using (44), Lemma 9, Lemma 10(b), and (42). □

It must be now remarked that, at the corresponding point in the analysis of [27]—for the case with no setup delays $\phi_i \equiv 1$—one could establish the positivity of the marginal work metric, i.e., $g_{(a^-,i)}^{S_0 \oplus S_1} > 0$ for $(a^-, i) \in \mathcal{Y}$, $S_0 \oplus S_1 \in \mathcal{F}$, which is the first PCL-indexability condition and it implies the less stringent condition (i) in Theorem 1(a). However, here, it is apparent from Lemma 11(c) that, for $i \in S_0$, $g_{(1,i)}^{S_0 \oplus S_1}$ can be negative for $\beta$ that is close to 1. This is why we cannot use here the same line of argument that is given in [27] to show indexability.

As mentioned above, we will use, instead, for such a purpose, Theorem 1(a). The following result shows that condition (i) in that theorem holds for the model of concern.

**Lemma 12.** *For $S_0 \oplus S_1 \in \mathcal{F}$,*

$$g_{(a^-,i)}^{S_0 \oplus S_1} > 0, \quad (a^-, i) \in S_0 \oplus S_1, S_0 \oplus S_1 \setminus \{(a^-, i)\} \in \mathcal{F}$$
$$g_{(a^-,i)}^{S_0 \oplus S_1} > 0, \quad (a^-, i) \in \mathcal{Y} \setminus S_0 \oplus S_1, S_0 \oplus S_1 \cup \{(a^-, i)\} \in \mathcal{F}.$$

**Proof.** First, consider the case $S_0 \oplus S_1 = \emptyset \oplus \emptyset$. Then, using Lemma 11(a–d), along with $g_i^{\emptyset} \equiv 1$, gives that, for $i \in \mathcal{X}$,

$$g_{(1,i)}^{\emptyset\oplus\emptyset} = g_i^{\emptyset} = 1 > 0, \quad g_{(0,i)}^{\emptyset\oplus\emptyset} = \frac{1-\phi_i}{1-\beta} + g_i^{\emptyset} = \frac{1-\phi_i}{1-\beta} + 1 > 0.$$

Now, consider the case $S_0 \oplus S_1 = \mathcal{X} \oplus \mathcal{X} = \mathcal{Y}$. Then, again using Lemma 11(a–d) along with $g_i^{\mathcal{X}} \equiv 1$ gives that, for $i \in \mathcal{X}$,

$$g_{(1,i)}^{\mathcal{X}\oplus\mathcal{X}} = \frac{1-\beta\phi_i}{1-\beta}\left(g_i^{\mathcal{X}} - \beta\frac{1-\phi_i}{1-\beta\phi_i}\right) = 1 > 0, \quad g_{(0,i)}^{\mathcal{X}\oplus\mathcal{X}} = 1 - \phi_i + \phi_i g_i^{\mathcal{X}} = 1 > 0.$$

Finally, consider $S_0 \oplus S_1 \in \mathcal{F}$, which is different from $\emptyset \oplus \emptyset$ and $\mathcal{X} \oplus \mathcal{X}$. Then, Lemma 11 and (35) imply that it could only happen that marginal work metric $g_{(a^-,i)}^{S_0\oplus S_1}$ be negative if $a^- = 1$ and $i \in S_0$. However, such a case is not included in the required conditions, since $(1,i) \in S_0 \oplus S_1$ (due to $S_0 \subseteq S_1$), yet $S_0 \oplus S_1 \setminus \{(1,i)\} = S_0 \oplus (S_1 \setminus \{i\}) \notin \mathcal{F}$ (since $i \in S_0 \not\subseteq S_1 \setminus \{i\}$). This completes the proof. □

We are now ready to deploy Theorem 1(a) in the present model.

**Proposition 1.** *The present restless bandit model is $\mathcal{F}$-indexable and Algorithm 3 computes its Whittle index.*

**Proof.** Lemmas 8 and 12 show that conditions (i) and (ii) in Theorem 1(a) hold, respectively, which implies the result. □

*4.3. The AT Index Is the Whittle Index*

We next use the results above in order to prove the identity between the Whittle index and the AT index. We will reformulate the AT index formulae in (7)–(8) while using active sets $S \subseteq \mathcal{X}$, rather than stopping times $\tau$. Thus, we can reformulate the continuation and switching AT indices, as

$$\lambda_{(1,i)}^{\mathrm{AT}} \triangleq \max_{S\subseteq\mathcal{X}:\, i\in S} \frac{F_i^S}{G_i^S}, \tag{45}$$

and

$$\lambda_{(0,i)}^{\mathrm{AT}} \triangleq \max_{S\subseteq\mathcal{X}:\, i\in S} \frac{-c_i + \phi_i F_i^S}{\frac{1-\phi_i}{1-\beta} + \phi_i G_i^S}. \tag{46}$$

Recall that we denote the Whittle index by $\lambda_{(a^-,i)}^*$.

**Proposition 2.** *For $i \in \mathcal{X}$, $\lambda_{(1,i)}^* = \lambda_{(1,i)}^{\mathrm{AT}}$ and $\lambda_{(0,i)}^* = \lambda_{(0,i)}^{\mathrm{AT}}$.*

**Proof.** We start by showing that $\lambda_{(1,i)}^* = \lambda_{(1,i)}^{\mathrm{AT}}$, while using the equivalences

$$\begin{aligned}
\lambda \geqslant \lambda_{(1,i)}^* &\iff \text{resting the project in } (1,i) \text{ is optimal for problem (33)} \\
&\iff 0 \geqslant \max_{S_0\subseteq S_1\subseteq\mathcal{X}:\, i\in S_1} F_{(1,i)}^{S_0\oplus S_1} - \lambda G_{(1,i)}^{S_0\oplus S_1} \\
&\iff \lambda \geqslant \max_{S_0\subseteq S_1\subseteq\mathcal{X}:\, i\in S_1} \frac{F_{(1,i)}^{S_0\oplus S_1}}{G_{(1,i)}^{S_0\oplus S_1}} \\
&\iff \lambda \geqslant \max_{i\in S_1\subseteq\mathcal{X}} \frac{F_i^{S_1}}{G_i^{S_1}} = \lambda_{(1,i)}^{\mathrm{AT}},
\end{aligned}$$

drawing on the project's $\mathcal{F}$-indexability (Proposition 1), and so, if resting the project iin $(1,i)$ is optimal, then resting it in $(0,i)$ is also optimal, together with Lemmas 10(b) and 14(b).

We next prove that $\lambda^*_{(0,i)} = \lambda^{\mathrm{AT}}_{(0,i)}$, through the chain of equivalences

$$\begin{aligned}
\lambda \geqslant \lambda^*_{(0,i)} &\iff \text{resting the project in } (0,i) \text{ is optimal for (33)} \\
&\iff 0 \geqslant \max_{S_0 \subseteq S_1 \subseteq \mathcal{X}:\, i \in S_0} F^{S_0 \oplus S_1}_{(0,i)} - \lambda G^{S_0 \oplus S_1}_{(0,i)} \\
&\iff \lambda \geqslant \max_{S_0 \subseteq S_1 \subseteq \mathcal{X}:\, i \in S_0} \frac{F^{S_0 \oplus S_1}_{(0,i)}}{G^{S_0 \oplus S_1}_{(0,i)}} \\
&\iff \lambda \geqslant \max_{S_1 \subseteq \mathcal{X}:\, i \in S_1} \frac{-c_i + \phi_i F_i^{S_1}}{\dfrac{1-\phi_i}{1-\beta} + \phi_i G_i^{S_1}} = \lambda^{\mathrm{AT}}_{(0,i)},
\end{aligned}$$

drawing on the result that the project is $\mathcal{F}$-indexable, together with Lemmas 10(c) and 14(c). □

*4.4. Reward Metric Analysis*

We proceed by considering how to calculate the reward and marginal reward metrics $F^{S_0 \oplus S_1}_{(a^-,i)}$ and $f^{S_0 \oplus S_1}_{(a^-,i)}$, by relating them to the metrics $F^S_i$ and $f^S_i$ for the corresponding non-restless project with no setup penalties.

For every active set $S \subseteq \mathcal{X}$, the reward metric $F^S_i$ is determined by the evaluation equations

$$F^S_i = \begin{cases} R_i + \beta \sum_{j \in S} p_{ij} F^S_j & \text{if } i \in S \\ 0 & \text{otherwise,} \end{cases} \tag{47}$$

and the marginal reward metric is given by

$$f^S_i \triangleq F^{\langle 1,S \rangle}_i - F^{\langle 0,S \rangle}_i = R_i + \beta \sum_{j \in S} p_{ij} F^S_j - \beta F^S_i = \begin{cases} (1-\beta) F^S_i & \text{if } i \in S \\ R_i + \beta \sum_{j \in S} p_{ij} F^S_j & \text{otherwise.} \end{cases} \tag{48}$$

Going back to the semi-Markov restless bandit reformulation, the following result shows the evaluation equations for the reward metrics $F^{S_0 \oplus S_1}_{(a^-,i)}$, for an active set $S_0 \oplus S_1 \in \mathcal{F}$.

**Lemma 13.**

$$F^{S_0 \oplus S_1}_{(a^-,i)} = \begin{cases} R_i + \beta \sum_{j \in \mathcal{X}} p_{ij} F^{S_0 \oplus S_1}_{(1,j)} & \text{if } a^- = 1, i \in S_1 \\ -c_i + \phi_i \big( R_i + \beta \sum_{j \in \mathcal{X}} p_{ij} F^{S_0 \oplus S_1}_{(1,j)} \big) & \text{if } a^- = 0, i \in S_0 \\ \beta F^{S_0 \oplus S_1}_{(0,i)} & \text{otherwise.} \end{cases}$$

The following result formulates the reward metric $F^{S_0 \oplus S_1}_{(a^-,i)}$, in terms of the $F^S_i$'s.

