INFORMATIONAL ASPECTS OF DECENTRALIZED RESOURCE ALLOCATION

by

Takashi Ishikida

Memorandum No. UCB/ERL/IGCT M92/60

1 May 1992

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

INFORMATIONAL ASPECTS OF DECENTRALIZED RESOURCE ALLOCATION

by

Takashi Ishikida

Memorandum No. UCB/ERL/IGCT M92/60

1 May 1992

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Informational Aspects of Decentralized Resource Allocation[1]

Takashi Ishikida
Department of Industrial Engineering and Operations Research
University of California, Berkeley, CA 94720
May 1, 1992

Abstract

In this thesis decentralized resource allocation is formulated and analyzed from the viewpoint of economic theory of mechanism. Allocation problems are expressed as constrained optimization problems, where each resource user's valuation of the resources is private knowledge and the goal is to maximize the aggregate valuation of all users. Since the relevant information is distributed among resource users, messages need to be exchanged between them so that the goal can be accomplished.

Mechanism theory is applied to determine the minimum requirement for information-carrying capacity—the minimum size of message space in the language of mechanism theory. A single-stage deterministic allocation problem is analyzed first. Then the effects of uncertainty and intertemporality in users' valuation on the size of information-carrying capacity are examined through examples from electric power pricing and assignment of a digital communication link. The main result is that the information-carrying capacity needs to be large enough to accommodate prices for commodities in the sense of the Arrow-Debreu economy in order to accomplish the goal.

As a special case of assignment of a digital communication link, a multi-armed bandit problem is studied. A new proof of the optimality of the (Gittins) index rule is obtained. The asymptotic optimality of the index rule under the average reward criterion is also derived.

[1]Work supported by NSF grant IRI-8902813

Contents

Chapter 1

Introduction

1.1 Decentralized Procedure

Imagine a system consisting of many participants such as our economic system where the individual characteristics of its participants are, to a certain degree, kept private. When such a system faces a certain task, the communication of private information and the coordination of actions among its participants are essential to the fulfillment of the task. A typical task is the efficient allocation of a limited resource: the system may be a university and the resource may be processors in its computer center, a community and its telecommunication networks, the world economy and the oil supply, and so on. Another example is the coordination of parallel computations; a number of processors work on related subproblems and coordination of the actions of the individual processor is achieved by passing messages among them.

Situations like these call for procedures which take account of the distributed nature of the relevant information. Unfortunately, many of procedures proposed in the control theoretic literature for analyzing systems and designing control strategies are based on the presupposition of *centrality*: it is implicitly assumed that all the information in the system is available to a central decision making body and it performs all the necessary calculations.[1] These procedures are not directly applicable to situations with distributed information.

Let us consider a task of a system which is to be accomplished at a specified time (in the future). A formal approach to the solution of this task needs to include the

[1]Refer Sandel *et al.*[21] for a survey of decentralized control methods for large scale systems.

specification of

1. a well-defined goal of the system,
2. the initial distribution of the information relevant to accomplish the goal,
3. possible actions.

A *decentralized* procedure (to solve the task) involves two phases:

1. the communication phase in which messages are exchanged among the participants,
2. the action phase in which the 'consensus' or the terminal message is translated into actions.

The design of procedures includes the choice of

1. type and size of messages to be exchanged,
2. method of communication,
3. rule to terminate the communication phase.

In an economic system, a specification of behavioral assumptions regarding participants' emissions of and responses to the chosen messages will be needed as well. Incentive compatibility in the game theoretic framework is an important issue there.

In economics, the procedures just described above are often termed *adjustment processes*[10]. However, we note that the communication phase in an adjustment process is not a real-time dynamical system but what is in economics called a tâtonnement process.

1.2 Communication vs. Computation

Since the aim is to design a procedure, it is desirable to have criteria to compare the performance of different procedures. One of the criteria studied by economists is the 'size' of messages that must be exchanged according to a procedure. For example, the price of a commodity announced by a seller and the expressed demand of the consumers for the commodity constitute a two-dimensional message space. The size of the message space represents a part of the cost of communication; namely, the channel capacity or information-carrying capacity required by the procedure, supposing the participants are transmitting their messages *simultaneously*. Considerable research has been conducted to find a lower bound on the size of the message space for a goal-realizing procedure; in particular, Pareto optimal allocation of resources in the exchange economy has been under extensive study. These analyses are indifferent to how the communications are conducted. They are sensitive only to the resulting terminal messages, called *equilibrium messages*. This static form of a procedure is often called a *mechanism* by economists.

An obvious drawback of the these analysis is that they ignore the computational efforts necessary to find the terminal messages. There appears to be a tradeoff between the size of message space and the computational efforts in realizing the goal of the system. A unifying qualitative measure has yet to be defined.[2]

1.3 Constraints on Communication Capability

Designers of decentralized procedures may not have the liberty of selecting the medium of communication and building the necessary capacities. Rather they may themselves confined to the existing medium and capacities. When this is the case, a model of the system should in principle include the specification of types and sizes of feasible messages just as it includes a specification of feasible actions. This is an issue which does not appear in the design of centralized procedures, but it bears practical importance when we consider the implementation of a procedure. Even when the designers have freedom to build a new capacity, it will be impractical to have huge message spaces. We can imagine the situation that the system has to settle for a slightly lesser goal to keep the burdens of communications and associated information processing manageable. In Chapter 4, we study the pricing schemes for the allocation of electric power under uncertainty. We will see that when the consumers of electric power are *inflexible*, the efficient allocation requires the communication of the distribution function of the underlying stochastic events. When the distribution is continuous, communicating a finite approximation of it will be a practical alternative, provided the resulting efficiency loss is within the reasonable range.

[2]See the discussion in Mount and Reiter[16].

1.4 Message Exchanges over Time

The importance of limited communication capability is even more pronounced when the accomplishment of the goal requires coordinating the participants' actions over long periods of time. It is possible to regard the issue of communication just as in a single period case; namely, to communicate all the relevant information at the beginning of the periods (and do nothing after that). Indeed, as dynamic programming methodology suggests, it is necessary to do so in general. In Chapter 5, we will see an example of a goal for which all the relevant information has to be revealed at the beginning of the periods to accomplish the goal.[3] However, such an approach will most likely require impractically huge communication capability. An alternative will be to relax the goal and spread message exchanges over time.

1.5 Scope of this Thesis

This thesis studies informational aspects of resource allocation problems. Allocation problems are formulated as (constrained) optimization problems. Each participant's valuation over allocation patterns is regarded as private knowledge. Optimality conditions of the problems turn out to be the key to the design of the decentralized procedures. The communication issue is addressed. Lower bounds on the sizes of messages spaces are derived. 'Price mechanisms' are of particular interest. Computational aspects, regrettably, are not covered. Game theoretic issues concerning participant behavior are not addressed.

In Chapter 2, the economic theory of 'allocation mechanisms' is reviewed to the extent needed in this thesis. A class of mechanisms which have a convenient form for the study of price mechanisms is introduced.

In Chapter 3, a deterministic resource allocation problem is formulated as a convex program. The minimum size of message space for a goal-realizing mechanism is found.

In Chapter 4, pricing schemes for allocation of electric power are studied as an example of resource allocation under uncertainty. Uncertain events such as a generator failure or a sudden burst of consumer demand will cause a shortage of supply and the need for rationing. A two-stage recourse model is employed to formulate the interruption cost of consumers whose demands are rationed.

[3]In the same chapter, we also see a problem which allows us to spread message exchanges over time.

In Chapter 5, assignment of a 'digital pipe' is studied in deterministic setting. A digital pipe is a communication link that connects a source and a destination and can transport one fixed-length packet per unit time. Several users are to share the pipe over a fixed time interval. Users' valuations of the use of the pipe are intertemporal. The possibility of spreading message exchanges over time is discussed.

In Chapter 6, a multi-armed bandit problem is studied. This is a special case of the assignment of the digital pipe in a stochastic setting. A new proof of the Gittins index rule is given. The case where the system admits arrivals of new pipe users is also studied.

Concluding remarks and suggestions for future work are made in Chapter 7.

Chapter 2

Allocation Mechanisms

In § 2.1 the economic theory of 'allocation mechanisms' is reviewed to the extent needed in this thesis. An extensive survey of this subject is found in Hurwicz[10]. Refer also to Reiter[19].

In § 2.2 a class of allocation mechanisms with message spaces in a form convenient for the study of 'price mechanisms' is introduced.

2.1 Review of Mechanism Theory

Since we study decentralized resource allocation problems posed as (constrained) optimization problems, an example of an allocation problem is given below, and the basic terminology, concepts, and proof techniques are illustrated through the example.

2.1.1 Example of a decentralized allocation problem

Imagine that a company consists of N divisions which share K resources among them, and a resource management division. The goal of the company is to allocate resources among its divisions so that its profit is maximized. Only the resource management division knows the amounts of the resources available for the company. Each of the other divisions knows the profit it can make as a function of the amounts of the resources allocated to it. The profit function of each division is assumed to be independent of the allocation to the other divisions.[1] A division does not know the profit functions of the other divisions, nor what the resource management division does.

[1] A more general problem will be studied in Chapter 3.

Let us index the resource management division by $N+1$ and the other divisions by 1 through N. Let the profit function of division j be denoted by $u^j(\cdot)$ and the amount of resource i available to the company by b_i. Also let y_i^j be the amount of resource i allocated to division j. With this notation, the goal of the company is

P:

$$\begin{array}{lll} \max & \sum_{j=1}^N u^j(y_1^j, y_2^j, \ldots, y_K^j) & \\ \text{sub. to} & \sum_{j=1}^N y_i^j \leq b_i, & i = 1, 2, \ldots, K \\ & y_i^j \geq 0, & j = 1, \ldots, N,\ i = 1, \ldots, K. \end{array}$$

We assume

- the profit functions are concave and continuously differentiable,
- an optimal solution of problem **P** exists.

2.1.2 Terminology

We consider a task of someone who is asked to design a decentralized procedure for the company. It should be kept in mind that the designer's task is not just to find an optimal solution for a particular problem instance, but to devise an algorithm which works for different problem instances of the same sort.[2]

Environment: 'Characteristics' of participant j, say e^j, are called a *local environment* of participant j. The set of possible local environments of j is denoted by E^j. The *(system) environment* is a tuple consisting of local environments of all the participants in the system and denoted by e. The set of possible system environments is denoted by E.

In our example, the local environment of the resource management division is a vector of the amounts of the resources available to the company, i.e., $e^{N+1} = b := (b_1, b_2, \ldots, b_K)$,[3] and the local environments of the other divisions are their profit functions, i.e., $e^j := u^j(\cdot)$. The sets of local environments are

$$\begin{aligned} E^j &:= \text{the set of differentiable convex functions on } R_+^K, \quad j = 1, 2, \ldots, N, \\ E^{N+1} &:= R_+^K, \end{aligned}$$

[2]See the distinction made between an *instance of optimization problem* and an *optimization problem* in Papadimitriou and Steiglitz[18].

[3]Throughout this thesis "$A := B$ " stands for "A is defined as B".

and the set of system environments is $E := E^1 \times \cdots \times E^N \times E^{N+1}$.

Action space: A set of possible actions by the system is called an *action space* or an *outcome space*, and denoted by $\mathcal{A}$.

In our example, the feasible region of problem $\mathbf{P}$ is the action space.

Goal correspondence: The relation between environments and desired actions of the system is represented as a point-to-set mapping. This point-to-set mapping is called a *goal correspondence* or a *performance standard*, and denoted by F. We will refer to it simply as a goal.

In our example, the goal is the relation between the problem instance and its corresponding optimal solutions. That is, denoting the set of optimal allocations for problem instance $e := (u^1, \ldots, u^N, b)$ by argmax $\mathbf{P}(e)$, the goal $F : E \twoheadrightarrow \mathcal{A}$ is defined by

$$F(e) := \operatorname{argmax} \mathbf{P}(e).$$

The terms so far introduced are needed to describe the system. We move on to introduce (static) procedures, to be called mechanisms. Definitions of the components of a mechanism follow.

Mechanism: A triplet $\langle \mathcal{M}, \mu, h \rangle$ of a message space $\mathcal{M}$, an equilibrium correspondence μ, and an outcome function h is called a *mechanism* in equilibrium correspondence form. A triplet $\langle \mathcal{M}, g, h \rangle$ of a message space $\mathcal{M}$, a verification function g, and an outcome function h is called a mechanism in verification function form.

Message space and its size: A set of messages chosen for communication by the designer is called a *message space* and denoted by $\mathcal{M}$.

Throughout this thesis, message spaces are taken to be subsets of real vector spaces. The *size of a message space* $\mathcal{M}$ is defined as the dimension of the smallest real vector space in which there is an open set W such that $\mathcal{M} \subseteq W$. It is denoted by $\dim \mathcal{M}$. For a discussion of more general message spaces and their sizes, refer to Hurwicz[10] and the references therein.

As mentioned in Chapter 1, there is no specification of dynamics (neither real-time nor tâtonnement) in the 'static equilibrium framework' of mechanisms. What is specified are terminal messages.

Equilibrium messages: Terminal messages of the communication phase of a procedure are called *(joint) equilibrium messages.*

Equilibrium correspondence: The relation between environments and a mechanism's equilibrium messages is represented as a point-to-set mapping. This mapping is called an *(joint) equilibrium correspondence*, and denoted by $\mu(: E \twoheadrightarrow \mathcal{M})$. The *individual* or *(coordinate) equilibrium correspondence* of participant j, $\mu^j : E^j \twoheadrightarrow \mathcal{M}$, represents a relation between the local environments of j and terminal messages emitted or accepted by j.

To capture the private nature of the initial distribution of the information, we require the following **privacy-preserving** property on the equilibrium correspondence:

$$\mu(e) = \bigcap_j \mu^j(e^j), \ \forall e \in E, \tag{2.1}$$

where the intersection is taken over all the participants of the system.

Intuitively, the equilibrium messages are the messages accepted by every participant. This notion of consensus may be better captured through the use of *verification functions.*

Verification function: An alternative way of specifying terminal messages acceptable to participant j is to introduce a (vector-valued) function $g^j : \mathcal{M} \times E^j \rightarrow R^{n_j}$ such that

$$g^j(m, e^j) = 0 \iff m \in \mu^j(e^j), \ \forall e^j \in E^j, \ m \in \mathcal{M}.$$

The function g^j is called a *verification function* or an *agreement function of participant j.* Participant j answers 'yes' or returns value 0, when a message *announced publicly* is to its liking. Collectively, the individual verification functions is called the *verification function of the system*, and denoted by $g := (g^1, g^2, \ldots)$.

Outcome function: A function which translates messages into actions is called an *outcome function*, and denoted by $h(: \mathcal{M} \rightarrow \mathcal{A})$.

A mechanism is said to **realize the goal** F over the system environment E if

$$h(\mu(e)) \subseteq F(e), \ \forall e \in E. \tag{2.2}$$

The *direct revelation mechanism* is a trivial example of a goal-realizing mechanism in which every participant reveals its characteristics, and a centralized procedure is devised. In our example, the direct revelation mechanism consists of

$$\mathcal{M} \ := \ E,$$

$$\mu^j := E^1 \times \cdots \times E^{j-1} \times \{u^j(\cdot)\} \times E^{j+1} \times \cdots \times E^N \times E^{N+1}, \; j = 1, 2, \ldots, N,$$
$$\mu^{N+1} := E^1 \times \cdots \times E^N \times \{b\},$$

and a centralized algorithm for solving problem **P** as an outcome function. Notice this mechanism requires an infinite dimensional message space. It is natural to ask if there is a goal-realizing mechanism with a smaller message space; and if there is, what a lower bound of the size of message spaces is. For our example, we will see that $K(N+1)$ is such a lower bound. The following subsection introduces a proof technique to show it.

2.1.3 Uniqueness property

There is a useful proof technique to show the minimum size of a message space of a goal-realizing mechanism. Since the technique is used later, it is outlined here. We follow Hurwicz[10].

A key step is to choose a subset E^* in the set of system environments E such that

1. E^* has the target size (a candidate for a lower bound on a message space of a goal-realizing mechanism),
2. an equilibrium correspondence μ (of any goal-realizing mechanism) has a single-valued inverse on $\mu(E^*)$, i.e.,

$$\mu(e) \cap \mu(\tilde{e}) \neq \emptyset \Longrightarrow e = \tilde{e}, \quad \text{for all } e, \tilde{e} \in E^*. \tag{2.3}$$

By definition, when restricted to $\mu(E^*)$, μ^{-1} is onto. Therefore, naively speaking, the size of $\mu(E^*)$ is at least as large as that of E^*. Since $\mu(E^*) \subseteq \mathcal{M}$, the size of the message space $\mathcal{M}$ is also at least as large. Hence, the size of E^* gives a lower bound for the size of a message space of a goal-realizing mechanism. If we can show that there is a goal-realizing mechanism with a message space of the target dimension, then the lower bound is indeed the minimum size of a message space.

Rigorous arguments require a certain *regularity condition* on either equilibrium correspondences (or verification functions) or their inverses. Regularity conditions are introduced to prohibit the use of dimension-increasing mappings such as Peano's space filling curves. This will be discussed further in the following subsection.

Since every equilibrium correspondence must have a single-valued inverse on E^*, it is convenient to have a sufficient condition for the single-valuedness of μ^{-1} in terms of the goal F. The following condition provides just that.

Uniqueness property: The subset $E^*(\subseteq E)$ is said to have the *uniqueness property with respect to the goal F*, if and only if

$$\text{for all } e, \tilde{e} \in E^*, \text{ if } \exists a \in \mathcal{A} \text{ such that } a \in F(e) \cap F(\tilde{e}) \cap (\bigcap_j F(\tilde{e} \otimes_j e)), \text{ then } e = \tilde{e}, \quad (2.4)$$

where $\tilde{e} \otimes_j e := (e^1, \cdots, e^{j-1}, \tilde{e}^j, e^{j+1}, \cdots)$, and the last intersection is taken over all the participants of the system.[4] □

Note that the $\tilde{e} \otimes_j e$'s are not required to stay in E^*.

Lemma 2.1.1 *Let $\langle \mathcal{M}, \mu, h \rangle$ be a privacy-preserving mechanism realizing the goal F over E. Let $E^* \subseteq E$ be a subset having the uniqueness property with respect to F. Then the inverse of μ is single-valued on $\mu(E^*)$. In other words, (2.4) implies (2.3).*

Proof Let $\langle \mathcal{M}, \mu, h \rangle$ be as stated in the lemma. Let e and $\tilde{e}$ be in E^*. Assume $m \in \mu(e) \cap \mu(\tilde{e})$. Then by the privacy-preserving property (2.1),

$$m \in \mu^j(e^j) \cap \mu^j(\tilde{e}^j), \ \forall j.$$

Thus

$$m \in \mu(\tilde{e} \otimes_j e), \ \forall j.$$

Since $\langle \mathcal{M}, \mu, h \rangle$ realizes the goal, by (2.2),

$$h(m) \in F(e) \cap F(\tilde{e}) \cap (\bigcap_j F(\tilde{e} \otimes_j e)).$$

But then by the uniqueness property (2.4), $e = \tilde{e}$ as desired. □

Remark The uniqueness property can be strengthened. In the stronger form, the conclusion of (2.4) holds when we intersect with $\bigcap_j F(e \otimes_j \tilde{e})$ the value taken by F for any two environments in which some components are from e and the rest from $\tilde{e}$. We use this stronger version freely.

Example: We apply the proof technique to our example in § 2.1.1 to show $K(N+1)$ is a lower bound for the size of a message space of a goal-realizing mechanism.

As the first step, it will be shown for $K = 1$.

As shown above, it comes down to the choice of E^*. Consider profit functions of the form:

$$u^j(y) = 2\alpha^j \sqrt{y}, \text{ where } \alpha^j > 0.$$

[4] We follow the notation of Mount and Reiter[15]. See their paper for details of this *crossing condition.*

We abuse notation and write

$$E^{*j} := \{\alpha^j \in R_{++}\},\ j = 1, 2, \ldots, N,$$

rather than writing $E^{*j} = \{u^j(\cdot) = 2\alpha^j\sqrt{\cdot}|\alpha^j \in R_{++}\}$.

$$E^{*N+1} := R_{++}.$$

And $E^* := E^{*1} \times E^{*N} \times E^{*N+1}$. Notice E^* has the dimension $N+1$.

Lemma 2.1.2 *As defined above, E^* has the uniqueness property with respect to the goal F.*

Proof Since the derivatives of u^j's tend to infinity as their arguments tend to 0, the resource should be exhausted, and both the allocation and the associated Lagrange multiplier are strictly positive at the optimal for every problem instance from E^*.

Let $e := (\alpha^1, \ldots, \alpha^N, b), \tilde{e} := (\tilde{\alpha}^1, \ldots, \tilde{\alpha}^N, \tilde{b}) \in E^*$ be such that $y := (y^1, \ldots, y^N) \in F(e) \cap F(\tilde{e}) \cap (\bigcap_{j=1}^{N+1} F(\tilde{e} \otimes_j e))$. Then by the optimality condition for problem $\mathbf{P}(e)$,

$$\frac{\alpha^1}{\sqrt{y^1}} = \frac{\alpha^2}{\sqrt{y^2}} = \cdots = \frac{\alpha^N}{\sqrt{y^N}},$$

$$\sum_{j=1}^{N} y^j = b.$$

Since $y \in F(\tilde{e} \otimes_{N+1} e)$, y is optimal for $\mathbf{P}(\alpha^1, \ldots, \alpha^N, \tilde{b})$. Thus $\tilde{b} = \sum_{j=1}^{N} y^j = b$. For $j = 1$, since $y \in F(\tilde{e} \otimes_1 e)$, $\tilde{\alpha}^1/\sqrt{y^1} = \alpha^2/\sqrt{y^2} = \alpha^1/\sqrt{y^1}$. Thus $\tilde{\alpha}^1 = \alpha^1$. Similar arguments show that $\tilde{\alpha}^j = \alpha^j$ for $j = 2, 3, \ldots, N$. Thus $\tilde{e} = e$ as desired. □

Now for $K \geq 2$, consider the profit function of the form

$$u^j(y_1^j, \ldots, y_K^j) = \sum_{i=1}^{K} 2\alpha_i^j \sqrt{y_i^j},\ \text{where } \alpha_i^j > 0,\ i = 1, 2, \ldots, K.$$

We can choose E^* as follows

$$\begin{aligned} E^{*j} &:= \{(\alpha_1^j, \ldots, \alpha_K^j) \in R_{++}^K\},\ j = 1, 2, \ldots, N, \\ E^{*N+1} &:= R_{++}^K, \\ E^* &:= E^{*1} \times \cdots \times E^{*N} \times E^{*N+1}. \end{aligned}$$

Since the problem can be decomposed into K-subproblems of the type involving a single resource that we just analyzed, we can see that E^* has the uniqueness property.

