Adaptive Greedy Rules for Dynamic and Stochastic Resource Capacity Allocation Problems *

Peter Jacko
BCAM, Spain

February 3, 2010

Abstract

In this paper we briefly present a novel dynamic and stochastic model of resource allocation that generalizes a variety of problems addressed in the literature and we outline a unified methodology for designing adaptive greedy rules. Such rules are important in practice, since they may provide an easy-to-interpret and easy-to-implement solution to problems that are intractable for optimal solution due to the curse of dimensionality, or they embody an elegant optimal solution in some problems with simpler structure. We bridge the methodological gap between static/deterministic optimization and dynamic/stochastic optimization by stressing the connection between the classic knapsack problem and a group of related problems in management and stochastic scheduling unified by our model.

1 Introduction

Consider a collection of competitors with resource capacity demands that in aggregate are, at least at some time instants, beyond the available resource capacity. Suppose that to each competitor we can assign a price, independent of other competitors, that measures the efficiency of attaining a joint goal if a particular amount of resource capacity is allocated to her at a given moment. We are interested in designing and evaluating rules based on these prices and the resource capacity demands which accomplish the goal of good resource capacity allocation, whatever the "good" may mean.

In the special case when the competitors and the resource capacity are static (and therefore deterministic), this becomes the well-known *knapsack problem*. Indeed, this is the problem of determining the most valuable knapsack capacity allocation to a collection of competing items with given prices and knapsack capacity demands (weights). The knapsack problem is NP-hard to solve optimally, but a simple greedy rule was proposed by Dantzig (1957):

*This work has been supported by the Comunidad Autonoma de Madrid and the Universidad Carlos III de Madrid through the joint grant CCG08UC3M/ESP-4162. This is an author's preprint of the article with the following reference: Jacko, P. (2009). Adaptive Greedy Rules for Dynamic and Stochastic Resource Capacity Allocation Problems, Medium for Econometric Applications 17(4):10–16, Available online at http://www.met-online.nl .

Price/demand rule: *Allocate the capacity to the items with the highest price/demand ratios.*

In the special case when these competing items have equal capacity demands, the price/demand rule is optimal and reduces to:

Price rule: *Allocate the capacity to the items with the highest prices.*

In this work we allow for dynamically and stochastically evolving competitors. Therefore, we shift our focus to greedy rules that are *adaptive*, i.e., to rules which generalize the price/demand rule that is static. Several questions must be addressed in such a setting:

(i) [Economic question] For a given joint goal, is it possible to define dynamic quantities for each competitor that can be interpreted as prices? And if yes,

(ii) [Algorithmic question] How to calculate such prices quickly?

(iii) [Mathematical question] Under what conditions is there a greedy rule that achieves optimal resource capacity allocation?

(iv) [Experimental question] If greedy rules are not optimal, how close to optimality do they come? And how do they compare to alternative rules?

When the competitors are dynamic (even if they are non-stochastic), such a problem is PSPACE-hard (Papadimitriou and Tsitsiklis, 1999). This *curse of dimensionality* justifies the interest in greedy rules, since optimal solutions for high-dimensional problems appearing in the real world are unlikely to be obtained. In addition, this result implies that we can expect that optimality of greedy rules will occur only in problems with largely restricted dynamics. Such is, as the following example illustrates, the case with the $c\mu$-rule when several customers are competing for a single server.

Example 1 (Job Sequencing Problem: Statement)**.** Consider $K - 1 \geq 1$ customers (jobs) waiting for service at a server that can serve one customer at a time. Let $1/\mu_k > 0$ be the expected service time of job k and let $c_k > 0$ be the holding cost per period incurred for customer k waiting. The server can also be left idle, denoting this option by $k = K$, or allocated to a customer with a completed job. Thus, these K options are competitors and the task is to decide to which option the server should be allocated.