**Lemma 14.** *For $S_0 \oplus S_1 \in \mathcal{F}$:*

(a) $F^{S_0 \oplus S_1}_{(a^-,i)} = 0 = F^{S_1}_i$, *for* $a^- \in \{0,1\}, i \in \mathcal{X} \setminus S_1$.

(b) $F^{S_0 \oplus S_1}_{(1,i)} = F^{S_1}_i$, *for* $i \in S_1$.

(c) $F^{S_0 \oplus S_1}_{(0,i)} = -c_i + \phi_i F^{S_1}_i$, *for* $i \in S_0$.

(d) $F^{S_0 \oplus S_1}_{(0,i)} = 0 = F^{S_0}_i$, *for* $i \in S_1 \setminus S_0$.

**Proof.** (a) This part follows from the definition of $S_0 \oplus S_1$.

(b) We have, for $i \in S_1$,

$$F_{(1,i)}^{S_0 \oplus S_1} = R_i + \beta \sum_{j \in S_1} p_{ij} F_{(1,j)}^{S_0 \oplus S_1} + \beta \sum_{j \in \mathcal{X} \setminus S_1} p_{ij} F_{(1,j)}^{S_0 \oplus S_1} = R_i + \beta \sum_{j \in S_1} p_{ij} F_{(1,j)}^{S_0 \oplus S_1},$$

while using Lemma 13 and part (a). Thus, the $F_{(1,i)}^{S_0 \oplus S_1}$'s, for $i \in S_1$, satisfy (47), which yields the result.

(c) We can write, for $i \in S_0$,

$$F_{(0,i)}^{S_0 \oplus S_1} = -c_i + \phi_i \big( R_i + \beta \sum_{j \in S_1} p_{ij} F_{(1,j)}^{S_0 \oplus S_1} \big) = -c_i + \phi_i F_i^{S_1},$$

using parts (a, b), Lemma 13, and (47).

(d) The result follows from the definition of $S_0 \oplus S_1$. □

Concerning the marginal reward metric $f_{(a^-,i)}^{S_0 \oplus S_1}$, we obtain, from (37) and Lemma 13, that

$$\begin{aligned} f_{(1,i)}^{S_0 \oplus S_1} &= R_i + \beta \sum_{j \in \mathcal{X}} p_{ij} F_{(1,j)}^{S_0 \oplus S_1} - \beta F_{(0,i)}^{S_0 \oplus S_1} \\ f_{(0,i)}^{S_0 \oplus S_1} &= -c_i + \phi_i \big( R_i + \beta \sum_{j \in \mathcal{X}} p_{ij} F_{(1,j)}^{S_0 \oplus S_1} \big) - \beta F_{(0,i)}^{S_0 \oplus S_1}. \end{aligned} \tag{49}$$

The following result represents the marginal reward $f_{(a^-,i)}^{S_0 \oplus S_1}$ in terms of the $f_j^S$.

**Lemma 15.** *For $S_0 \oplus S_1 \in \mathcal{F}$:*

(a) $f_{(1,i)}^{S_0 \oplus S_1} = f_i^{S_1}$, *for* $i \in \mathcal{X} \setminus S_1$.

(b) $f_{(0,i)}^{S_0 \oplus S_1} = -c_i + f_i^{S_1}$, *for* $i \in \mathcal{X} \setminus S_1$.

(c) $f_{(1,i)}^{S_0 \oplus S_1} = \beta c_i + \dfrac{1 - \beta \phi_i}{1 - \beta} f_i^{S_1}$, *for* $i \in S_0$.

(d) $f_{(0,i)}^{S_0 \oplus S_1} = -(1 - \beta) c_i + \phi_i f_i^{S_1}$, *for* $i \in S_0$.

(e) $f_{(1,i)}^{S_0 \oplus S_1} = \dfrac{f_i^{S_1}}{1 - \beta}$, *for* $i \in S_1 \setminus S_0$.

(f) $f_{(0,i)}^{S_0 \oplus S_1} = -c_i + \phi_i \dfrac{f_i^{S_1}}{1 - \beta}$, *for* $i \in S_1 \setminus S_0$.

**Proof.** (a) We have, for $i \in \mathcal{X} \setminus S_1$,

$$f_{(1,i)}^{S_0 \oplus S_1} = R_i + \beta \sum_{j \in \mathcal{X}} p_{ij} F_{(1,j)}^{S_0 \oplus S_1} - F_{(1,i)}^{S_0 \oplus S_1} = R_i + \beta \sum_{j \in S_1} p_{ij} F_j^{S_1} = f_i^{S_1},$$

using (49), Lemmas 13 and 14(a,b), (47), and (48).

(b) We can write, for $i \in \mathcal{X} \setminus S_1$,

$$\begin{aligned} f_{(0,i)}^{S_0 \oplus S_1} &= -c_i + \phi_i \big( 1 + \beta \sum_{j \in \mathcal{X}} p_{ij} F_{(1,j)}^{S_0 \oplus S_1} \big) - \beta F_{(0,j)}^{S_0 \oplus S_1} \\ &= -c_i + \phi_i \big( 1 + \beta \sum_{j \in S_1} p_{ij} F_j^{S_1} \big) = -c_i + \phi_i f_i^{S_1}, \end{aligned}$$

using (49), (48), and Lemma 14(a,b).

(c) We have, for $i \in S_0$,

$$f_{(1,i)}^{S_0 \oplus S_1} = F_{(1,i)}^{S_0 \oplus S_1} - \beta F_{(0,i)}^{S_0 \oplus S_1} = F_i^{S_1} - \beta\big(-c_i + \phi_i F_i^{S_1}\big)$$
$$= \beta c_i + (1 - \beta\phi_i) F_i^{S_1} = \beta c_i + \frac{1 - \beta\phi_i}{1 - \beta} f_i^{S_1},$$

using (49), $S_0 \subseteq S_1$, Lemmas 13 and 14(b,c), and (48).

(d) We can write, for $i \in S_0$,

$$f_{(0,i)}^{S_0 \oplus S_1} = -c_i + \phi_i F_{(1,i)}^{S_0 \oplus S_1} - \beta F_{(0,i)}^{S_0 \oplus S_1} = -c_i + \phi_i F_i^{S_1} - \beta\big(-c_i + \phi_i F_i^{S_1}\big)$$
$$= -(1 - \beta)c_i + \phi_i(1 - \beta) F_i^{S_1} = -(1 - \beta)c_i + \phi_i f_i^{S_1},$$

while using Lemmas 13 and 14(b,c), $S_0 \subseteq S_1$, and (48).

(e) We have, for $i \in S_1 \setminus S_0$,

$$f_{(1,i)}^{S_0 \oplus S_1} = F_{(1,i)}^{S_0 \oplus S_1} - \beta F_{(0,i)}^{S_0 \oplus S_1} = F_i^{S_1} = \frac{f_i^{S_1}}{1 - \beta},$$

using (49), Lemmas 13 and 14(d), and (48).

(f) We obtain, for $i \in S_1 \setminus S_0$,

$$f_{(0,i)}^{S_0 \oplus S_1} = -c_i + \phi_i\big(R_i + \beta \sum_{j \in N} p_{ij} F_{(1,j)}^{S_0 \oplus S_1}\big) - \beta F_{(0,i)}^{S_0 \oplus S_1} = -c_i + \phi_i F_i^{S_1} = -c_i + \phi_i \frac{f_i^{S_1}}{1 - \beta},$$

using (49), Lemmas 13 and 14(b), and (48). This completes the proof. □

## 5. Designing an Efficient Two-Stage Index Algorithm

This section draws on the above in order to develop an efficient index algorithm, which exploits special structure to simplify the one-stage adaptive-greedy algorithm in Algorithm 3, by *decoupling* the calculation of the continuation and switching indices into a two-stage method, for which an efficient implementation is provided.

### *5.1. Marginal Productivity Metric Analysis*

We start by addressing the calculation of required marginal productivity metrics $\lambda_{(a^-,i)}^{S_0 \oplus S_1}$ in (38), also by relating them to metrics $\lambda_i^S$ for the corresponding non-restless project without setup penalties, which are given by

$$\lambda_i^S \triangleq \frac{f_i^S}{g_i^S}, \quad i \in \mathcal{X}, S \subseteq \mathcal{X}. \tag{50}$$

The next result represents $\lambda_{(a^-,i)}^{S_0 \oplus S_1}$ in terms of the $\lambda_j^S$.

**Lemma 16.** *For $S_0 \oplus S_1 \in \mathcal{F}$:*

(a) $\lambda_{(1,i)}^{S_0 \oplus S_1} = \lambda_i^{S_1}$, *for* $i \in \mathcal{X} \setminus S_1$.

(b) $\lambda_{(0,i)}^{S_0 \oplus S_1} = \dfrac{-c_i + f_i^{S_1}}{\frac{1-\phi_i}{1-\beta} + g_i^{S_1}} = \dfrac{g_i^{S_1}}{\frac{1-\phi_i}{1-\beta} + g_i^{S_1}}\big(\lambda_i^{S_1} - \dfrac{c_i}{g_i^{S_1}}\big)$, *for* $i \in \mathcal{X} \setminus S_1$.