2.1.4 Regularity condition

We turn to the discussion of regularity conditions briefly mentioned in the previous subsection.

It is possible to 'smuggle' two or more variables by encoding them in the value of one variable, for example, by the use of a space-filling curve. When such encoding is allowed, the notion of the size of a message space or information-carrying capacity becomes ambiguous. A regularity condition on equilibrium correspondence is needed to avoid the possibility of smuggling of information.

Let E^* be a subset (in some real vector space) having the uniqueness property. One way of ruling out smuggling is to force a mechanism to have an equilibrium correspondence μ satisfying the following conditions:

1. there is an open set $W \subseteq E^*$, such that there is a continuous selection $m : W \rightarrow \mu(E^*)$, i.e., there is a continuous function m such that $m(e) \in \mu(e), \forall e \in W$,
2. μ^{-1} is continuous on $\mu(W)$.

Then, since μ^{-1} is single-valued on $\mu(W) \subseteq \mu(E^*)$, $m^{-1} = \mu^{-1}$ on $\mu(W)$, and it follows that W and $\mu(W)$ are homeomorphic. In real vector spaces, this implies

$$\dim E^* = \dim W = \dim \mu(W) \leq \dim \mu(E^*) \leq \dim \mathcal{M},$$

and we will have the desired inequality for the size of the message space. The first requirement on μ is known as spot-threadedness. The precise definition will be given shortly.

In order to state the minimality results with sufficient rigor, the relevant definitions, a lemma, and a proposition are cited below from Hurwicz[10].

The idea of using homeomorphism to compare the 'size' of spaces can be extended to the comparison of the size of more general topological spaces. Since generalization is fairly straightforward in many cases that we will study, the relevant definitions are included.

Fréchet size: A topological space X is said to have *Fréchet size* at least as great as a topological space Y (written $X \geq^F Y$) if and only if there exists some subspace W of X such that W is homeomorphic to Y, i.e., Y can be 'embedded homeomorphically' in X.

Note that $\geq^F$ has the monotonicity property with respect to subspaces.

Similarity property: A topological space X has the *similarity property* if and only if every open set W in X has a subset W' which, in the relative topology, is homeomorphic to X.

Spot-threadedness: A correspondence $\Phi : X \twoheadrightarrow Y$ between two topological spaces is *spot-threaded* (with W as a *spot-domain*) if and only if there is an open set W in X and a continuous function (*spot selection*) $\phi : W \to Y$ such that $\phi(x) \in \Phi(x)$ for all x in W.

Lemma 2.1.3 *Let X and Y be topological spaces, with X having the similarity property. Let $\Phi : X \twoheadrightarrow Y$ be a spot-threaded injective correspondence (i.e., Φ^{-1} is single-valued) with a spot-domain W. Then*

$$Y \geq^F X$$

if either of the following two conditions is satisfied:

1. *both X and Y are Hausdorff and X is locally compact; or*
2. *the inverse function $\Phi^{-1} : \Phi(X) \to X$ is continuous on $\Phi(W)$.*

Refer to Hurwicz[10] for the proof of the lemma. We just remark that the first condition is used to imply the second by appealing to the fact that a one-to-one, onto, and continuous function from a compact space to a Hausdorff space is a homeomorphism (see for example, Armstrong[1], Theorem 3.7).

In the context of mechanism theory, E^* plays the role of X, μ that of Φ, and $\mathcal{M}$ that of Y in the lemma above. The following proposition is immediate from the lemma.

Proposition 2.1.1 *Let E be a topological space of environments and $\langle \mathcal{M}, \mu, h \rangle$ a mechanism on E. Let E^* be a subspace of E having the similarity property and let the restriction of μ to E^* be injective and spot-threaded (with a spot-domain $W \subseteq E^*$). Then*

$$\mathcal{M} \geq^F E^*$$

if either of the following two conditions is satisfied:

1. *both $\mathcal{M}$ and E^* are Hausdorff, and E^* is locally compact; or*
2. *μ^{-1} is continuous on $\mu(W)$.*

If the subspace E^* is homeomorphic to the real vector space of finite dimension n, the above conclusion becomes $\mathcal{M} \geq^F R^n$.

Note that if E^* has the uniqueness property with respect to a goal F and the mechanism is goal-realizing, the injectiveness of μ is automatic and the proposition gives the desired size inequality. When $\mathcal{M}$ is a subset of a real vector space, $\mathcal{M} \geq^F R^n$ means $\dim \mathcal{M} \geq n$.

We refer to the condition on μ in Proposition 2.1.1 as the regularity condition in this thesis. Only regular mechanisms are considered throughout this thesis and they are referred simply as mechanisms.

Other ways of comparing size of the topological spaces and associated regularity conditions appeared in the literature. Walker[28] clarifies the relationships among them.

Example: We impose the regularity condition introduced above on mechanisms, and we will see that the minimum size of a message space of a goal-realizing mechanism is $K(N+1)$ for our example in §2.1.1.

As before we examine the case with $K = 1$. The extension to the cases with larger K is straightforward.

First, we introduce a *price mechanism* in equilibrium correspondence form. The price mechanism is goal-realizing and has a message space of the target dimension, $N + 1$.

– *Price mechanism,* $\langle \mathcal{M}, \mu, h \rangle$:

A message is a pair comprising a price vector of the resource $p \in R_+$ and an allocation vector $y := (y^1, \ldots, y^N) \in R_+^N$. Let Λ be the set of price vectors, and $\mathcal{A}$ the set of allocations (i.e., the action space). Then the message space is

$$\mathcal{M} := \Lambda \times \mathcal{A} \subseteq R^{N+1}.$$

This is an example of the product form message space we examine in the next section. We endow $\mathcal{M}$ with the usual topology of R^{N+1}.

The equilibrium correspondence of each division is designed to check the optimality condition of problem **P**, i.e.,

$$\begin{aligned} \mu^j(u^j(\cdot)) &:= \{(p,y) \in \mathcal{M} | \frac{du^j}{dy^j}(y^j) - p \leq 0, \text{ and } y^j(\frac{du^j}{dy^j}(y^j) - p) = 0\}, \ j = 1, \ldots, N, \\ \mu^{N+1}(b) &:= \{(p,y) \in \mathcal{M} | \sum_{j=1}^{N} y^j \leq b, \text{ and } p(b - \sum_{j=1}^{N} y^j) = 0\}. \end{aligned}$$

Roughly speaking, division $j(\neq N+1)$ maximizes its own profit at the given price p, and checks its own resource utilization. The resource management division checks the resource utilization of the system.

The outcome function projects a message on the space of allocation, i.e.,

$$h(p,y) := y.$$

Defined as above, it is clear that the price mechanism realizes the goal.

Let E^* be as defined in §2.1.3. A topology on this set has to be specified. We endow each of $E^{*j} = R_{++}$ with the usual topology, and E^* with the product topology.

Restricted to E^*, $(p,y) \in \mu(e)$ implies that (p,y) solves

$$\begin{aligned} g^j((p,y),\alpha^j) &:= \frac{\alpha^j}{\sqrt{y^j}} - p = 0, \; j = 1,\ldots,N, \\ g^{N+1}((p,y),b) &:= b - \sum_{j=1}^{N} y^j = 0. \end{aligned}$$

It is easily seen that the pair of the optimal solution and the associated Lagrange multiplier for problem $\mathrm{P}(\alpha^1,\ldots,\alpha^N,b)$ is the unique solution of $g^j((p,y),e^j) = 0, j = 1,\ldots,N+1$. Therefore μ is a singleton, and thus a function on E^*. Moreover, it can be shown that $\mu(\cdot)$ is continuous on some open subset W of E^*. (Apply the Implicit Function Theorem to the system of equations $g((p,y),e) = 0$. In §3.1.3, a detailed proof for a similar problem will be given.) Hence, μ restricted to E^* is spot-threaded. By arguing in a similar manner for a general K and appealing to Proposition 2.1.1, we have shown

Theorem 2.1.1 *Let E be a topological space of environments containing E^* as a subspace in its relative topology. Let $\langle \mathcal{M},\mu,h\rangle$ be any mechanism realizing goal on E such that μ is spot-threaded on E^* and $\mathcal{M}$ is Hausdorff, then $\mathcal{M} \geq^F R^{K(N+1)}$.*

Remarks:

1. Since we are only interested in message spaces in real vector spaces, we suppress topologies associated with them in the rest of this thesis. Likewise topologies on 'test classes', E^*'s, will be understood.

The results about lower bounds on message space size will be phrased loosely as "A goal-realizing mechanism has a message space of dimension at least $K(N+1)$."

2. Ideally, we want the topology on E^* to be the relative topology inherited from E or its equivalent. For our example in this section, we could take the E^{N+1} to be some compact subset of R^N, say $[0,B]^K$, where B is a sufficiently large positive number representing a conceivable upper bound for the available amount of any of K resources, and could take $E^j, j = 1, \ldots, N$ to be a set of the profit functions defined on that compact set. If we would endow E^j with the uniform topology, i.e., $\|u^j\| := \sup_{e^{N+1} \in E^{N+1}} u^j(e^{N+1})$, then the relative topology on E^* would be equivalent to that in the example above.

2.2 Message Spaces of Product Form

In the previous section, we saw an example of a price mechanism. Since price mechanisms play a central role in subsequent chapters, it is worthwhile to study the form of message spaces which best accommodates price mechanisms.

We may break down the communication process of the price mechanism of the previous section into the following two steps:

1. The resource management division sets prices for the resources, and transmits them to the other divisions.

2. Given the price, each division finds consumption levels of the resources which maximize its own profit, and transmits them to the resource management division.

In this story, a common K-dimensional price vector is sent to each division in the step 1, and then N different K-dimensional allocation vectors are sent back to the resource management division. In this process, a channel capacity large enough to carry K-dimensional vectors is needed between the resource division and each of the other N divisions. It would be nice if we could single out this type of communication requirement (dimension K in this example) from the capacity requirement we discussed in the previous section—the total capacity requirement for the system (dimension $K(N+1)$).

One motivation to study this type of capacity requirement is its potential role in the design of algorithms. In the two steps above, when the equilibrium message is exchanged, this communication process terminates without any iterations. Nonetheless this form suggests an algorithm or a tâtonnement process which takes the system to the equilibrium messages starting with an arbitrary initial message; namely, as step 3, the resource division checks the 'supply-demand balance' and adjusts the price, and the system repeats the three steps until equilibrium is reached. Throughout this iterative process, K-dimensional messages will be communicated at each message exchange.

In §2.2.1, the notion of message spaces of level sets is introduced in the attempt to formalize the capacity requirement mentioned above.

In §2.2.2, the idea developed in §2.2.1 is applied to the pure exchange economy setting.

2.2.1 The message space of a level set

Let E, F, and $\mathcal{A}$ be a set of environments, a system goal, and an action space respectively. We assume $F(e) \neq \emptyset$ for all $e \in E$, and $\mathcal{A} = F(E)$, i.e., every environment has an optimal action and every action in the action space is optimal for some environment.

Level set: A *level set of an action $a \in \mathcal{A}$ under the goal F* is the inverse image of a and denoted by $F^{-1}(a)$, i.e.,

$$F^{-1}(a) := \{e \in E | a \in F(e)\}, \ a \in \mathcal{A}.$$

Let the collection of the level sets be denoted by E/F, i.e., $E/F := \{F^{-1}(a) | a \in \mathcal{A}\}$. Clearly, $E = \bigcup_{a \in \mathcal{A}} F^{-1}(a)$. Note that when F is a function (rather than a correspondence), E/F is a partition of E. With this notation, the goal may be restated as:

> Given an environment e, find its correct level set $F^{-1}(a) \in E/F$, i.e., such that $e \in F^{-1}(a)$.

A centralized goal-realizing procedure would find the correct level set $F^{-1}(a)$ and take the action a. If a decentralized goal-realizing mechanism could assign one message to each level set just as a centralized procedure does, it would be, naively speaking, most efficient. For such an (imaginary) mechanism, the message space size would be that of $F(E)$ (or E/F, with the proper topology). But such efficiency is highly unlikely to be attainable.

We may ask the following questions:

1. Given a level set $F^{-1}(a)$ and a goal-realizing (decentralized) mechanism $\langle \mathcal{M}, \mu, h \rangle$, how many different messages are needed on this subset of E for the mechanism to realize the goal?

2. Is there a 'minimum' message space required for the task above regardless of what mechanisms we use?

These are the same kind of questions which prompted the search for a lower bound on the message space size, only this time we are looking at a particular level set rather than the entire set of environments. We can proceed to determine a lower bound in much the same way. Let $\langle \mathcal{M}, \mu, h \rangle$ be a mechanism.

Message space of a level set and its size: The *message space of level set* of action a is defined as the image of level set of a under an equilibrium correspondence μ, and is denoted by $\mathcal{M}_a$, i.e.,

$$\mathcal{M}_a := \mu(F^{-1}(a)), \ a \in \mathcal{A}.$$

Its size is defined as the dimension of the smallest real vector space in which there is an open set W such that $\mathcal{M}_a \subseteq W$. It is denoted by $\dim \mathcal{M}_a$.

Informationally maximal level set: A level set whose message space has the largest dimension among $\mathcal{M}_a$ is called an *informationally maximal level set* under the mechanism $\langle \mathcal{M}, \mu, h \rangle$. It will simply be called a maximal level set.

The following corollary of Lemma 2.1.1 is of frequent use later.

Corollary 2.2.1 *Let $\langle \mathcal{M}, \mu, h \rangle$ be a privacy-preserving mechanism realizing F over E. Fix a in $\mathcal{A}$. Let $E^* \subseteq F^{-1}(a)$ be such that*

$$\text{for all } \ e, \tilde{e} \in E^*, \ \text{if } \tilde{e} \otimes_j e \in F^{-1}(a), \forall j, \ \text{then } \ e = \tilde{e}.$$

Then the inverse of μ is single-valued on $\mu(E^*)$.

Proof This is a special case of Lemma 2.1.1. □

For any choice of a or $F^{-1}(a)$, if there is an E^* satisfying the condition of the corollary, then

$$\dim E^* \leq \dim \mu(E^*) \leq \dim \mu(F^{-1}(a)) = \dim \mathcal{M}_a,$$

provided the regularity condition is met. Thus $\dim E^*$ provides a lower bound for the size of message space of a maximal level set. If we can find a mechanism which has a message space of a maximal level set of this size, then the lower bound is tight, and $\dim E^*$ is the minimum dimension required for any goal-realizing mechanism.

Example: We will see that K, the number of the resources involved, is the minimum size of a message space of a maximal level set, and the price mechanism has this dimension for our example in §2.1.1.

Choice of E^* is the key step. As before we restrict the profit functions to be of the form

$$u^j(y_1^j, \ldots, y_K^j) = \sum_{i=1}^{K} 2\alpha_i^j \sqrt{y_i^j}, \text{ where } \alpha_i^j > 0,\ i = 1, \ldots, K.$$

Again $\alpha^j := (\alpha_1^j, \ldots, \alpha_K^j)$ is understood to represent the profit function of division j by abuse of notation.

Let 1_K denote the K-vector whose components are all 1. E^* is defined by

$$E^* := \{(\alpha, \alpha, \ldots, \alpha, N1_K) | \alpha \in R_{++}^K\}.$$

Thus in this subset, every division has the same profit function, and the amount of each resource is fixed at N. E^* has dimension K.

By the concavity of the profit function and the symmetry of the problem, it is clear that allocating resources evenly to each division is optimal. Let this allocation be denoted by 1_{KN}. With this notation, $E^* \subseteq F^{-1}(1_{KN})$.

Lemma 2.2.1 *As defined above, E^* has the uniqueness property.*

Proof Let $e := (z, z, \ldots, z, N1_K), \tilde{e} := (w, w, \ldots, w, N1_K) \in E^*$ such that $\tilde{e} \otimes_j e \in F^{-1}(1_{KN})$ for $j = 1, \ldots, N$. We need to show $z = w$. Since the problem instances $e, \tilde{e}$, and $\tilde{e} \otimes_j e$ are all separable with respect to resources, it suffices to argue for a fixed resource. We choose resource 1.

Assume $z_1 > w_1$. Consider the problem instance $\tilde{e} \otimes_1 e = (w, z, \cdots, z, N1_K)$. It is clear that allocating resource 1 evenly is *not* optimal, and hence $\tilde{e} \otimes_1 e \notin F^{-1}(1_{KN})$, which is a contradiction. Similarly we can rule out $z_1 < w_1$. Thus $z_1 = w_1$. □

We have established that K is a lower bound on the size of maximal level sets. Let us examine the price mechanism in §2.1. As noted there, it has a product form message space. Since it is goal-realizing, given any allocation y, the message space of the level set of y is included in $R_+^K \times \{y\}$.[5] Thus the price mechanism has a K-dimensional message space of a maximal level set. We have shown

Theorem 2.2.1 *The minimum size of a message space of a maximal revel set is K, i.e., the number of the resources involved.*

[5]Strictly speaking, there is a possibility of an environment belonging to two or more level sets. For such an environment, the outcome function h provides a 'tie-breaking' rule. Depending on a tie-breaking rule which a particular price mechanism employs, the allocation part of a message may not be a but another allocation. Such ambiguity may be removed by incorporating a tie-breaking rule in the goal and thereby making it a function rather than a correspondence.

We proceed to define message spaces of the product form.

Observe that message spaces of different level sets cannot share the same message except possibly a message corresponding to environments in the intersection of these level sets. Thus, when a goal is a function, the collection of message spaces of level sets $\{\mathcal{M}_a | a \in F(E)\}$ defines a partition of $\mu(E)$. Let us go back to the algorithm story. Imagine that an equilibrium message of a mechanism $\langle \mathcal{M}, \mu, h \rangle$ is reached by running an associated algorithm. Given an environment, the algorithm will visit messages from different $\mathcal{M}_a$'s (or participants emit messages from different $\mathcal{M}_a$'s) in an effort to identify the correct equilibrium message and the associated action.

Consider the smallest real vector space in which there is an open set Λ such that $\mathcal{M}_a \subseteq \Lambda$ for all $a \in F(E)$. In our scenario, it is necessary to have enough information-carrying capacity to accommodate Λ. We may just as well install this set Λ as an information-carrying capacity. Then, in effect, the implementation of the mechanism through an algorithm takes a message space of $\Lambda \times F(E)(\supseteq \mu(E))$ or more. The dimension of a message space of a maximal level set under this mechanism captures the dimension of Λ.

We call a message space of the form $\Lambda \times \mathcal{A}$ as a *message space of the product form.* If Λ has the minimum size of a message space of a maximal level set, then a goal-realizing mechanism with a message space $\Lambda \times F(E)$ will have the minimum message space size among mechanisms with the product form message spaces.

We note that the price mechanism in our example has the minimum message space not only among mechanisms with the product form message spaces but also among any goal-realizing mechanisms. It appears that this is not in general the case. We suspect that the minimality of a product form mechanism hinges upon what we call *agentwise separability* in Chapter 3.

The use of the product form $\Lambda \times F(E)$ or $\Lambda \times \mathcal{A}$ has an intuitive appeal for the problems we will face in this thesis. Λ has an interpretation as a space of 'price' and, $\mathcal{A}$ as a space of 'consumer demands' or a commodity space in terms of the market economy. Or they can be interpreted as a space of 'dual variables' or 'Lagrange multipliers', and a space of 'primal variables' in terms of optimization theory.

The notion of size of a message space of a maximal level set proves to be useful when $\mathcal{A}$ is a discrete set as will be the case in Chapter 5. In such instances, we cannot speak of dimension of $\mathcal{A}$, yet we can speak of dimension of Λ.

We apply the technique developed here to the pure exchange economy, where the theory of mechanism initiated, in order to single out the dimension of 'price', and close the chapter.

2.2.2 Minimum size of message space of maximal level set in pure exchange economy

The problem instance is briefly described. Refer, for example, to Varian[26] for more details.

Imagine a system consists of N consumers, each of whom holds some initial bundle of $K+1$ commodities. The preference of each consumer is represented by the *utility function* over the commodity bundles. The goal of the system is to find a 'trade' which is 'individually rational' and 'Pareto optimal'. The meaning of these terms will be made precise shortly.

An initial endowment of commodities 1 through K of consumer j is denoted by $x^j := (x_1^j, x_2^j, \ldots, x_K^j) \in R_+^K$, and that of commodity $K+1$ by $y^j \in R_+$. The latter acts as the 'numeraire commodity'. A utility function of consumer j is denoted by $u^j : R_+^K \times R_+ \to R_+$. Utility functions of consumers are assumed to be concave and differentiable. The set of local environments of consumer j is defined by

$$E^j := \{e^j := ((x^j, y^j), u^j)\},$$

and the set of system environments $E := E^1 \times E^2 \times \cdots \times E^N$.