The joint goal is to minimize the aggregate expected holding cost over an infinite horizon. It turns out that, in several model variants, the following greedy rule applied while there are waiting customers attains such a goal:

$c\mu$-rule: *Allocate the server to the waiting customer with the highest value $c_k\mu_k$.*

Such a quantity measures the expected savings on holding costs per expected service time, or the efficiency of attaining the goal, if customer k is served. Thus, the $c\mu$-rule allocates the server to the customer who contributes most efficiently to minimization of the aggregate expected holding cost.

• • •

From this example we can learn several properties we can expect when considering more complex problems in the framework of Markov decision processes (MDPs). In general, the price, if exists, is state- and action-dependent and it may not be defined for some state/action combinations (here, the price is undefined for the jobs already completed). The concept of "state" must contain all the information relevant for the resource capacity allocation decision (for instance, whether the customer is still waiting). Further, whenever defined, the price measures the *opportunity cost* of two allowable levels of resource capacity allocation, which is a well-known result in economics. Indeed, a "correct" price must take into account both the value of the best action and the value of the second-best action. Finally, in order to be comparable across competitors, the prices must be in the same units; the price is calculated per unit of expected *resource capacity consumption*.

In the following section we present an MDP formulation of the *Dynamic and Stochastic Resource Capacity Allocation Problem* (DSRCAP). This novel general problem builds on several special cases of increasing complexity that have been addressed in the literature since 1950's, such as the job sequencing problem, the multi-armed bandit problem, the multi-armed restless bandit problem, and a large number of their special cases and extensions. For description of these problems and their applications see, for instance, Gittins (1989); Jun (2004); Sundaram (2005); Niño-Mora (2007); McCall and McCall (2007); Hero et al. (2008); Jacko (2009).

In this paper we strongly digress from the jargon existing in the literature, which seems to become abundant and confusing with the increasing number of extensions and variants of the problem. Instead, we use terminology which is expected to be found natural and intuitive by practitioners and researchers in the areas of business and economics, thus providing a way for an effective application of the methodology described in the following sections. This is an abridged version of a survey article in preparation.

2 DSRCAP: MDP Formulation

In this section we present a discrete-time MDP formulation of the dynamic and stochastic resource capacity allocation problem.[1] Consider the time slotted into time epochs $t \in \mathcal{T} := \{0, 1, 2, \dots\}$ at which decisions can be made. Time epoch t corresponds to the beginning of time period t. We consider the problem over an infinite horizon, as this may cover also a finite horizon problem if properly defined.

Suppose that there are $K \geq 1$ (integer) competitors, labeled by $k \in \mathcal{K}$, competing for a resource divisible into $W \geq 1$ (integer) units. We call W the resource *capacity*. We assume that the resource is *fully regenerative*, i.e., its full capacity is repetitively available at every time epoch t. The capacity not used at a given epoch is lost, i.e., the resource capacity is *nonmarketable*.

[1] A continuous time model is also possible. However, for the continuous-time MDP (i.e., when all the inter-decision times are exponentially distributed) and the semi-Markov decision processes formulations, the standard uniformization technique (see Puterman, 2005, Chapter 11) can be used to reformulate it as a discrete-time MDP model covered by our setting.

2.1 Competitors

Every competitor can be allocated any non-negative integer number of capacity units not exceeding the given resource capacity W. We denote by $\mathcal{A} := \{0, 1, \ldots, W\}$ the *action space*, i.e., the set of allowable levels of capacity allocation. This action space is the same for every competitor k.

Each competitor k is defined independently of other competitors as the tuple

$$\left(\mathcal{N}_k, (\boldsymbol{W}_k^a)_{a\in\mathcal{A}}, (\boldsymbol{R}_k^a)_{a\in\mathcal{A}}, (\boldsymbol{P}_k^a)_{a\in\mathcal{A}}\right),$$

