General Notions of Indexability – Applications and Challenges

Kevin Glazebrook*, David Hodge$^{\phi}$ and Chris Kirkbride*

*Department of Management Science, Lancaster University, UK.
$^{\phi}$School of Mathematical Sciences, University of Nottingham, UK.

Introduction

Focus of Study

Models and methods for the dynamic allocation of a single key resource among a collection of stochastic projects which are competing for it.

Goal of Study

Explore extensions to the index policies of Gittins (1979) and Whittle (1988) which have proved so successful in very simple bandit problems.

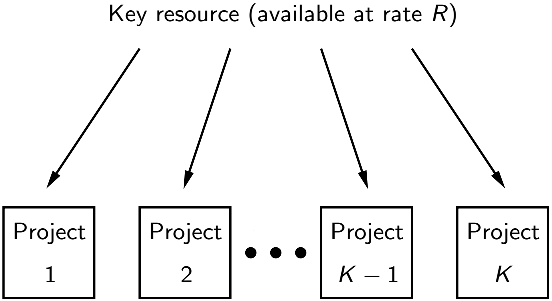
Central Ideas

Development of appropriate measures of the cost effectiveness of allocating resource at a given level to a project in a given state. Development of heuristic policies for resource allocation using such measures.

Methodology

Follow Whittle (1988) in achieving a project-based decomposition of the problem via the development of a Lagrangian relaxation.

Introduction (2)



- Projects evolve over time under the action of the key resource, earning returns (or incurring costs) as they go.
- Decision-maker reviews the distribution of the resource whenever a project changes state.
- Inventory Management; Queueing Control; Asset Management; Machine Maintenance; Military Logistics.

A Model for Dynamic Resource Allocation

- K stochastic reward generating/cost incurring projects are driven by the application of some divisible resource.
- At each decision epoch (state transition) an action $\mathbf{a} = (a_1, a_2, \ldots, a_K)$ is applied to the system.
- Admissible actions:

  $$A = \left[\mathbf{a}; a_k \in \{0, 1, \ldots, R\}, 1 \leq k \leq K, \text{ and } \sum_{k=1}^{K} a_k \leq R\right].$$
- System state: $\mathbf{x} = (x_1, x_2, \ldots, x_K) \in \mathbb{N}^K$.
- Project k: $\begin{cases} \text{Reward rate earned,} & d_k(x_k), \\ \text{Transition rates,} & q_k\left(x_k' | a_k, x_k\right) \end{cases}$
- Want a policy for resource allocation to maximise the average return per unit time from all projects.

Key Steps to Index Policies (1)

Optimisation Goal:

$$D^{opt} = \max_{\mathbf{u}} \sum_{k=1}^{K} D_k(\mathbf{u}) \qquad \text{(admissible policies)}$$

$$\downarrow$$

Lagrangian Relaxation (LR):

$$D(W) = \max_{\mathbf{u}} \sum_{k=1}^{K} \{D_k(\mathbf{u}) - WR_k(\mathbf{u})\} + WR$$

(constraint $\sum a_k \leq R$ abandoned)

$$D(W) \geq D^{opt}, \;\; W \in \mathbb{R}^+$$

$$\downarrow$$

Projectwise Decomposition:

$$D(W) = \sum_{k=1}^{K} D_k(W) + WR, \quad \text{where}$$

$$D_k(W) = \max_{u_k} \{D_k(u_k) - WR_k(u_k)\} \qquad \text{(problem } P(k, W)\text{)}$$

Key Steps to Index Policies (2)

(Full) Indexability:

Project k is **fully indexable** if there exist stationary policies $\{u_k(W); W \in \mathbb{R}^+\}$ such that

(a) $u_k(W)$ is optimal for $P(k, W)$, and

(b) $u_k(x_k, W)$ is decreasing in $W \ \forall \ x_k$

$$\downarrow$$

Indicies:

If project k is **fully indexable**, define **indices**

$$W_k(a_k, x_k) = \inf\{W; u_k(x_k, W) \leq a_k\} \qquad \text{(index as fair charge)}$$

$$\downarrow$$

Index Solution to LR:

If all K projects are **fully indexable** the above Lagrangian Relaxation is solved by the policy $\mathbf{u}(W)$ such that $\forall \ \mathbf{x}$

$$\mathbf{u}(W, \mathbf{x}) = \mathbf{a} \iff W_k(a_k - 1, x_k) > W \geq W_k(a_k, x_k), \ \forall \ k.$$

In words: accumulate resource at each project until the fair charge for adding further resource falls below the prevailing charge W.

