05031 – Extended Abstract
Marginal Productivity Index Policies for Scheduling Restless Bandits with Switching Penalties
— Dagstuhl Seminar —

José Niño-Mora*

Universidad Carlos III de Madrid, Department of Statistics
E-28911 Leganés (Madrid), Av. Universidad 30, Spain
jnimora@alum.mit.edu

Abstract. We address the dynamic scheduling problem for discrete-state restless bandits, where sequence-independent setup penalties (costs or delays) are incurred when starting work on a project. We reformulate such problems as restless bandit problems without setup penalties, and then deploy the theory of marginal productivity indices (MPIs) and partial conservation laws (PCLs) we have introduced and developed in recent work, building on and extending previous work by Gittins (1979) and Whittle (1988). As a result, we obtain new dynamic index policies for scheduling restless bandits with setup penalties.

Keywords. restless bandits, switching penalties, setups, index policies, conservation laws

1 Introduction

The *restless bandit problem (RBP)*, in its basic version, concerns the optimal dynamic allocation of effort to a collection of stochastic projects $k \in \mathbb{K} \triangleq \{1,\ldots,K\}$. Project k is modelled as a discrete-time Markov decision process (MDP), moving over the finite state space N_k, and involving binary actions $a_k = 1$ (work/active) or $a_k = 0$ (rest/passive). Its dynamics are given by one-period state-transition probability matrices $p_k^{a_k}(i_k, j_k)$, and its one-period costs are given by state- and action-dependent cost rate function $h_k^{a_k}(i_k)$. Costs are discounted over time at the geometric rate $0 < \beta < 1$. We denote by $X_k(t)$ and $a_k(t)$ the project's state and action at period $t = 0, 1, \ldots$. The state space N_k is partitioned into the set $N_k^{\{0,1\}}$ of *controllable states*, where there is an effective choice of action, and the set $N_k^{\{0\}}$ of *uncontrollable states*, where the passive action must be taken.

* This research has been supported in part by the Spanish Ministry of Education & Science under a *Ramón y Cajal Investigator Award* and project MTM2004-02334, and by the European Union's Network of Excellence Euro-NGI. The work has been presented at the Dagstuhl Seminar on Algorithms for Optimization with Incomplete Information, January 16–21, 2005.

Dagstuhl Seminar Proceedings 05031
Algorithms for Optimization with Incomplete Information
http://drops.dagstuhl.de/opus/volltexte/2005/64

At each time period, at most one of the projects can be worked on, among those occupying a controllable state. We thus have the sample-path service capacity constraint

$$\sum_{k\in\mathbb{K}} a_k(t) \leq 1, \quad t \geq 0.$$

Actions are dynamically prescribed by adoption of a *scheduling policy* π, chosen from the class $\mathit{\Pi}$ of admissible policies, which are only required to be nonanticipative and to satisfy the service capacity constraint.

The RBP is to find an admissible policy minimizing the expected total discounted value of costs accrued over an infinite horizon:

$$\min_{\pi\in\mathit{\Pi}} \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t h_k^{a_k(t)}(X_k(t))\right]. \tag{1}$$

Problem (1) has a great modelling power, yet in counterpart it is computationally intractable (P-space hard; see [1]). Yet, the underlying assumption in problem (1) that it is costless to switch between projects is not realistic in a variety of applications, where such switching involving significant costs and/or delays. As a result, optimal or near-optimal policies for problem (1) may involve a high frequency of switching which is not acceptable in such applications. This fact motivates the research interest in extensions of the RBP which incorporate switching penalties.

We focus in this paper on the case of positive sequence-independent setup costs. Thus, we consider that the first period project k is worked on, a setup cost $d_k > 0$ is incurred. In such setting, our goal is to design a well-grounded and tractable *index policy*.

In such extended model, index policies are based on definition of an index $\nu_k\colon \{0,1\}\times N_k^{\{0,1\}} \to \mathbb{R}$, which is a real function of the project's *augmented state* (a_k^-, i_k). Here, a_k^- represents the previous action taken at the project. The resulting index policy prescribes to work at each time on the project with current larger index value, among those occupying controllable states, if any. This raises the research issue of How to define and construct well-grounded index functions for a restless bandit project?

A powerful approach to index design, introduced in [2], was extended in [3] and recently developed in [4,5,6]. In short, such approach defines a *marginal productivity index (MPI)*, which has a sound economic interpretation, as it measures the marginal productivity of work at every state. The MPI exists in models that satisfy the economic law of diminishing marginal returns to effort. Hence, MPI-based resource allocation policies seek to dynamically allocate a resource to its currently more productive use, using the MPI as a proxy measure of the true marginal productivity of effort.