Imagine the consumers are exchanging commodities among themselves. A *trade* is the net gain in amounts of commodities after the exchanges. A trade of agent j is denoted by $(\Delta x^j, \Delta y^j)$. When agent j has a net increase in the amount of commodity k after the exchange, we take $\Delta x_k^j > 0$, and in case of a net decrease $\Delta x_k^j < 0$. Similarly for the numeraire commodity. Given an initial endowment $\{(x^j, y^j), j = 1, 2, \ldots, N\}$, $-x^j \leq \Delta x^j \leq \sum_{n \neq j} x^n$ and $y^j \leq \Delta y^j \leq \sum_{n \neq j} y^n$. The entire trade for the system is denoted by $(\Delta x, \Delta y)$ for brevity. The action space $\mathcal{A}$ is the set of trades.

The goal of the system F is to find a trade $(\Delta x, \Delta y)$ which satisfies the following two conditions:

1. *Individual rationality:* every consumer is at least as well-off after the exchange as before the exchange, i.e.,

$$u^j(x^j + \Delta x^j, y^j + \Delta y^j) \geq u^j(x^j, y^j), \; j = 1, 2 \ldots, N.$$

2. *Pareto optimality:* there is no trade which makes at least one consumer strictly well-off than under $(\Delta x, \Delta y)$ without making at least one consumer strictly worse off, i.e., there exists no $(\hat{\Delta x}, \hat{\Delta y}) \neq (\Delta x, \Delta y)$ such that

$$u^j(x^j + \hat{\Delta x}^j, y^j + \hat{\Delta y}^j) \geq u^j(x^j + \Delta x^j, y^j + \Delta y^j), \; j = 1, 2, \ldots, N,$$

with at least one strict inequality.

It is well-known that the minimum dimension of a message space of a goal-realizing mechanism for this goal[6] is KN. We now proceed to show that K, the number of commodities not counting the numeraire commodity, gives a lower bound for size of a message space of a maximal level set.

Consider the following subset E^* in E:

Every agent has the same initial endowment and utility function. Moreover, the initial endowment of the numeraire commodity is fixed at $y_0 > 0$, and the utility function is given by

$$u^j(x^j, y^j) := \sum_{k=1}^{K} \sqrt{x_k^j} + y^j.$$

Thus the only variable characteristics are the levels of (common) initial endowments of commodities 1 through K, and E^* has dimension K.

Let 'no trade' be denoted by $(0,0)$. Clearly, no trade realizes the goal on E^*, i.e., $E^* \subseteq F^{-1}(0,0)$.

Lemma 2.2.2 *E^* has the uniqueness property with respect to the goal F.*

Proof We abuse notation to denote $e \in E^*$ by $(z, z, \ldots, z)$, where $z \in R_+^L$ represents the levels of common initial endowments of commodities 1 through K.

Let $e := (z, z, \ldots, z), \tilde{e} := (w, w, \ldots, w) \in E^*$ be such that $\tilde{e} \otimes_j e \in F^{-1}(0,0)$,for $j = 1, 2, \ldots, N$. We need to show that $w = z$.

First we assume $w_k > z_k$ for some k, and show that it leads to $\tilde{e} \otimes_1 e = (w, z, \ldots, z) \notin F^{-1}(0,0)$, which is a contradiction.

Consider the following trade:

From agent 1 to agent 2: amount $\Delta x_k := \frac{1}{2}(w_k - z_k)$ of commodity k.
From agent 2 to agent 1: amount $\Delta y := \sqrt{\frac{1}{2}(w_k + z_k)} - \sqrt{z_k}$ of the numeraire commodity, i.e., the payment for commodity k received.
No other net exchanges.

[6]Strictly speaking, the minimality results are obtained for the *interior-valued* goal correspondence.

We assume y_0 is large enough so that the full payment can be made. Agent 2's utility is unchanged after the trade, since Δy is chosen to offset the agent 2's utility gain from the increase in commodity k, i.e.,

$$\begin{aligned} u^2(z_1,\ldots,z_{k-1},z_k+\Delta x_k,z_{k+1},\ldots,z_L,y_o-\Delta y) - u^2(z_1,z_2,\ldots,z_L,y_o) \\ &= \sqrt{\frac{1}{2}(w_k+z_k)}+y_0-\Delta y-(\sqrt{z_k}+y_o) \\ &= 0. \end{aligned}$$

But agent 1's utility increases, because of the strict concavity of the utility function with respect to commodity k. To see that

$$\begin{aligned} u^1(w_1,\ldots,w_{k-1},w_k-\Delta x_k,w_{k+1},\ldots,w_L,y_o+\Delta y) - u^1(w_1,w_2,\ldots,w_L,y_o) \\ &= \sqrt{\frac{1}{2}(w_k+z_k)}+y_0+\Delta y-\left(\sqrt{w_k}+y_o\right) \\ &= \Delta y-\left(\sqrt{w_k}-\sqrt{\frac{1}{2}(w_k+z_k)}\right) \\ &= \left(\sqrt{\frac{1}{2}(w_k+z_k)}-\sqrt{z_k}\right)-\left(\sqrt{w_k}-\sqrt{\frac{1}{2}(w_k+z_k)}\right) \\ &= 2\sqrt{\frac{1}{2}(w_k+z_k)}-\left(\sqrt{z_k}+\sqrt{w_k}\right) \\ &> 0. \end{aligned}$$

The last inequality is due to the strict concavity of the square root function. But this means 'no trade', $(0,0)$, is not Pareto optimal. Thus, we have shown $w_k \le z_k$.

To show $w_k \ge z_k$, we need only to reverse the direction of the exchange and repeat the argument. □

Clearly, the proof above is applicable to any N greater than or equal to 2. Appealing to Proposition 2.1.1, we have established

Proposition 2.2.1 *A goal-realizing mechanism has a message space of a maximal level set at least as large as dimension K.*

Remark We remark that $F(E) \subseteq \mathcal{A}$ has dimension $K(N-1)$. It is known that for an individually rational and Pareto optimal (interior) trade $(\Delta x, \Delta y)$, there exists a 'price vector' $p := (p_1, p_2, \ldots, p_K)$ such that

$$\frac{\partial u^j/\partial x_k^j(x^j+\Delta x^j, y^j+\Delta y^j)}{\partial u^j/\partial y^j(x^j+\Delta x^j, y^j+\Delta y^j)} = p_k,\ k=1,\ldots,K,$$

and

$$\sum_{k=1}^{K} p_k \Delta x_k^j + \Delta y^j = 0,$$

for all consumers. Combined with $\sum_{j=1}^{N} \Delta x^j = 0$ and $\sum_{j=1}^{N} \Delta y^j = 0$, we can see that only $K(N-1)$ of trades need to be specified. The argument may be made rigorous by appealing to the Implicit Function Theorem.

Chapter 3

A Deterministic Resource Allocation Problem

In this chapter, a deterministic resource allocation problem is formulated as a convex program. The problem involves the optimal allocation of K resources among N participants. The objective function is the aggregate utility of the participants.

Our aim is to find the minimum size of a message space of a goal-realizing mechanism.

In § 3.1, a certain class of convex programming problems, which we call *agentwise separable convex programming* (ASCP), is studied. Many deterministic resource allocation problems belong to this class. The main results are that a message space large enough to accommodate both an allocation and the associated Lagrange multipliers is needed, and that the price mechanism has the minimum size message space.

In § 3.2, linear programming is discussed as a special case of ASCP.

In § 3.3, a nonseparable example is examined. It is shown that a nonseparable problem requires a larger message space than its separable counterpart.

3.1 Agentwise Separable Convex Programming

In §3.1.1 a problem instance is stated.

In § 3.1.2 the optimality conditions of the (centralized) problem are stated, and the price mechanism is described.

In § 3.1.3 it is shown that the price mechanism has the minimum size message space among goal-realizing mechanisms.

In § 3.1.4 the minimum size of a message space of a maximal level set is obtained.

3.1.1 Problem Instance

Imagine once again a company consisting of N divisions which share K resources among them, and a resource management division. The company is to decide 'activity levels' of each division so that its profit is maximized while meeting the resource constraints. The resource division knows the resources available for the company. Each of the other divisions knows the profit and the consumption levels of the resources as functions of its own activity levels. We assume the profit function and consumption functions of each division are independent of the activities of the other divisions. We call this situation *agentwise separable.*

Let us index the resource management division by $N + 1$ and the other divisions by 1 through N. The activity levels of division j is denoted by $x^j \in R^{n^j}$, its profit function by f_0^j, and the consumption function of resource i by f_i^j. The amount of resource i available to the company is b_i. With this notation, the company's problem is:

P:

$$\begin{array}{lll} \max & \sum_{j=1}^{N} f_0^j(x^j) & \\ \text{sub. to} & \sum_{j=1}^{N} f_i^j(x^j) \le b_i, & i = 1, 2, \ldots, K, \\ & x^j \ge 0, & j = 1, 2, \ldots, N. \end{array}$$

Assumption 3.1.1 *We assume*

- *The feasible region is nonempty and has an interior point.*
- *The profit functions are concave and the consumption functions are convex. All are differentiable.*
- *An optimal solution exists.*

The assumption guarantees the existence of an optimal solution which satisfies the Karush-Kuhn-Tucker (KKT) conditions. The KKT conditions will be stated in the next subsection. They suggest the price mechanism as a goal-realizing mechanism.

Local environments, action space, and goal for this problem are now specified. The sets of local environments are defined by

$$\begin{aligned} E^j &:= \{e^j := (f_0^j, f_1^1, \ldots, f_K^j)\},\ j = 1, 2, \ldots, N, \\ E^{N+1} &:= \{e^{N+1} := (b_1, b_2, \ldots, b_K) \in R_+^K\}. \end{aligned}$$

The set of system environments is $E := E^1 \times \cdots \times E^N \times E^{N+1}$. Its generic element is denoted by $e := (e^1, e^2, \ldots, e^{N+1})$.

The action space $\mathcal{A}$ is the set of feasible allocations of the resources (*not the space of feasible activity levels*[1]). The allocation to division j is denoted by $y^j := (y_1^j, y_2^j, \ldots, y_K^j)$, where y_i^j is the amount of resource i allocated to division j.

The goal F is to find an optimal allocation of resources. The set of optimal allocations for a problem instance specified by e is denoted by $F(e)$.

3.1.2 The KKT conditions and the price mechanism

Let λ_i be the Lagrange multiplier to the constraint for resource i, and let $\lambda := (\lambda_1, \lambda_2, \ldots, \lambda_K)$. The Lagrangian for the company's problem **P** is

$$L(x^1, \ldots, x^N, \lambda) := \sum_{j=1}^{N} f_0^j(x^j) - \sum_{i=1}^{K} \lambda_i [\sum_{j=1}^{N} f_i^j(x^j) - b_i].$$

The KKT conditions for problem **P** are:

$$x_n^j \geq 0,\ \frac{\partial L}{\partial x_n^j} \leq 0, \text{ and } x_n^j \frac{\partial L}{\partial x_n^j} = 0,\ j = 1, \ldots, N,\ n = 1, \ldots, n^j, \tag{3.1}$$

i.e.,

$$\begin{aligned} x_n^j \geq 0 \quad &\text{and} \quad \frac{\partial f_0^j}{\partial x_n^j}(x^j) - \sum_{i=1}^{K} \lambda_i \frac{\partial f_i^j}{\partial x_n^j} \leq 0, \\ x_n^j > 0 \quad &\Rightarrow \quad \frac{\partial f_0^j}{\partial x_n^j}(x^j) - \sum_{i=1}^{K} \lambda_i \frac{\partial f_i^j}{\partial x_n^j}(x^j) = 0, \\ \frac{\partial f_0^j}{\partial x_n^j}(x^j) - \sum_{i=1}^{K} \lambda_i \frac{\partial f_i^j}{\partial x_n^j}(x^j) < 0 \quad &\Rightarrow \quad x_n^j = 0, \end{aligned}$$

and

$$\lambda_i \geq 0,\ \frac{\partial L}{\partial \lambda_i} \geq 0, \text{ and } \lambda_i \frac{\partial L}{\partial \lambda_i} = 0,\ i = 1, \ldots, K, \tag{3.2}$$

[1] We can take the space of feasible activity levels as an action space, say $\tilde{\mathcal{A}}$, and the goal as finding optimal activity levels, say $\tilde{F}$. It turns out that $\tilde{F}(E) \subseteq \tilde{\mathcal{A}}$ has the same dimension as $F(E) \subseteq \mathcal{A}$ for ASCP.

i.e.,

$$\lambda_i \geq 0 \quad \text{and} \quad \sum_{j=1}^{N} f_i^j(x^j) \leq b_i,$$

$$\lambda_i > 0 \quad \Rightarrow \quad \sum_{j=1}^{N} f_i^j(x^j) = b_i,$$

$$\sum_{j=1}^{N} f_i^j(x^j) < b_i \quad \Rightarrow \quad \lambda_i = 0.$$

Let us consider a property of an optimal solution. Given any optimal allocation $y := (y^1, \ldots, y^N)$, the agentwise separability implies that division j's problem
$\mathbf{P}^j(y^j)$:

$$\begin{aligned} \max \quad & f_0^j(x^j) \\ \text{sub. to} \quad & f_k^j(x^j) \leq y_k^j, \quad k = 1, 2, \ldots, K, \\ & x^j \geq 0, \end{aligned}$$

is optimized.

Conversely, the KKT conditions for these N-subproblems suggest that given an allocation y, if the optimal solutions of these subproblems share the common Lagrange multipliers and the complimentary slackness condition for the resource utilization of the system is met, then such an allocation is optimal.

The price mechanism stated below (in equilibrium correspondence form) utilizes this property of the optimal solution.

Price mechanism, $\langle \mathcal{M}, \mu, h \rangle$:
Announced publicly is a pair comprising a 'price vector' $p = (p_1, p_2, \ldots, p_K) \in R_+^K$ and an 'allocation' $y = (y^1, y^2, \ldots, y^N) \in R_+^{KN}$. Thus

$$\mathcal{M} := R_+^K \times R_+^{KN},$$

and it has dimension $K(N+1)$.

The equilibrium correspondence of division $j(\neq N+1)$ is given as follows:
The first step is to find an optimal solution to the profit maximization problem
$\mathbf{P}^j(p)$:

$$\max_{x^j \geq 0} f_0^j(x^j) - \sum_{i=1}^{K} p_i f_i^j(x^j).$$

Note this optimization involves only the local characteristics of division j. Let $x^j(p)$ denote an optimal solution to problem $\mathbf{P}^j(p)$. The equilibrium correspondence of division j is defined by

$$\mu^j(e^j) := \left\{ (p,y) \in \mathcal{M} \middle| \begin{array}{l} \exists x^j(p) \text{ such that} \\ f_i^j(x^j(p)) \le y_i^j \text{ and } p_i[y_i^j - f_i^j(x^j(p))] = 0, \quad i = 1,2,\ldots,K \end{array} \right\} \tag{3.3}$$

It may be interpreted that given price p, division j maximizes its own profit and checks the resource utilization of its own.

The resource division checks the systemwide resource utilization or the complementary slackness condition (of the system). Its equilibrium correspondence is defined by

$$\mu^{N+1} := \{(p,y) \in \mathcal{M} | \sum_{j=1}^{N} y_i^j \le b_i \text{ and } p_i(b_i - \sum_{j=1}^{N} y_i^j) = 0, \; i = 1,2,\ldots,K\}. \tag{3.4}$$

The outcome function is given by

$$h(p,y) := y.$$

Lemma 3.1.1 *The price mechanism realizes the goal F.*

Proof Since $x^j(p)$ is optimal for problem $\mathbf{P}^j(p)$, it satisfies

$$\frac{\partial f_0^j}{\partial x_n^j}(x^j(p)) - \sum_{i=1}^{K} p_i \frac{\partial f_i^j}{\partial x_n^j}(x^j(p)) \le 0, \quad \text{and} \quad x_n^j(p)[\frac{\partial f_0^j}{\partial x_n^j}(x^j(p)) - \sum_{i=1}^{K} p_i \frac{\partial f_i^j}{\partial x_n^j}(x^j(p))] = 0,$$

for $n = 1,2,\ldots,n^j$. Together with (3.3) and (3.4), this implies that p and $(x^1(p),\ldots,x^N(p))$ satisfy the KKT conditions of the system problem. □

3.1.3 Minimum size of a message space

In this subsection, we will find a subset E^* of dimension $K(N+1)$ which has the uniqueness property with respect to the goal. It will be shown that the equilibrium correspondence of the price mechanism is spot-threaded on E^*.

We start from the following subset (which includes E^*) of E:

- Each division has only one activity level to decide, i.e., $n^j = 1$ for $j = 1,2,\ldots,N$. Thus in the remainder of this subsection, x^j is a scalar, and we let $x := (x^1,x^2,\ldots,x^N)$.

- All divisions have the same profit function $f_0^j(z) := 2\sqrt{z}$, for $j = 1, 2, \ldots, N$.
- The consumption functions are linear, i.e., $f_i^j(z) := a_i^j z$ for $i = 1, 2, \ldots, K$, $j = 1, 2, \ldots, N$. We take a_i^j to be nonnegative.

We denote the $(K \times N)$-matrix with entries a_i^j's (coefficient matrix of activity levels) by A, its jth column by A^j, and the resource K-vector by b. By abuse of notation, the local environment of division $j(\neq N+1)$ is represented by A^j. The local environment of division $N+1$ is denoted by b as before. We denote the set of these environments by $\tilde{E}$, i.e.,

$$\tilde{E} := \{(A, b) \in R_+^{KN} \times R_+^K\}.$$

Problem **P** now becomes
P(A, b):

$$\begin{array}{rl} \max & \sum_{j=1}^{N} 2\sqrt{x^j} \\ \text{sub. to} & Ax \leq b \\ & x \geq 0. \end{array}$$

Let y^j be a feasible allocation to division j. Then the optimal activity level for division j under this allocation is $x^j = \min\{y_i^j / a_i^j | a_i^j \neq 0,\ i = 1, 2, \ldots, K\}$ (by Assumption 3.1.1, the optimal value is bounded). Thus finding optimal activity levels is no harder than finding an optimal allocation. Conversely, given optimal activity level x^j, an optimal allocation to j is found as $y^j = A^j x^j$, though an optimal allocation need not be unique when some of the constraints are not binding. If we can find a subset of $\tilde{E}$ on which all the constraints are tight at the optimal solution, then the optimal allocation and the optimal activity levels will have one-to-one correspondence on such a subset.

With this in mind, we study optimal activity levels, rather than optimal allocations. Let $\tilde{F}$ be the goal which asks for any optimal activity levels.

We will find an open set E^* in $\tilde{E}$ which has the uniqueness property with respect to goal $\tilde{F}$. This set is chosen so that both the optimal solution and the associated Lagrange multipliers are *strictly* positive for every problem instance from this set. Then E^* has the uniqueness property with respect to goal F, too.

Let λ be the Lagrange multiplier.

Lemma 3.1.2 *There exists an open set $E^* \subseteq \tilde{E}$ such that the optimal allocation x, and the associated Lagrange multiplier λ are strictly positive for all $(A,b) \in E^*$. The set E^* has the uniqueness property with respect to goal $\tilde{F}$.*

Proof First we assume the existence of E^* with stated properties, and prove that it has the uniqueness property with respect to $\tilde{F}$.

Since the Lagrange multipliers are strictly positive, the constraints are binding, i.e.,

$$\sum_{j=1}^{N} a_i^j x^j = b_i, \ i = 1,2,\ldots,K,$$

for all problem instances $(A,b) \in E^*$.

Let $e := (A^1,\ldots,A^N,b), \tilde{e} := (\tilde{A}^1,\ldots,\tilde{A}^N,\tilde{b}) \in E^*$ be such that there exists x with

$$x \in \tilde{F}(e) \cap \tilde{F}(\tilde{e}) \cap (\bigcap_{j=1}^{N+1} \tilde{F}(\tilde{e} \otimes_j e)).$$

We need to show $\tilde{A} = A$ and $\tilde{b} = b$.

First we will show $\tilde{b} = b$. Since $x \in \tilde{F}(e)$, $Ax = b$. Since $x \in \tilde{F}(\tilde{e} \otimes_{N+1} e) = \tilde{F}(A^1,\ldots,A^N,\tilde{b})$, $Ax = \tilde{b}$. Thus $\tilde{b} = b$.

Similarly for $j \neq N+1$, since $x \in \tilde{F}(e)$ and $x \in \tilde{F}(\tilde{e} \otimes_j e)$, and x is strictly positive,

$$a_i^j = \tilde{a}_i^j = \frac{1}{x^j}(b_i - \sum_{k \neq j} a_i^k x^k), \ i = 1,2,\ldots,K.$$

Thus E^* has the uniqueness property with respect to $\tilde{F}$.

It remains to show the existence of E^*. Let q and r be nonnegative integers such that

$$N = qK + r, \ \ 0 \leq q, \ \ 0 \leq r < K.$$

Let 1_n be the n-dimensional row vector whose components are all 1. Let a $(K \times N)$-matrix A^* be

$$A^* := \begin{pmatrix} 1_{q+1} & & & & & \\ & \ddots & & & 0 & \\ & & 1_{q+1} & & & \\ & & & 1_q & & \\ & 0 & & & \ddots & \\ & & & & & 1_q \end{pmatrix}$$

where rows 1 through r have 1_{q+1} as their entries, and rows $r+1$ through K have 1_q as their entries. Also let

$$b^* := 1_K.$$

We will show that we can take the open set E^* around (A^*, b^*). It is easy to see that (A^*, b^*) satisfies the desired property; namely, both the optimal solution and the associated multipliers are strictly positive for problem $\mathbf{P}(A^*, b^*)$. In fact, the optimal solution and the associated multipliers are:

$$\begin{aligned} x^{*j} &= \tfrac{1}{q+1}, && \text{for } j = 1, 2, \ldots, (q+1)r, \\ x^{*j} &= \tfrac{1}{q}, && \text{for } j = (q+1)r+1, \ldots, N, \\ \lambda_i^* &= \tfrac{1}{\sqrt{q+1}}, && \text{for } i = 1, 2, \ldots, r, \\ \lambda_i^* &= \tfrac{1}{\sqrt{q}}, && \text{for } i = r+1, \ldots, K. \end{aligned}$$

They are unique.

