Optimal Stopping of Markov Chain and Three Abstract Optimization Problems

Isaac M. Sonin

Dept. of Mathematics and Statistics, Univ. of North Carolina at Charlotte,

Charlotte, NC 28223, USA

May 24, 2010

Abstract

There is a well known connection between three problems related to Optimal Stopping of Markov Chain and the equality of three corresponding indices: the classical Gittins index in the Ratio Maximization Problem, the Kathehakis-Veinot index in a Restart Problem, and Whittle index in a family of Retirement Problems.

In [13] these three problems and these three indices were generalized in such a way that it become possible to use the State Elimination algorithm [11] to calculate this common Generalized Gittins index α.

The main goal of this note is to demonstrate that the equality of these (generalized) indices is a special case of a more general relation between three simple abstract optimization problems.

Key words: Gittins index, Markov chain, Optimal stopping, The State Elimination Algorithm.

1 Introduction.

There is a well known connection between three problems related to Optimal Stopping (OS) of Markov Chain (MC), the ratio (cycle) maximization, the Kathehakis-Veinot Restart Problem, Whittle family of Retirement Problems, and the equality of three corresponding indices: the classical Gittins index, the Kathehakis-Veinot (KV) index, and the Whittle index.

In [13] these three problems and corresponding indices were generalized in such a way that it become possible to use the so called State Elimination (SE) algorithm developed earlier by author to solve OS of MC to calculate this common index α. This generalization also sheds a new light on a meaning of this index, which we call the Generalized Gittins index (GGI), and relates this index to an index introduced earlier by Mitten in [8] and thus to such papers as [5], [3] and [9], for which Mitten's paper was a starting point. In these papers GGI plays an important in the description of optimal strategy.

The main goal of this note is to demonstrate that the equality of these indices is a special case of a similar equality for three simple *abstract optimization* problems. By an abstract optimization problem we mean a problem with maximization over an abstract set of indices U without any specifics about this set. There is no doubt that the relationship between these problems was used in optimization theory before, on different occasions in specific problems, but we fail to find a general statement of this kind in the vast literature on optimization.

In the next section 2 we briefly repeat the main statements of [13], in section 3 we describe three abstract optimization problems and prove our main result - Theorem 3. At the end of this section we discuss the reduction of OS problems to abstract optimization problems. In section 4 we discuss the possible generalization of OS problem to more general - continue, quit, restart (CQR) probability model, which will be treated in forthcoming paper [14], and present some relevant open problems.

The author would like to thank Robert Andersen, Joseph Quinn and Ernst Presman who read the first version of this paper and made valuable comments, and two anonymous referees for helpful suggestions.

2 Three Classical Indices and their Generalizations

Let us recall some useful facts related to classical Gittins index (GI) $\gamma(x)$. A special case of a general Markov Decision model (see e.g. [4] is a model of OS of MC specified by a tuple $M = (X, P, c(x), g(x), \beta)$, where X is a countable state space, $P = \{p(x, y)\}$ is stochastic (transition) matrix, $c(x)$ is a *one-step reward* function, $g(x)$ is a *terminal reward* function, both can be positive or negative and β is a discount factor, $\beta = const, 0 < \beta \leq 1$. If the reward function g is absent, or equivalently if $g = -\infty$, we call such model a (*Markov) Reward Model.*

Given a reward model M and point $x \in X$, the classical *Gittins index,* $\gamma(x)$, see [6], as well as e.g. [15] and [1], is defined as the *maximum of the expected discounted total reward* during the interval $[0, \tau)$ *per unit of expected discounted time* for the Markov chain starting from x, i.e.

$$\gamma(x) = \sup_{\tau>0} \frac{E_x \sum_{n=0}^{\tau-1} \beta^n c(Z_n)}{E_x \sum_{n=0}^{\tau-1} \beta^n} = (1-\beta) \sup_{\tau>0} \frac{E_x \sum_{n=0}^{\tau-1} \beta^n c(Z_n)}{1 - E_x \beta^\tau}, \tag{1}$$

where $0 < \beta < 1$, τ is a stopping time, $\tau > 0$, and a trivial equality $(1-\beta)\sum_{n=0}^{k-1} \beta^n = 1-\beta^k$ is used to obtain the second equality in (1). In other words, a decision maker (DM) has two actions available at each state - to *continue* or to *stop*, and her goal is to maximize the ratio in (1). Without loss of generality we can consider as stopping times only the moments of a first visit to sets $G \subset X, x \notin G$. The GI index plays an important role in the theory of Multi-armed bandit problems with *independent* arms but it also appears naturally in many other problems of stochastic optimization.