(c) $\lambda_{(1,i)}^{S_0 \oplus S_1} = \dfrac{\beta c_i + \frac{1-\beta\phi_i}{1-\beta} f_i^{S_1}}{\frac{1-\beta\phi_i}{1-\beta}\big(g_i^{S_1} - \beta\frac{1-\phi_i}{1-\beta\phi_i}\big)} = \dfrac{g_i^{S_1}}{g_i^{S_1} - \beta\frac{1-\phi_i}{1-\beta\phi_i}}\big(\lambda_i^{S_1} + \dfrac{\beta(1-\beta)}{1-\beta\phi_i}\dfrac{c_i}{g_i^{S_1}}\big)$, *for* $i \in S_0$ *such that* $g_i^{S_1} \neq \beta\frac{1-\phi_i}{1-\beta\phi_i}$.

(d) $\lambda_{(0,i)}^{S_0 \oplus S_1} = \dfrac{-(1-\beta)c_i + \phi_i f_i^{S_1}}{1 - \phi_i + \phi_i g_i^{S_1}} = \dfrac{-(1-\beta)c_i + \phi_i g_i^{S_1} \lambda_i^{S_1}}{1 - \phi_i + \phi_i g_i^{S_1}}$, *for* $i \in S_0$.

(e) $\lambda_{(1,i)}^{S_0\oplus S_1} = \lambda_i^{S_1}$, *for* $i \in S_1 \setminus S_0$.

(f) $\lambda_{(0,i)}^{S_0\oplus S_1} = \lambda_i^{S_1} - \dfrac{(1-\beta)c_i + (1-\phi_i)\lambda_i^{S_1}}{1-\phi_i+\phi_i g_i^{S_1}}$, $i \in S_1 \setminus S_0$.

**Proof.** All of the parts follow readily from (50), (38), and Lemmas 11 and 15. □

*5.2. Simplified Version of the Index Algorithm*

Using the above results allows for us to give a simplified and more explicit version of the index algorithm $\mathrm{AG}_{\mathcal{F}}$ in Algorithm 3, which is given in Algorithm 4. In it, we draw on Lemma 16(b,d) to formulate marginal productivity rates $\lambda_{(a^-,i)}^{S_0\oplus S_1}$ in terms of the $g_j^S$ and $\lambda_j^S$. Thus, the $g_j^{(k_1-1)}$ and $\lambda_j^{(k_1-1)}$ in the algorithm correspond to $g_{(1,j)}^{(k_0-1,k_1-1)}$ and $\lambda_{(1,j)}^{(k_0-1,k_1-1)}$, respectively. Further, we use $\lambda_{(0,j)}^{(0,k_1-1)}$ (which denotes $\lambda_{(0,j)}^{S_0^0\oplus S_1^{k_1-1}}$) in place of $\lambda_{(0,j)}^{(k_0-1,k_1-1)}$, drawing on Lemma 16(d). Note that such simplifications achieve significant savings in computer memory, since storage of quantities $\lambda_j^{(k_1-1)}$ and $\lambda_{(0,j)}^{(0,k_1-1)}$ entail one less dimension than storing of the $\lambda_{(1,j)}^{(k_0-1,k_1-1)}$ and $\lambda_{(0,j)}^{(k_0-1,k_1-1)}$.

**Algorithm 4:** Simplified version of index algorithm $\mathrm{AG}_{\mathcal{F}}$.

**Output:** $\{(0,i_0^{k_0}), \lambda^*_{(0,i_0^{k_0})}\}_{k_0=1}^n, \{(1,i_1^{k_1}), \lambda^*_{(1,i_1^{k_1})}\}_{k_1=1}^n$

$S_0^0 := \emptyset$; $S_1^0 := \emptyset$; $k_0 := 1$; $k_1 := 1$; **compute** $\{(g_i^{(0)}, \lambda_i^{(0)}) : i \in \mathcal{X}\}$
**while** $k_0 + k_1 \leqslant 2n+2$ **do**
  **if** $k_1 \leqslant n$ **choose** $j_1^{\max} \in \arg\max\{\lambda_j^{(k_1-1)} : j \in \mathcal{X} \setminus S_1^{k_1-1}\}$
  $\lambda_{(0,j)}^{(0,k_1-1)} := \lambda_j^{(k_1-1)} - \dfrac{(1-\beta)c_j + (1-\phi_j)\lambda_j^{(k_1-1)}}{1-\phi_j+\phi_j g_j^{(k_1-1)}}$, $j \in S_1^{k_1-1} \setminus S_0^{k_0-1}$
  **if** $k_0 < k_1$ **choose** $j_0^{\max} \in \arg\max\{\lambda_{(0,j)}^{(0,k_1-1)} : j \in S_1^{k_1-1} \setminus S_0^{k_0-1}\}$
  **if** $k_1 = n+1$ **or** $\{k_0 < k_1 \leqslant n$ **and** $\lambda_{j_1^{\max}}^{(k_1-1)} < \lambda_{j_0^{\max}}^{(0,k_1-1)}\}$
    $i_0^{k_0} := j_0^{\max}$; $\lambda^*_{(0,i_0^{k_0})} := \lambda_{(0,i_0^{k_0})}^{(0,k_1-1)}$; $S_0^{k_0} := S_0^{k_0-1} \cup \{i_0^{k_0}\}$; $k_0 := k_0+1$
  **else**
    $i_1^{k_1} := j_1^{\max}$; $\lambda^*_{i_1^{k_1}} := \lambda_{i_1^{k_1}}^{(k_1-1)}$; $S_1^{k_1} := S_1^{k_1-1} \cup \{i_1^{k_1}\}$; $k_1 := k_1+1$
    **compute** $\{(g_i^{(k_1)}, \lambda_i^{(k_1)}) : i \in \mathcal{X}\}$
  **end** { if }
**end** { while }

*5.3. Two-Stage Implementation of the Index Algorithm*

We next proceed to still further simplify the index algorithm in Algorithm 4, by decoupling it into two successive algorithms. The first stage of such a scheme computes the continuation index $\lambda^*_{(1,i)}$, which we saw above is just the Gittins index $\lambda_i^*$. We will need additional quantities as input to the second stage: the $g_j^{(k_1)}$ and $\lambda_j^{(k_1)}$ appearing in Algorithm 4.

In order to obtain such an index and the required additional quantities, consider the algorithmic scheme $\mathrm{AG}^1$ in Algorithm 5, which is a variant of that in [8], reformulated as in [28]. For implementations, we can use algorithms that are provided in the latter paper, in particular the *fast-pivoting algorithm with extended output*, which has an $(4/3)n^3 + O(n^2)$ arithmetic-operation count.

**Algorithm 5:** Gittins-index algorithmic scheme $\mathrm{AG}^1$.

**Output:** $\{i_1^{k_1}\}_{k_1=1}^n$, $\{\lambda_j^* : j \in \mathcal{X}\}$, $\{(g_j^{(k_1)}, \lambda_j^{(k_1)}) : j \in S_1^{k_1}\}_{k_1=1}^n$

**set** $S_1^0 := \emptyset$; **compute** $\{(g_i^{(0)}, \lambda_i^{(0)}) : i \in \mathcal{X}\}$
**for** $k_1 := 1$ **to** $n$ **do**
  **choose** $i_1^{k_1} \in \arg\max \{\lambda_i^{(k_1-1)} : i \in \mathcal{X} \setminus S_1^{k_1-1}\}$
  $\lambda^*_{i_1^{k_1}} := \lambda^{(k_1-1)}_{i_1^{k_1}}$; $S_1^{k_1} := S_1^{k_1-1} \cup \{i_1^{k_1}\}$
  **compute** $\{(g_i^{(k_1)}, \lambda_i^{(k_1)}) : i \in \mathcal{X}\}$
**end**

We next address the computation of the switching index in the second stage, once the Gittins index and required extra quantities have been computed. Consider algorithm $\mathrm{AG}^0$ that is given in Algorithm 6, whose input is the output of algorithm $\mathrm{AG}^1$, and which returns a sequence of all the states $i_0^{k_0}$ in $\mathcal{X}$, together with index values $\lambda^*_{(0,i_0^{k_0})}$. Note that such an algorithm is formulated in a form applying to the case of concern herein, with a positive setup delay at every state $j$, so $\phi_j < 1$.

**Algorithm 6:** Switching-index algorithm $\mathrm{AG}^0$.

**ALGORITHM** $\mathrm{AG}^0$:
**Input:** $\{i_1^{k_1}\}_{k_1=1}^n$, $\{\lambda_j^* : j \in \mathcal{X}\}$, $\{(g_j^{(k_1)}, \lambda_j^{(k_1)}) : j \in S_1^{k_1}\}_{k_1=1}^n$
**Output:** $\{i_0^{k_0}\}_{k_0=1}^n$, $\{\lambda^*_{(0,j)} : j \in \mathcal{X}\}$

$\widehat{c}_j := \dfrac{1-\beta}{1-\phi_j} c_j$, $j \in \mathcal{X}$; $z_j = \phi_j/(1-\phi_j)$; $S_0^0 := \emptyset$; $S_1^0 := \emptyset$; $k_0 := 0$
**for** $k_1 := 1$ **to** $n$ **do**
  $S_1^{k_1} := S_1^{k_1-1} \cup \{i_1^{k_1}\}$; AUGMENT$_1$ := `false`
  $\lambda^{(0,k_1)}_{(0,j)} := \lambda_j^{(k_1-1)} - \dfrac{\widehat{c}_j + \lambda_j^{(k_1-1)}}{1 + z_j g_j^{(k_1-1)}}$, $j \in S_1^{k_1} \setminus S_0^{k_0}$
  **while** $k_0 < k_1$ **and not**(AUGMENT$_1$) **do**
    **choose** $j_0^{\max} \in \arg\max \{\lambda^{(0,k_1)}_{(0,j)} : j \in S_1^{k_1} \setminus S_0^{k_0}\}$
    **if** $k_1 = n$ **or** $\lambda^*_{i_1^{k_1}} < \lambda^{(0,k_1)}_{(0,j_0^{\max})}$
      $i_0^{k_0+1} := j_0^{\max}$; $\lambda^*_{(0,i_0^{k_0+1})} := \lambda^{(0,k_1)}_{(0,i_0^{k_0+1})}$
      $S_0^{k_0+1} := S_0^{k_0} \cup \{i_0^{k_0+1}\}$; $k_0 := k_0 + 1$
    **else**
      AUGMENT$_1$ := `true`
    **end** { if }
  **end** { while }
**end** { for }

We have the following result.