In [7], we deploy such approach to design and construct new dynamic MPI policies for scheduling restless bandits with setup costs. The approach is further extended to address the case of setup delays. Here we outline the approach and main results.

2 Indexability and the MPI

2.1 Indexable projects

We review in this section the definition, interpretation and construction of the discounted MPI, as it applies to a generic project $k \in \mathbb{K}$. We now consider the subsystem obtained by considering a generic class k *in isolation*. To reduce notational clutter, the project label k is dropped in what follows. Thus, e.g. now Π denotes the space of admissible service control policies for operating the class' subsystem, prescribing the action $a(t) \in \{0, 1\}$ to be taken at each decision epoch.

To evaluate the value of costs incurred under a policy $\pi \in \Pi$, when starting at state $i \in N$, we use the *discounted cost measure*

$$f^{\pi}(i) \triangleq \mathbb{E}_i^{\pi}\left[\sum_{t=0}^{\infty} h(X(t))\beta^t \, dt\right].$$

We further evaluate the work expended, by the *discounted work measure*

$$g^{\pi}(i) \triangleq \mathbb{E}_i^{\pi}\left[\sum_{t=0}^{\infty} a(t)\beta^t \, dt\right],$$

To avoid technical issues arising from the choice of initial state, we consider this to be drawn from a distribution having a positive probability mass function $p(i) > 0$ for $i \in N$. We denote the resulting cost and work measures by f^{π} and g^{π}.

Suppose now that the server is paid a *wage* of $\nu \in$ € per unit work performed. We will address the project's ν*-wage subproblem*

$$\min_{\pi \in \Pi} f^{\pi} + \nu g^{\pi}, \tag{2}$$

which is to find an admissible policy minimizing the value of its holding and working costs.

To solve problem (2), we postulate (and then establish) that its optimal policies are of *threshold* type; namely, that there exists a state ordering $m_0, m_1, \ldots, m_n$, with

$$N^{\{0\}} = \{m_0\} = \{0\} \quad \text{and} \quad N^{\{0,1\}} = \{m_1, m_2, \ldots, m_n\}, \tag{3}$$

such that the policies prescribe to work when the project' state lies "above" —relative to such ordering— a threshold state, and to rest otherwise. We represent a threshold policy by its *active-state set*. These are of the form

$$S(m_i) \triangleq \begin{cases} \{m_{i+1}, \ldots, m_n\} & \text{if } 0 \leq i < n \\ \emptyset & \text{if } i = n. \end{cases} \tag{4}$$

giving the nested *active-state set family*

$$\mathscr{F} \triangleq \{S(m_0), S(m_1), \ldots, S(m_n)\}.$$

We will henceforth refer to such policies as $\mathscr{F}$*-policies*, writing e.g. $f^{S(i)}$, $g^{\alpha, S(i)}$.

We next define a key property of the project based on the structure of optimal policies for problem (2) as the prevailing wage $\nu \in \mathbb{R}$ varies. We say the project is $\mathscr{F}$-*indexable* (under the discounted criterion) if there exists an *index* $\nu^*\colon N^{\{0,1\}} \to \mathbb{R}$ which is nondecreasing along the state ordering, i.e.

$$\nu^*(m_1) \leq \cdots \leq \nu^*(m_n), \tag{5}$$

such that, for any $0 < i < n$, the $S(m_i)$-*active policy* is optimal for problem (2) iff $\nu^*(m_i) \leq \nu \leq \nu^*(m_{i+1})$. We then term ν^* the project's *discounted MPI*.

When it exists, the MPI gives an intuitively appealing rule to solve problem (2): it is optimal to work in a state $i \in N^{\{0,1\}}$ iff the latter's MPI value does not exceed the prevailing wage, i.e. $\nu(i) \leq \nu$. This suggests, drawing on the economic theory of resource allocation, that $\nu^*(i)$ must measure the *marginal productivity of work at state* i. Such is indeed the case, as established in [6]. In that paper, we further prove the result that the project is $\mathscr{F}$-indexable iff it obeys the economic *law of diminishing marginal returns (to work)*, consistently with $\mathscr{F}$-policies. Namely, if one considers the *achievable work-cost performance region* spanned by points (g^π, f^π) as π ranges over $\mathit{\Pi}$, its lower boundary (*efficient frontier*) is the piecewise linear and convex function obtained by linear interpolation on points $(g^{S(m_i)}, f^{S(m_i)})$, for $i \in N$. The discounted MPI thus has the evaluation

$$\nu^*(m_i) = \frac{f^{S(m_i)} - f^{S(m_{i-1})}}{g^{S(m_{i-1})} - g^{(m_i)}}, \quad 1 \leq i \leq n. \tag{6}$$

2.2 PCL-indexability conditions and MPI calculation

We will not discuss here the PCL framework. For our present purpose, it will suffice to formulate next the relevant *PCL-indexability* conditions that need to be checked to ensure indexability.