An interesting interpretation of the GI, the so called *Restart in State* interpretation, was given by Kathehakis and Veinot in [7]. Given a reward model M, let us consider a *family* of Markov decision models indexed by a *fixed* initial point $s \in X$, where a DM has two following actions available at each state x, - to *continue* or to *return* (restart) to state s and continue from there. In other words, MC (Z_n) starting from a point s after a positive stopping time $\tau > 0$ can be restarted at the same point s, and so on. It is convenient to assume that a new "cycle" starts *instantly* at the moment of restart. It means that though a decision to restart is made at stopping time τ at some point

x, different from point s, but a position y at moment $\tau + 1$ is defined by transition probabilities $p(s, y)$ as if at moment τ the decision to continue from s also has been made.

Let $h(x|s)$ denote the supremum over all strategies of the expected total discounted reward on the infinite time interval in this model with an initial point x, and restart point s. Using the standard results of Markov Decision Processes theory, Kathehakis and Veinot proved that function $h(x|s)$ satisfies the equality

$$h(x|s) = \sup_{\tau>0} E_x[\sum_{n=0}^{\tau-1} \beta^n c(Z_n) + \beta^\tau h(s)], \tag{2}$$

and $\gamma(s) = (1 - \beta)h(s)$, where by definition $h(s) = h(s|s)$. We call index $h(s)$ a KV index. This index can be defined for any point $x \in X$, so we use also notation $h(x)$.

Another important interpretation of the GI, the so called *Retirement Process* formulation was provided by Whittle in [16]. Given a reward model M, he introduced the parametric family of OS models $M(k) = (X, P, c(x), k, \beta)$, where parameter k is a real number and the terminal reward function $g(x) = k$ for all $x \in X$. Denote $v(x, k)$ the value function for such a model, i.e. $v(x, k) = \sup_{\tau \geq 0} E_x[\sum_{n=0}^{\tau-1} \beta^n c(Z_n) + \beta^\tau k]$, and denote Whittle index $w(x) = \inf\{k : v(x, k) = k\}$. Since $\beta < 1$, for sufficiently large k it is optimal to stop immediately and $v(x, k) = k$. Thus $w(x) < \infty$. The results of Whittle imply that $v(x, k) = k$ for $k \geq w(x)$, $v(x, k) > k$ for $k < w(x)$, and $w(x) = h(x)$. Thus the following theorem holds

Theorem 1. *The three indices defined for a reward model* $M = (X, P, c(x), \beta), 0 < \beta < 1$, *coincide, i.e.* $h(x) = w(x) = \gamma(x)/(1 - \beta), x \in X$.

The main goal of [13] was to present a simple and transparent algorithm to calculate this common index. This algorithm is based on a general so called State Elimination (SE) algorithm developed by author for the problem of OS of MC and described in [10] and [11], see also [12]. To apply this algorithm it is necessary to replace a constant discount factor β by a variable "survival" probability $\beta(x)$ because after the first recursive step a discount factor is not a constant anymore. So by necessity a more general model was considered and the classical GI $\gamma(x)$ was replaced by a *generalized* Gittins Index (GGI) $\alpha(x)$ as follows.

It is well-known that the optimizations problems, such as described above, with an explicit discount factor β, are equivalent to problems where a state space is complemented by an absorbing point e and the initial transition probabilities are modified. The probability of entering an absorbing point e in one step for any state $x \neq e$ (probability of termination) is equal to $1-\beta$ and all other initial transition probabilities are multiplied by β. In other words, β is the probability of "survival", i.e. nontermination. In the sequel we consider a *reward model with termination* $M = (X, P, c(x), \beta(x))$, where we assume from the beginning that the state space X contains an absorbing point e, $p(e, e) = 1$, the function $\beta(x)$ is the probability of "survival" at point x, so $1 - \beta(x) = p(x, e)$ is the probability of termination. Function $\beta(x)$ can be a constant or variable. To simplify presentation we will assume that $\beta(x) < 1$ though this assumption can be weakened. Strictly speaking the function $\beta(x)$ is completely specified by a new transition matrix P but we include $\beta(x)$ in the tuple M to stress the presence of e and $\beta(x)$. From now on notation E_x, P_x and (Z_n) are referred to such model and survival probabilities $\beta(\cdot)$ now are automatically included under the signs P_x and E_x. We also assume that at an absorbing state $c(e) = 0$.