**Proposition 3.** *Algorithm* $\mathrm{AG}^0$ *computes index* $\lambda^*_{(0,i)}$ *in no more than* $(5/2)n^2 + O(n)$ *operations.*

**Proof.** The fact that algorithm $\mathrm{AG}^0$ calculates the $\lambda^*_{(0,i)}$ follows by noting that we have obtained it from algorithm $\mathrm{AG}_{\mathcal{F}}$ in Algorithm 4 simply by decoupling the calculation of the $\lambda^*_{(0,i)}$ and the $\lambda^*_{(1,i)} = \lambda^*_i$.

As for the algorithm's arithmetic-operation count, it is dominated by the statements

$$\lambda_{(0,j)}^{(0,k_1)} := \lambda_j^{(k_1-1)} - \frac{\widehat{c}_j + \lambda_j^{(k_1-1)}}{1 + z_j g_j^{(k_1-1)}}, \quad j \in S_1^{k_1} \setminus S_0^{k_0},$$

for $k_1 = 2, \ldots, n+1$, each of which performs no more than $5k_1$ operations. This gives the maximum stated operation count. □

## 6. How Does the Index Depend on Switching Penalties?

We next present and discuss properties on the index dependence on the switching penalties, when considering the case where the latter are constant across states: $c_i \equiv c$, $d_i \equiv d$ and $\phi_i \equiv \phi$ for $i \in \mathcal{X}$. The notation below will make explicit the prevailing penalties, writing $\lambda_{(1,i)}^*(d,\psi)$, and $\lambda_{(0,i)}^*(c,d,\phi,\psi)$.

We write, as $\lambda_i^* \geqslant 0$, the Gittins index, and as $F_i^S \geqslant 0$, the reward metric of the original project with no switching penalties. We will draw on the following expression for the switching index:

$$\lambda_{(0,i)}^*(c,d,\phi,\psi) = \max_{S \subseteq \mathcal{X}:\ i \in S} H\big(c,d,\phi,\psi,F_i^S,G_i^S\big), \tag{51}$$

where

$$H(c,d,\phi,\psi,F,G) \triangleq \frac{-(c+\phi d) + \phi\big(F + (1-\beta)dG\big)}{\dfrac{1-\phi\psi}{1-\beta} + \phi\psi G}.$$

Note that (51) uses the transformation that is considered in Section 2.1, together with the switching-index formulation in (46), while using the result that the original non-restless project's reward metric with transformed rewards $\widetilde{R}_j = (R_j + (1-\beta)d)/\psi$, for $j \in \mathcal{X}$, is $\widetilde{F}_i^S = (F_i^S + (1-\beta)dG_i^S)/\psi$.

We will further use the following preliminary result.

**Lemma 17.**

(a) *If $S \subset S' \subseteq \mathcal{X}$, then $F_i^S \leqslant F_i^{S'}$ and $G_i^S \leqslant G_i^{S'}$.*

(b) *If $d + \psi c \geqslant \phi\psi F_i^{\mathcal{X}}$, then $H(c,d,\phi,\psi,F,G)$ is monotone increasing in $F$ and in $G$.*

**Proof.** (a) The results follows from the interpretation of work and reward metrics, using Assumption 1(ii) for the latter.

(b) This part follows from the following results:

$$\frac{\partial}{\partial F}H(c,d,\phi,\psi,F,G) = \frac{\phi}{\frac{1-\phi\psi}{1-\beta} + \phi\psi G} > 0 \quad \text{and} \quad \frac{\partial}{\partial G}H(c,d,\phi,\psi,F,G) = \phi\frac{d + \psi c - \phi\psi F}{\big(\frac{1-\phi\psi}{1-\beta} + \phi\psi G\big)^2} > 0.$$

□

We have the following result.

**Proposition 4.**

(a) $\lambda_{(1,i)}^*(d,\psi) = (\lambda_i^* + (1-\beta)d)/\psi$.

(b) *If $d + \psi c \geqslant \phi\psi F_i^{\mathcal{X}}$, then $\lambda_{(0,i)}^* = \phi\lambda_i^{\mathcal{X}} - (1-\beta)c$.*

(c) $\lambda_{(0,i)}^*(c,d,\phi,\psi)$ *is convex and piecewise linear in $(c,d)$, decreasing in $c$ and non-increasing in $d$.*

(d) *For $d + \psi c \geqslant \phi\psi F_i^{\mathcal{X}}$, or for $c,d \geqslant 0$ small enough and $R_i > 0$, or for $c = d = 0$, $\lambda_{(0,i)}^*(c,d,\phi,\psi)$ is convex and non-decreasing in $\phi$ and in $\psi$.*

(e) $\lim_{\phi \searrow 0} \lambda_{(0,i)}^*(c,d,\phi,\psi) = -(1-\beta)c$.

(f) $\lambda_{(0,i)}^*(c,d,\phi,\psi) = \phi\lambda_i^N - (1-\beta)c + O(\psi^2)$, *as $\psi \searrow 0$.*

**Proof.** (a) The result follows from noting that $\lambda^*_{(1,i)}(d,\psi)$ is the Gittins index of the project with modified active rewards $\widetilde{R}_j = (R_j + (1-\beta)d)/\psi$ (cf. Section 2.1), which is related to the project Gittins index $\lambda^*_i$ (with unmodified rewards $R_j$) by the stated expression.

(b) Using Lemma 17(b) and $\lambda^{\mathcal{X}}_i = (1-\beta)F^{\mathcal{X}}_i$, we obtain

$$\lambda^*_{(0,i)}(c,d,\phi,\psi) = \max_{(F,G)\in[0,F^{\mathcal{X}}_i]\times[0,G^{\mathcal{X}}_i]} H(c,d,\phi,\psi,F,G) = H(c,d,\phi,\psi,F^{\mathcal{X}}_i,G^{\mathcal{X}}_i) = \phi\lambda^{\mathcal{X}}_i - (1-\beta)c.$$

(c) The result follows by noting that (51) formulates $\lambda^*_{(0,i)}(c,d,\phi,\psi)$ as the maximum of linear functions in $(c,d)$ that decrease in $c$ and are non-increasing in $d$.

(d) Concerning the dependence on $\phi$, when $d+\psi c \geqslant \phi\psi F^{\mathcal{X}}_i$ the result follows by (b). Furthermore,

$$\frac{\partial}{\partial\phi}H(c,d,\phi,\psi,F^S_i,G^S_i) = (1-\beta)\frac{F^S_i - \big(1-(1-\beta)G^S_i\big)(d+\psi c)}{\big(1-\phi\psi\big(1-(1-\beta)G^S_i\big)\big)^2} \geqslant 0$$

$$\frac{\partial^2}{\partial\phi^2}H(c,d,\phi,\psi,F^S_i,G^S_i), = \frac{2(1-\beta)\big(1-(1-\beta)G^S_i\big)\psi}{\big(1-\phi\psi\big(1-(1-\beta)G^S_i\big)\big)^3}\big(F^S_i - \big(1-(1-\beta)G^S_i\big)(d+\psi c)\big) \geqslant 0,$$

where the inequalities hold for $c,d$ small enough, using that $R_i > 0$ so that $F^S_i > 0$, and for $c = d = 0$. Hence, $\lambda^*_{(0,i)}(c,d,\phi,\psi)$ is a maximum of convex non-decreasing functions, which is also convex non-decreasing.

The same argument can be applied to dependence on $\psi$, while using that

$$\frac{\partial}{\partial\psi}H(c,d,\phi,\psi,F^S_i,G^S_i) = \frac{(1-\beta)\big(1-(1-\beta)G^S_i\big)\phi}{\big(1-\phi\psi\big(1-(1-\beta)G^S_i\big)\big)^2}\big(\phi F^S_i - c - \big(1-\big(1-\beta\big)G^S_i\big)\phi d\big)$$

$$\frac{\partial^2}{\partial\psi^2}H(c,d,\phi,\psi,F^S_i,G^S_i) = \frac{2(1-\beta)\big(1-(1-\beta)G^S_i\big)^2\phi^2}{\big(1-\phi\psi\big(1-(1-\beta)G^S_i\big)\big)^3}\big(\phi F^S_i - c - \big(1-\big(1-\beta\big)G^S_i\big)\phi d\big).$$

Parts (e) and (f) follow straightforwardly. □

We conjecture that Lemma 4(c) should hold without the qualifications considered above.