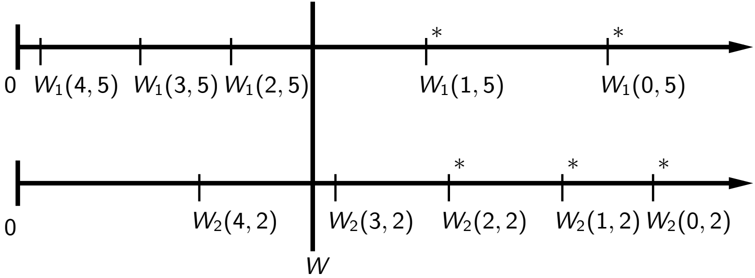
Key Steps to Index Policies (3)

Index heuristic for the original problem:

Increase resource levels at the projects in decreasing order of the appropriate indicies/fair charges until the resource constraint is violated.

Example:

Let $K = 2,\ \mathbf{x} = (5, 2)$



Optimal action for Lagrangian Relaxation: $(a_1, a_2) = (2, 4)$.

If $R = 5$, greedy index heuristic $*$ chooses $(a_1, a_2) = (2, 3)$.

Asymptotically Optimal Performance

An asymptotic framework:

n **fully indexable** bandits, total resource constraint $n\alpha R$ for $\alpha \in (0,1)$ (scaling with n, as $n \to \infty$)

Index solution to LR is still optimal.

Asymptotic result:

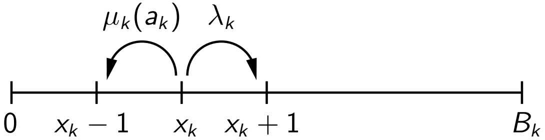
Under mild irreducibility conditions, and good behaviour of solutions to a deterministic fluid-limit differential equation. Index heuristic for the original problem yields asymptotically optimal reward per bandit – the same reward per bandit as the index solution to the LR.

Small state bandits:

In the case of two- and three-state bandits the differential equation is well behaved.

A Queueing Control Example

- A team of R servers provides service at K stations. Station k has finite waiting room of size B_k. Completed services at station k earn a return d_k.
- How to dynamically allocate the R servers among the stations to maximise the aggregate return rate?
- Dynamics at station k with a_k servers:



Service rate $\mu_k(a_k)$ is strictly increasing and strictly concave in a_k.

Reward rate: $d_k(x_k) = d_k \lambda_k I(x_k < B_k)$

- Station k is **fully indexable**.

Numerical Results

Analysis of problems with $K = 2,\ R = 25,\ d_1 = d_2 = 1,$

$$\mu_k(a_k) = a_k \mu_k (a_k + \nu_k)^{-1},\ k = 1,\ 2$$

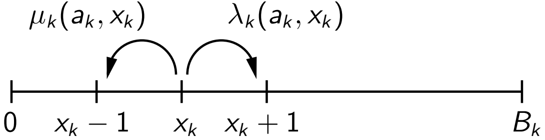
and a range of choices for $\lambda_1,\ \lambda_2,\ \mu_1,\ \mu_2,\ \nu_1,\ \nu_2,\ B_1,\ B_2.$

	MIN	LQ	MED	UQ	MAX	N
Greedy Index	0.0023	0.0148	0.0235	0.0336	0.1199	5250
Optimum Static	17.9544	23.4628	25.4526	27.4720	33.8567	5250

Percentage reward rate deficit compared to optimum

A Model for Asset Management ("Spinning Plates")

- Resource is available at rate R to support the performance of K reward generating assets. In the absence of investment, the reward-earning capability of an asset deteriorates.
- Dynamics of asset k under resource level a_k:



  Enhancement rate $\lambda_k(a_k, x_k)$ is strictly increasing and strictly concave in a_k $\forall$ x_k.

  Deterioration rate $\mu_k(a_k, x_k)$ is strictly decreasing and strictly convex in a_k $\forall$ x_k.

  Return rate $d_k(x_k)$ is increasing in x_k.

- Asset k is **fully indexable**.

Numerical Results

Analysis of 2,000 problems with

$$K = 2,\ R = 5,\ B_1 = B_2 = 10,\ d_k(x_k) = x_k(x_k + 1)^{-1},\ k = 1,\ 2$$

$$\lambda(a_k, x_k) = a_k(a_k + \phi_k)^{-1},$$
$$\mu_k(a_k, x_k) = \phi_k(a_k + \phi_k)^{-1}\eta_k,\ k = 1,\ 2$$

and a range of choices for $\phi_1,\ \phi_2,\ \eta_1,\ \eta_2$.

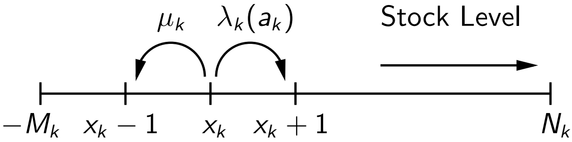
	MIN	LQ	MED	UQ	MAX
Index	0.0000	0.1482	0.6752	1.0751	1.9082
Static	0.0719	3.7812	6.1724	7.4822	13.6966
Myopic	0.0027	4.7774	16.7270	26.5042	39.3193

Percentage return rate deficit compared to optimum

What happens when indexability fails to hold?