The numerator in (1), in the presence of an absorbing state e, equals to $E_x \sum_{n=0}^{\tau-1} c(Z_n)$, where now, given a subset $G \subset X, x \notin G$, $\tau = \min(n : Z_n \in G \cup e)$. Such equality holds independently of whether $\beta(x)$ is a constant or variable. Let us denote this numerator by $R^\tau(x)$. The denominator in the last expression in (1), in the presence of an absorbing state e, when $\beta = const$, equals to $P_x(Z_\tau = e)$. In general case, when $\beta(x)$ can be variable, we denote $P_x(Z_\tau = e)$ by $Q^\tau(x)$, the *probability of termination* on $[0, \tau)$, and we *define* the *Generalized GI* (GGI), $\alpha(x)$ for a model with termination as

$$\alpha(x) = \sup_{\tau>0} \frac{R^\tau(x)}{Q^\tau(x)}, \tag{3}$$

i.e. $\alpha(x)$ is the *maximum discounted total reward per chance of termination.*

If $\beta(x) = const = \beta$ then the second equality in (1) obviously implies that $\gamma(x) = (1-\beta)\alpha(x)$. The crucial point however is that, if $\beta(x)$ is not a constant, then the latter equality, or some kind of proportionality, can not be preserved anymore, even if the definition of $\gamma(x)$ is correspondingly modified, i.e. $E_x \sum_{n=0}^{\tau-1} \beta^n$ is replaced by $E_x \sum_{n=0}^{\tau-1} I_{\neq e}(Z_n)$.

In other words, the expected time of survival till termination is proportional to the probability of termination only if $\beta(x) = const$. Thus, in the general case, the *proportionality of the two indices $\gamma(x)$ and $\alpha(x)$ as functions of x completely disappears.* At the same time, for a reward model with termination a (generalized) KV index $h(x)$, and a (generalized) Whitlle index $w(x)$ can be defined in an absolutely similar way as above. This means that the sum in (2) with new P and E and modified τ has a form $E_x[\sum_{n=0}^{\tau-1} c(Z_n) + h(s)]$ and value function in Whittle model $v(x,k) = \sup_{\tau \geq 0} E_x[\sum_{n=0}^{\tau-1} c(Z_n) + k]$. As Theorem 2 shows below, *the equality $\alpha(x) = w(x) = h(x)$ is preserved* ! This means that the "true meaning" of the Gittins index is given by the expression in (3) and not in (1) !

Theorem 2 ([13]). *The three indices defined for a reward model with termination $M = (X, P, c(x), \beta(x))$ coincide, i.e. $\alpha(x) = h(x) = w(x)$.*

As a result of this theorem any of three problems can be used as a basis to calculate $\alpha(x)$ but the most convenient is the Whittle family of OS models $M(k)$, because the problem of calculation $v(x,k)$ for a particular k can be reduced to solving stopping problems using the State Elimination algorithm. The corresponding algorithm described in [13], calculates sequentially the index $\alpha(x)$ for all points $x \in X$ in an order which we do not know in advance. If the goal is to find $\alpha(s)$ for a particular s, and X is a finite set then we know only that $\alpha(s)$ will be obtained at some stage. We also can apply this algorithm to some cases of countable X.

Without describing this algorithm in detail, we give two simple examples with calculations to illustrate the difference between constant and variable β and we use the latter example later to illustrate our main result about the equivalence of the three optimization problems.

Example 1. State set $X = \{1,2\}, c(x) = x, x = 1,2, p(1,1) = \frac{2}{3}, p(1,2) = \frac{1}{3}, p(2,1) = p(2,2) = \frac{1}{2}$. Let $\beta = const, 0 < \beta < 1$. First we introduce an absorbing state e, so $X_1 = \{1,2,e\}, c_1(x) = x, x = 1,2, c_1(e) = 0$. Then the transition probabilities are modified as follows: $p_1(1,1) = \beta\frac{2}{3}$, $p_1(1,2) = \beta\frac{1}{3}, p_1(2,1) = p(2,2) = \beta\frac{1}{2}, p_1(1,e) = p_1(2,e) = 1 - \beta(1) = 1 - \beta(2) = 1 - \beta$. According to [13] we need to calculate the function $d_1(x) = c_1(x)/(1 - \beta(x))$ and $\alpha(z) = d_1(z)$ if the maximum of the function $d_1(x)$ is obtained at z. We have $d_1(1) = 1/(1-\beta)$ and $d_1(2) = 2/(1-\beta)$. Thus $\alpha(2) = 2/(1-\beta)$ and the point $z = 2$ must be "eliminated". It means that the state space X_1 is reduced to $X_2 = X_1 \setminus z = \{1, e\}$ and new transition probabilities $p_2(x, y)$ and new cost function $c_2(x)$ must be recalculated by the formulas

$$p_2(x, y) = p_1(x, y) + p_1(x, z)n_1(z)p_1(z, y), \quad c_2(x) = c_1(x) + p_1(x, z)n_1(z)c_1(z), \tag{4}$$