We utilize the KKT conditions to complete the proof. Let $l : R^{KN} \times R^K \times R^N \times R^K \rightarrow R^N \times R^K$ be defined by

$$\begin{aligned} l_j(A, b, x, \lambda) &:= \sum_{k=1}^{K} \lambda_k a_k^j - \frac{1}{\sqrt{x^j}}, \; j = 1, 2, \ldots, N, \\ l_{N+i}(A, b, x, \lambda) &:= \sum_{n=1}^{N} a_i^n x^n - b_i, i = 1, 2, \ldots, K. \end{aligned}$$

The KKT conditions say that the strict positivity of the optimal solution and the associated multipliers imply that l evaluated at the problem instance and the solution is 0.

Provided x is positive,

$$(D_x l, D_\lambda l) = \begin{pmatrix} \frac{1}{2}(x^1)^{-3/2} & & 0 & \\ & \ddots & & A^T \\ 0 & & \frac{1}{2}(x^N)^{-3/2} & \\ & A & & 0 \end{pmatrix}$$

where A^T is the transpose of A. It is easy to see that $(D_x l, D_\lambda l)$ evaluated at $(A^*, b^*, x^*, \lambda^*)$ is nonsingular. Applying the Implicit Function Theorem, we see that there is an open set $\hat{E} \ni (A^*, b^*)$ in $R^{KN} \times R^K$ and a unique differentiable function $G : \hat{E} \rightarrow R^N \times R^K$ such that $l(A, b, G(A, b)) = 0$ for $(A, b) \in \hat{E}$. Since $G(A^*, b^*) = (x^*, \lambda^*) > 0$, we can take an open set $E^* \subseteq \hat{E}$ so that $G(A, b) > 0$ for $(A, b) \in E^*$. Thus we have shown the existence of the desired subset. □

Corollary 3.1.1 *As defined above, E^* has the uniqueness property with respect to goal F.*

Proof Let $x^j(A,b), y^j(A,b)$ be the optimal activity level and optimal allocation respectively at the problem instance $(A,b) \in E^*$. Since constraints are tight, $y^j(A,b) = A^j x^j(A,b)$. □

Note that the price mechanism for goal F is easily modified to obtain a goal-realizing mechanism for $\tilde{F}$. We can replace the outcome function $h(p,y) = p$ by $\tilde{h}(p,y) := (x^1(p), \ldots, x^N(p))$.[2] Thus the following spot-threadedness result is applicable for both version of mechanisms.

Lemma 3.1.3 *The equilibrium correspondence μ is spot-threaded on E^*.*

Proof On E^* the optimal solution and the associated Lagrange multipliers are unique, and with the notation in the proof of the previous lemma, $(x(A,b), \lambda(A,b)) = G(A,b)$. Hence the equilibrium message $m(A,b) = (\lambda(A,b), Ax(A,b))$ is a singleton. Thus continuity follows. □

By appealing to Proposition 2.1.1, we have, for both F and $\tilde{F}$:

Theorem 3.1.1 *The minimum size of a message space of a goal realizing mechanism is $K(N+1)$, and the price mechanism has a message space of the minimum size.*

Remark: It is not hard to directly prove that μ^{-1} restricted to E^* is continuous in this case. At the optimal, we have

$$\begin{aligned} \frac{1}{\sqrt{x^j}} - \sum_{i=1}^{K} \lambda_i A_i^j &= 0 \\ y^j - A^j x^j &= 0 \end{aligned}$$

Thus (x^j, A^j) is a continuous function of (y, λ). Then $b = Ax$ is also a continuous function of (y, λ).

Corollaries of the theorem are stated below.

Corollary 3.1.2 *Consider problem* **P** *with additional 'local' constraints of the form $k_m^j(x^j) \leq 0$, where $k_m^j(\cdot)$'s are convex and differentiable. Under Assumption 3.1.1, the results of the previous theorem hold.*

[2] It may be necessary to specify a tie-breaking rule among the optimizers of problem $\mathbf{P}^j(p)$. However, it can be 'privately' done.

Proof Let $k^j_m(x^j) \le 0, m = 1,\ldots,m^j$, be division j's private constraints. Let the associated Lagrange multipliers be $\mu^j := (\mu^j_1,\ldots,\mu^j_{m^j})$. Then the KKT conditions are

$$\frac{\partial f^j_0}{\partial x^j_n}(x^j) - \sum_{i=1}^{K} \lambda_i \frac{\partial f^j_i}{\partial x^j_n}(x^j) - \sum_{m=1}^{M^j} \mu^j_m \frac{\partial k^j_m}{\partial x^j_n}(x^j) \le 0,$$

$$x^j[\frac{\partial f^j_0}{\partial x^j_n}(x^j) - \sum_{i=1}^{K} \lambda_i \frac{\partial f^j_i}{\partial x^j_n}(x^j) - \sum_{m=1}^{M^j} \mu^j_m \frac{\partial k^j_m}{\partial x^j_n}(x^j)] = 0, \quad j = 1,\ldots,N,\ n = 1,2,\ldots,n^j,$$

$$\lambda_i[\sum_{j=1}^{N} f^j_i(x^j) - b_i] = 0, \qquad i = 1,\ldots,K,$$

$$\mu^j_m[k^j_m(x^j)] = 0, \quad j = 1,\ldots,N,\ m = 1,2,\ldots,m^j,$$

along with the feasibility of activity levels and nonnegativity of the multipliers.

The price mechanism is modified slightly. Division $j(\ne N+1)$ solves a constrained profit maximization problem

P$^j(p)$:

$$\max_{x^j \ge 0} \quad f^j_0(x^j) - \textstyle\sum_{i=1}^{K} p_i f^j_i(x^j)$$
$$\text{sub. to} \qquad k^j_m(x^j) \le 0, \qquad m = 1,2,\ldots,m^j.$$

In the definition (3.3) of the equilibrium correspondence μ^j, we take $x^j(p)$ to be an optimizer of this problem. It is easy to see this modified version of the price mechanism realizes the goal.

We can use the same E^* for the uniqueness argument. □

Corollary 3.1.3 *When the resource availability, b, is public knowledge, the minimum size of a message space of a goal realizing mechanism is KN.*

Proof Division N can take over the resource division's task by modifying the price mechanism slightly.

Announced publicly is $(p, y^1, y^2, \ldots, y^{N-1}) \in R^{KN}$. Division N computes y^N by $y^N = b - \sum_{j=1}^{N-1} y^j$. It carries out the profit maximization with price p, and compare the resulting optimal consumption levels of the resources with this y^N. It also checks the systemwide resource utilization previously done by the resource division. The other divisions respond as before. This mechanism realizes the goal.

The subset E^* with b fixed, say at b^*, in the proof of Lemma 3.1.2 has the uniqueness property with respect to the goal. □

Corollary 3.1.4 *When some or all of the inequality constraints are replaced by equality constraints in problem P, the minimum size of a message space remains the same, provided that there exists an optimal solution which satisfies the KKT conditions. Also the nonnegativity constraints on activity levels can be removed without affecting the minimum size.*

Proof The price mechanism can be modified to accommodate the KKT conditions with the equality constraints without increasing the size of messages announced. The subset E^* has the uniqueness property with equality constraints as well, since E^* is so chosen to begin with. A similar argument holds for the case of unrestricted activity levels. □

3.1.4 Minimum size of a message space of a maximal level set

The price mechanism of § 3.1.2 has the product form $\Lambda \times \mathcal{A}$, with Λ as the space of price vectors. One may suspect that this Λ has the minimum size of a message space of a maximal level set. We verify it here.

We consider the same subset of environments $\tilde{E} = \{A, b\}$ as in § 3.1.3. We take our goal $\tilde{F}$ to be finding optimal activity levels.

Let a be a K-dimensional positive column vector. The subset E^* is defined by

$$E^* := \{(a, a, \ldots, a, Na) | a \in R^K_{++}\},$$

namely, every division has the same activity coefficients, and the amount of resource available is N times that coefficients. E^* has dimension K. Because the profit functions are concave and divisions are identical, allocating resources evenly among divisions 1 through N is optimal. In other words, optimal activity levels are $x^j = 1$ for all j. We denote this activity level by 1_N. With this notation $E^* \subseteq F^{-1}(1_N)$.

Lemma 3.1.4 *As defined above, E^* has the uniqueness property with respect to goal $\tilde{F}$.*

Proof Let $e := (z, z, \ldots, z, Nz), \tilde{e} := (w, w, \ldots, w, Nw) \in E^*$ be such that

$$\tilde{e} \otimes_j e \in \tilde{F}^{-1}(1_N), \; j = 1, 2, \ldots, N + 1.$$

We will show $w = z$.

Since $\tilde{e} \otimes_{N+1} e = (z, z, \ldots, z, Nw) \in F^{-1}(1_N)$, the feasibility of 1_N implies $Nz \leq Nw$ and hence $z \leq w$ (i.e, $z_i \leq w_i,\ i = 1, 2, \ldots, K$). Similary, $\tilde{e} \otimes_1 e = (w, z, \ldots, z, Nz) \in F^{-1}(1_N)$ implies $w + (N-1)z \leq Nz$, and hence $w \leq z$. □

Since the price mechanism has K-dimensional message spaces of maximal level sets, it follows

Theorem 3.1.2 *The minimum size of a message space of a maximal level set is K, the number of resources. The message spaces of level sets of the price mechanism has the minimum size.*

3.2 Linear Programming

Linear programming (LP) is a special case of the agentwise separable convex programming. In § 3.1, we found that linearity of the constraints in ASCP does not reduce the minimum size of a message space. However, a linear objective function along with linear constraints does reduce the size. It is because LP has a basic optimal solution.

We state the problem instance, and then obtain the minimum size of message spaces of goal-realizing mechanisms.

3.2.1 Problem instance

The company's problem is restated. In this section, we assume

$$N \geq K,$$

i.e., the number of divisions is at least as large as the number of the resources. We seek optimal activity levels in this section, since that is the more LP customary.

In this section, we take *all vectors as column vectors.* A row vector is denoted as the transpose of a column vector.

Let A^j be $(K \times n^j)$-coefficient matrix of activity levels $x^j \in R^{n^j}$ of division j. The profit function of division j is given by $f_0^j(x^j) := c^{jT} x^j$, where $c^{jT} := (c_1^j, \ldots, c_{n^j}^j)$. With this notation, the company's problem is

LP:

$$\begin{array}{ll} \max & \sum_{j=1}^{N} c^{jT} x^j \\ \text{sub. to} & \sum_{j=1}^{N} A^j x^j \leq b, \\ & x^j \geq 0, \qquad j = 1, 2, \ldots, N. \end{array}$$

We assume that an optimal solution exists.

The sets of local environments are

$$\begin{aligned} E^j &:= \{(c^j, A^j) \in R^{n^j} \times R^{K \times n^j}\},\ j = 1, 2, \ldots, N, \\ E^{N+1} &:= \{b \in R^K\}. \end{aligned}$$

The action space $\mathcal{A}$ is defined as the set of the feasible solutions of problem LP. The goal $\tilde{F}$ is to find optimal activity levels.

3.2.2 Price mechanism

We modify the price mechanism of § 3.1 in order to take advantage of LP. Note that in a basic feasible solution (or at an extreme point of the polyhedron defined by the constraints), at most K divisions receive the resources, and LP has a basic optimal solution. Thus we need only specify K bundles of the resources rather than N bundles needed for ASCP, *provided the recipients of these K bundles are also specified.* We introduce an *allocation indicator* for that purpose. It is defined as an one-to-one function from the set of indices of K resource bundles $\{1, 2, \ldots, K\}$ into the set of the indices of divisions $\{1, 2, \ldots, N\}$. The set of allocation indicators are denoted by Π, and its generic element by π. Note Π is a finite set with $N!/K!(N-K)!$ elements.

Price mechanism:

Announced publicly is a triple consisting of a price vector $p \in R_+^K$, an allocation indicator π, and a collection of K resource bundles $y := (y^{(1)}, y^{(2)}, \ldots, y^{(K)})$. A resource bundle $y^{(k)}$ is allocated to division $\pi(k)$. We do not exclude the possibility that $y^{(k)} = 0$ for some k. Thus the message space is

$$\mathcal{M} := R_+^K \times \{\Pi \times R_+^{K^2}\}.$$

Note $\Pi \times R_+^{K^2} \subseteq \mathcal{A}$, so $\mathcal{M}$ is of the product form. This is the first example of the message space which has a discrete set in its specification. We will see other examples in Chapter 5.

The equilibrium correspondence of division j ($\neq N+1$) is defined in a manner similar to the ASCP case. Let $z \in R_+^K$. The first step is to solve

$\mathrm{LP}^j(z)$:

$$\max\{c^{jT} x^j | A^j x^j \leq z,\ x^j \geq 0\}.$$

Let $x(z)$ be an optimal solution of $\mathrm{LP}^j(z)$. Then, $\mu^j(c^j, A^j)$ is defined as the set of messages $(p, \pi, y) \in \mathcal{M}$ which satisfy

1. if $\pi^{-1}(j) = \emptyset$, then $y^j = 0$ and $c^{jT} - p^T A^j \leq 0$, and

2. if $\pi^{-1}(j) \neq \emptyset$, then there exists $x^j(y^{(\pi^{-1}(j))})$ such that

   (a) $c^{jT} - p^T A^j \leq 0$ and $x^j(y^{(\pi^{-1}(j))})[c^{jT} - p^T A^j] = 0$, and

   (b) $A^j x^j(y^{(\pi^{-1}(j))}) \leq y^{(\pi^{-1}(j))}$ and $p^T[y^{(\pi^{-1}(j))} - A^j x^j(y^{(\pi^{-1}(j))})] = 0$.

The equilibrium correspondence of the resource division is defined by

$$\mu^{N+1}(b) := \{(p, \pi, y) \in \mathcal{M} | \sum_{k=1}^{K} y^{(k)} \leq b \text{ and } p^T[b - \sum_{k=1}^{K} y^{(k)}] = 0\}.$$

The outcome function $\tilde{h}$ is defined by

$$\begin{aligned} x^j(p, \pi, y) &:= x^j(y^{(\pi^{-1}(j))}) \text{ for } j \in \pi(\{1, \ldots, K\}), \\ x^j(p, \pi, y) &:= 0, \text{ otherwise.} \end{aligned}$$

Lemma 3.2.1 *The price mechanism realizes goal $\tilde{F}$.*

Proof The allocation $\tilde{h}(p, \pi, y)$ is a feasible solution of **LP**. Also p is a feasible solution of its dual. Together, they satisfy the complementary slackness conditions. Thus, $\tilde{h}(p, \pi, y)$ is an optimal solution. □

Proposition 3.2.1 *The size of the message space of a goal-realizing mechanism is at least as large as $K(K+1)$.*

Proof We will show that there exists E^* of dimension $K(K+1)$ and which has the uniqueness property with respect to goal $\tilde{F}$. As before, we want both the optimal activity levels and the associated multipliers to be strictly positive for problem instances from E^*. The same argument as in the proof of Lemma 3.1.2 leads to the uniqueness property.

We consider the case $N = K$.

As in the proof of Lemma 3.1.2, we consider the case when each division has only one activity levels to decide. Thus A^j is a column vector. Furthermore we take $c = b$. Therefore, this subset of environments is specified by $(A^1, \ldots, A^K, b)$ and has dimension $K(K+1)$. When A is nonsingular, it is clear that both optimal solution and the associated multipliers are given by $A^{-1}b$. Now take A to be the $(K \times K)$-identity matrix I_K and $b = 1_K$. Then, the optimal solution is $x = 1_K$, which is strictly positive. We can take an open set E^* in a neighborhood of $(I_K, 1_k)$ so that A is nonsingular and the optimal solution for problem instance (A, b) is strictly positive for every (A, b) in E^*. □

Corollaries similar to those of Theorem 3.1.1 can be derived.

We state the minimum size of a message space of a maximal level set.

Proposition 3.2.2 *The minimum size of a message space of a maximal level set of a goal-realizing mechanism is K, the number of the resources.*

Proof The proof of Lemma 3.1.4 is applicable for the uniqueness argument. The price mechanism has a message space of a maximal level set of dimension K. □

3.3 Nonseparable Example

When a convex program is not agentwise separable, the minimum size of a message space required to realize the goal (finding an optimal allocation or optimal activity levels) is in general larger than its separable counterpart—a problem with the same number of resources and the same number of participants. In this section, we show an example of such a convex program.

We consider a resource allocation problem that involves one resource, its management division, and two divisions sharing the resource. The instance is similar to the example in Chapter 2 except two divisions incur a nonseparable joint cost between them. We consider a specific joint cost function of the form

$$k(x^1, x^2) := \frac{1}{2}(k^1(x^1) + k^2(x^2))^2.$$

We assume that k^j's are nonnegative, convex, and twice differentiable. We also assume that the profit function is concave and differentiable as before. The company's problem is
P:

$$\begin{array}{ll} \max & f^1(x^1) + f^2(x^2) - \frac{1}{2}(k^1(x^1) + k^2(x^2))^2 \\ \text{sub. to} & x^1 + x^2 \le b \\ & x^1 \ge 0, \ x^2 \ge 0. \end{array}$$

It is easy to see that the objective function is concave.

Sets of local environments are

$$E^1 := \{f^1, k^1\},\ E^2 := \{f^2, k^2\},\ \text{and } E^3 := \{b \in R_+\}.$$

The action space $\mathcal{A}$ is the set of feasible allocation, and the goal F is to find an optimal allocation.

Let λ be the Lagrange multiplier to the resource constraint. The KKT conditions include

$$x^1[\frac{df^1}{dx^1}(x^1) - \frac{dk^1}{dx^1}(x^1)\{k^1(x^1) + k^2(x^2)\} - \lambda] = 0, \tag{3.5}$$

$$x^2[\frac{df^2}{dx^2}(x^2) - \frac{dk^2}{dx^2}(x^2)\{k^1(x^1) + k^2(x^2)\} - \lambda] = 0, \tag{3.6}$$

$$\lambda[x^1 + x^2 - b] = 0. \tag{3.7}$$

The price mechanism described in Chapter 2 does not work, because the first condition includes the term $k^2(x^2)$ which is unknown to division 1, and also the second condition includes the term $k^1(x^1)$ which is unknown to division 2.

The KKT conditions suggest that we need at least a 4-dimensional message apace, which accommodates x^1, x^2, λ, and a part of marginal 'social cost' $k^1(x^1) + k^2(x^2)$, while it is sufficient to have a 3-dimensional message space for its ASCP counterpart.

We proceed to show that dimension 4 is indeed a lower bound, by choosing a subset E^* of that dimension which has the uniqueness property.

In the rest of this section, we use a subscript to index the divisions in order to avoid a confusion between them and the exponents for power.

Profit functions are restricted to be of the form

$$f_j(x_j) = 2z_j\sqrt{x_j},\ j = 1, 2,$$

and contributions to the joint cost to be of the form

$$k_j(x_j) = w_j x_j,\ j = 1, 2,$$

where z_j, w_j are strictly positive. Hence the subsets of these instances are

$$\tilde{E}_1 = \tilde{E}_2 = R^2_{++},\ \text{ and } \tilde{E}_3 = R_{++}.$$

The Lagrangian for the problem is

$$L(x_1, x_2, \lambda) := 2z_1\sqrt{x_1} + 2z_2\sqrt{x_2} - \frac{1}{2}(w_1x_1 + w_2x_2)^2 - \lambda(x_1 + x_2 - s)$$

Note that a profit function has a derivative which tends to infinity as its argument tends to 0, while a cost function has a derivative which tends to 0 as its argument tends to 0. Thus x_1 and x_2 is strictly positive at the optimal allocation. By the KKT conditions (3.5) and (3.6), we have

$$\lambda = \frac{z_1}{\sqrt{x_1}} - w_1(w_1x_1 + w_2x_2)$$
$$\lambda = \frac{z_2}{\sqrt{x_2}} - w_2(w_1x_1 + w_2x_2)$$

By inspection, we see that for z_j large and w_j small, λ is also strictly positive. We take a subenvironment which yields strictly positive multiplier. We take

$$E^* := \{e = ((z_1, w), (z_2, w), b) | 1 < z_1 < z_2,\ 0 < w < 1,\ 0 < b < \frac{1}{2}\}.$$

Note the coefficients in the cost terms are taken to be equal. This subset has dimension 4.

Proposition 3.3.1 *E^* has the uniqueness property with respect to the goal.*

Proof Let $e := ((z_1, w), (z_2, w), b), \tilde{e} := ((\tilde{z}_1, \tilde{w}), (\tilde{z}_2, \tilde{w}), \tilde{b}) \in E^*$ be such that there exists $x := (x_1, x_2)$ which satisfies

$$x \in F(e) \cap F(\tilde{e}) \cap (\bigcap_{j=1}^{3} F(\tilde{e} \otimes_j e)).$$

We will show that $\tilde{e} = e$. Because of the choice of E^*, all problem instances defined by $e, \tilde{e}, \tilde{e} \otimes_j e$ yield strictly positive multipliers at the optimal.

Since the resource constraint is binding, it is immediate from $x \in F(e)$ and $x \in F(\tilde{e} \otimes_3 e)$ that $\tilde{b} = b$.