Now, consider the following examples to illustrate the results above . The first example concerns a three-state project with no setdown penalties or setup costs, setup delay transform $\phi$, $\beta = 0.95$,

$$\mathbf{R} = \begin{bmatrix} 0.7221 \\ 0.9685 \\ 0.1557 \end{bmatrix} \quad \text{and} \quad \mathbf{P} = \begin{bmatrix} 0.8061 & 0.1574 & 0.0365 \\ 0.1957 & 0.0067 & 0.7976 \\ 0.1378 & 0.5959 & 0.2663 \end{bmatrix}.$$

Figure 2 plots the project's switching index for each of the three states versus $1-\phi$. Note that each of the lines shown corresponds to one of the project states. The plot agrees with Proposition 4(d, e). It also illustrates that the relative ordering of states that is induced by the switching index can vary with $\phi$.

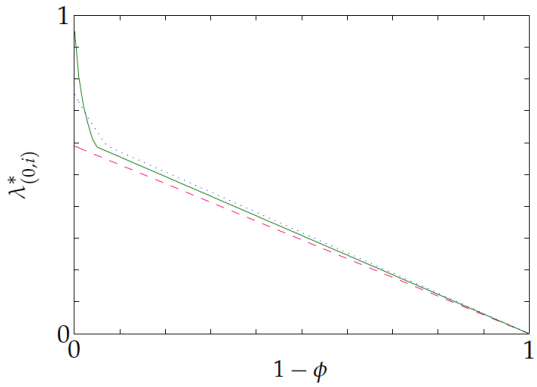


**Figure 2.** Switching index versus setup delay transform.

The following example is based on the same project, but with no setup delays and with setdown delay transform $\psi$. Figure 3 displays the continuation and switching indices for each of the three states versus $1-\psi$. Note that each of the lines shown corresponds to one of the project states. The plots agree with Proposition 4(a,d,f). Note that the continuation index $\lambda^*_{(1,i)}(d,\psi)$ increases to infinity as $\psi$ vanishes, as the incentive of sticking to a project increases steeply as the setdown delay becomes larger. The plot for the switching index further shows that the relative ordering of states can vary with $\psi$.

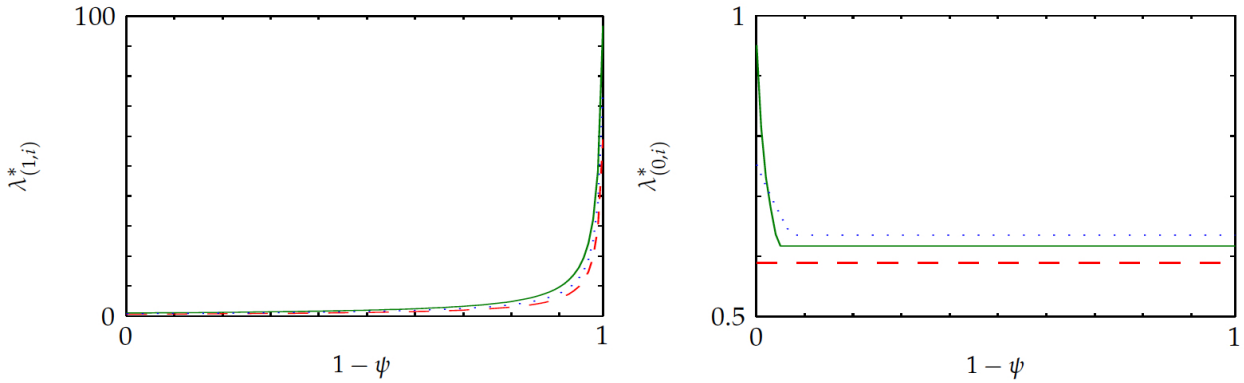


**Figure 3.** Continuation and switching indices versus setdown delay transform.

## 7. Numerical Study

We next report on the results of a numerical study, which is based on MATLAB implementations of the algorithms that are discussed here developed by the author.

The first experiment addressed the runtime of the decoupled index computing method. A random project instance with setup delays and costs was randomly generated for each of the following numbers of states: $n = 500, 1000, \ldots, 5000$. For each such $n$, the time to compute the continuation index and required extra quantities while using the fast-pivoting algorithm with extended output in [28] was recorded, as well as the time for computing the switching index by algorithm $\mathrm{AG}^0$, and the time for jointly computing both indices by using the simplex-based implementation that is given in [49] of the adaptive-greedy algorithm $\mathrm{AG}_{\mathcal{F}}$. This experiment was run on a 2.8 GHz PC with 4 GB of memory.

Figure 4 shows the results. The left pane plots total runtimes (measured in hours) to compute both indices versus $n$. Red squares represent the $\mathrm{AG}_{\mathcal{F}}$ joint-computing scheme, and blue circles represent the two-stage scheme. We see that the latter attained approximately a fourfold speed-up over the former. The right pane plots runtimes (measured in seconds), for the switching index algorithm versus the number of states $n$. The timescale change from hours to seconds highlights the order-of-magnitude speed-up attained.

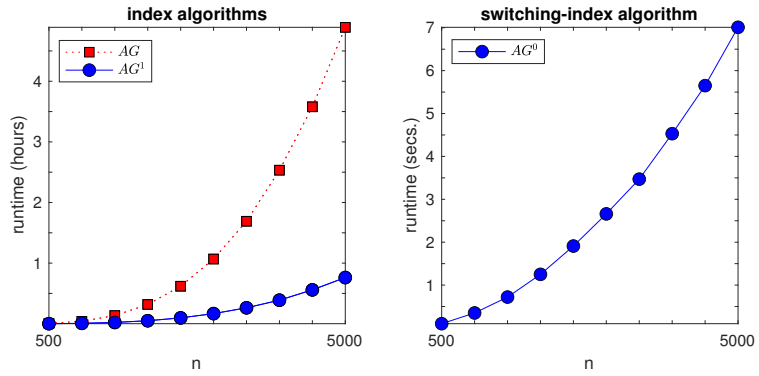


**Figure 4.** Exp. 1: Runtimes of index algorithms.

The following experiments were designed in order to evaluate the average relative performance of the Whittle index policy in randomly generated two- and three-project instances, both versus the optimal problem value, and versus the benchmark Gittins index policy, which does not take setups into account. For each problem instance, the optimal value was calculated by solving with the CPLEX LP solver the LP formulation of the DP optimality equations. The Whittle index and benchmark scheduling policies were evaluated by solving, with MATLAB, the appropriate systems of linear evaluation equations.

The second experiment was designed to assess the dependence of the relative performance of Whittle's index policy for two-project instances on a constant setup-time transform $\phi$ and discount factor $\beta$—with no setdown penalties. A sample of 100 randomly generated instances with 10-state projects was obtained with MATLAB. In each instance, the parameters for each project were independently drawn: transition probabilities (by scaling a matrix with uniform entries) and uniform (between 0 and 1) active rewards. For every instance $k = 1, \ldots, 100$ and parameters $(\phi, \beta) \in [0.5, 0.99] \times [0.5, 0.95]$—with a 0.1 grid—the optimal value $V^{(k),\mathrm{opt}}$ and the values of the Whittle index ($V^{(k),\mathrm{W}}$) and benchmark ($V^{(k),\mathrm{bench}}$) policies were calculated, together with the relative optimality of the Whittle index policy $\Delta^{(k),\mathrm{W}} \triangleq 100(V^{(k),\mathrm{opt}} - V^{(k),\mathrm{W}})/|V^{(k),\mathrm{opt}}|$, and the optimality-gap ratio of the Whittle index over the benchmark policy $\rho^{(k),\mathrm{W,bench}} \triangleq 100(V^{(k),\mathrm{W}} - V^{(k),\mathrm{opt}})/(V^{(k),\mathrm{bench}} - V^{(k),\mathrm{opt}})$. The latter were then averaged over the 100 instances for each $(c, \beta)$ pair, in order to obtain the average values $\Delta^{\mathrm{W}}$ and $\rho^{\mathrm{W,bench}}$.

Values $V^{(k),\mathrm{opt}}$, $V^{(k),\mathrm{W}}$ and $V^{(k),\mathrm{bench}}$ were computed, as follows. The corresponding value functions $V^{(k),\mathrm{opt}}_{((a_1^-,i_1),(a_2^-,i_2))}$, $V^{(k),\mathrm{W}}_{((a_1^-,i_1),(a_2^-,i_2))}$ and $V^{(k),\mathrm{bench}}_{((a_1^-,i_1),(a_2^-,i_2))}$ were calculated. Subsequently, the values were calculated when considering that both projects start out being passive, as

$$V^{(k),\pi} \triangleq \frac{1}{n^2} \sum_{i_1,i_2 \in N} V^{(k),\pi}_{((0,i_1),(0,i_2))}, \quad \pi \in \{\mathrm{opt}, \mathrm{W}, \mathrm{bench}\}. \tag{52}$$

Figure 5 displays, in its left pane, the relative gap $\Delta^{\mathrm{W}}$ versus $\phi$—note the inverted $\phi$-axis used throughout—for multiple $\beta$, while using cubic interpolation. The gap starts at 0 as $\phi$ approaches 1 (as the optimal policy is then obtained), and then grows up to a maximum, which is below 0.18%, and then decreases to 0 as $\phi$ gets smaller. That pattern agrees with intuition: for small enough $\phi$, both the optimal and Whittle index policies initially pick a project and stick to it. Because the best such project can be determined by single-project evaluations, the Whittle index policy will correctly choose it. The right pane shows that $\Delta^{\mathrm{W}}$ is not monotonic in $\beta$, as it is increasing for small $\beta$ and then decreases for $\beta$ closer to 1. Hence, in the left pane, the higher peaks typically correspond to larger values of $\beta$.