where

- $\mathcal{N}_k$ is the *state space*, i.e., a finite set of possible states competitor k can occupy;
- $\boldsymbol{W}_k^a := \left(W_{k,n}^a\right)_{n\in\mathcal{N}_k}$, where $W_{k,n}^a$ is the expected one-period capacity consumption, or *work* required by competitor k at state n if action a is decided at the beginning of a period;
- $\boldsymbol{R}_k^a := \left(R_{k,n}^a\right)_{n\in\mathcal{N}_k}$, where $R_{k,n}^a$ is the expected one-period *reward* earned by competitor k at state n if action a is decided at the beginning of a period;
- $\boldsymbol{P}_k^a := \left(p_{k,n,m}^a\right)_{n,m\in\mathcal{N}_k}$ is the competitor-k stationary one-period *state-transition probability matrix* if action a is decided at the beginning of a period, i.e., $p_{k,n,m}^a$ is the probability of moving to state m from state n under action a.

The dynamics of competitor k is thus captured by the *state process* $X_k(\cdot)$ and the *action process* $a_k(\cdot)$, which correspond to state $X_k(t) \in \mathcal{N}_k$ and action $a_k(t) \in \mathcal{A}$, respectively, at all time epochs $t \in \mathcal{T}$. As a result of deciding action $a_k(t)$ in state $X_k(t)$ at time epoch t, competitor k consumes (possibly a part of) the allocated capacity, earns the reward, and evolves its state for the time epoch $t+1$. It is natural to require that the capacity consumed is nonnegative and not greater than the capacity allocated, therefore we assume $0 \le W_{k,n}^a \le a$. To avoid technical difficulties we will also assume that $R_{k,n}^a$ is bounded.

Note that we have the same action space $\mathcal{A}$ available at every state, which assures a technically useful property that $\boldsymbol{W}_k^a, \boldsymbol{R}_k^a, \boldsymbol{P}_k^a$ are defined in the same dimensions under any $a \in \mathcal{A}$. Though this may appear at first glance as a limitation on the applicability of the model, in fact the opposite is true. Notice that in order to effectively restrict the number of allowable actions at certain states we can define some of the actions as duplicates by having the same one-period consequences, i.e. actions a and b are *duplicates* at state n of competitor k if and only if $W_{k,n}^a = W_{k,n}^b, R_{k,n}^a = R_{k,n}^b$, and $p_{k,n,m}^a = p_{k,n,m}^b$ for all m. We will usually define non-useful actions as duplicates of the action $a = 0$.

Let $\mathcal{A}_{k,n} := \mathcal{A}\backslash\{a : W_{k,n}^a = W_{k,n}^0, R_{k,n}^a = R_{k,n}^0,$ and $p_{k,n,m}^a = p_{k,n,m}^0$ for all $m \in \mathcal{N}_k\}$ be the set of all the *positive* allowable capacity allocations for competitor k and its state $n \in \mathcal{N}_k$. That is, in $\mathcal{A}_{k,n}$ we have removed the zero capacity allocation (i.e., action $a = 0$) and all its duplicates from the action set of state

n. If $a \notin \mathcal{A}_{k,n}$, then we say that competitor k at its state $n \in \mathcal{N}_k$ is *uncontrollable* by action a. Further, we say that competitor k at its state $n \in \mathcal{N}_k$ is *totally uncontrollable*, if it is uncontrollable by all actions $a \in \mathcal{A}$ (i.e., if all the actions $1, 2, \ldots, W$ are defined as duplicates of the zero capacity allocation 0, and thus $\mathcal{A}_{k,n} = \emptyset$).

In some cases it is useful to assume that competitor $k = K$ is the *static κ-priced competitor*, with a single state (therefore static) and with price κ per allocated capacity unit. I.e., such a competitor k is defined by $\mathcal{N}_k := \{0\}, W^a_{k,0} := a, R^a_{k,0} := \kappa a, p^a_{k,0,0} := 1$ for all $a \in \mathcal{A}$. As we will see later, the role of this competitor is to cut-off capacity allocation to competitors whenever they are priced below κ.

Example 2 (Job Sequencing Problem: MDP Formulation)**.** Consider the discrete-time variant of the job sequencing problem described in Example 1, in which μ_k is interpreted as the probability that the service of job k is completed within one period. Recall that there are $K - 1$ customers with jobs waiting at the beginning (i.e., at time epoch $t = 0$), and the idling option $k = K$.