From the KKT conditions for the problem instance e, we have

$$x_1 = \frac{sz_1^2}{z_1^2 + z_2^2},\ x_2 = \frac{sz_2^2}{z_1^2 + z_2^2}.$$

Similarly, from the problem instance $\tilde{e}$,

$$x_1 = \frac{s\tilde{z}_1^2}{\tilde{z}_1^2 + \tilde{z}_2^2},\ x_2 = \frac{s\tilde{z}_2^2}{\tilde{z}_1^2 + \tilde{z}_2^2}.$$

Thus we have

$$(\tilde{z}_1, \tilde{z}_2) = \alpha(z_1, z_2), \tag{3.8}$$

The KKT conditions for problem instance $\tilde{e} \otimes_1 e$ is, denoting the multiplier by μ,

$$\mu = \frac{\tilde{z}_1}{\sqrt{x_1}} - \tilde{w}(\tilde{w}x_1 + wx_2) \tag{3.9}$$

$$\mu = \frac{z_2}{\sqrt{x_2}} - w(\tilde{w}x_1 + wx_2) \tag{3.10}$$

from which, we get

$$\mu(w - \tilde{w}) = \frac{w\tilde{z}_1}{\sqrt{x_1}} - \frac{\tilde{w}z_2}{\sqrt{x_2}}.$$

Substituting values of x_1, x_2 and (3.8) into the above, and setting $c := \frac{\sqrt{s}}{\sqrt{z_1^2+z_2^2}}$, we have

$$\mu(w - \tilde{w}) = (wk - \tilde{w})/c. \tag{3.11}$$

The KKT conditions for problem instance $\tilde{e} \otimes_2 e$ is, denoting the multiplier by ν,

$$\nu = \frac{z_1}{\sqrt{x_1}} - w(wx_1 + \tilde{w}x_2) \tag{3.12}$$

$$\nu = \frac{\tilde{z}_2}{\sqrt{x_2}} - \tilde{w}(wx_1 + \tilde{w}x_2) \tag{3.13}$$

from which, we get

$$\nu(w - \tilde{w}) = (wk - \tilde{w})/c. \tag{3.14}$$

By (3.11) and (3.14), we have

$$(\mu - \nu)(w - \tilde{w}) = 0.$$

Suppose $\mu - \nu = 0$. Then by (3.9) and (3.12) with $\tilde{z}_1 = \alpha z_1$,

$$\alpha - 1 = (\tilde{w}^2 - w^2)cx_1.$$

Similarly by (3.10) and (3.13) with $\tilde{z}_2 = \alpha z_2$,

$$\alpha - 1 = (\tilde{w}^2 - w^2)cx_2.$$

Therefore,

$$(\tilde{w}^2 - w^2)(x_1 - x_2) = 0.$$

By our assumption, $x_1 - x_2 = c^2(z_1^2 - z_2^2) \neq 0$. This proves $\tilde{w} = w$.

Lastly, by substituting $\tilde{w} = w$ into (3.11), we have $\alpha = 1$. In view of (3.8), this completes the proof. □

The minimum size of a message space of a maximal level set also increases.

Proposition 3.3.2 *Size of a message space of a maximal level set of a goal-realizing mechanism is at least 2.*

Proof It can be shown that a subset

$$\{((z,w),(\sqrt{2}z,w),\sqrt{3})|z,w\in R_{++}\}\subseteq F^{-1}(1/\sqrt{3},2/\sqrt{3})$$

has the uniqueness property. □

Remark The situation here is similar to a case of 'market failure' under the presence of an externality in welfare economics. The increase in the minimum size of a message space of a maximal level set may be compared to the need for a Pigouvian tax/subsidy, or a Coasian bargaining or a creation of market for a public *bad.*

Under the presence of join costs or externalities, the need of assumptions on participant behavior is more pronounced. Under the agentwise separability, a price mechanism can be thought of as an incentive-compatible scheme. Here, what is meant by 'incentive compatible' is not clear without behavioral assumptions. In this thesis, each participant is assumed to cooperate in order to accomplish a system goal. A participant in ASCP may be understood as a coalition of individuals whose activities affect each other in a nonseparable manner. Within a coalition information is centralized.

Chapter 4

Pricing of Electric Power under Uncertainty

4.1 Introduction

Allocation of resources is often done through various pricing schemes. Prices may be set by 'market force', or by regulatory bodies acting as caretakers of society's welfare. In either case, prices signal the value of resources to the system, and evoke responses from its members.

In terms of mechanism design, we can regard the efficient allocation of the resources as a goal of the system, each participant's valuation of resources as private knowledge, an allocation of resources as an action, and the prices as a message made public for the sake of realizing the goal. An important additional requirement on mechanism design is incentive compatibility.

We have already seen an example of a price mechanism in Chapter 3, though the emphasis there was to show the minimum size of the message space needed to solve convex programming problems. In this chapter, we add an element of uncertainty to a resource allocation problem and explore its effect on decentralized mechanisms, in particular, on price mechanisms and their variants. An allocation problem is posed as a two-stage stochastic recourse model, and as such it has a flavor of sequential decision-making.

Our problem is presented as allocation of electric power, partly because we can reexamine existing work on electric power allocation from the view point of mechanism design. But more importantly, we believe the recourse model provides a convenient framework to analyze and design a pricing scheme for electric power allocation, and we wish to show the need to start from the model.

In § 4.2 some pricing schemes from the literature are briefly reviewed. Among these are Brown and Johnson[3], Crew and Kleindorfer[6], Caramanis *et al.*[4], and Chao and Wilson[5].

In §4.3 the spot pricing scheme and the priority service scheme are described. Consumers are assumed to be 'flexible'—capable of adjusting their consumption levels to the price without experiencing any inconvenience. A nested structure of the priority service is shown.

In § 4.4 the allocation problem is formulated as a two-stage recourse model in order to capture inflexibility of consumers. A price forecast mechanism is described and shown to have a message space of the minimum size among goal-realizing mechanisms with product form message spaces. The implication is that the probability distribution of the underlying stochastic events needs to be made public in one way or other.

In §4.5 the pricing schemes from the literature are applied to our two-stage recourse model and compared.

The results of this chapter are summarized in §4.6.

4.2 Review of Pricing Schemes

When there is no uncertainty in a system, the efficient allocation of electric power will be achieved by equating marginal value of demand for energy to marginal cost of energy supply. When allocation takes place over time, it would be ideal to update price in 'real time' so that the price could signal the value where marginal demand and cost are equated, provided there are available means to communicate pricing information in real time and the means to respond to it automatically. Vickrey[27] argued the effectiveness of this scheme and called it 'responsive price'[1]. (It should be noted that his argument was not limited to a deterministic setting, in fact it was the consideration of uncertainty in the system which

[1] When valuations of power of suppliers and consumers show intertemporal dependency, the notion of spot price is not clear. In order to retain the property that the information it provides is sufficient to obtain an optimal allocation, the spot price must be the Arrow-Debreu equilibrium price. The dimension of (the space of the Arrow-Debreu equilibrium) prices can be enormous; and such a price no longer has the properties implicit in the true 'spot price'.

prompted his argument.) In this scheme, the price acts as an indirect means to control demand levels. The system will incur less frequent blackouts, less fuel consumption, lower level of optimal total capacity, and other benefits. Unfortunately, the infrastructure required for this scheme is not yet generally available, and a price is predetermined in practice.

In a real life system, there will always be stochastic elements: fluctuation of power supply due to generation failures, changes in consumer demand due to weather, and so on. In a stochastic setting, the effectiveness of allocation is often measured by value of expected social welfare it achieves. It is conceivable that the uncertainty complicates the pricing schemes for effective allocation.

Caramanis *et al.*[4] proposed 'optimal spot pricing'[2] which can be thought of as an elaboration of Vickrey's idea. They recognized the need to communicate the probability distribution of future spot prices for an effective allocation. They also proposed a pricing scheme which combines the spot price and the 'predetermined' price. This scheme will also appear in Section 4.5, where it is applied to the two-stage model we study in Section 4.4. For now, we move on to the schemes proposed by Brown and Johnson[3], and Crew and Kleindorfer[6].

The schemes involve a 'future price' and 'rationing'; a price is announced in period 0 (today) for energy to be delivered in period 1 (tomorrow), and excess demand is rationed if demand exceeds supply in period 1.

In their rather controversial yet seminal paper[3], Brown and Johnson study the case of random demand and deterministic supply. A simplified version of their work is described to illustrate the idea. We ignore the long run marginal costs and associated consideration of capital investment. Supply cost is assumed non-random. Demand at time 1 is random. A social planner knows the aggregate demand curve for each contingency and its probability distribution, as well as the supply curve. The planner's task is to set a single price at time 0 to maximize expected social welfare (aggregate utility of consumers minus cost of supply).

Brown and Johnson show that the maximum expected social welfare is achieved by setting the price at the value where expected aggregate demand curve meets the supply curve, *assuming the demands with lower marginal utility levels are rationed* when demand exceeds supply.

[2] Again, this is an equilibrium price system in the Arrow-Debreu sense.

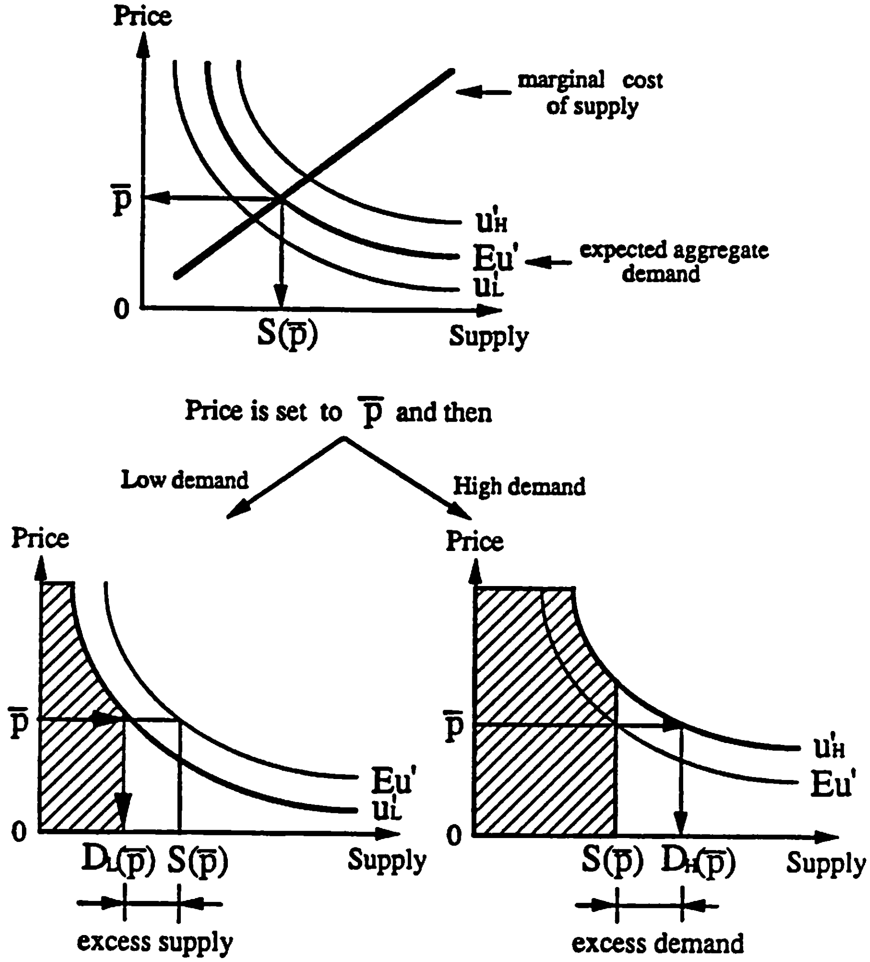

Figure 4.1. Brown & Johnson scheme

Figure 4.1 illustrates the scheme for the case when there are only two contingencies, high and low, with equal probability.

Turvey[24] criticized this scheme for failing to recognize the cost of rationing: consumers whose demands are curtailed are likely to suffer the loss due to the rationing. Crew and Kleindorfer[6] took up Turvey's point and included rationing cost as a function of excess demand in the social welfare.

From the viewpoint of mechanism design, a more pointed criticism involves the communication required by optimal rationing. Let us examine the Brown and Johnson scheme using Figure 4.2.

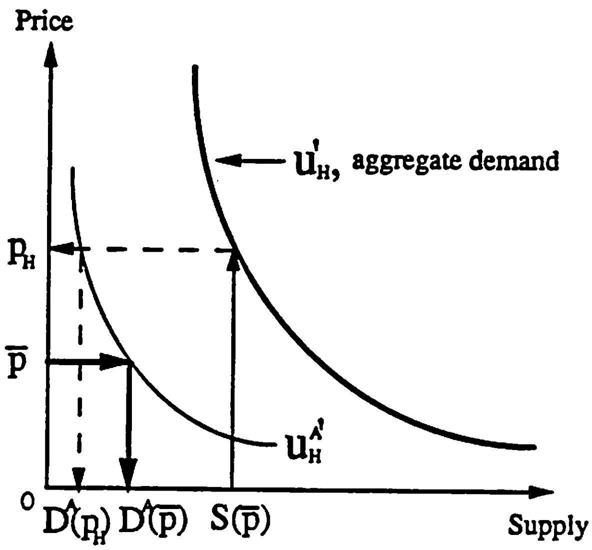

Figure 4.2. $D^A(p_H)$ is unknown to the social planner

The demand curve of consumer A, $u^{A'}_H$, is drawn as well as the aggregate demand curve, u'_H. What is known to the social planner is the entire u'_H-curve and A's response to the predetermined price $\bar{p}$, i.e., $D^A(\bar{p})$. But the assumed rationing requires knowing $D^A(p_H)$, which is not available to the social planner. The inadequacy of the information obtainable from a single predetermined price undermines their results. Crew and Kleindorfer's analysis suffers the same drawback.

The issue raised above presents a problem whenever allocation schemes employ a combination of future prices and rationing. Efficient rationing requires more information than can be extracted by a single price. The question rises: "How much more information?" The theory of mechanism can answer this question when it is formulated properly. We will see the answer in Sections 4.3 and 4.4.

There is another type of interruptible (or rationing) scheme proposed by Oren *et al.*[17] and Chao and Wilson[5], which they term 'priority service'. In their formulation, each unit of valuation of electric power is regarded as a decision maker responding to the priority service contract. Because of that formulation, it is not immediately clear what kind of consumer model is assumed, and how proposed contracts are interpreted by a consumer of the usual kind. We address these points in Section 4.3.

Many pricing schemes are proposed, but unfortunately, it is not easy to compare these schemes because we lack a basic model or a common ground to which we can apply different schemes and compare them. The recourse model described in Section 4.4 provides such common ground.

4.3 Single-stage Model

This section serves two purposes. One aim is to prepare for the analysis of the two-stage model discussed in the next section. The difference between the two models has to do with the 'flexibility' of consumers. Here consumers are assumed to be flexible (allowing costless rationing, if necessary). The precise meaning of flexibility will be given in § 4.3.1. The spot pricing scheme is described and shown to have the minimize message space in § 4.3.2.

The second goal is to understand a consumer model underlying the priority service proposed by Chao and Wilson[5], which is the subject of § 4.3.3.

4.3.1 Problem instance

Since the purpose of this chapter is more to extract the essence of various pricing schemes than to propose a new pricing scheme, problems in the simplest form are considered; network constraints, capacity constraints on lines and generation equipment are assumed not to be binding (simply put, ignored), and the network is assumed lossless. We assume the system consists of a supplier, who is a social planner as well, and N consumers. The supply is costless. In short, the problem is much like the allocation problem in Chapter 2, except that there is uncertainty in the system, such as fluctuation of supply and changes in consumer's valuation of power due to various causes.

We assume that the underlying probability space consists of a finite number of events (or contingencies), $\omega_1, \omega_2, \ldots, \omega_K$. The probability distribution of the events is denoted by $q := (q_1, q_2, \ldots, q_K)$, where q_k is the probability that event ω_k occurs. Allocation takes place at time 1 after everyone in the system observes the outcome of the stochastic event.

Consumers are indexed by numbers 1 through N, and the supplier by $N+1$. Consumers are assumed to be *flexible*, meaning that they can respond to any event without experiencing any inconvenience. Consumer j's valuation of power under event ω_k is represented by a *utility function* $u_k^j(\cdot)$. The supply under ω_k is denoted by S_k. The goal of the system is to find an optimal (contingency-dependent) allocation which maximizes the expected aggregate utility. Our allocation problem is

P1:

$$\begin{array}{lll} \max & \sum_{k=1}^{K} q_k \sum_{j=1}^{N} u_k^j(z_k^j) & \\ \text{sub. to} & \sum_{j=1}^{N} z_k^j \leq S_k, & k = 1,2,\ldots,K, \\ & z_k^j \geq 0, & j = 1,2,\ldots,N,\ k = 1,2,\ldots,K. \end{array}$$

We assume that the supplier alone knows the probability distribution q. The assumption makes sense in the main case considered later, with deterministic demand and random supply.

Let Σ^K be the K-dimensional simplex, i.e., $\Sigma^K := \{q \in R_+^K | \sum_{i=1}^K q_i = 1\}$.

The sets of local environments are

$$\begin{aligned} E^j &:= \{(u_k^j(\cdot), k = 1,2,\ldots,K)\},\ j = 1,\ldots,N, \\ E^{N+1} &:= \{(q,S) | q \in \Sigma^K,\ S := (S_1, S_2, \ldots, S_K) \in R_{++}^K\} \end{aligned}$$

We assume the $u_k^j(\cdot)$ are concave and differentiable.

The action space $\mathcal{A}$ is defined as the set of feasible allocations.

The goal F is to find an optimal allocation *for every contingency.*

4.3.2 Spot price

Let us consider the minimum size of message spaces of goal-realizing mechanisms. Note that Theorem 2.1.1 is not directly applicable because of the presence of q in the objective function of the individual profit maximization problem.

However, it is clear that a (centralized) optimal solution is obtained by solving K-different problems, one for each contingency,

P1$_k$:

$$\begin{array}{lll} \max & \sum_{j=1}^{N} u_k^j(z_k^j) & \\ \text{sub. to} & \sum_{j=1}^{N} z_k^j \leq S_k, & \\ & z_k^j \geq 0, & j = 1,2,\ldots,N. \end{array}$$

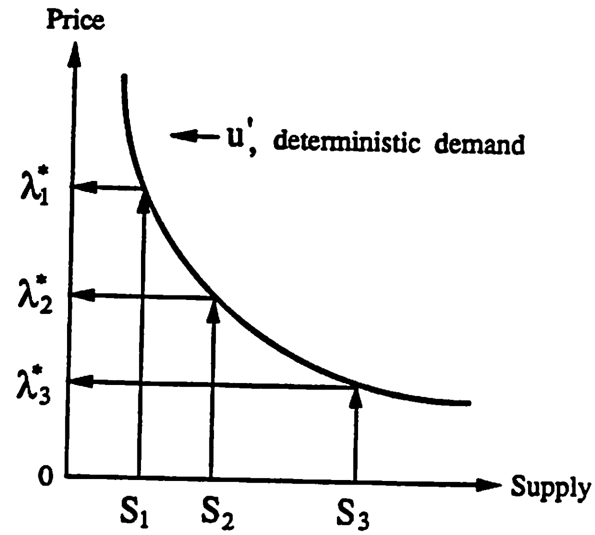

Figure 4.3. Spot Price

For these subproblems, Theorem 2.1.1 applies, and we see that the minimum size is $N+1$ for each one of them. Thus for problem **P1**, we find that $K(N+1)$ is the minimum size of message spaces.

Rather than finding an optimal solution for every contingency, our goal may be finding an optimal allocation for any particular event that happened to prevail at time 1. For this goal the minimum size of a message space is $N+1$.

As often done in the literature, let us assume that the supplier knows the aggregate utility $u_k(\cdot) := \sum_{j=1}^{N} u_k^j(\cdot)$ for each contingency ω_k. Then the supplier can set the price for commodity k (contingency commodity) at the point where the aggregate demand curve meets the vertical line S_k (see Figure 4.3). We call this price the *spot price* (at contingency ω_k). Applied to this simple setting, this scheme was proposed by Caramanis *et al.*[4]. The spot price here is the Lagrange multiplier λ_k^* for the resource constraint associated with the optimal solution in problem $\mathbf{P1}_k$. Note that the probability distribution q is immaterial. We summarize the scheme.

Spot pricing scheme

At time 0; no action.

At time 1; from the supplier to the consumers.

After observing the outcome of the random event, the supplier announces the spot price.

At time 1; from the consumers to the supplier.

A consumer responds to the spot price and demands the amount of energy he desires. He solves

P1$_k^j(\lambda_k^*)$:

$$\max_{z_k^j \geq 0} u_k^j(z_k^j) - \lambda_k^* z_k^j.$$

4.3.3 Priority service

The idea of priority service proposed by Chao and Wilson[5] is best illustrated in a situation with deterministic demand and stochastic supply. We consider this situation until otherwise stated. The difference between the spot pricing scheme and the priority service scheme becomes clear when we understand the respective commodity spaces. As noted above the spot price scheme regards energy supplies under different contingencies as different commodities. The priority service organizes a commodity space in a different way. Assume that contingencies are indexed so that

$$0 < S_1 < S_2 < \cdots < S_K. \tag{4.1}$$

We can think of amount S_1 of electric power as most 'reliable' supply (with supply probability 1), the amount $S_2 - S_1$ as the second most reliable supply (with supply probability $1 - q_1$), and so on. This way of viewing supply contingencies leads to selling different levels of *reliability*[3] as commodities. The *priority service contract* described below is designed to create a market for these commodities. See also Figure 4.4.

Priority service

At time 0; from the supplier to the consumers.

Contracts which specify the reliability level of delivery of unit amount of energy, say 1 kWh, are offered. We denote reliability levels by $(r_1, r_2, \ldots, r_K)$. (We will see shortly that K contracts are enough to sustain the optimal allocation.) The contract with reliability level r_i is priced at p_i. Notice that the price is for the contract and not for the delivered energy, that is, what the consumer pays will not be refunded in case of nondelivery.

[3]In the literature of electric power systems, the term 'reliability' is used in the context of *system security*. Here we use the term reliability to denote the tail distribution of electric power supply.