Figure 6 shows similar plots for the optimality-gap ratio $\rho^{\mathrm{W,bench}}$ of the Whittle index over the benchmark policy. They highlight that the average optimality gap for the Whittle index policy remains below 45% of that for the benchmark policy. The left pane shows that the ratio vanishes for $\phi$ that is small enough, as the Whittle index policy is then optimal.

Additionally, the right pane shows that the ratio is increasing with $\beta$. Thus, in the left pane, for fixed $\phi$, higher values correspond to larger $\beta$.

The third experiment was similar in nature as the previous one, but, when considering instead a constant setup delay $T$ for each project, $\phi = \beta^T$. Figures 7 and 8 show the results, which highlight that Whittle's index policy was optimal for $T \geqslant 2$, its relative optimality gap did not exceed 0.06%, and it substantially outperformed the benchmark Gittins-index policy, as the optimality-gap ratio stays below 2%.

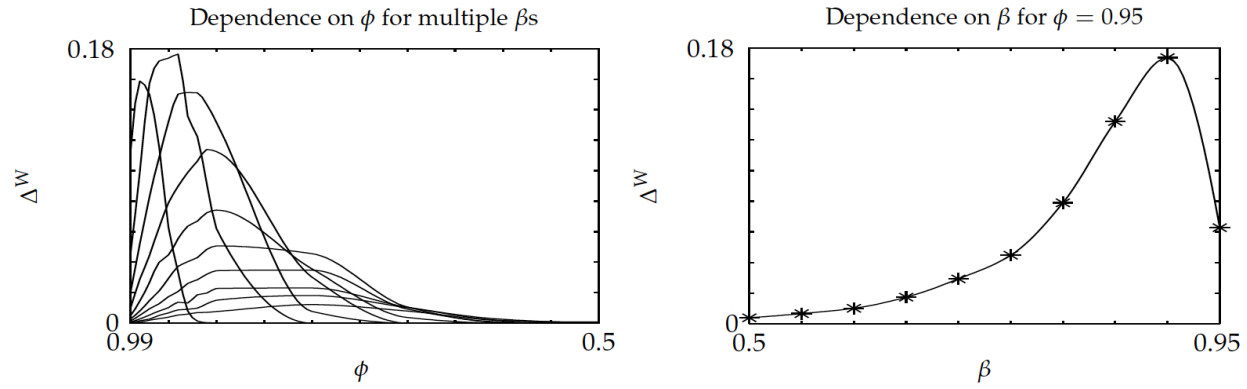


**Figure 5.** Exp. 2: Average optimality gap (%) of Whittle's index policy.

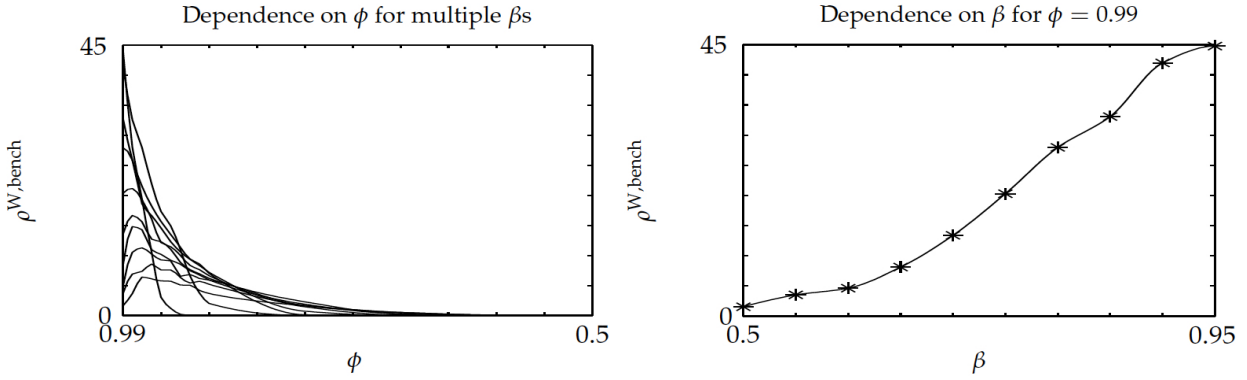


**Figure 6.** Exp. 2: Average optimality-gap ratio (%) of Whittle's index policy over the benchmark policy.

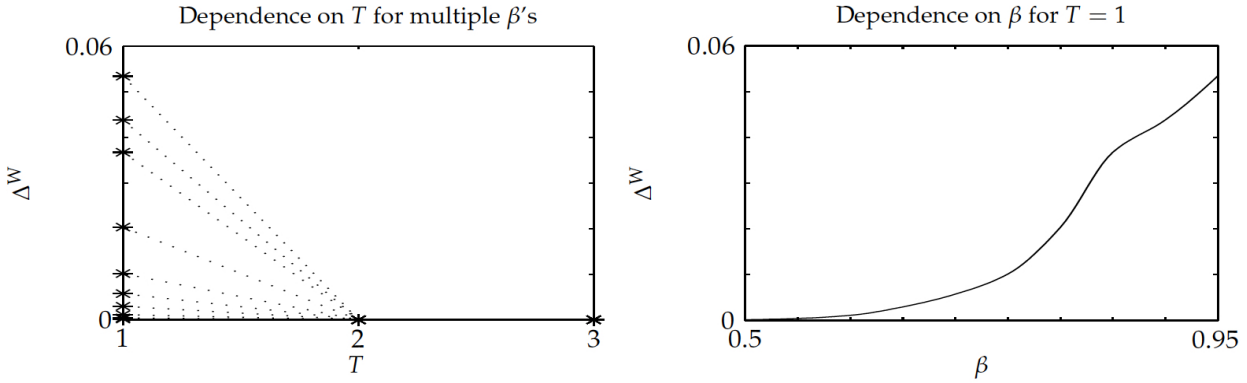


**Figure 7.** Exp. 3: Average optimality gap (%) of Whittle's index policy.

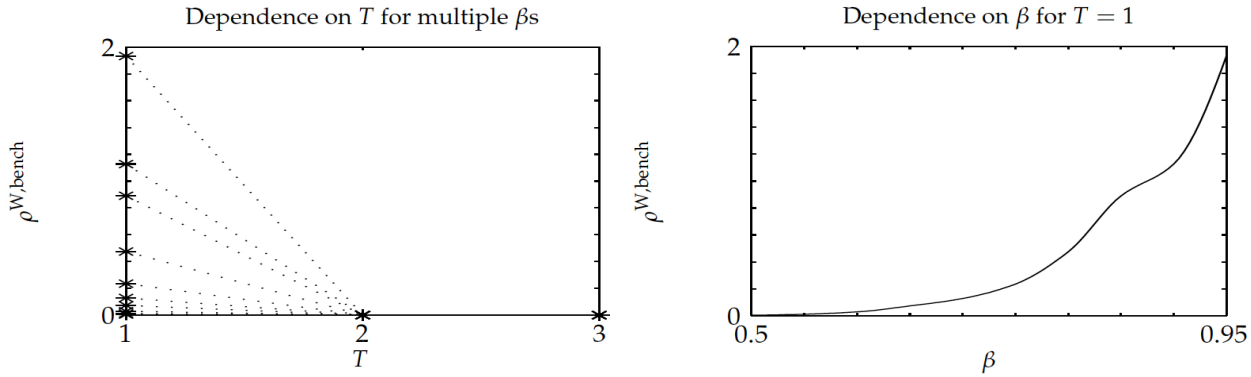


**Figure 8.** Exp. 3: Average optimality-gap ratio (%) of Whittle's index over benchmark policy.

The fourth experiment addressed the effect of asymmetric (and constant) setup delay transforms, with these varying over the range $(\phi_1, \phi_2) \in [0.8, 0.99]^2$, in two-project instances with discount factor $\beta = 0.9$. In the left contour plot in Figure 9 it is shown that the average relative optimality gap of Whittle's index policy, $\Delta^{\mathrm{W}}$, reaches a maximum of approximately 0.14%, and it vanishes as both $\phi_1$ and $\phi_2$ get close to unity, and as either of them becomes small enough. The right contour plot shows that the optimality-gap ratio $\rho^{\mathrm{W}}$ reaches the maximum values of nearly 50%, then vanishing as either $\phi_1$ or $\phi_2$ becomes sufficiently small.

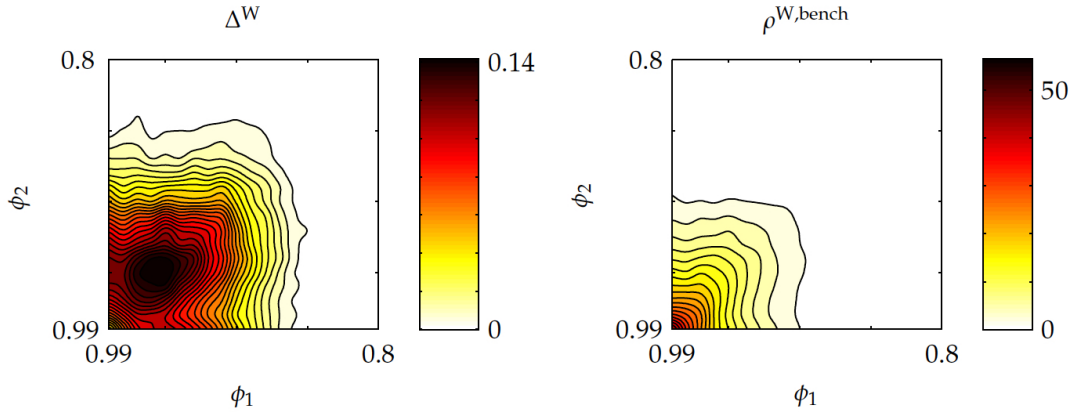


**Figure 9.** Exp. 4: Average relative performance (%) of Whittle's index policy versus $(\phi_1, \phi_2)$, for $\beta = 0.9$.