If the server is preemptive (i.e., the service of a customer can be interrupted at any time epoch even if not completed), then all its capacity $W = 1$ is available at every time epoch, and therefore it is fully regenerative. We have the action space $\mathcal{A} := \{0, 1\}$, where action 0 means allocating zero capacity (i.e., "not serving"), and action 1 means allocating full capacity (i.e., "serving").

Thus, we define job $k \leq K - 1$ with

- state space $\mathcal{N}_k := \{\text{'completed'}, \text{'waiting'}\}$;
- expected one-period works

$$W^1_{k,\text{'completed'}} := 1, \qquad W^1_{k,\text{'waiting'}} := 1,$$
$$W^0_{k,\text{'completed'}} := 0, \qquad W^0_{k,\text{'waiting'}} := 0;$$

- expected one-period rewards, i.e., the negative of the holding cost expected to be paid at the next time epoch,

$$R^1_{k,\text{'completed'}} := 0, \qquad R^1_{k,\text{'waiting'}} := -c_k \cdot (1 - \mu_k) - 0 \cdot \mu_k,$$
$$R^0_{k,\text{'completed'}} := 0, \qquad R^0_{k,\text{'waiting'}} := -c_k;$$

- one-period state-transition probability matrices

$$\boldsymbol{P}^1_k := \begin{array}{r} \\ \text{'completed'} \\ \text{'waiting'} \end{array} \begin{array}{c} \begin{array}{cc} \text{'completed'} & \text{'waiting'} \end{array} \\ \begin{pmatrix} 1 & 0 \\ \mu_k & 1 - \mu_k \end{pmatrix} \end{array},$$

$$\boldsymbol{P}^0_k := \begin{array}{r} \\ \text{'completed'} \\ \text{'waiting'} \end{array} \begin{array}{c} \begin{array}{cc} \text{'completed'} & \text{'waiting'} \end{array} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{array}.$$

Notice that when the server is allocated to customer k, its whole capacity is allocated ($a = 1 = W$). Further, all the states are controllable, since action 1 has different one-period consequences than action 0. The idling option $k = K$ is defined as the static 0-priced competitor.

• • •

2.2 A Unified Optimization Criterion

Before describing the problem we first define an averaging operator that will allow us to discuss the infinite-horizon problem under the traditional β-discounted criterion and the time-average criterion in parallel. Let $\Pi_{X,a}$ be the set of all the policies that for each time epoch t decide (possibly *randomized*) action $a(t)$ based only on the state-process history $X(0), X(1), \ldots, X(t)$ and on the action-process history $a(0), a(1), \ldots, a(t-1)$ (i.e., *non-anticipative*). Let $\mathbb{E}_\tau^\pi$ denote the expectation over the state process $X(\cdot)$ and over the action process $a(\cdot)$, conditioned on the state-process history $X(0), X(1), \ldots, X(\tau)$ and on policy π.

Consider any expected one-period quantity $Q_{X(t)}^{a(t)}$ that depends on state $X(t)$ and on action $a(t)$ at any time epoch t. For any policy $\pi \in \Pi_{X,a}$, any initial time epoch $\tau \in \mathcal{T}$, and any *discount factor* $0 \leq \beta \leq 1$ we define the infinite-horizon *β-average quantity* as[2]

$$\mathbb{B}_\tau^\pi \left[Q_{X(\cdot)}^{a(\cdot)}, \beta, \infty \right] := \lim_{T \to \infty} \frac{\sum_{t=\tau}^{T-1} \beta^{t-\tau} \, \mathbb{E}_\tau^\pi \left[Q_{X(t)}^{a(t)} \right]}{\sum_{t=\tau}^{T-1} \beta^{t-\tau}}. \tag{1}$$

Thus, when $\beta = 1$, the problem is formulated under the *time-average criterion*, whereas when $0 < \beta < 1$, the problem is considered under the *β-discounted criterion* (scaled by $1-\beta$). The remaining case when $\beta = 0$ reduces to a static problem and hence is considered in order to define a *myopic policy*. In the following we consider the discount factor β to be fixed and the horizon to be infinite, therefore we omit them in the notation and write briefly $\mathbb{B}_\tau^\pi \left[Q_{X(\cdot)}^{a(\cdot)} \right]$.