where $n_1(z) = 1/(1 - p_1(z, z))$. Such transformation describes MC (Z_n) and related costs during its visits to the set X_2. Using these formulas, we obtain $p_2(1, 1) = \beta_2(1) = \beta(4-\beta)/3(2-\beta)$, $p_2(1, e) = 1-\beta_2(1) = (6-\beta)(1-\beta)/3(2-\beta)$, $c_2(1) = (6+\beta)/3(2-\beta)$. Now the maximum of the function $d_2(x) = c_2(x)/(1-\beta_2(x))$ gives the next value of α. In our case $\alpha(1) = (6+\beta)/(6-\beta)(1-\beta)$. Correspondingly the classical GI $\gamma(x) = (1-\beta)\alpha(x)$ takes values: $\alpha(1) = (6 + \beta)/(6 - \beta)$ and $\alpha(2) = 2$.

Now let us consider the case of variable $\beta(x)$.

Example 2. We keep all the parameters as in Example 1 execept that the constant value β is replaced by $\beta(1) = \frac{2}{3}, \beta(2) = \frac{1}{2}$. Then $p_1(1, 1) = \frac{4}{9}, p_1(1, 2) = \frac{2}{9}, p_1(1, e) = \frac{3}{9}$, $p_1(2, 1) = p_1(2, 2) = \frac{1}{4}, p_1(2, e) = \frac{1}{2}$. Then $d_1(1) = 1/(1-\frac{2}{3}) = 3$ and $d_1(2) = 2/(1-\frac{1}{2}) = 4$. Therefore $\alpha(2) = 4$ and the point $z = 2$ must be "eliminated". Using formulas (4), we obtain $\beta_2(1) = p_2(1, 1) = \frac{14}{27}$ and $c_2(1) = \frac{43}{27}$. Then $\alpha(2) = \frac{43}{27}/(1 - \frac{14}{27}) = \frac{43}{13} \approx 3.3077$. Note that $\alpha(2)/\alpha(1) = \frac{52}{43} \approx 1.2093$, whereas in example 1, the ratio $\alpha(2)/\alpha(1) = 2(6 - \beta)/(6 + \beta) > \frac{52}{43}$ for all $\beta, 0 < \beta < 1$.

3 Three abstract optimization problems

The common part of all three problems described above is a maximization over the set of all positive stopping times τ, or equivalently over all partitions of the state set X into two sets, continuation and stopping (restart) regions. This is a special case of a very general situation.

First we present three abstract optimization problems 1, 2 and 3.

Suppose there is an abstract index set U, and $A = \{a_u\}$ and $B = \{b_u\}$ be two sets of real numbers indexed by the elements of U. Suppose that an assumption **U** holds,

$$-\infty < a_u \leq a < \infty, \; 0 < b \leq b_u \leq 1. \;\; (\mathbf{U})$$

We assume also that $b_u < 1$ for at least one u. In all three problems a DM knows sets U, A and B.

Problem 1. Restart Problem. Find solution(s) of the equation

$$h = \sup_{u \in U} [a_u + (1 - b_u)h] \equiv H(h). \tag{5}$$

It is easy to see that equation (5) is a Bellman (optimality) equation for the "value of the game", i.e. the supremum over all possible strategies, in the following optimization problem. There are two equivalent interpretations of this problem. In both cases set U represents a set of available actions, which we call "buttons" (arms). A DM can select one of them and push (test). She obtains a *reward* a_u and according to the first interpretation with *probability* b_u the game is *terminated*, and with complimentary probability $1 - b_u$ she is again in an initial situation, i.e. she can select any button and push. Her goal is to maximize the total (undiscounted) reward.

According to the second interpretation the game is continued sequentially without possibility of random termination, but the value $1 - b_u$ is now not a probability but a discount factor applied to the future rewards after a button u was used at the first step.

Our second optimization problem is

Problem 2. Ratio (cycle) Problem. Find

$$\alpha = \sup_{u \in U} \frac{a_u}{b_u}. \tag{6}$$

The interpretation of this problem is straightforward: a DM can push some button u only once and her goal is to maximize the ratio in (6), the one step reward per "chance of termination". Since the game is terminated after the first push anyway, $1/b_u$ is a "multiplicator" applied to a "direct" reward a_u.

In the sequel we shall use shorthand notation $a \vee b$ for $\max(a, b)$. Let $H(k)$ be a function defined in the right side of (5).

Problem 3. A Parametric Family of Retirement Problems. Find w, defined as follows: given parameter $k, -\infty < k < \infty$, let

$$v(k) = k \vee H(k), \quad w = \inf\{k : v(k) = k\}. \tag{7}$$

In this problem, given number k, a DM has the following one step choice: to obtain k immediately or to push some button u once and then to obtain a reward a_u and additionally with probability $1 - b_u$ to obtain k, and with complimentary probability to obtain zero.