The fifth experiment studied the effect of state-dependent setup delay parameters $\phi_i$, as the discount factor is changed. Uniform[0.9, 1] i.i.d. state-dependent setup costs were randomly generated for every instance. The left pane shown in Figure 10 displays the average relative optimality gap versus the discount factor, showing that such a gap stays below 0.14%. The right pane highlights that the average optimality-gap ratio $\rho^{\mathrm{W,bench}}$ stays below 20%.

The sixth experiment considered the relative performance of Whittle's index policy on three-project instances in terms of a setup delay parameter $\phi$ and discount factor, while using a random sample of 100 instances of three eight-state projects. For each instance, the parameters varied over the range $(\phi, \beta) \in [0.5, 0.99] \times [0.5, 0.95]$. The results are displayed in Figures 11 and 12, which are the counterparts of Figures 5 and 6. Comparing Figures 5 and 11 shows a slight degradation of performance for Whittle's index policy in the latter, although the average gap $\Delta^{\mathrm{W}}$ stays small, beneath 0.25%. Comparing Figures 6 and 12 shows similar values for the ratio $\rho^{\mathrm{W,bench}}$.

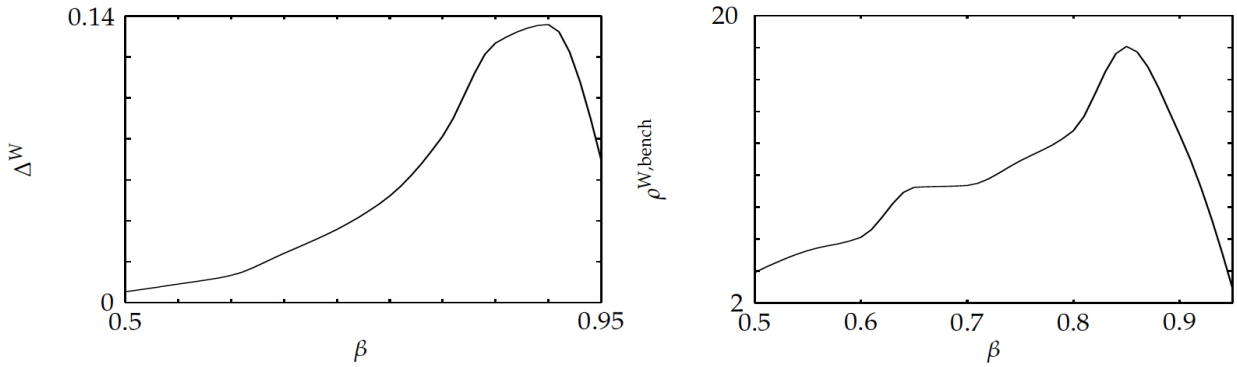


**Figure 10.** Exp. 5: Average relative performance (%) of Whittle's index policy with state-dependent setup delays.

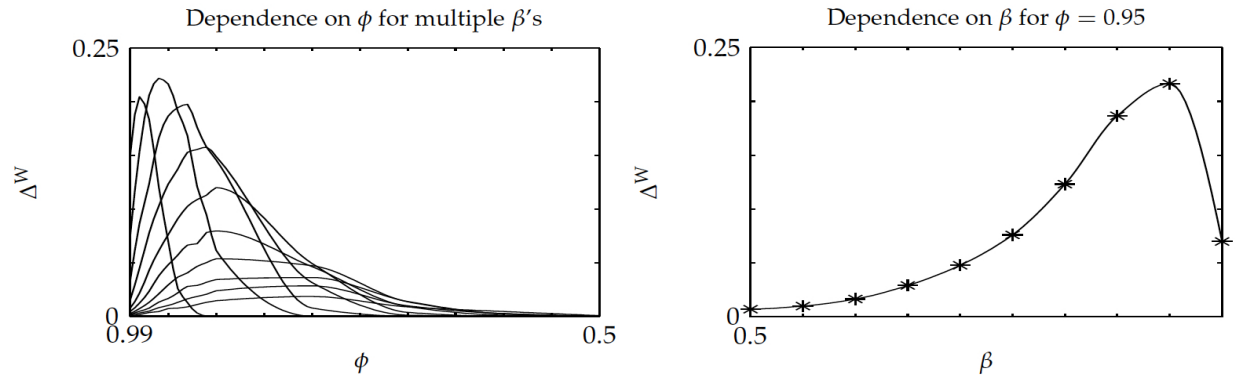


**Figure 11.** Exp. 6: Version of Figure 5 for three-project instances.

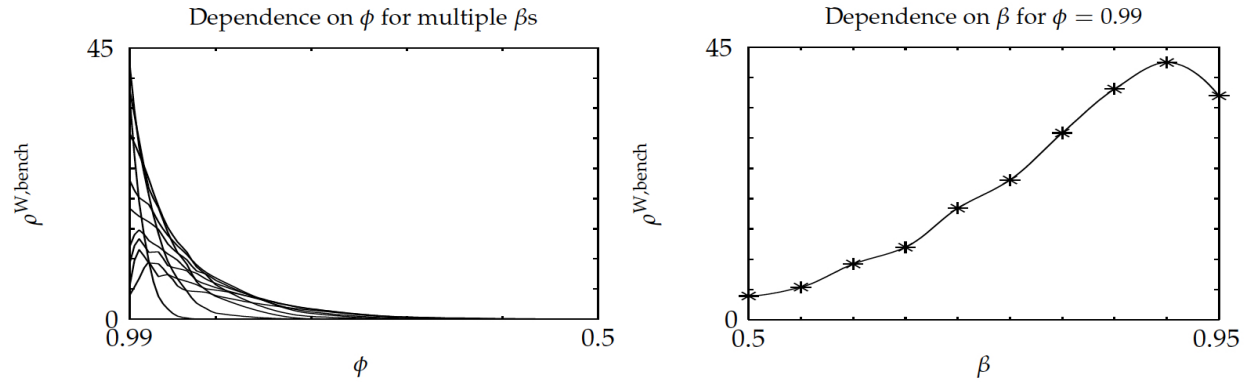


**Figure 12.** Exp. 6: Version of Figure 6 for three-project instances.

## 8. Conclusions

Bandit models with switching penalties are relevant for a wide variety of applications. Computing optimal policies is generally intractable, which motivates the search for simple policies that can be implemented in practice and perform well. Index policies are an appealing class of policies, which have been proposed for such problems. Yet, while algorithms are given in [10,27] for computing the Asawa and Teneketzis index for a bandit with switching costs only, no algorithms have been given in the literature in order to compute the extension of such an index for bandits with switching penalties that incorporate switching delays. This paper presents the first such algorithm. It further provides evidence in a numerical study that the resulting index policy is nearly optimal across the instances considered. This work could be extended in several directions, including developing specialized algorithms for computing the index, in particular, models that arise in applications.

**Funding:** This research has been developed over a number of years, and has been funded in part by the Spanish Government under grants MEC MTM2004-02334 and MTM2007-63140, and PID2019-109196GB-I00 / AEI / 10.13039/501100011033. This work has also been funded in part by the *Comunidad de Madrid* in the setting of the multi-year agreement with Universidad Carlos III de Madrid within the line of activity "*Excelencia para el Profesorado Universitario*", in the framework of the V Regional Plan of Scientific Research and Technological Innovation 2016–2020.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data have not been made publicly available because the article describes in full detail how they can be generated by computer simulation and computational experiments.

**Acknowledgments:** The author has presented a preliminary version of this work at ValueTools '07, the Second International Conference on Performance Evaluation Methodologies and Tools, which appears in abridged form in the online proceedings [51]. A preliminary version was also posted as the working paper [52].