2.3 Optimization Problem

Now we describe in more detail the problem we consider and formulate it below. Let $\Pi_{\boldsymbol{X},\boldsymbol{a}}$ be the space of randomized and non-anticipative policies depending on the joint state-process $\boldsymbol{X}(\cdot) := (X_k(\cdot))_{k \in \mathcal{K}}$ and deciding the joint action-process $\boldsymbol{a}(\cdot) := (a_k(\cdot))_{k \in \mathcal{K}}$, i.e., $\Pi_{\boldsymbol{X},\boldsymbol{a}}$ is the *joint policy space*.

Suppose that the capacity W must be *exhaustively* allocated to competitors in $\mathcal{K}$ at every time epoch.[3] Suppose further that at every time epoch, apart from earning the expected one-period rewards $R_{k,n}^a$ for allocated capacity, we also earn a reward ε per unit of the *latent capacity*, i.e., the capacity that is allocated but not consumed by competitors in $\mathcal{K}$. Note that this happens only when the expected one-period capacity consumption is lower than the allocated capacity, i.e., $W_{k,n}^a < a$, since this implies that the (actual) one-period capacity consumption is strictly lower than the allocated capacity with a probability greater than zero.

[2] For definiteness, we consider $\beta^0 = 1$ for $\beta = 0$.

[3] One could consider the problem variant in which the capacity is permitted to be allocated partially, while we earn a reward κ per unit of the capacity that is *not* allocated to competitors in $\mathcal{K}$. However, we can easily transform this problem variant into the problem with exhaustive capacity allocation, if we assume that competitor $k = K$ is an artificially introduced static κ-priced competitor, which represents idle capacity.

For any discount factor β, the problem is to find a joint policy $\boldsymbol{\pi}$ maximizing the *objective* given by the β-average of the sum of aggregate reward and latent-capacity reward starting from the initial time epoch 0, subject to the family of *sample path* allocation constraints, i.e.,

$$\max_{\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}} \mathbb{B}_0^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}}\left(R^{a_k(\cdot)}_{k,X_k(\cdot)}+\varepsilon\left(a_k(\cdot)-W^{a_k(\cdot)}_{k,X_k(\cdot)}\right)\right)\right], \tag{2}$$

$$\text{subject to } \mathbb{E}_t^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}} a_k(t)\right] = W, \text{ for all } t\in\mathcal{T}. \tag{3}$$

Note that the sample-path constraint could equivalently be expressed as $\sum_{k\in\mathcal{K}} a_k(t) = W$ for all $t\in\mathcal{T}$ under policy $\boldsymbol{\pi}$ and for any possible joint state-process history $\boldsymbol{X}(0),\boldsymbol{X}(1),\ldots,\boldsymbol{X}(t)$.

3 DSRCAP: Relaxations and Decomposition

We first reformulate problem (2)-(3) in a more comfortable way, defining

$$\widetilde{R}^a_{k,n} := R^a_{k,n}+\varepsilon\left(a-W^a_{k,n}\right), \tag{4}$$

$$\widetilde{W}^a_{k,n} := W^a_{k,n}+\left(a-W^a_{k,n}\right)=a, \tag{5}$$

which allows to rewrite the problem as

$$\max_{\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}} \mathbb{B}_0^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}}\widetilde{R}^{a_k(\cdot)}_{k,X_k(\cdot)}\right], \tag{P}$$

$$\text{subject to } \mathbb{E}_t^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}}\widetilde{W}^{a_k(t)}_{k,X_k(t)}\right] = W, \text{ for all } t\in\mathcal{T}.$$

3.1 Relaxations

The epoch-t constraint in (P) is conditioned on information available at time t, therefore it implies the *epoch-t expected* capacity consumption constraint,

$$\mathbb{E}_0^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}}\widetilde{W}^{a_k(t)}_{k,X_k(t)}\right] = W, \text{ for all } t\in\mathcal{T} \tag{6}$$

requiring that the capacity be consumed at every time epoch if conditioned on information available at time 0.