Theorem 3. a) *Solution h of equation* (5) *is finite and unique*;

b) $h = \alpha = w$;

c) *the optimal index, or an optimizing sequence for any of the three problems is the optimal index (an optimizing sequence) for the other two problems.*

A bit later we present Propositions 2 and 3 which describe in detail the properties of functions $H(k)$ and $v(k)$ and explains the appearance of one more indicator (index), but to prove Theorem 3 we need only a simple

Proposition 1. a) *Functions $H(k)$ and $v(k), -\infty < k < \infty$, are nondecreasing, continuous, and convex (concave up)*;

b) *index $w < \infty$, and function $v(k) = k > H(k)$ for all $k > w$, and $v(k) = H(k) > k$ for all $k < w$.*

Proof. a) follows directly from the definition of $H(k)$ and $v(k)$. b) The assumption U implies that the slope of function $H(k)$ is bounded by $(1 - b)$. Therefore $H(k) < k$ for large k, $w < \infty$ and inequalities in b) hold.

Proof of Theorem 3. Assumption (**U**) implies that $\alpha \leq a/b < \infty$. The definition of α implies that for any $u \in U$, $\alpha = a_u + (1 - b_u)\alpha + \varepsilon_u b_u$, where $\varepsilon_u \geq 0$ and that there is a sequence u_n such that in a corresponding equality $\varepsilon_n \to 0$. In particular it is possible that all (a_n, b_n) coincide and $\varepsilon_n = 0$. The first equality implies that $\alpha \geq H(\alpha)$. The second relationship implies that $\alpha = \lim_n(a_n + (1 - b_n)\alpha) \leq H(\alpha)$. Therefore $\alpha = H(\alpha)$.

If h is a solution of the equation (5) then $h = H(h) = a_u + (1 - b_u)h + \delta_u b_u$, with $\delta_u \geq 0$ for any u and there is a sequence (a_n, b_n) such that in a corresponding equality

$\delta_n \to 0$. The first equality implies that $h = a_u/b_u + \delta_u/b_u$ and hence $h \geq \alpha$. The second relationship, together with an assumption $0 < b \leq b_u$, implies that $h = \lim_n a_n/b_n \leq \alpha$. Therefore $\alpha = h$ and h is a unique solution of (5).

To prove the equality $\alpha = w$, note that by Proposition 1, if $k \geq w$ then $v(k) = k \geq H(k) \geq a_u + (1 - b_u)k$ for all u and hence $k \geq a_u/b_u$ for all u, i.e. $k \geq \alpha$. Since this is true for any $k \geq w$ we obtain $w \geq \alpha$. If $k < w$ then $k < H(k)$ and hence there is u such that $k < a_u + (1 - b_u)k$. Therefore $k < a_u/b_u$ and $k \leq \alpha$. Since this is true for any $k < w$ we obtain that then $w \leq \alpha$. Thus $\alpha = w$. Point c) of Theorem 3 can be easily obtained using standard reasoning.

Remark. As two simple examples below show, we can not skip the inequality $0 < b \leq b_u$ or to remove the inequalities $a_u \leq a < \infty, b_u \leq 1 < \infty$ in assumption (**U**).

1) Let $U = \{1, 2, ...\}, a_n = 2/n, b_n = 1/n$. Then $\alpha = 2$ but it is easy to check that any $h \geq 2$ is a solution of the equation (2).

2) Let $U = \{1, 2, ...\}, a_n = 2n + \delta_n, b_n = n$ and $\lim_n \delta_n = \infty$, $\lim_n \delta_n/n = 0$. Then $\alpha = 2$ but it is easy to check that the equation (2) has no solutions.

Let us consider one more problem initially analyzed by Mitten in one page paper [8], where index α plays an important role.

Problem 4. Suppose that a DM has to solve the optimization problem similar to one in Problem 1 with sequential selection of buttons with only one distinction - every button can be used at most once.

The Mitten's result essentially can be described as

Theorem 4. *Suppose that there is a sequence of indices u_n such that after the reordering $\alpha_1 = \frac{a_1}{b_2} \geq \alpha_2 = \frac{a_2}{b_2} \geq ... \geq \frac{a_u}{b_u}$ for each $u \in U$ not in this sequence. Then to push buttons in the order $1, 2, ...$ is an optimal strategy.*

See the brief discussion of this problem and its relation to the GI and GGI in [13].

Now we present Proposition 2 and 3.