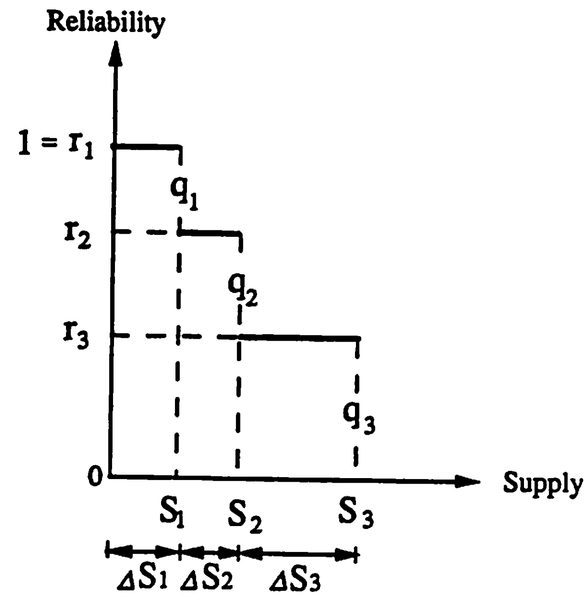

Figure 4.4. Commodities of priority service

We refer to the pair (r_i, p_i) as contract i, and the collection $\{(r_i, p_i), i = 1, 2, \ldots, K\}$ as the priority service price menu, or simply, the price menu.

The reliability levels characterizing contracts are *not independent* of each other but *nested* as follows:

Assume the contracts are indexed so that

$$r_1 > r_2 > \cdots > r_K. \tag{4.2}$$

Let y_i^j be units of contract i purchased by user j at time 0. Let Y^j be the total amount delivered at time 1 under this menu. Then $(r_1, r_2, \ldots, r_K)$ is the tail distribution of Y^j, i.e.,

$$\text{Prob}\{Y^j \geq \sum_{i=1}^{k} y_i^j\} = r_k, \; k = 1, 2, \ldots, K, \tag{4.3}$$

or, more specifically,

$$Y^j = \begin{cases} y_1^j + y_2^j + \cdots + y_{K-1}^j + y_K^j, & \text{with prob.} \quad r_K =: \Delta r_K, \\ y_1^j + y_2^j + \cdots + y_{K-1}^j, & \text{with prob.} \quad r_{K-1} - r_K =: \Delta r_{K-1}, \\ \vdots & \vdots \\ y_1^j + y_2^j, & \text{with prob.} \quad r_2 - r_3 =: \Delta r_2, \\ y_1^j, & \text{with prob.} \quad r_1 - r_2 =: \Delta r_1. \end{cases}$$

This may be summarized by saying that the energy associated with a lower reliability level will be delivered only after energy associated with higher reliability levels is delivered.

The reason for the nesting is to guarantee the highest quality service (the highest reliability level) is always received by those who value it most.

From a consumer's point of view, 1 kWh of delivered power associated with a lower reliability level is just as good as 1 kWh of delivered power associated with a higher reliability level, but knowing the nested structure of the price menu is essential to her decision-making.

At time 0; from the consumers to the supplier.
A consumer purchases the contracts as she needs. We denote consumer j's purchase of contract i by y_i^j as above. It will be determined by solving
$\mathbf{P1}^j(r)$:

$$\begin{array}{lll} \max & \sum_{i=1}^K \Delta r_i u^j(\sum_{k=1}^i y_k^j) - \sum_{i=1}^K p_i y_i^j, & \\ \text{sub. to} & y_i^j \geq 0, & i = 1,2,\ldots,K. \end{array}$$

The subscripts indicating contingencies are dropped from the utility functions since they are assumed to be deterministic.

At time 1:
The supplier will fill the contracts starting with the highest reliability and going down till the supply is exhausted.

The task of the supplier/social planner is to design an optimal menu, that is, a menu that induces an optimal allocation through the procedures described above. It is necessary to recover the tail distribution of the system supply by adding up the tail distributions of the consumers. Thus, it is natural to set reliability by

$$r_i := \sum_{k=i}^K q_k, \; k = 1,2,\ldots,K, \text{ and hence } \Delta r_i = q_i. \tag{4.4}$$

Note this choice of reliability is consistent with (4.1) and (4.2).

Let $u^{j\prime}(\cdot)$ be the derivative of a utility function $u^j(\cdot)$. The KKT conditions for problem $\mathbf{P1}^j(r)$ include

$$y_i^j > 0 \;\Rightarrow\; \sum_{k=i}^K q_k u^{j\prime}(\sum_{l=1}^k y_l^j) - p_i = 0,$$

$$\sum_{k=i}^{K} q_k u^{j\prime}(\sum_{l=1}^{k} y_l^j) < p_i \quad \Rightarrow \quad y_i^j = 0.$$

Let $z^{*j} := (z_1^{*j}, \ldots, z_K^{*j})$ and λ_k^* be the optimal solution and the associated Lagrange multipliers of problem $\mathbf{P1}_k$. Suppose that this scheme induces the optimal allocation and customer j receives energy under every contingency. Checking against the optimal solution of $\mathbf{P1}_k$, we are led to

$$u^{j\prime}(\sum_{l=1}^{k} y_l^j) = \lambda_k^*,$$

and

$$y_i^j = z_{i+1}^{*j} - z_i^{*j} =: \Delta z_i^{*j}, \text{ for } i = 0, 1, \ldots, K-1,$$

where $z_0^{*j} \equiv 0$. This will be satisfied by setting

$$p_i = \sum_{k=i}^{K} q_k \lambda_k^*. \tag{4.5}$$

Price p_i of contract i can be interpreted as the expected marginal utility from this contract, since the energy will be delivered for the contingencies $k \geq i$ under this contract.

Since p_i is independent of consumer j, this price menu induces an optimal allocation. We have seen

Proposition 4.3.1 *When the reliabilities and prices are set by (4.4) and (4.5), priority service induces the optimal allocation.*

Priority service can be combined with 'insurance' or 'payback' in case of nondelivery. Let c_k be the cash paid back per unit of nondelivered energy associated with reliability level r_k. Let the price of this contract be denoted by $p_k(c_k)$. Thus the price menu is $\{(r_k, c_k, p_k(c_k)), k = 1, \ldots, K\}$.

The consumer j's problem now is

$$\max_{y_k^j \geq 0} \sum_{i=1}^{K} q_i [u^j(\sum_{k=1}^{i} y_i^j) + \sum_{k=i+1}^{K} c_k y_k^j] - \sum_{i=1}^{K} p_i(c_i) y_i^j.$$

By proceeding in a similar manner, we see that if the price is set by

$$p_i(c_i) := p_i + c_i \sum_{k=1}^{i-1} q_k,$$

i.e., the price of the priority service without payback plus expected payback from the contract i, this price menu induces the optimal allocation.

When the payback is set equal to the marginal utility, i.e., $c_k := \lambda_k^*$, it may be thought of as a fair premium. The scheme of 'callable forward contracts' proposed by Gedra and Varaiya[7] uses this payback.

Remarks

1. A reliability level of, say 90%, may be understood to mean that when the allocation scenario repeats (independently) over time, the energy associated with the contract will be delivered 9 times out of 10.

2. Let us consider the expected revenue of the supplier under the spot price. Let $\Delta S_k := S_k - S_{k-1}$, with $S_0 \equiv 0$. The expected revenue is given by

$$\begin{aligned} E[\text{Revenue}] &= q_1\lambda_1^* S_1 + q_2\lambda_2^* S_2 + \cdots + q_K\lambda_K^* S_K \\ &= (q_1\lambda_1^* + q_2\lambda_2^* + \cdots + q_K\lambda_K^*)\Delta S_1 + (q_2\lambda_2^* + \cdots + q_K\lambda_K^*)\Delta S_2 + \cdots \\ &\quad + q_K\lambda_K^*\Delta S_K \\ &= p_1\Delta S_1 + p_2\Delta S_2 + \cdots + p_K\Delta S_K. \end{aligned}$$

The last term is the revenue under priority service. Note, however, the difference between the two. Under priority service, this amount is always collected by the supplier, while it is the expected revenue under spot pricing.

3. Note that the priority service communicates the probability distribution as reliability levels. Its message space has dimension $K(N+1)+(K-1)$. We can extend the argument to the case where the supply has a continuous distribution (or a density function). It should be noted that in that case the priority service asks for complete revelation of the demand curve of each consumer. It induces 'truth telling'. Once we realize that, the efficiency results found in the literature of the priority service can be anticipated.

4. When consumer demand is random, the priority service scheme appears to lack a consumer model consistent with its nested structure (or its commodity space). Let us consider the two contingency case. Let (z_1^{*j}, z_2^{*j}) be the optimal allocation to consumer j. In the case of deterministic demand and random supply discussed above ($S_1 < S_2$), we

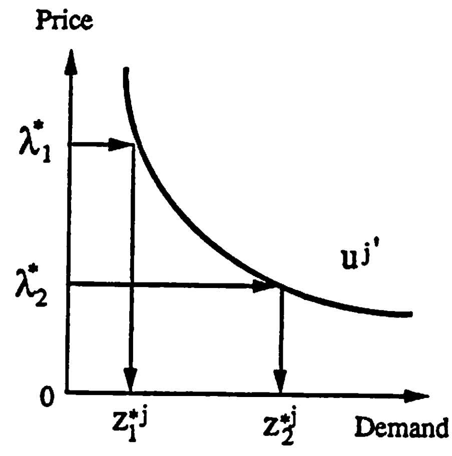

Figure 4.5a. Deterministic demand

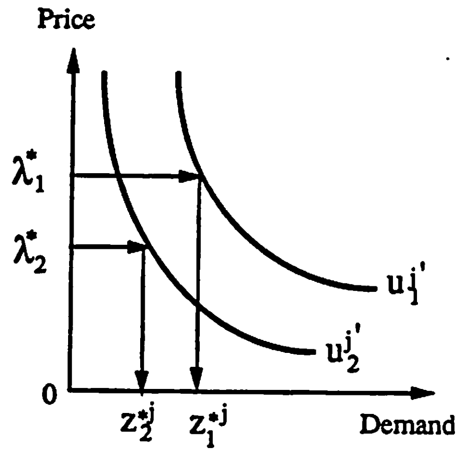

Figure 4.5b. Random demand

know $z_2^{*j} \geq z_1^{*j}$ for *all* consumers (see Figure 4.5a). But when demand is random this will not be guaranteed. For some consumer, it could well be $z_2^{*j} < z_1^{*j}$ (see Figure 4.5b), which cannot be sustained by priority service contracts.

Thus priority service is pricing with a one-dimensional characteristic, and it has its limitation. The claimed general applicability of priority service hinges upon the presupposition of

> "Each consumer can freely choose from the menu any priority option and assign it to any increment of his consumption. Therefore, without loss of generality, each consumer can be simply characterized by a single unit of demand and the associated willingness-to-pay[5] . . ."

However, the example above shows that there are cases in which a consumer (in the usual sense, as represented by its utility function) cannot identify itself as a collection of 'the single units' which supposedly responds to the priority service menu.

5. Consider the case when the allocation scenario repeats over time. In this context, priority service can be regarded as an 'off-line' procedure and spot pricing as an 'on-line' procedure. Priority service requires a one-time message exchange at the beginning of the allocation interval, while spot pricing requires a 'real-time' communications infrastructure. The choice of a pricing scheme should take into account the transaction cost incurred by information exchange as well as the allocation efficiency. In the absence of a real-time communications infrastructure, priority service can be a practical alternative to spot pricing.

See Chao and Wilson[5] for a discussion of implementation issues of priority service.

We have considered flexible consumers (and hence costless rationing). In the next section, we take up Turvey's point, and consider the costs of interruption.

4.4 Two-stage Recourse Model

The consumers are assumed to make commitments at time 0 based on their knowledge of the underlying stochastic events. Interruption costs of consumers are modeled as the penalties associated with the *recourse actions* they take at time 1.

In § 4.4.1 the model is described. In § 4.4.2 the KKT conditions for the problem are stated. In § 4.4.3 the price forecast mechanism is presented and shown to have the minimum size of a message space of a maximal level set among goal-realizing mechanisms. The implication is that when consumers are not flexible, a price forecast of some sort is necessary for efficient allocation.

4.4.1 Problem instance

As in the previous section, we assume that a system consists of a power supplier, whom we regard as a social planner, and N consumers. The network constraints, losses in the networks, and the other capacity constraints are ignored. There are stochastic elements in the system. Supply of energy becomes available at time 1, and hence actual allocation of power takes place at the time. At time 0, the probability distribution of the underlying stochastic events is known to the supplier.

When all consumers are flexible, the stochastic elements can be suppressed as in the spot pricing scheme of the previous section. However, when not all consumers are flexible, this wait-and-see approach results in a worse allocation than the well-planned allocation made at time 0. In practice, some of the consumers need to plan their consumption levels ahead of the actual time of consumption. A factory owner may need to commit other resources, such as manpower, at time 0 for activities at time 1. Her commitment could depend on the amount of power she expects to receive.

This 'inflexibility' of consumers is formulated in a recourse model. At time 0, each user plans to consume an amount x of energy and makes a commitment based on it. We refer to this planned consumption level as a *commitment*. After observing the event occurrence at time 1, she takes a recourse action z, a deviation from commitment x. Thus the commitment x is contingency-independent, while the recourse action z is contingency-dependent, say $z(\omega)$. The recourse action $z(\omega)$ can be either positive or negative. The initial commitment x is restricted to be nonnegative. Also, the actual allocation $x - z(\omega)$ at time 1 is assumed to be nonnegative. Each consumer is assumed to have a valuation on the pair $(x, z(\omega))$. The task of the supplier/social panner is to maximize expected aggregate valuation.

The problem is stated in a general format first. As in the previous section, a finite sample space is considered. Dependency on contingencies is denoted by subscripts. Let user j's value function under event ω_k be $f_k^j(x^j, z_k^j)$. Our optimization problem is:

P2:

$$
\begin{array}{lll}
\max\limits_{x,z} & \sum_{k=1}^{K} q_k \sum_{j=1}^{N} f_k^j(x^j, z_k^j) & \\
\text{sub. to} & \sum_{j=1}^{N}(x^j - z_k^j) \leq S_k, & k = 1, \ldots, K, \\
& x^j \geq 0, & j = 1, 2, \ldots, N, \\
& x^j - z_k^j \geq 0, & j = 1, 2, \ldots, N, \; k = 1, \ldots, K.
\end{array}
$$

Assumption 4.4.1 *Assume that*

- *the value functions f_k^j are concave in (x^j, z_k^j) with partial derivatives $\partial f_k^j / \partial x^j$, $\partial f_k^j / \partial z_k^j$ continuous except at a finite number of points.*
- *an optimal solution which satisfies the KKT conditions exists.*

The set of local environments are

$$
\begin{aligned}
E^j &:= \{f_k^j, k = 1, \ldots, K\}, \; j = 1, \ldots, N, \\
E^{N+1} &:= \{(q, S) | q \in \Sigma^K, \; S \in R_+^K\}.
\end{aligned}
$$

The set of system environments is $E := E^1 \times \cdots \times E^N \times E^{N+1}$ and its generic element is denoted by e as usual.

The action space is the set of feasible solutions $\mathcal{A} = \mathcal{A}(0) \times \mathcal{A}(1)$, where $\mathcal{A}(0)$ is the space of commitments (time 0 decision), and $\mathcal{A}(1)$ is the space of recourse actions (time 1 decision).

The goal F is to find an optimal solution, i.e.,

$$F(e) := \text{argmax } \mathbf{P2}(e).$$

4.4.2 The KKT conditions

First we regard this problem as a centralized problem, and derive the optimality conditions. Refer to Rockafellar and Wets[20] for an analysis of a a multi-stage recourse model with a more general sample space.

Let $q_k\lambda_k$ be the Lagrange multiplier to the resource constraint for the kth contingency. Also let $q_k\mu_k^j$ be the multiplier to the nonnegativity constraint on $x^j - z_k^j$. We also use the abbreviated notation E_kX_k for $\sum_{k=1}^K q_kX_k$. The Lagrangian is defined by

$$L(x,z,\lambda,\mu) := E_k[\sum_{j=1}^N f_k^j(x^j,z_k^j) - \lambda_k\{\sum_{j=1}^N(x^j - z_k^j) - S_k\} + \sum_{j=1}^N \mu_k^j(x^j - z_k^j)]$$

Under the assumption above, the KKT conditions for optimality are:

$$x^j \geq 0,\ \frac{\partial L}{\partial x^j} \leq 0, \text{ and } x^j\frac{\partial L}{\partial x^j} = 0,\ j = 1,2,\ldots,N,$$

thus,

$$x^j > 0 \ \Rightarrow\ E_k\frac{\partial f_k^j}{\partial x^j}(x^j,z_k^j) = E_k\lambda_k - E_k\mu_k^j \tag{4.6}$$

$$E_k\frac{\partial f_k^j}{\partial x^j}(x^j,z_k^j) < E_k\lambda_k - E_k\mu_k^j \ \Rightarrow\ x^j = 0 \tag{4.7}$$

and

$$\frac{\partial L}{\partial z_k^j} = 0,\ j = 1,2,\ldots,N,$$

i.e.,

$$-\frac{\partial f_k^j}{\partial z_k^j}(x^j,z_k^j) = \lambda_k - \mu_k^j, \tag{4.8}$$

and

$$\lambda_k \geq 0,\ \frac{\partial L}{\partial \lambda_k} \geq 0, \text{ and } \lambda_k[\sum_{j=1}^N(x^j - z_k^j) - S_k] = 0,\ k = 1,2,\ldots,K,$$

thus

$$\lambda_k > 0 \ \Rightarrow\ \sum_{j=1}^N(x^j - z_k^j) = S_k \tag{4.9}$$

$$\sum_{j=1}^N(x^j - z_k^j) < S_k \ \Rightarrow\ \lambda_k = 0, \tag{4.10}$$

and

$$\mu_k^j \geq 0,\ \frac{\partial L}{\partial \mu_k^j}, \text{ and } \mu_k^j(x^j - z_k^j) = 0,\ j = 1,2,\ldots,N,\ k = 1,2,\ldots,K,$$

thus

$$\mu_k^j > 0 \Rightarrow z_k^j = x^j \tag{4.11}$$

$$z_k^j < x^j \Rightarrow \mu_k^j = 0 \tag{4.12}$$

The conditions (4.8), (4.11), and (4.12) can be combined to form

$$x^j - z_k^j > 0 \Rightarrow -\frac{\partial f_k^j}{\partial z_k^j}(x^j, z_k^j) = \lambda_k \tag{4.13}$$

$$-\frac{\partial f_k^j}{\partial z_k^j}(x^j, z_k^j) < \lambda_k \Rightarrow x^j - z_k^j = 0 \tag{4.14}$$

(4.13) says that when actual allocation is positive, the scarcity cost is equal to the marginal valuation with respect to the recourse action.

The conditions (4.6) and (4.8) can combined to form

$$x^j > 0 \Rightarrow E_k \frac{\partial f_k^j}{\partial x^j}(x^j, z_k^j) = -E_k \frac{\partial f_k^j}{\partial z_k^j}(x^j, z_k^j), \tag{4.15}$$

$$E_k \frac{\partial f_k^j}{\partial x^j}(x^j, z_k^j) < -E_k \frac{\partial f_k^j}{\partial z_k^j}(x^j, z_k^j) \Rightarrow x^j = 0. \tag{4.16}$$

That is, when the optimal commitment is positive, the expected marginal utility gain with respect to the commitment and that with respect to the recourse action are equal.

4.4.3 Price forecast mechanism and the minimum size of message space

If the probability distribution of the underlying stochastic events were public knowledge, then the price mechanism would work. The results from Chapter 3 would apply, and the minimum size of message spaces would be $K(N+1)$ and the minimum size of a message space of a maximal level set would be K.

In the price forecast scheme, the supplier announces the probability distribution, and proceeds as in the price mechanism. Thus the price forecast mechanism would use an additional message of dimension $K-1$.

Price forecast mechanism

Announced publicly are a triple consisting of a probability distribution $q \in \Sigma^K$, a price vector $p \in R_+^K$, and an allocation $y := (y^1, \ldots, y^N) \in R_+^{KN}$. A price vector $p := (p_1, \ldots, p_K)$ is a list of contingency prices, i.e., p_i is the price of energy under contingency i. An allocation to consumer j, $y^j := (y_1^j, \ldots, y_K^j)$, is a vector of the amounts of energy allocated to consumer j at time 1 under respective contingencies. Thus the message space of this mechanism is

$$\mathcal{M} := \Sigma^K \times R_+^K \times R_+^{KN}.$$

Given the probability distribution q, each consumer can consider $\sum_{i=1}^K q_i f_i^j$ as her (deterministic) utility function. Each consumer follows the step of the price mechanism in Section 3.1.2.

Since it reduces to an ASCP in Chapter 3, this mechanism realizes the goal.

Thus we have a goal realizing mechanism with a $2K - 1$ dimensional space Λ. From the form of the KKT conditions, it appears that this Λ has the minimum size (of a message space of a maximal level set). We will verify this conjecture. The key step is the choice of the subset E^* of environments we use for the uniqueness property.

Since we compare a few allocation mechanisms for a special form of value functions in the next section, the form is described here. The subset E^* is chosen to fit the problem instances generated by them.

The value functions are restricted to be deterministic and to take the following form:

$$f_k^j(x^j, z_k^j) := u^j(x^j - z_k^j) - l^j(z_k^j) \tag{4.17}$$

The first term $u^j(\cdot)$ is interpreted as the utility over the actual allocation, and the second term $l^j(\cdot)$ as the penalty associated with the recourse action. We denote the actual allocation $x^j - z_k^j$ by y_k^j.

As we have seen, the trick is to choose E^* so that all resource constraints are tight (strictly positive multipliers) and nonnegativity constraints on decision variables are not binding at the optimal solution.