Finally, as proposed in Whittle (1988), we may require this constraint to hold only on β-average, as the *β-average capacity consumption constraint*

$$\mathbb{B}_0^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}}\widetilde{W}^{a_k(\cdot)}_{k,X_k(\cdot)}\right] = \mathbb{B}_0^{\boldsymbol{\pi}}\left[W\right]. \tag{7}$$

Noticing that $\mathbb{B}_0^{\boldsymbol{\pi}}[W] = W$, we obtain the following *Whittle relaxation* of problem (P),

$$\max_{\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}} \mathbb{B}_0^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}} \widetilde{R}_{k,X_k(\cdot)}^{a_k(\cdot)}\right] \tag{P$^{\mathrm{W}}$}$$
$$\text{subject to } \mathbb{B}_0^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}} \widetilde{W}_{k,X_k(\cdot)}^{a_k(\cdot)}\right] = W.$$

The Whittle relaxation (P$^{\mathrm{W}}$) can be approached by traditional Lagrangian methods, introducing a Lagrangian parameter, say ν, to dualize the constraint, obtaining thus the following Lagrangian relaxation,

$$\max_{\boldsymbol{\pi}\in\Pi_{\boldsymbol{X},\boldsymbol{a}}} \mathbb{B}_0^{\boldsymbol{\pi}}\left[\sum_{k\in\mathcal{K}} \left(\widetilde{R}_{k,X_k(\cdot)}^{a_k(\cdot)} - \nu\widetilde{W}_{k,X_k(\cdot)}^{a_k(\cdot)}\right)\right] + \nu W. \tag{P$^{\mathrm{L}}_{\nu}$}$$

Note finally that by the definition of relaxation, (P$^{\mathrm{L}}_{\nu}$) for every ν provides an upper bound for the optimal value of both problem (P$^{\mathrm{W}}$) and problem (P).

3.2 Decomposition into Single-Competitor Subproblems

We now set out to decompose the optimization problem (P$^{\mathrm{L}}_{\nu}$) as it is standard for Lagrangian relaxations, considering ν as a parameter. Notice that any joint policy $\boldsymbol{\pi} \in \Pi_{\boldsymbol{X},\boldsymbol{a}}$ defines a set of single-competitor policies $\widetilde{\pi}_k$ for all $k \in \mathcal{K}$, where $\widetilde{\pi}_k$ is a randomized and non-anticipative policy depending on the *joint* state-process $\boldsymbol{X}(\cdot)$ and deciding the *competitor-k* action-process $a_k(\cdot)$. We will write $\widetilde{\pi}_k \in \Pi_{\boldsymbol{X},a_k}$. We will therefore study the competitor-k subproblem

$$\max_{\widetilde{\pi}_k\in\Pi_{\boldsymbol{X},a_k}} \mathbb{B}_0^{\widetilde{\pi}_k}\left[\widetilde{R}_{k,X_k(\cdot)}^{a_k(\cdot)} - \nu\widetilde{W}_{k,X_k(\cdot)}^{a_k(\cdot)}\right]. \tag{8}$$

4 DSRCAP: Solution via Prices and Greedy Rules

4.1 Prices and Solution to the Single-Competitor Subproblem

We now want to attach to each competitor $k \in \mathcal{K}$ a set of prices $v_{k,n}^a$, independent of other competitors, and defined for each state $n \in \mathcal{N}_k$ and each positive allowable capacity allocation $a \in \mathcal{A}_{k,n}$. We want price $v_{k,n}^a$ to measure the efficiency rate of attaining the joint goal of maximizing the aggregate β-average reward in (P) if competitor k at its state n is allocated a capacity units at the current time epoch.