Proposition 2. *Function $H(k)$ is either strictly increasing for all k, or there are finite c and d, such that $H(k) = c$ on interval $(-\infty, d)$ and strictly increasing on (d, ∞).*

Proof. If $H(k)$ is not strictly increasing for all k then $H(k_1) = H(k_2) = c < \infty$ for some c and some $k_1 < k_2$. Since $a_u + (1 - b_u)k_2 = a_u + (1 - b_u)k_1 + (1 - b_u)(k_2 - k_1)$ the former equality implies that there is a sequence of indices u_n such that $\lim_n a_n = c$ and $\lim_n b_n = 1$. Then obviously $H(k) = c$ for all $k \le k_2$ and therefore there is $d = \sup\{k : H(k) = c\}$. Since H is unbounded we have $d < \infty$. Then $c < H(k)$ for all $k > d$, and for any sequence of indices u_n such that $\lim_n b_n = 1$ we have $\sup a_n \le c$. If function $H(k)$ is strictly increasing for all k, we set $d = -\infty$. If $-\infty < d$ then it is convenient to assume that set U is complemented by an extra index e such that $a_e = c$ and $b_e = 1$. Proposition 2 is proved.

If $-\infty < d$, then in Problem 3 we can be interested also to find t defined as

$$t = \sup\{k : v(k) = c\}. \tag{8}$$

In other words $t = min(d, w)$. The properties of function $H(k)$ described in Proposition 1 and 2 and definitions of w and t imply that indices t and w satisfy inequalities

$$-\infty < w, \quad -\infty \le t \le w < \infty, \text{and } t \le d < \infty. \tag{9}$$

The properties of function $v(k)$ are described in Proposition 3.

Proposition 3. *Function $v(k)$ satisfies $v(w) = H(w) = w$, $v(k) = k > H(k)$ for all $k > w$, $v(k) = H(k) = c > k$ for all $k \le t$, and $v(k) = H(k) > c \vee k$ for all $t < k < w$.*

Proof. The definition of $H(k)$ and its continuity imply that $H(w) = w$. If $H(k) \le k$ for some k, then by definition of $H(k)$ for all u we have $a_u + (1 - b_u)k \le k$, which is equivalent to $a_u/b_u \le k < s$ for all $s > k$. Hence $a_u + (1 - b_u)s = a_u + (1 - b_u)k + a_u + (1 - b_u)(s - k) \le k + (1 - b_u)(s - k) < s$ for all u and $H(s) < s$ for all such s. This implies that $w \le k$ and $H(k) < v(k) = k$ for all $k > w$, and that $v(k) = H(k) > k$ for all $k < w$. The continuity of $v(k)$ implies that $v(w) = w$ and if $-\infty < t$ then $v(t) = c$ for all $k \le t$. The definitions of w and t imply that if $t < w$ then $v(k) = H(k) > c \vee k$ for all $t < k < w$. Proposition 3 is proved.

Theorem 3 shows the equivalence of three abstract problems but leaves an open question which of them should be solved. Probably, there is no general answer to this question. It is possible that in some situations Problems 1 will be the easiest, and in some other - Problem 2. At the same time Problem 3 provides the most general approach since its solution breaks up in two stages: a solution for a particular k and finding w. Exactly such situation occurs in Markov reward model and three related indices. Let us show formally how the three problems described in sections 1 and 2 can be presented as abstract problems.

Given a reward with termination $M = (X, P, c(x), \beta(x))$, and an initial point x, let us define the set $U = \{u\} = \{$ set of all Markov moments $\tau > 0\}$, $\tau = \tau_G = \min(n : Z_n \in G \cup e), G \subset X, x \notin G$. The rewards a_u and probabilities b_u we define as $a_u = R^\tau(x) = E_x \sum_{n=0}^{\tau-1} c(Z_n)$, the total expected reward till moment τ, and $Q^\tau(x) = P_x(Z_\tau = e)$, the probability of termination on $[0, \tau)$. These are quantities participating in (3). Then function $H(k)$ coincides with $\sup_{\tau>0} E_x g(Z_\tau)$, where $g(x) = k$. Respectively $v(x|k) = k \vee H(k) = \sup_\tau E_x g(Z_\tau)$, i.e. $v(x|k)$ is the value function in an OS for MC in model $M(k)$.