Given an action $a \in \{0,1\}$ and an active-state set $S \in \mathscr{F}$, denote by $\langle a, S\rangle$ the policy that takes action a in the initial *period* (between decision epochs), and adopts the S-active policy thereafter. For every controllable state $i \in N^{\{0,1\}}$ and set $S \in \mathscr{F}$, define the *discounted* (i,S)*-marginal workload*

$$w^S(i) \triangleq g^{\langle 1,S\rangle}(i) - g^{\langle 0,S\rangle}(i). \tag{7}$$

Notice that $w^S(i)$ measures the marginal increment in work expended which results from having the server work instead of rest in the initial period, given the S-active policy is used thereafter.

We analogously define the *discounted* (i,S)*-marginal cost*

$$c^S(i) \triangleq f^{\langle 0,S\rangle}(i) - f^{\langle 1,S\rangle}(i). \tag{8}$$

Thus, $c^S(i)$ measures the marginal increment in cost incurred which results from having the server rest instead of work in the initial period, given the S-active policy is used thereafter.

Define now the *discounted* (i, S)*-marginal productivity rate* by

$$\nu^S(i) \triangleq \frac{c^S(i)}{w^S(i)}, \tag{9}$$

provided the denominator does not vanish. Finally, we define *index* $\nu^*\colon N^{\{0,1\}} \to \mathbb{R}$ by

$$\nu^*(m_i) \triangleq \nu^{S(m_{i-1})}(m_i), \quad 1 \leq i \leq n. \tag{10}$$

Let us now say that the class is *PCL*$(\mathscr{F})$*-indexable*, relative to the discounted criterion, if the following conditions hold:

(i) Positive marginal workloads: $w^S(i) > 0$, for $i \in N^{\{0,1\}}, S \in \mathscr{F}$.
(ii) Nondecreasing index: (5) holds.

The key result we will use to establish indexability, proven in [4,5,6] in increasingly general settings, is the following.

Theorem 1. *A PCL*$(\mathscr{F})$*-indexable class is* $\mathscr{F}$*-indexable, with MPI* $\nu^*(\cdot)$.

3 PCL-indexability for a project with setup costs

Consider now a restless bandit project, as above, but with the additional feature of incorporating a setup cost $d > 0$. In such setting, we define

$$\widehat{\mathscr{F}} \triangleq \left\{(S^0, S^1) \triangleq \{0\} \times S^0 \cup \{1\} \times S^1 \colon S^0 \subseteq S^1, S^0, S^1 \in \mathscr{F}\right\}$$

We can prove the following result.

Theorem 2. *Under the work regularity conditions for the project without setup costs, it holds that*

$$w^{(S^0,S^1)}(a, i) > 0, \quad (a, i) \in \{0, 1\} \times N^{\{0,1\}}, (S^0, S^1) \in \widehat{\mathscr{F}}.$$

Our main result is the following.

Theorem 3. *If the project without setup costs is PCL*$(\mathscr{F})$*-indexable, then the project with setup costs is PCL*$(\widehat{\mathscr{F}})$*-indexable.*

As a result of Theorem 3, we can construct a well-defined MPI by applying the adaptive-greedy algorithm introduced in [4,5].

Such results can be extended to the case of setup delays. See [7] for the complete paper.

References

1. Papadimitriou, C.H., Tsitsiklis, J.N.: The complexity of optimal queuing network control. Math. Oper. Res. **24** (1999) 293–305
2. Gittins, J.: Bandit processes and dynamic allocation indices. J. Roy. Statist. Soc. Ser. B **41** (1979) 148–177 With discussion.
3. Whittle, P.: Restless bandits: Activity allocation in a changing world. In Gani, J., ed.: A Celebration of Applied Probability. *J. Appl. Probab. Special Vol. 25A*. Applied Probability Trust (1988) 287–298
4. Niño-Mora, J.: Restless bandits, partial conservation laws and indexability. Adv. in Appl. Probab. **33** (2001) 76–98
5. Niño-Mora, J.: Dynamic allocation indices for restless projects and queueing admission control: a polyhedral approach. Math. Program. **93** (2002) 361–413
6. Niño-Mora, J.: Restless bandit marginal productivity indices, diminishing returns and optimal control of make-to-order/make-to-stock $M/G/1$ queues. Technical report, Department of Statistics, Universidad Carlos III de Madrid (2004) Conditionally accepted in *Math. Oper. Res.*
7. Niño-Mora, J.: Marginal productivity index policies for scheduling restless bandits with switching penalties. Technical report, Department of Statistics, Universidad Carlos III de Madrid (2005)