The existence of prices and their efficient computation is a complex issue and is left out of this paper due to the space restrictions. For problems in which the only positive allowable capacity allocation is that of one capacity unit, i.e., $\mathcal{A}_{k,n} \subseteq \{1\}$ for all competitors k and all states n, the concept of price reduces to that of the *marginal productivity index*, well developed in the framework of restless bandits and surveyed in Niño-Mora (2007). The case when there is a single positive allowable capacity allocation (but not necessarily that of one capacity unit), i.e., $|\mathcal{A}_{k,n}| \leq 1$ for all competitors k and all states n, was treated in Jacko (2009, Section 5), and is closely related to the classic knapsack problem. For more general settings see Weber (2007); Niño-Mora (2008).

4.2 Greedy Rules for DSRCAP

If all the prices exist, then at every time epoch t we need to solve the following generalized knapsack problem, denoting the actual state of every competitor k by $n_k := X_k(t)$:

$$\begin{aligned} \max_{\boldsymbol{z}} \quad & \sum_{k\in\mathcal{K}, a\in\mathcal{A}_{k,n_k}} v^a_{k,n_k} z_{k,a} \\ \text{subject to} \quad & \sum_{k\in\mathcal{K}, a\in\mathcal{A}_{k,n_k}} a z_{k,a} = W \qquad\qquad\qquad\qquad (\text{GKP}) \\ & \sum_{a\in\mathcal{A}_{k,n_k}} z_{k,a} \le 1 \qquad\qquad \text{for all } k\in\mathcal{K}, \\ & z_{k,a} \in \{0,1\} \qquad\qquad \text{for all } k\in\mathcal{K}, a\in\mathcal{A}_{k,n_k} \end{aligned}$$

where $\boldsymbol{z} := (z_{k,a})_{k\in\mathcal{K}, a\in\mathcal{A}_{k,n_k}}$ is the (t-dependent) vector of binary decision variables denoting whether competitor k is allocated amount a of capacity W. The first constraint represents the exhaustive capacity allocation, and the second one means that at most a unique level of positive allowable capacity allocation must be decided for each competitor. Then, the resource capacity allocation to competitor k at time epoch t is given by

$$a_k(t) := \sum_{a\in\mathcal{A}_{k,n_k}} a z_{k,a} \qquad\qquad \text{for all } k\in\mathcal{K}. \qquad (9)$$

Note that (9) together with the first constraint in (GKP) assure the sample path capacity allocation constraint (3).

Here is where the greedy rules arise, since the price/demand rule for the knapsack problem applies to the generalized knapsack problem as well; one only needs to assure that a unique level of positive allowable capacity allocation is chosen for every competitor. In our case, the capacity demands are given by positive allowable actions $a \in \mathcal{A}_{k,X_k(t)}$, so that the following adaptive greedy rule can be considered:

General rule: *Allocate the capacity at time t to competitors with the highest value* $\max\{v^a_{k,X_k(t)}/a : a \in \mathcal{A}_{k,X_k(t)}\}$.

This value represents the highest capacity allocation efficiency for a given competitor. The above rule reduces to the index rule, well known in the restless bandit literature, as described in the example below.

Index rule: *Allocate the resource at time t to a competitor with the highest value* $v^1_{k,X_k(t)}$.