Now we will show the value $\alpha = h = w$ in the example 2. In this example the set U consists of all possible stopping times τ corresponding to possible subsets G of a state space $X = \{1, 2, e\}$. Since each G should contain e, we have only four subsets $G_1 = \{1, e\}, G_2 = \{2, e\}, G_3 = \{1, 2, e\}$ and $G_4 = \{e\}$ and correspondingly four possible stopping times $\tau_i, i = 1, 2, 3, 4$. For each τ the expression for $R^\tau(x)$ in (3) can be calculated by the equality $R^\tau(x) = \sum_{n=0}^{\infty} \sum_{y \notin G} P_x(Z_n = y)c(y)$. Correspondingly $Q^\tau(x) = \sum_{n=0}^{\infty} \sum_{y \notin G} P_x(Z_n = y)p(y, e)$, where x is an initial point. Let us denote $R^\tau(1) = a_i, Q^\tau(1) = b_i$ for $\tau = \tau_i, i = 1, ..., 4$. Using the transition probabilities of example 2, we have $a_1 = 1 + \frac{2}{9}(2 + \frac{1}{4}2 + (\frac{1}{4})^2 2 + ...) = \frac{43}{27}$, $b_1 = \frac{3}{9} + \frac{2}{9}(\frac{1}{2} + \frac{1}{4}\frac{1}{2} + (\frac{1}{4})^2 \frac{1}{2} + ...) = \frac{13}{27}$ and $a_1/b_1 = \frac{43}{13}$. It can be checked similarly that $a_2 = \frac{9}{5}, b_2 = \frac{3}{5}$, $a_2/b_2 = a_3/b_3 = 3$ and $a_4/b_4 = \frac{43}{13}$. Thus the maximum of the ratio in Problem 2 is obtained on u_1 and u_4 and equal to $\alpha = \frac{43}{13}$. Correspondingly the equation (5) in Problem 1 has a solution $h = \frac{43}{13} = \frac{43}{27} + (1 - \frac{13}{27})\frac{43}{13}$. Similar calculations for the initial point 2 give $\alpha = 4$ and equation $4 = 2 + \frac{1}{2}4$. For Problem 3 function $H(k)$ is a piecewise linear function with two linear parts and the equality in Theorem 3 is easily checked.

This almost trivial example shows also that the equivalence of the three problems does not lend itself to the solution of these problems. The set of all partitions of X, which gives the size of the set U, grows exponentially with $|X| = n$ but the algorithm in [13] to calculate GGI is polynomial with complexity of order n^3.

4 CQR Model and Open Problems

A restart in s Markov model can be naturally generalized into the following model which we call *continue-quit-restart* (CQR) model. It is specified by a tuple $M = (X, B, P, A(x), c, q, r_i(x))$, where X is a countable state space, $B = \{s_1, ..., s_m\}$ is a *set of restart points*, a subset of a state space X, $P = \{p(x, y)\}$ is a stochastic matrix. At each state x a set of available actions is $A(x) = \{c, q, r_i, i = 1, .., m\}$, continue, quit, and restart to point $s_i, i = 1, .., m$. A reward function $r(x, a)$ is specified by particular functions $c(x), q(x)$ and $r_i(x), i = 1, 2, ..., m$. If an action c, "continue" is selected then $r(x, c) = c(x)$ and transition to a new state occurs according to transition probabilities $p(x, y)$, if an action q, "quit" is selected then $r(x, q) = q(x)$ and transition to an absorbing state e occurs with probability one, if an action r_i, "restart to state s_i " is selected then $r(x, r_i) = r_i(x)$ and transition to a state s_i occurs with probability one. If set B consists of one point s we obtain model similar to KV model but with variable discount rate $\beta(x)$ and fees for quit and restart, $q(x)$ and $r(x)$. The study of this problem and algorithm for its solution are given in [14]. A strategy in this problem is a partition of set X into three regions: continue, quit, restart.

If there are more than one restart point $B = \{s_1, ..., s_m)$, then we have the following optimization problem. Index set U ={set of strategies}, where each strategy $\mathbf{u}$ is now a partition of a set X into $m + 1$ sets, $X = S_c \cup S_q \cup_{i=1,...,m} S_i$. Given strategy $\mathbf{u}$ a vector $\mathbf{a}_u = (a_u(i), i = 1, ..., m)$ and (sub)stochastic matrix $\mathbf{B}_u$ are defined as follows: $a_u(i)$ equals to the expected total reward up to moment of termination starting at point s_i. The moment of termination is a moment of quit or restart or hitting the absorbing state e. The total reward includes the accumulated sum of current rewards plus the reward for quit or restart; an element $b_u(i, j)$ of a matrix $\mathbf{B}_u$ is a probability of return to state s_j starting from s_i using strategy $\mathbf{u}$. Then set $A = \{\mathbf{a}_u\}$, is a set of vectors $\mathbf{a}_u \in R^m$, and set $B = \{\mathbf{B}_u\}$ is a set of (sub)stochastic $m \times m$ matrices both indexed by the elements of U. This model suggests the following *multidimensional abstract optimization* problem: given an abstract set of indices U and sets $A = \{\mathbf{a}_u\}$, $\mathbf{a}_u \in R^m$, and $B = \{\mathbf{B}_u\}, \mathbf{B}_u$ a stochastic matrix, to find a solution (an equilibrium point) of an equation

$$\mathbf{h} = \sup_u [\mathbf{a}_u + \mathbf{B}_u \mathbf{h}], \tag{10}$$

where the supremum over a multi-dimensional vector can be understood as a maximum norm or some other norm.