Let us examine the KKT conditions for the problem at hand. Let the derivatives of $u^j(\cdot)$ and $l^j(\cdot)$ be denoted by $u^{j\prime}(\cdot)$ and $l^{j\prime}(\cdot)$ respectively. Since $\partial f_k^j/\partial x^j = u^{j\prime}$, and $-\partial f_k^j/\partial z_k^j = u^{j\prime} + l^{j\prime}$, (4.15) becomes

$$x^j > 0 \Rightarrow E_k l^{j\prime}(z_k^j) = 0. \tag{4.18}$$

Also (4.13) becomes

$$z_k^j < x^j \Rightarrow u^{j\prime}(y_k^j) + l^{j\prime}(z_k^j) = \lambda_k \tag{4.19}$$

We want optimal solutions such that

$$x^j > 0, z_k^j < x^j, \forall k,\ j, \text{ and } \lambda_k > 0, \forall k. \tag{4.20}$$

Now assume every consumer shares the same utility and penalty functions. Then, because of the concavity of the objective function, there is an optimal solution such that every consumer has identical commitment and recourse. For such an instance, (4.18) becomes, letting $s_k := S_k/N$,

$$\sum_{k=1}^{K} q_k l^{j\prime}(x^j - s_k) = 0. \tag{4.21}$$

This equation allows us to determine the optimal commitment x^j and subsequently optimal recourse actions as $z_k^j = x^j - s_k$. E^* is chosen to satisfy conditions (4.20).

From our earlier experience in ASCP, we can expect dimension K to come out of the K equations in (4.19). We pull another $K - 1$ dimensions out of a single equation (4.21).

Observe that (4.21) is essentially the computation of the inner product. We will digress a little and study the size of message space required for the computation of the inner product.[4] It will help to understand the rather odd-looking choice of E^*.

Computation of expectation as inner product of two vectors

The underlying sample space is finite with K events. Participant 1 knows the probability distribution $q := (q_1, \ldots, q_K)$, and participant 2 knows the values of a random variable X expressed as a vector $(X_1, \ldots, X_K)$.

The sets of local environments are

$$E^1 := \Sigma_K \text{ and } E^2 := R^K.$$

Goal $f : E^1 \times E^2 \to R$ is to compute the expectation, i.e.,

$$f(q, X) := \sum_{k=1}^{K} q_k X_k.$$

The action space is the range R of f.

If participant 1 would communicate the distribution q, a $(K-1)$-dimensional message, participant 2 would compute the inner product and communicate it back to participant 1. This takes K-dimensional message space and $(K - 1)$-dimensional message space of a maximal level set. We show that this is the best possible.

[4]See also Hurwicz[10].

Lemma 4.4.1 *The minimum size of a message space of a maximal level set is $K-1$.*

Proof Let subset E^* be defined by

$$E^* := \{(q,X) \in \Sigma^K \times R^K | (q_1,\ldots,q_{K-1},q_K),(q_1,\ldots,q_{K-1},-\frac{\sum_{k=1}^{K-1} q_k^2}{q_K}), q>0\}.$$

Clearly, $E^* \subseteq f^{-1}(0)$, and it has dimension $K-1$. We will show that E^* has the uniqueness property.

Let $e := (q,X), \tilde{e} := (\tilde{q},\tilde{X}) \in E^*$ be such that $\tilde{e} \otimes_1 e \in f^{-1}(0)$ and $\tilde{e} \otimes_2 e \in f^{-1}(0)$. Since $\tilde{e} \otimes_1 e \in f^{-1}(0)$, we have

$$\sum_{k=1}^{K-1} \tilde{q}_k q_k - \tilde{q}_K (\sum_{k=1}^{K-1} q_k^2)/q_K = 0. \tag{4.22}$$

Similarly, from $\tilde{e} \otimes_2 e \in f^{-1}(0)$,

$$\sum_{k=1}^{K-1} q_k \tilde{q}_k - q_K (\sum_{k=1}^{K-1} \tilde{q}_k^2)/\tilde{q}_K = 0. \tag{4.23}$$

These two equations together yield

$$(\sum_{k=1}^{K-1} q_k \tilde{q}_k)^2 = (\sum_{k=1}^{K-1} q_k^2)(\sum_{k=1}^{K-1} \tilde{q}_k^2).$$

The Cauchy-Schwarz inequality says $LHS \leq RHS$ in the above with equality only when $\tilde{q}_k = \alpha q_k, k = 1,2,\ldots,K-1$ for some scalar α. Since the q_k are positive, $\alpha > 0$. By substituting $\tilde{q}_k = \alpha q_k$ into (4.23), we obtain $\tilde{q}_K = \alpha q_K$. Since $\sum_{k=1}^K \tilde{q}_k = \sum_{k=1}^K q_k = 1$, it follows $\alpha = 1$, and $\tilde{q} = q$ as desired. □

We are ready to define E^*. We may assume $0 < S_1 < S_2 < \cdots < S_K$. E^* is chosen so that the optimal solution to problem instances from this set is

$$\begin{aligned} x^j = x^* &:= s_K - 1, & j &= 1,\ldots,K, \\ z_k^j = z_k^* &:= x^* - s_k, & j &= 1,\ldots,N,\ k = 1,\ldots,K-1, \\ z_K^j = z_K^* &:= -1, & j &= 1,\ldots,K. \end{aligned}$$

In the above, z_K^* is (arbitrarily) fixed at -1, and the rest are determined so that the resource constraints are tight. We denote this solution by (x^*, z^*).

We start with the set of local environments of the supplier, E^{*N+1}, defined as follows:

- Let levels of supply be denoted by $S := (S_1, S_2, \ldots, S_K)$. Let us use $s := (s_1, s_2, \ldots, s_K) := S/N$ since this is more convenient in the subsequent definitions and analysis.

  Supply s (hence S) is fixed at levels which satisfy

$$0 < s_1 < s_2 < \cdots < s_K, \text{ and } s_K > s_{K-1} + 2$$

- Probability distributions are restricted to the following subset of Σ^K.

$$Q := \{q \in \Sigma^K | q_1 \geq q_2 \geq \ldots \geq q_K > \epsilon\}$$

  where ϵ is a prespecified small positive number less than $1/K$.

- $E^{*N+1} := Q \times \{S\}$. It has dimension of $K - 1$.

Consumers in E^* are taken to be identical. We specify the types of utilities and penalties in this subset. Utility functions are restricted to be piecewise linear; let $d_k := \frac{1}{2}(s_{k-1} + s_k)$, for $k = 1, 2, \ldots, K$ with $s_0 \equiv -s_1$. Also $d_{K+1} \equiv +\infty$.

$$u(y) := a_k y + b_k, \quad d_k \leq y < d_{k+1}, \; k = 1, 2, \ldots, K \tag{4.24}$$

where $a := (a_1, a_2, \ldots, a_K)$ satisfies

$$a_1 \geq a_2 \geq \cdots \geq a_K \geq 1/\epsilon \tag{4.25}$$

We denote this set of a by A. A has dimension K. The constants b_k are chosen to make $u(\cdot)$ continuous, namely

$$b_1 := 0, \text{ and } b_k := b_{k-1} + d_{k-1}(a_{k-1} - a_k), \; k = 2, 3, \ldots, K.$$

Loss functions are restricted to be piecewise linear for the positive value of z and quadratic for negative values of z:

$$l(z) := \begin{cases} \frac{1}{2}\frac{\sum_{k=1}^{K-1} q_k^2}{q_K} z^2, & z \leq 0, \\ q_{K-1} z, & 0 < z \leq x^* - d_{K-1} \\ q_k z + c_k, & x^* - d_{k+1} < z \leq x^* - d_k, k = 1, 2, \ldots, K-2 \end{cases} \tag{4.26}$$

where the constants c_k are defined to make $l(\cdot)$ continuous, namely

$$c_{K-1} := 0, \text{ and } \; c_{k-1} := c_k + (x^* - d_k)(q_k - q_{k-1}), k = K-2, K-3, \cdots, 1.$$

We abuse notation and identify the element of E^* with a and q, i.e.,

$$E^* := \{e := (a,q) | a \in A, q \in Q\}.$$

We proceed to show that E^* has the uniqueness property with respect to the goal.

Lemma 4.4.2 *All problem instances in E^* have (x^*, z^*) as an optimal solution.*

Proof We check the the KKT conditions with this solution.

$$\begin{aligned} u'(x^* - z_k^*) = u'(s_k) = a_k, \quad & k = 1,2,\ldots,K, \\ l'(z_k^*) = l'(x^* - s_k) = q_k, \quad & k = 1,2,\ldots,K-1, \\ l'(z_K^*) = l'(-1) = -(\sum_{k=1}^{K} q_i^2)/q_K & \end{aligned}$$

Since $x^* - z_k^* = s_k > 0$, we have by (4.19),

$$\lambda_k = a_k + q_k > 0, \ k = 1,2,\ldots,K-1,$$

and

$$\lambda_K = a_K - (\sum_{k=1}^{K-1} q_k^2)/q_K > a_K - 1/\epsilon \geq 0.$$

Also

$$E_k l'(z_k^*) = \sum_{k=1}^{K-1} q_k^2 - q_K \frac{\sum_{k=1}^{K-1} q_k^2}{q_K} = 0$$

Clearly all resource constraints are binding. Therefore the KKT conditions are satisfied. □

Proposition 4.4.1 *As defined above, $E^* \subseteq F^{-1}(x^*, z^*)$ has the uniqueness property with respect to goal F. Hence, a goal-realizing mechanism has a message space of a maximal level set of dimension at least $(2K-1)$.*

Proof Let $e = (a,q), \tilde{e} = (\tilde{a}, \tilde{q}) \in E^*$ be such that $\tilde{e} \otimes_j e \in F^{-1}(x^*, z^*)$, and $e \otimes_j \tilde{e} \in F^{-1}(x^*, z^*), j = 1,2,\ldots N+1$. We need to show that $\tilde{a} = a$ and $\tilde{q} = q$.

First we will show $\tilde{q} = q$. Since $\tilde{e} \otimes_{N+1} e \in F^{-1}(x^*, z^*)$, it has to satisfy the KKT conditions with (x^*, z^*). Following the steps in the proof of the previous lemma, we see that

$$\sum_{k=1}^{K} \tilde{q}_k l'(z_k^*) = \sum_{k=1}^{K-1} \tilde{q}_k q_k - \tilde{q}_K (\sum_{k=1}^{K-1} q_k^2)/q_K = 0. \tag{4.27}$$

Similarly, from $e \otimes_{N+1} \tilde{e}$, we have

$$\sum_{k=1}^{K-1} q_k \tilde{q}_k - q_K(\sum_{k=1}^{K-1} \tilde{q}_k^2)/\tilde{q}_K = 0. \tag{4.28}$$

By the same arguments as in the proof of Lemma 4.4.1, we see $\tilde{q} = q$.

Next we will show $\tilde{a} = a$. Since $\tilde{e} \otimes_1 e \in F^{-1}(x^*, z^*)$, we have, $\lambda_k = \tilde{a}_k + \tilde{q}_k$ for consumer 1 and $\lambda_k = a_k + q_k$ for consumer 2 for all k. In view of $\tilde{q} = q$, we have $\tilde{a} = a$. □

Remarks

1. In the example above, at an interior solution, the scarcity cost is equal to the sum of the marginal utility gain and the marginal avoided loss. Also, the expected scarcity cost is equal to the expected marginal utility gain.

2. It is intuitively clear that commitment and recourse actions have to be jointly determined in general. In fact, it takes at least $(2K-1)$-dimensional message space of a maximal level set to determine the commitment x alone.

Let $\bar{F}^{-1}(x^*)$ denote the set of environments such that $(x^j = x^*, j = 1, \ldots, N)$ is a correct commitment at time 0. Clearly, $E^* \in \bar{F}(x^*)$.

Let $e := (a, q), \tilde{e} := (\tilde{a}, \tilde{q}) \in E^*$ be such that $\tilde{e} \otimes_j e \in F^{-1}(x^*, z^*)$, $j = 1, 2, \ldots N+1$. Consider problem instance $\tilde{e} \otimes_{N+1} e$. Since consumers are identical, the concavity of the objective function implies the symmetric solution. Thus we can consider the constraints of the form

$$x^* - z_K \le s_k.$$

Also it is clear that at the optimal solution,

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_K,$$

again by concavity (and $S_1 \le S_2 \le \cdots \le S_K$). If we show that $\lambda_K > 0$, then the resource constraints are tight, and we are back to the case examined above.

Since $x^* - z_K = s_K - 1 - z_k \le s_K$, $z_K \ge -1$ at the optimal solution. Assume $z_K > -1$. Then

$$u'(x^* - z_K) + l'(z_K) = a_K + \frac{\sum_{i=1}^{K-1} q_i^2}{q_K} z_K \le \lambda_K = 0.$$

(λ_K is 0, since the constraint is not binding.)

But $a_K + \frac{\sum_{i=1}^{K-1} q_i^2}{q_K} z^K > a_K - \frac{1}{\epsilon} > 0$. Thus $z_K = -1$, and $\lambda_K = a_K - \frac{\sum_{i=1}^{K-1} q_i^2}{q_K} > 0$.

4.5 Comparison of Pricing Schemes

In this section, we compare various pricing schemes in the context of the two stage recourse model introduced in the previous section. We consider the following special case of the recourse model.

P3:

$$\begin{array}{llll} \max & \sum_{k=1}^{K} q_k \sum_{j=1}^{N} u^j(x^j - z_k^j) - l^j(z_k^j) & & \\ \text{sub. to} & \sum_{j=1}^{N}(x^j - z_k^j) \leq S_k, & & k = 1,2,\ldots,K, \\ & x^j \geq 0, & & j = 1,2,\ldots,N, \\ & x^j - z_k^j \geq 0, & & j = 1,2,\ldots,N,\ k = 1,2,\ldots,K \end{array}$$

The benefit from the consumption of energy depends only on the actual energy allocated. The penalty associated with the recourse action depends only on the amount adjusted. We allow negative values of z_k^j, i.e., consumers can consume more than they planned at a cost. We assume that $u^j(\cdot)$'s are concave, $l^j(\cdot)$ convex, both positive and $u^j(0) = 0,\ l^j(0) = 0$. We denote the respective derivatives by $u^{j\prime}(\cdot)$ and $l^{j\prime}(\cdot)$. Flexible consumers have $l(\cdot) \equiv 0$.

A problem similar to this is studied by Tan and Varaiya[23]. One of their models deals with a mass of identical (infinitesimal) consumers indexed by $j \in [0,1]$. In their formulation, they restrict the recourse action to be of an 'all-or-nothing' type, i.e., $z_k^j \in \{0, x^j\}$. These extra constraints impose the need to differentiate identical consumers into several groups. They show when there are K contingencies in supply, it suffices to differentiate the population into at most K groups. They also show that the optimal allocation can be sustained by implementing reliability-based contracts.

Price forecast scheme:

This is the modification of the price forecast mechanism in the previous section.

At time 0; from the supplier to the consumers.

Price forecast $\{(\lambda_k^*, q_k), k = 1,2,\ldots,K\}$ is announced, where the $q_k\lambda_k^*$ are the Lagrange multipliers associated with the optimal solution of **P3**. We note that λ^* can be obtained from knowledge of the aggregate utility and the aggregate loss.

At time 0; the consumer problem.

Consumer j decides her optimal commitment level along with recourse actions. She solves

P3$^j(\lambda^*)$:

$$\begin{array}{ll} \max & \sum_{k=1}^K q_k[u^j(x^j - z_k^j) - l^j(z_k^j) - \lambda_k^*(x^j - z_k^j)] \\ \text{sub. to} & x^j \geq 0,\ x^j - z_k^j \geq 0,\ k = 1, 2, \ldots, K. \end{array}$$

At time 1; from the supplier to the consumers.

After observing the event occurrence, the supplier announces the price.

At time 1; from the consumers to the supplier.

Consumers adjust to the price and demand the energy.

Priority service

Priority service can also be implemented for this problem. Contracts $\{(r_k, p_k), k = 1, 2, \ldots, K\}$ are interpreted as in Section 4.3. We assume the contingencies are indexed so that $S_1 < S_2 < \cdots < S_K$ as before. Also, we set $r_k := \sum_{i=k}^K q_i,\ k = 1, 2, \ldots, K$.

At time 0; from the supplier to the consumers.

Contracts $\{(r_k, p_k)\}$ are offered.

At time 0; from consumers to the supplier.

Consumers purchase the contracts. Let $y^j := (y_1^j, y_2^j, \ldots, y_K^j)$ be consumer j's purchase. It will be determined by solving

P3$^j(r)$:

$$\begin{array}{ll} \max & \sum_{i=1}^K q_i[u^j(\sum_{k=1}^i y_k^j) - l^j(x^j - \sum_{k=1}^i y_k^j)] - \sum_{i=1}^K p_i y_i \\ \text{sub. to} & y_k^j \geq 0,\ k = 1, 2, \ldots, K \end{array}$$

The KKT conditions include

$$\begin{aligned} x^j[\sum_{i=1}^K l^{j\prime}(x^j - \sum_{k=1}^i y_k^j)] &= 0 \\ y_i^j[\sum_{k=i}^K q_k\{u^{j\prime}(\sum_{k=1}^i y_k^j) + l^{j\prime}(x^j - \sum_{k=1}^i y_k^j)\} - p_i] &= 0 \end{aligned}$$

As before, by setting $p_k := \sum_{i=k}^K \lambda_k^*,\ k = 1, 2, \ldots, K$, the optimal allocation will be sustained.

The same remark as in Section 4.3 applies for priority service. When the consumers' valuations are arbitrary functions of (x, z_k), there is no consumer model consistent with the nested structure of the contracts.

Spot price with predetermined price

Caramanis *et al.*[4] and Schweppe *et al.*[22] propose the use of the predetermined prices along with spot prices. The idea is to group the consumers according to their communication capabilities and/or needs. They develop the scheme for a multi-stage problem, but applied to our two stage model, and stripped of the network and other constraints, it may be stated as follows.

A *predetermined price* is denoted by π. This is the price announced at time 0, and there is a group of consumers who respond to this price by maximizing their utility function $u^j(\cdot)$ by solving

$$\max_{x^j \geq 0} u^j(x^j) - \pi x^j.$$

We denote this group by J. There is another group of consumers who respond to the *spot price* at time 1. The spot prices are contingency-dependent. We denote this group by I. Their response is given just as the consumers in group J, i.e., under contingency ω_k, they solve

$$\max_{x_k^i \geq 0} u^i(x_k^i) - \pi_k x_k^i.$$

If the aggregate demand exceeds supply at time 1, the excess demand of consumers in group J will be rationed and they experience the loss according to the amount rationed. Thus this is a special case of the recourse model we have studied. Their model may be stated as

$$\begin{array}{rl} \max\limits_{\pi,\pi_1,\ldots,\pi_K} & \sum_{k=1}^K q_k[\sum_{j\in J}\{u^j(x^j(\pi)) - l^j(z_k^j)\} + \sum_{i\in I} u^i(x^i(\pi_k))] \\ \text{sub. to} & \sum_{j\in J}(x^j(\pi) - z_k^j) + \sum_{i\in I} x^i(\pi_k) \leq S_k, \ k = 1,2,\ldots,K, \\ & z_k^i \geq 0, \ i \in I, \ k = 1,2,\ldots,K, \end{array}$$

where $x^j(\pi)$ and $x^i(\pi_k)$ are the maximizers of the respective consumer problems above and z_k^j represents amount rationed.

This model of group J consumers' response is rather unnatural. Since these consumers suffer from rationing, they are the ones who would benefit from the price forecast. The supplier can improve the expected social welfare by announcing the price forecast rather than announcing a single predetermined price.

4.6 Concluding Remarks

We have seen that a model which distinguishes individual participants of a system clarifies the communication issues associated with decentralized resource allocation processes.

The recourse model studied in this chapter is a natural starting point for studying various pricing schemes. The different commodity spaces can be used to sustain an efficient allocation. The choice of the commodity space will depend on the context at hand.

When the consumers are inflexible, some form of a price forecast improves expected social welfare. Any scheme which announces a single price will not be efficient.

We have ignored the network constraints and the other constraints for simplicity of exposition. A full-blown version of the recourse model is found in Kaye *et al.*[11]. The focus of their paper is to assure system security through pricing. They termed the recourse actions by participants of the system as *contingency offerings.*

Considerations of system security, network constraints, and losses in the system inevitably make the problem agentwise *non*separable. The implication is that a uniform price cannot sustain the optimal allocation simply because it does not stimulate enough message exchanges.

That leads to 'individualized prices' in which the participants pay more or less than the average according to their contribution to the nonseparable (or joint) costs or the nonseparable constraints affecting the social welfare. Kaye *et al.*[11]'s pricing scheme may be viewed as inducing a Nash equilibrium allocation for a game played by the participants of the system.

It is essential to have the knowledge of the underlying stochastic events for allocating resources efficiently. This information will in general be distributed among the participants of the system. Thus the exchange of information on stochastic events will be as important as exchanging information about the willingness to pay and the marginal cost of supply which have been emphasized in previous studies.

There is need to implement a procedure which extracts the private information about the uncertainty.

Chapter 5

Assignment of Digital Pipe

5.1 Introduction

Defining and studying 'good coordination' of digital communication networks is important for the formulation of policy regarding public carriers. As a first step, we study assignment problems of a data pipe shared by many users. A data pipe, which we call a *digital pipe*, is a communication link connecting two points, a source and a destination, and through which data are sent. We assume data are gathered into *packets* by each user. The pipe can send only one packet per time period. The pipe users take turns to use the pipe to send their packets from the source.

Each user is assumed to have a *utility* over her own assignment pattern, i.e., the time periods her packets are sent. Her utility is independent of the other users' assignment patterns. Given an overall assignment pattern, we call the sum of the utilities of users the *aggregate utility* under the assignment. The pipe owner's task is to choose an assignment which maximizes the aggregate utility.

We introduce some notation. Assume there are N (> 1) users, and that the pipe is available for T time periods. Let $[y^j(t), t = 1, 2, \ldots, T,\ j = 1, 2, \ldots, N]$ be an assignment, where $y^j(t)$ takes values 0 or 1, and $y^j(t) = 1$ indicates that user j is assigned the pipe at time t. The utility of user j is $U^j : \{0,1\}^T \rightarrow R_+$. With this notation, the task of pipe owner/social planner is

P:

$$\max_{[y^j(t)]} \ \sum_{j=1}^N U^j(y^j(1), y^j(2), \cdots, y^j(T))$$

$$\text{sub. to} \qquad \sum_{j=1}^{N} y^j(t) \leq 1, \qquad t = 1,2,\ldots,T$$
$$y^j(t) \in \{0,1\}, \qquad t = 1,2,\ldots,T,\ j = 1,2,\ldots,N.$$

The utility of each user is private information, i.e., the set of local environments of user j is

$$E^j := \{U | U : \{0,1\}^T \rightarrow R_+\}.$$

The set of system environments is $E := E^1 \times \cdots \times E^N$.