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Gittins, J.C. *Multi-Armed Bandit Allocation Indices*; Wiley: Chichester, UK, 1989.
2. Gittins, J.C.; Jones, D.M. A dynamic allocation index for the sequential design of experiments. In *Progress in Statistics* (*Eur. Meeting of Statisticians, Budapest, 1972*); Gani, J., Sarkadi, K., Vincze, I., Eds.; North-Holland: Amsterdam, The Netherlands, 1974; pp. 241–266.
3. Gittins, J.C. Bandit processes and dynamic allocation indices. *J. R. Statist. Soc. Ser. B* **1979**, *41*, 148–177. [CrossRef]
4. Whittle, P. Multi-armed bandits and the Gittins index. *J. R. Statist. Soc. Ser. B* **1980**, *42*, 143–149. [CrossRef]
5. Weber, R. On the Gittins index for multiarmed bandits. *Ann. Appl. Probab.* **1992**, *2*, 1024–1033. [CrossRef]
6. Bertsimas, D.; Niño-Mora, J. Conservation laws, extended polymatroids and multiarmed bandit problems; a polyhedral approach to indexable systems. *Math. Oper. Res.* **1996**, *21*, 257–306. [CrossRef]
7. Bellman, R. A problem in the sequential design of experiments. *Sankhyā* **1956**, *16*, 221–229.
8. Varaiya, P.P.; Walrand, J.C.; Buyukkoc, C. Extensions of the multiarmed bandit problem: The discounted case. *IEEE Trans. Automat. Control* **1985**, *30*, 426–439. [CrossRef]
9. Banks, J.S.; Sundaram, R.K. Switching costs and the Gittins index. *Econometrica* **1994**, *62*, 687–694. [CrossRef]
10. Asawa, M.; Teneketzis, D. Multi-armed bandits with switching penalties. *IEEE Trans. Automat. Control* **1996**, *41*, 328–348. [CrossRef]
11. Jun, T.S. Survey on the bandit problem with switching costs. *De Econ.* **2004**, *152*, 513–541. [CrossRef]
12. Agrawal, R.; Hegde, M.V.; Teneketzis, D. Asymptotically efficient adaptive allocation rules for the multiarmed bandit problem with switching cost. *IEEE Trans. Automat. Control* **1988**, *33*, 899–906. [CrossRef]
13. Van Oyen, M.P.; Pandelis, D.G.; Teneketzis, D. Optimality of index policies for stochastic scheduling with switching penalties. *J. Appl. Probab.* **1992**, *29*, 957–966. [CrossRef]
14. Bergemann, D.; Valimaki, J. Stationary multi-choice bandit problems. *J. Econ. Dyn. Control* **2001**, *25*, 1585–1594. [CrossRef]
15. Sundaram, R.K. Generalized bandit problems. In *Social Choice and Strategic Decisions*; Austen-Smith, D., Duggan, J., Eds.; Studies in Choice and Welfare; Springer: Berlin, Germany, 2005; pp. 131–162.
16. Arlotto, A.; Chick, S.E.; Gans, N. Optimal hiring and retention policies for heterogeneous workers who learn. *Manag. Sci.* **2014**, *60*, 110–129. [CrossRef]
17. Hauser, J.R.; Liberali, G.; Urban, G. Website morphing 2.0: Switching costs, partial exposure, random exit, and when to morph. *Manag. Sci.* **2014**, *60*, 1594–1616. [CrossRef]
18. Liberali, G.B.; Hauser, J.R.; Urban, G.L. Morphing theory and application. In *Handbook of Marketing Decision Models*; Wierenga, B., van der Lans, R., Eds.; International Series in Operations Research & Management Science; Springer: Cham, Switzerland, 2017; Chapter 18. Volume 254, pp. 531–562.
19. Lin, S.; Zhang, J.J.; Hauser, J.R. Learning from experience, simply. *Mark. Sci.* **2015**, *34*, 1–19. [CrossRef]
20. Huang, J.; Gan, X.; Feng, X. Multi-armed bandit based opportunistic channel access: A consideration of switch cost. In Proceedings of the IEEE International Conference on Communications—Ad-hoc and Sensor Networking Symposium, Budapest, Hungary, 9–13 June 2013; pp. 1651–1655.
21. Qin, Z.Q.; Wang, J.L.; Chen, J.; Sun, Y.M.; Du, Z.Y.; Xu, Y.H. Opportunistic channel access with repetition time diversity and switching cost: A block multi-armed bandit approach. *Wirel. Netw.* **2018**, *24*, 1683–16977. [CrossRef]
22. McCardle, K.F.; Tsetlin, I.; Winkler, R.L. When to abandon a research project and search for a new one. *Oper. Res.* **2018**, *66*, 799–813. [CrossRef]

23. Savelov, M.P. Gittins index for simple family of Markov bandit processes with switching cost and no discounting. *Theory Probab. Appl.* **2019**, *64*, 355–364. [CrossRef]
24. Dusonchet, F.; Hongler, M.O. Optimal hysteresis for a class of deterministic deteriorating two-armed bandit problem with switching costs. *Automatica* **2003**, *39*, 1947–1955. [CrossRef]
25. Dusonchet, F.; Hongler, M.O. Priority index heuristic for multi-armed bandit problems with set-up costs and/or set-up time delays. *Int. J. Comput. Integr. Manuf.* **2006**, *19*, 210–219. [CrossRef]
26. Mason, A.J.; Anderson, E.J. Minimizing flow time on a single machine with job classes and setup times. *Nav. Res. Logist.* **1991**, *64*, 333–350. [CrossRef]
27. Niño-Mora, J. A faster index algorithm and a computational study for bandits with switching costs. *INFORMS J. Comput.* **2008**, *20*, 255–269. [CrossRef]
28. Niño-Mora, J. A $(2/3)n^3$ fast-pivoting algorithm for the Gittins index and optimal stopping of a Markov chain. *INFORMS J. Comput.* **2007**, *19*, 596–606. [CrossRef]
29. Whittle, P. Restless bandits: Activity allocation in a changing world. *J. Appl. Probab.* **1988**, *25A*, 287–298. [CrossRef]
30. Niño-Mora, J. Restless bandits, partial conservation laws and indexability. *Adv. Appl. Probab.* **2001**, *33*, 76–98. [CrossRef]
31. Niño-Mora, J. Dynamic allocation indices for restless projects and queueing admission control: A polyhedral approach. *Math. Program.* **2002**, *93*, 361–413. [CrossRef]
32. Niño-Mora, J. Restless bandit marginal productivity indices, diminishing returns and optimal control of make-to-order/make-to-stock M/G/1 queues. *Math. Oper. Res.* **2006**, *31*, 50–84. [CrossRef]
33. Niño-Mora, J. A verification theorem for threshold-indexability of real-state discounted restless bandits. *Math. Oper. Res.* **2020**, *45*, 465–496. [CrossRef]
34. Niño-Mora, J. Dynamic priority allocation via restless bandit marginal productivity indices. *Top* **2007**, *15*, 161–198. [CrossRef]
35. Papadimitriou, C.H.; Tsitsiklis, J.N. The complexity of optimal queuing network control. *Math. Oper. Res.* **1999**, *24*, 293–305. [CrossRef]
36. Qian, Y.; Zhang, C.; Krishnamachari, B.; Tambe, M. Restless poachers: Handling exploration-exploitation tradeoffs in security domains. In Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems, Singapore, 9–13 May 2016; ACM: New York, NY, USA, 2016; pp. 123–131.
37. Fu, J.; Moran, B.; Guo, J.; Wong, E.W.M.; Zukerman, M. Asymptotically optimal job assignment for energy-efficient processor-sharing server farms. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 4008–4023. [CrossRef]
38. Borkar, V.S.; Pattathil, S. Whittle indexability in egalitarian processor sharing systems. *Ann. Oper. Res.* **2017**, 1–21. [CrossRef]
39. Borkar, V.S.; Ravikumar, K.; Saboo, K. An index policy for dynamic pricing in cloud computing under price commitments. *Appl. Math.* **2017**, *44*, 215–245. [CrossRef]
40. Borkar, V.S.; Kasbekar, G.S.; Pattathil, S.; Shetty, P.Y. Opportunistic scheduling as restless bandits. *IEEE Trans. Control Netw. Syst.* **2018**, *5*, 1952–1961. [CrossRef]
41. Gerum, P.C.L.; Altay, A.; Baykal-Gursoy, M. Data-driven predictive maintenance scheduling policies for railways. *Transport. Res. Part C Emerg. Technol.* **2019**, *107*, 137–154. [CrossRef]
42. Abbou, A.; Makis, V. Group maintenance: A restless bandits approach. *INFORMS J. Comput.* **2019**, *31*, 719–731. [CrossRef]
43. Ayer, T.; Zhang, C.; Bonifonte, A.; Spaulding, A.C.; Chhatwal, J. Prioritizing hepatitis C treatment in US prisons. *Oper. Res.* **2019**, *67*, 853–873. [CrossRef]
44. Niño-Mora, J. Resource allocation and routing in parallel multi-server queues with abandonments for cloud profit maximization. *Comput. Oper. Res.* **2019**, *103*, 221–236. [CrossRef]
45. Fu, J.; Moran, B. Energy-efficient job-assignment policy with asymptotically guaranteed performance deviation. *IEEE/ACM Trans. Netw.* **2020**, *28*, 1325–1338. [CrossRef]
46. Hsu, Y.P.; Modiano, E.; Duan, L.J. Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals. *IEEE Trans. Mob. Comput.* **2020**, *19*, 2903–2915. [CrossRef]
47. Sun, J.Z.; Jiang, Z.Y.; Krishnamachari, B.; Zhou, S.; Niu, Z.S. Closed-form Whittle's index-enabled random access for timely status update. *IEEE Trans. Commun.* **2020**, *68*, 1538–1551. [CrossRef]
48. Li, D.; Ding, L.; Connor, S. When to switch? Index policies for resource scheduling in emergency response. *Prod. Oper. Manag.* **2020**, *29*, 241–262. [CrossRef]
49. Niño-Mora, J. A fast-pivoting algorithm for Whittle's restless bandit index. *Mathematics* **2020**, *8*, 2226. [CrossRef]
50. Yao, D.D. Comments on: "Dynamic priority allocation via restless bandit marginal productivity indices" [Top **15** (2007), no. 2, 161–198] by J. Niño-Mora. *Top* **2007**, *15*, 220–223. [CrossRef]
51. Niño-Mora, J. Computing an index policy for bandits with switching penalties. In Proceedings of the ValueTools '07, the Second International Conference on Performance Evaluation Methodologies and Tools, Nantes, France, 23–25 October 2007; ICST: Brussels, Belgium, 2007. Available online: https://dl.acm.org/doi/10.5555/1345263.1345361 (accessed on 29 December 2020).
52. Niño-Mora, J. *Two-Stage Index Computation for Bandits with Switching Penalties II: Switching Delays*; Working Paper 07-42, Statistics and Econometrics Series 10; Univ. Carlos III de Madrid: Madrid, Spain, 2007.