A Multi-Product Make-to-Stock Production System

- A resource, available at R, drives the manufacture of K products which are made to stock to meet exogenous demand.
- How to dynamically allocate the resource among the K products to minimise an aggregate of costs from lost sales, backorders and holding inventory?
- Dynamics for product k with resource level a_k:



Production rate $\lambda_k(a_k)$ is strictly increasing and strictly concave in a_k.
Cost rate

$$c_k(x_k) = \underset{\text{(holding)}}{h_k x_k^+} + \underset{\text{(backorder)}}{b_k x_k^-} + \underset{\text{(lost sales)}}{l_k \mu_k I(x_k = -M_k)}.$$

Typically $h_k \ll b_k \ll l_k \mu_k$.

- There exists $\hat{h}_k$ such that product k is **fully indexable** when $h_k \leq \hat{h}_k$. **Practical indexability**.

Problem set	Policies							
	Index		Static		OSPI		Myopic	
	Med	Max	Med	Max	Med	Max	Med	Max
Recip A	0.004	0.859	0.870	42.955	0.047	3.943	72.292	558.735
Recip B	0.239	4.055	25.513	122.785	2.311	11.536	347.324	>2000
Recip C	1.262	11.327	94.945	284.436	7.376	22.091	>2000	>20000
Power A	0	0.025	0.013	1.528	0	0.458	14.897	43.987
Power B	0.001	0.163	0.149	14.712	0.029	4.892	25.192	103.310
Power C	0.049	7.603	3.141	347.921	0.999	48.271	50.768	>2000
Log D	0.230	5.458	8.776	164.421	2.902	43.489	75.289	1181.374

Median and maximum percentage suboptimalities of five heuristic policies for three forms of production rate. There are 900 problems summarised in each row of the table.

Example

$K = 2,\ R = 25.$ Both products have
$M = N = 10,\ h = 0.025,\ b = 1.5,\ l = 200.$

Product 1: $\mu_1 = 1.576,\ \lambda_1(a) = 4.5a(a + 5.971)^{-1}$, is **not** fully indexable. We have

$$u_1(9, W) = \begin{cases} 0, & W \le 0.056, \\ 1, & W = 100, \\ 0, & W \ge 200 \end{cases} \quad \Longrightarrow$$

two natural values of fair charge $\tilde{W}_1(0, 9)$
(0.056,140)

Product 2: $\mu_2 = 1.046,\ \lambda_2(a) = 1.5a(a + 5.971)^{-1}$ **is** fully indexable.

Two natural candidate index heuristics:

$H1$ always uses the **higher** fair charge index

$H2$ always uses the **lower** fair charge index

$$C^{opt} = 4.333;$$

$$C^{H1} = 4.370 \text{ (0.86\% suboptimal)}; \quad C^{H2} = 4.926 \text{ (13.70\% suboptimal)}.$$

If μ_2 is significantly reduced, $H2$ outperforms $H1$.

- Fix $k,\ x_k$ and consider $u_k(x_k, W)$ as $W : \infty \searrow 0$.

Either: $u_k(x_k, W) : 0 \nearrow R$ as $W : \infty \searrow 0$. All $W_k(a_k, x_k)$ well defined.

Or: $u_k(x_k, W) : 0 \nearrow r \searrow 0$ as $W : \infty \searrow 0$. $W_k(a_k, x_k) = 0,\ a_k \geq r$ (**idling**)

and two natural values of the fair charges $\tilde{W}_k(a_k, x_k),\ a_k < r$.

- Define $H1,\ H2$ as on previous slide
- Write $C(W)$ for the value of the Lagrangian Relaxation.

$C^* = \min_{W \geq 0} C(W) = C(W^*)$, say, with $\mathbf{u}(W^*)$ the corresponding policy.

C^* is an easily computed lower bound for C^{opt}.
W^* is a measure of the underlying value placed on resource by the system.

- Key question: Which of $H1,\ H2$ is "closer to" $\mathbf{u}(W^*)$?

% subopt.	Dataset E H2'	OSPI	H1	Static	Dataset F H2'	OSPI	H1	Static
Min	0.034	0.108	6.425	26.491	0.001	0.016	123.708	269.711
LQ	0.557	2.583	45.938	103.908	0.057	7.276	155.009	333.037
Median	1.155	5.737	61.183	135.974	0.470	27.429	172.520	359.905
UQ	2.149	11.560	76.132	168.242	1.259	54.766	189.450	387.342
Max	8.887	54.747	134.928	279.051	4.672	257.294	282.338	496.718

Policies

Percentage suboptimalities of $H_2'(1/70)$, OSPI, H_1, and STAT, in a high h scenario. There are 900 problems for each dataset.