The natural questions here are: whether there is a solution $\mathbf{h}$ for the equation (10), and whether there exists an analog of the Gittins ratio.

A possible approach to a solution of (10) is to formulate an analog of the abstract Problem 3. Given the vector $\mathbf{k} \in R^m, \mathbf{k} = \{k_i, i = 1, ..., m\}$ we can introduce the function $v(\mathbf{k})$ as a solution to an optimization problem $v(\mathbf{k}) = \sup_u [\mathbf{a}_u + \mathbf{B}_u \mathbf{k}]$, and after that to try to use some recursive scheme of convergence of $v(\mathbf{k})$ to $\mathbf{h}$. It is an open problem to prove such convergence to an equilibrium point and to check whether such a point is unique. A similar idea was used in [2] for a specific problem.

Note also that an equation of type (10) appears naturally in many optimization problems related to renewal stochastic processes, and such equation can be defined not only for finite dimensional vectors but also for $\mathbf{h}$, $\mathbf{a}_u$ and $\mathbf{B}_u$of a more general kind.

Our final remark is that the idea of abstract optimization can be applied to a setting to formulate and prove a general theorem similar to the renowned Gittins theorem of optimality of a strategy based on the Gittins index in Multi-armed Bandit problem with independent arms. This is the subject of a forthcoming paper due to the author.

References

[1] Bank, P., El Karoui, N., *A stochastic representation theorem with applications to optimization and obstacle problems*, Ann. Probab. 32, no. 1B, (2004), pp. 1030–1067.

[2] Boyarchenko, S., Levendorski S., *Irreversible Decisions under Uncertainty: Optimal Stopping Made Easy*, Springer, 2007.

[3] Denardo, E., Rothblum, U., Van der Heyden, L., *Index policies for stochastic search in a forest with an application to R&D project management,* Math. Oper. Res., 29, no. 1, (2004), pp. 162–181.

[4] Feinberg, E., Shwartz, A., *Handbook of Markov decision processes*, Internat. Ser. Oper. Res. Management Sci., 40, Kluwer Acad. Publ., Boston, MA, 2002.

[5] Granot, D., Zuckerman, D., *Optimal sequencing and resource allocation in research and development projects,* Management Science, 37, (1991), pp. 140-156.

[6] Gittins, J. C., *Bandit processes and dynamic allocation indices,* J. Roy. Statist. Soc. Ser. B 41, no. 2, (1979), pp. 148–177.

[7] Katehakis, M., Veinott, A., *The multi-armed bandit problem: decomposition and computation,* Math. Oper. Res., 12, no. 2, (1987), pp. 262–268.

[8] Mitten, L., *An Analytic Solution to the Least Cost Testing Sequence Problem,* J. of Industr. Eng., 11, no. 1, (1960), pp. 17.

[9] Presman, E., Sonin I., *A Gittins Type Index Theorem for Randomly Evolving Graphs*, From Stochastic Calculus to Mathematical Finance. The Shiryaev Festschrift, Kabanov, Y; Lipster, R; Stoyanov, J (Eds.), Springer, XXXVIII, 2006, pp. 567–588.

[10] Sonin, I., *Two simple theorems in the problems of optimal stopping*, in Proc. INFORMS Appl. Prob. Conf., Atlanta, Georgia, 1995.

[11] Sonin, I., *The Elimination Algorithm for the Problem of Optimal Stopping,* Math. Meth. of Oper. Res., *(*1999), pp. 111-123.

[12] Sonin, I., *The State Reduction and related algorithms and their applications to the study of Markov chains, graph theory and the Optimal Stopping problem*, Advances in Mathematics, 145, (1999), pp. 159-188.

[13] Sonin I., *A Generalized Gittins Index for a Markov Chain and its Recursive Calculation,* Statistics & Probability Letters, 78, Issue 12, 1, (2008), pp. 1526-1533.

[14] Sonin, I., S. Steinberg, *Continue, Quit, Restart Probability Models,* manuscript, 2010.

[15] Varaiya, P., Walrand J., Buyukkoc C., *Extensions of the multiarmed bandit problem: the discounted case,* IEEE Trans. Autom. Control AC-30, (1984), pp. 26-439.

[16] Whittle, P., *Multi-armed Bandits and the Gittins Index,* J. Roy. Statist. Soc. Ser. B, 42(2), (1980), pp. 143-149.