Example 3. In the special case with capacity $W = 1$, the set $\mathcal{A}_{k,n}$ has at most one element ($a = 1$), so problem (GKP) reduces to the following capacity-1 knapsack problem to be solved at every time epoch t:

$$\begin{aligned} \max_{\boldsymbol{z}} \quad & \sum_{k\in\mathcal{K}_t} v^1_{k,n_k} z_{k,1} \\ \text{subject to} \quad & \sum_{k\in\mathcal{K}_t} z_{k,1} = 1 \qquad\qquad\qquad\qquad (\text{1KP}) \\ & z_{k,1} \in \{0,1\} \qquad\qquad \text{for all } k\in\mathcal{K}_t \end{aligned}$$

where $\mathcal{K}_t := \{k \in \mathcal{K} : 1 \in \mathcal{A}_{k,n_k}\}$ is the (t-dependent) set of all competitors whose actual state is not uncontrollable, and where $\boldsymbol{z} := (z_{k,1})_{k\in\mathcal{K}_t}$. Obviously, all the competitors $k \in \mathcal{K}_t$ are of equal demands, and therefore problem (1KP) is optimally solved by the price rule, known as the index rule in the restless bandit literature.

• • •

We remark that solution (9), in general, provides a heuristic to the intractable problem (P). Solution (9) was proven in the celebrated work of Gittins and Jones (1974) optimal for the multi-armed bandit problem, in which $W = 1$, $W^a_{k,n} = a$ for all k, n, a, $\boldsymbol{P}^0_k$ is an identity matrix, and the prices are defined as the *Gittins index* values. Such an optimality result also holds if (symmetric) competitors are allowed to appear randomly over time (Whittle, 1981). Finally, notice that the job sequencing problem stated in Example 1 is a special case of the multi-armed bandit problem, and therefore the $c\mu$-rule inherits optimality from the Gittins index rule.

About the Author

Peter Jacko is currently with the BCAM – Basque Center for Applied Mathematics, Spain. He obtained his Ph.D. in Business Administration and Quantitative Methods (2009) and D.E.A. in Statistics and Operations Research (2006) from the Universidad Carlos III de Madrid, Spain. He obtained his Mgr. (2003) and Bc. (2002) degrees in Mathematics from the Univerzita P. J. Šafárika v Košiciach, Slovakia. E-mail: `jacko@bcamath.org`.

References

Dantzig, G. B. (1957). Discrete-variable extremum problems. *Operations Research*, 5(2):266–277.

Gittins, J. C. (1989). *Multi-Armed Bandit Allocation Indices*. J. Wiley & Sons, New York.

Gittins, J. C. and Jones, D. M. (1974). A dynamic allocation index for the sequential design of experiments. In Gani, J., editor, *Progress in Statistics*, pages 241–266. North-Holland, Amsterdam.

Hero, A. O., Castañón, D. A., Cochran, D., and Kastella, K. (2008). *Foundations and Applications of Sensor Management*. Springer.

Jacko, P. (2009). *Marginal Productivity Index Policies for Dynamic Priority Allocation in Restless Bandit Models*. PhD thesis, Universidad Carlos III de Madrid. http://e-archivo.uc3m.es/bitstream/10016/5357/1/tesis_jacko_peter.pdf.

Jun, T. (2004). A survey on the bandit problem with switching costs. *De Economist*, 152(4):1–29.

McCall, B. P. and McCall, J. J. (2007). *The Economics of Search*. Routledge.

Niño-Mora, J. (2007). Dynamic priority allocation via restless bandit marginal productivity indices. *TOP*, 15(2):161–198.

Niño-Mora, J. (2008). An index policy for multiarmed multimode restless bandits. In *Proceedings of the 3rd International Conference on Performance Evaluation Methodologies and Tools*.

Papadimitriou, C. H. and Tsitsiklis, J. N. (1999). The complexity of optimal queueing network. *Mathematics of Operations Research*, 24(2):293–305.

Puterman, M. L. (2005). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Sundaram, R. K. (2005). Generalized bandit problems. In *Social Choice and Strategic Decisions*. Springer Berlin Heidelberg.

Weber, R. (2007). Comments on: Dynamic priority allocation via restless bandit marginal productivity indices. *TOP*, 15(2):211–216.

Whittle, P. (1981). Arm-acquiring bandits. *Annals of Probability*, 9(2):284–292.

Whittle, P. (1988). Restless bandits: Activity allocation in a changing world. *A Celebration of Applied Probability, J. Gani (Ed.), Journal of Applied Probability*, 25A:287–298.