The action space $\mathcal{A}$ is the set of feasible assignments of problem **P**, i.e.,

$$\mathcal{A} := \{[y^j(t)| \sum_{j=1}^{N} y^j(t) \leq 1, \forall t; y^j(t) \in \{0,1\}, \forall t, j\}.$$

The goal F of the system is to find an optimal assignment, i.e.,

$$F(U^1, \ldots, U^N) := \text{argmax}\ \mathbf{P}(U^1, \ldots, U^N).$$

Note the utility functions of pipe users are intertemporal. The possibility of spreading necessary message exchanges over time is of particular interest. In this chapter, we analyze two special cases of the problem which we call a *multi-armed bandit problem* and a *matching problem*. We shall show that a multi-armed bandit problem admits spreading of message exchanges, while a matching problem does not.

In § 5.2 a problem instance of a multi-armed bandit problem is described, and the minimum size of message space is obtained. In § 5.3 a matching problem is studied in a similar manner.

In § 5.4 the possibility of spreading message exchanges over time is discussed.

5.2 Multi-armed Bandit Problem

A special case of problem **P** is discussed. A stochastic version of this problem is well-known and called the multi-armed bandit problem. We borrow the name for our problem.

In §5.2.1, the problem instance is stated.

In §5.2.2, the problem is converted to a simpler but equivalent problem which in turn reduces to the problem of finding the ordering of T real numbers.

In §5.2.3, a lower bound for the size of a message space required for finding the ordering of real numbers is derived.

In §5.2.4, a goal-realizing mechanism, which we call a *sequential auction mechanism*, is presented. The possibility of spreading message exchanges is noted.

5.2.1 Problem instance

User j has $L^j(\geq 1)$ packets to send. Her utility increases in an incremental manner, that is, each packet sent adds to her total utility by a certain amount. However, there is a discounting associated with time which depreciates the value of the increment. The discounting is done in geometric manner and we assume that *the value of the* discount factor $\beta \in (0,1)$ *is common and known to all users (and the pipe owner).* This assumption is essential to our results. User j's prediscount increments are represented as a L^j-vector $Z^j := (Z^j(1), Z^j(2), \ldots, Z^j(L^j))$. When her lth packet is sent at time t, it adds $\beta^t Z^j(l)$ to user j's utility.

We assume that $T = \sum_{j=1}^N L^j$, that is, pipe is available for a long enough time to send all the packets in the system. This is another crucial assumption. When $T < \sum_{j=1}^N L^j$, our analysis below does not apply.

We also assume that $Z^j(l)$'s are nonnegative for simplicity. This is not essential to our results.

The set of local environments of users are

$$E^j := \{Z^j \in R_+^{L^j}\}, j = 1, \ldots, N.$$

The set of system environments is $E := E^1 \times \cdots \times E^N \subseteq R_+^T$.

We change the notation for an assignment from $[y^j(t)]$ to $[\Delta x^j(t)]$ in this section. This is to emphasize the cumulative nature of the utility at hand. The cumulative number of packets of user i sent by time t is denoted by $x^j(t)$, i.e., $x^j(t) = \sum_{s=1}^t \Delta x^j(s)$. We set $x^j(0) \equiv 0$ and $Z^j(0) \equiv 0$ for all users. With this notation, problem **P** reformulated as **P1**:

$$\begin{array}{lll} \max & \sum_{j=1}^N \sum_{t=1}^T \beta^t Z^j(x^j(t)) \Delta x^j(t) & \\ \text{sub. to} & \sum_{j=1}^N \Delta x^j(t) = 1, & t = 1, 2, \ldots, T \\ & \Delta x^j(t) \in \{0, 1\}, & t = 1, 2, \ldots, T,\ j = 1, 2, \ldots, N. \end{array}$$

5.2.2 Equivalent problem

We convert a vector of the incremental rewards Z^j to a new vector $\overline{Z}^j$ which has a convenient property for our analysis. The definition of $\overline{Z}^j$ will be given shortly, but its properties and their consequences are stated first.

Properties of $\overline{Z}^j$*:*

1. The conversion can be done in a private manner, i.e., $\overline{Z}^j$ is obtained independent of the other users' Z^i's.

2. The problem instances $\mathbf{P1}(Z^1, Z^2, \ldots, Z^N)$ and $\mathbf{P1}(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N)$ yield the same optimal assignments, i.e., an assignment $[\Delta x^j(t)]$ is optimal for $\mathbf{P1}(Z^1, Z^2, \ldots, Z^N)$ if and only if it is optimal for $\mathbf{P1}(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N)$.

3. $\overline{Z}^j(l)$ is nonincreasing in l, i.e., $\overline{Z}^j(1) \geq \overline{Z}^j(2) \geq \cdots \geq \overline{Z}^j(L^j)$.

Because of those properties of $\overline{Z}^j$'s, we need to consider only the problem instances with nonincreasing $Z^j(l)$'s.

When all Z^j's are nonincreasing, finding an optimal assignment is easy. The *myopic rule*, which assigns the pipe to a user with the largest immediate incremental reward, is optimal. If there is a tie in the immediate largest incremental reward, we can break the tie arbitrarily. It can also be shown that an optimal assignment is necessarily myopic. As a consequence, the problem reduces to finding the ordering of T nonnegative real numbers, i.e., the largest incremental reward receives assignment at time 1, the second largest at time 2, and so on.

Let us fix T, the total number of packets in the system. Then from the view point of information gathering, the worst case is when each user has one packet (so there are $N = T$ users in the system), since each reward is stored separately. The minimum size of a message space of a mechanism for finding the ordering of T real numbers, each number is privately known to its holder, coincides with the minimum size of message space of a goal-realizing mechanism for the multi-armed bandit problem. Finding the ordering of real numbers is the topic of the next subsection.

The conversion of Z^j to $\overline{Z}^j$ is done in two steps:

$$\nu^j(l+1) \;:=\; \max_{l<\tau\leq L^j} \frac{\sum_{s=l+1}^{\tau} \beta^s Z^j(s)}{\sum_{s=l+1}^{\tau} \beta^s}, \; l = 0, 1, \ldots, L^j - 1,$$

$$\overline{Z}^j(l) \;:=\; \min_{1\le s\le l} \nu^j(l),\; l=1,2,\ldots,L^j.$$

$\nu^j(l)$ is known as the Gittins index after Gittins who studied the stochastic version of this problem and proved the optimality of the celebrated index rule. (The index rule is to assign the pipe to the user with the largest current index.) We call $\overline{Z}^j$ a concave envelope of Z^j. From the definition, it is clear that $\overline{Z}^j$ is nonincreasing in l. For details, see Chapter 6.

5.2.3 Finding the ordering of N numbers

Imagine 3 persons each with a number. The goal is to find out who has the smallest, the middle, and the largest number respectively. The exact values of these numbers are not required. Once the middle number is identified, the rest is a matter of asking a series of yes-no-answer questions such as "Is you number larger than this?", "Smaller?", and so on. It appears that the problem is solved by a mechanism with 1 dimensional message space (of a maximal level set). In general, for $N = 2k$ and $2k+1$, a message space of k dimension is needed. It will be shown that we cannot realize the goal with a mechanism having a smaller message space.

We start by stating the problem instance, and then choose a subset which has the target dimension and the uniqueness property.

Problem instance

There are N participants each with a nonnegative number. The set of local environments of participant j is $E^j = R_+$.

The action space $\mathcal{A}$ is the set of permutations of $\{1,2,\ldots,N\}$, $a : \{1,2,\ldots,N\} \to \{1,2,\ldots,N\}$. We identify a with its value $\{a(1),a(2),\ldots,a(N)\}$; for example, the identity permutation will be denoted as $\{1,2,\ldots,N\}$. $\mathcal{A}$ has $N!$ elements.

The goal, $F : E \twoheadrightarrow \mathcal{A}$, is defined by

$$a \in F(e^1,e^2,\ldots,e^n) \iff e^{a(1)} \le e^{a(2)} \le \cdots \le e^{a(N)}.$$

A lower bound for size of a message space of a maximal level set

We assume $N = 2k$, and will show that k is a lower bound for the size of a message space of a maximal level set. For $N = 2k+1$, it will be clear that k is a lower bound as well. In the next subsection, it will be shown that k is the minimum size for both cases.

We consider the level set of the identity permutation $\{1,2,\ldots,N\} =: id$.

A subset $E^* \subseteq F^{-1}(id)$ is defined by

$$E^* := \{(z_1, z_1, z_2, z_2, \ldots, z_k, z_k) | 0 \leq z_1 \leq z_2 \leq \cdots \leq z_k\}. \tag{5.1}$$

It has dimension k. We proceed to show that it has the uniqueness property.

Lemma 5.2.1 *Defined as in (5.1), E^* has the uniqueness property with respect to goal F.*

Proof Let $z := (z_1, z_1, z_2, z_2, \ldots, z_k, z_k), w := (w_1, w_1, w_2, w_2, \ldots, w_k, w_k) \in E^*$ be such that $w \otimes_j z \in F^{-1}(id)$, $j = 1, 2, \ldots, N$.

Consider $w \otimes_{2i-1} z = (z_1, z_1, \ldots, z_{i-1}, z_{i-1}, w_i, z_i, z_{i+1}, z_{i+1}, \ldots, z_k, z_k)$ and $w \otimes_{2i} z = (z_1, z_1, \ldots, z_{i-1}, z_{i-1}, z_i, w_i, z_{i+1}, z_{i+1}, \ldots, z_k, z_k)$ for $i = 1, 2, \ldots, k$. Since $w \otimes_{2i-1} z \in F^{-1}(id)$, we have $w_i \leq z_i$. Also, since $w \otimes_{2i} z \in F^{-1}(id)$, we have $z_i \leq w_i$. Therefore, $w_i = z_i$ for $i = 1, 2, \ldots, k$. Thus $w = z$, as desired. □

5.2.4 Sequential auction mechanism

Let us go back to the multi-armed bandit problem. Let the total number of the packets be $T := 2k + 1$ ($k \geq 1$). Here is a goal-realizing mechanism with a message space of size k.

Sequential auction mechanism

Announced publicly are a pair comprising a vector of *k-winning bids (at even time periods)* $\lambda := (\lambda_1, \lambda_2, \ldots, \lambda_k)$ and an assignment $[\Delta x^j(t)]$. λ_i is nonincreasing in i, i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$. Let $\Lambda := \{\lambda \in R_+^k | \lambda_1 \geq \cdots \geq \lambda_k\}$. The message space is

$$\mathcal{M} := \Lambda \times \mathcal{A}.$$

Let $\tilde{\lambda} := (\tilde{\lambda}(1), \tilde{\lambda}(2), \ldots, \tilde{\lambda}(T))$ be $\tilde{\lambda} = (\lambda_1, \lambda_1, \lambda_2, \lambda_2, \ldots, \lambda_k, \lambda_k, 0)$. The vector $\tilde{\lambda}$ may be regarded as the rents of the pipe at respective time periods.

Pipe user j agrees to a message if whenever the pipe is assigned to her, her converted incremental reward $\overline{Z}^j(x^j(t))$ is greater than or equal to the rent $\tilde{\lambda}(t)$ at the time of the assingment. Otherwise she rejects the message. To guarantee the optimality, the rents for the even time periods should exactly equal the converted incremental rewards of those who are assigned the pipe. Thus the equilibrium correspondence of user j is

$$\mu^j := \left\{ (\lambda, [\Delta x^j(t)]) \in \mathcal{M} \left| \begin{array}{llll} \Delta x^j(t) = 1 & \Rightarrow & \overline{Z}^j(x^j(t)) = \tilde{\lambda}(t), & t = 2i, \quad i = 1, \ldots, k, \\ \Delta x^j(t) = 1 & \Rightarrow & \overline{Z}^j(x^j(t)) \geq \tilde{\lambda}(t), & t = 2i+1, \quad i = 0, 1, \ldots, k \end{array} \right. \right\}$$

The outcome function h is the projection of a message on the action space, i.e.,

$$h(\lambda, [\Delta x^j(t)]) := [\Delta x^j(t)].$$

Lemma 5.2.2 *The sequential auction mechanism realizes the goal.*

Proof Since $\tilde{\lambda}(t)$ is nonincreasing in t, the mechanism assigns the pipe in the decreasing order of the converted incremental rewards. Thus the assignment is optimal. □

Consider E^* in the previous section (translated into the problem instances of the multi-armed bandit problem). It is clear that the equilibrium correspondence of the sequential auction mechanism is a continuous function on E^*. Thus we have shown

Theorem 5.2.1 *When each user has one packet (hence the informationally worst case) and the total number of packets in the system is $T = 2k$, or $T = 2k + 1$, the minimum size of a message space of a maximal level set is k.*

The way the sequential auction mechanism works, it does not need to communicate k numbers $(\lambda_1, \ldots, \lambda_k)$ at once. At time 1, it suffices to communicate λ_1, and identify the largest and the second largest converted incremental rewards. Then at time 3, we can communicate λ_2, and identify the third and forth largest increments, and so on. This is a very attractive property, since we need one dimensional information-carrying capacity at each time. For very large T, a mechanism which exchanges all the necessary information at time 1 will not be implementable, whereas the sequential auction mechanism can be easily implemented.

Unfortunately, this property is due to a special structure of the multi-armed bandit problem, and is not generally available. In § 5.4, we discuss the structure of a problem which admits sequential message exchanges.

5.3 Matching Problem

We study another special case of problem **P** which we call a matching problem. In this problem, each pipe user has only one packet to send, and his utility is determined by when it is sent. Thus the problem can be seen as a matching of a user and a time frame.

Koopmans and Beckmann[12] called this 'the linear assignment problem'. They emphasized the use of 'price' in decentralized assignment. We follow their analysis closely.

In §5.3.1 the problem instance is stated.

In §5.3.2 a *price mechanism* is presented. It attaches a price to each time frame and hence it has a message space of dimension T.

In §5.3.3 it is shown that T is the minimum size of a message space of a maximal level set. The impossibility of spreading message exchanges over time is noted.

5.3.1 Problem instance

User j has one packet to send. His utility is represented by a T-vector $v^j := (v^j(1), v^j(2), \ldots, v^j(T))$, where $v^j(t)$ is the reward he receives when his packet is sent at time t. We assume that $v^j(t)$ is nonnegative. It is not essential for the minimality results.

The set of local environments of user j is

$$E^j := \{v^j \in R_+^T\}.$$

The set of the system environments is $E = E^1 \times \cdots \times E^N \subseteq R_+^{TN}$.

The action space and the goal are as defined in § 5.1.

We assume that there are more users (hence more packets) in the system than the available time periods, i.e.,

$$N > T.$$

Problem **P** is an integer linear program:

P2:

$$\begin{aligned}
\max_{[y^j(t)]} \quad & \textstyle\sum_{t=1}^{T} \sum_{j=1}^{N} v^j(t) y^j(t) \\
\text{sub. to} \quad & \textstyle\sum_{j=1}^{N} y^j(t) \le 1, \qquad t = 1, 2, \ldots, T \\
& \textstyle\sum_{t=1}^{T} y^j(t) \le 1, \qquad j = 1, 2, \ldots, N \\
& y^j(t) \in \{0, 1\}, \qquad t = 1, 2, \ldots, T,\ j = 1, 2, \ldots, N.
\end{aligned}$$

5.3.2 Price mechanism

A price mechanism is presented and shown to realize the goal. It is a variant of the price mechanisms appearing in Chapters 2 and 3. The optimality conditions of **P2** are the starting point of analysis. Let us take the dual of **P2**. Since the matrix defining constraints of **P2** is totally unimodular, we can replace $\{0, 1\}$ constraint by $[0, 1]$. Moreover, it is easy to see that the constraints giving the upper bound to the $y^j(t)$'s (i.e., $y^j(t) \le 1,\ \forall j$) are redundant. Thus, the following dual and the associated complementary slackness conditions may be used to argue optimality.

D2:

$$\begin{array}{lll} \min_{\lambda,\mu} & \sum_{t=1}^{T} \lambda(t) + \sum_{j=1}^{N} \mu^j & \\ \text{sub. to} & \lambda(t) + \mu^j \geq v^j(t), & t = 1,2,\ldots,T,\ j = 1,2,\ldots,N \\ & \lambda(t) \geq 0, & t = 1,2,\ldots,T \\ & \mu^j \geq 0, & j = 1,2,\ldots,N. \end{array}$$

The dual variable $\lambda(t)$ is the multiplier to the capacity constrains at time t, and μ^j is the multiplier to the constraint on user j's pipe usage.

The complimentary slackness conditions include

$$y^j(t) = 1 \quad \Rightarrow \quad \begin{cases} \mu^j = v^j(t) - \lambda(t) \\ y^j(s) = 0, \qquad \text{and} \quad \mu^j \geq v^j(s) - \lambda(s),\ s \neq t, \end{cases}$$

$$y^j(s) = 0, s = 1,2,\ldots,T \quad \Rightarrow \quad \mu^j = 0 \geq v^j(s) - \lambda(s),\ s = 1,2,\ldots,T$$

We can eliminate μ^j from the above to obtain:

The optimality condition for **P2**

An assignment $[y^j(t)]$ is optimal if and only if there is a nonnegative T-vector $\lambda := (\lambda(1), \lambda(2), \ldots, \lambda(T))$, such that

$$y^j(t) = 1 \quad \Rightarrow \quad v^j(t) - \lambda(t) \geq \max[0, \max_s(v^j(s) - \lambda(s))], \tag{5.2}$$

$$y^j(s) = 0, \forall s \quad \Rightarrow \quad v^j(s) - \lambda(s) \leq 0, \forall s. \tag{5.3}$$

This optimality condition suggests the following mechanism.

Price mechanism

Announced publicly is a pair comprising a price (or rent) vector $\lambda := (\lambda(1), \ldots, \lambda(T))$ and an assignment $[y^j(t)]$. Thus the message space of this mechanism is

$$\mathcal{M} := R_+^T \times \mathcal{A}.$$

User j considers $\lambda(t)$ as a rent of the pipe for time frame t, and maximizes his own profit over time. He agrees to a message if either (5.2) or (5.3) is satisfied. Otherwise, he rejects the message.

The outcome function is the projection of a message on the action space.

It is clear that the price mechanism realizes the goal.

5.3.3 Minimum size of message spaces

We will show that T, the number of time periods, is the minimum size of a message space of a maximal level set. Even though this is the size expected from the minimality results in Chapters 2 and 3, it does not follow from them. In those chapters, the proofs of the minimality results called for the optimal solution which spreads resources evenly to all participants, while here the resource is 'indivisible'.

We fix an assignment, and choose a subset E^* in its level set. E^* has the dimension T and the uniqueness property.

We choose an assignment which assigns the pipe to user 1 at time 1, user 2 at time 2, and so on. Users $T+1$ through N do not receive any assignment. Let $[\delta^j(t)]$ be this assignment, i.e.,

$$\delta^j(t) = \begin{cases} 1 & \text{if} \quad j = t, \\ 0 & \text{if} \quad j \neq t, \; t = 1, 2, \ldots, T. \end{cases}$$

Let $E^* \in F^{-1}([\delta^j(t)])$ be the set of environments in which every user has the same utility, i.e.,

$$E^* = \{(v, v, \ldots, v) \in E | v \in R_+^T\}, \tag{5.4}$$

where $v := (v(1), v(2), \ldots, v(T))$.

Proposition 5.3.1 *Defined as in (5.4), E^* has the uniqueness property with respect to goal F. Thus goal-realizing mechanisms have message spaces of at least dimension T.*

Proof Let $e := (z, \ldots, z)$, $\tilde{e} := (w, \ldots, w) \in E^*$ be such that $\tilde{e} \otimes_j e \in F^{-1}([\delta^j(t)])$, $j = 1, 2, \ldots, N$. We need to show that $w = z$.

First we will show that $w(1) \geq z(1)$. Consider $\tilde{e} \otimes_1 e = (w, z, z, \ldots, z)$. Assume $w(1) < z(1)$. Then the assignment which assigns the pipe to user N (who is not assigned to the pipe in $[\delta^j(t)]$) at time 1, and follows $[\delta^j(t)]$ for the rest of the time yields strictly better aggregate utility than $[\delta^j(t)]$. But this contradicts the optimality of $[\delta^j(t)]$ for problem instance $\tilde{e} \otimes_1 e$.

By a similar argument, we can show $w(1) \leq z(1)$, and hence $w(1) = z(1)$. Again by a similar argument, it can be shown that $w(t) = z(t)$, for all t. □

Unlike the multi-armed bandit problem, message exchanges cannot be spread over time without losing allocation efficiency in the matching problem. We will see that making a correct assignment at time 1 alone requires a message space of at least dimension T.

Let $F^{-1}(\delta^1(1))$ be the set of problem instances to which assigning the pipe to user 1 at time 1 is optimal. Clearly, $E^* \subseteq F^{-1}(\delta^1(1))$.

Let $e := (z, \ldots, z)$, $\tilde{e} := (w, \ldots, w) \in E^*$ be such that $\tilde{e} \otimes_j e, e \otimes_j \tilde{e} \in F^{-1}(\delta^1(1))$, $j = 1, 2, \ldots, N$. As in the proof of Proposition 5.3.1, we have $w(1) = z(1)$. Assume for $t \neq 1$, $w(t) > z(t)$, so $w(1) + z(t) < z(1) + w(t)$. Consider $\tilde{e} \otimes_1 e = (w, z, \ldots, z)$. By the inequality above, the assignment which assigns the pipe to user N at time 1, and to user 1 at time t, and assigns the pipe arbitrarily for the rest of the time outperforms the assignment which starts by assigning the pipe to user 1. But this is a contradiction, and hence $w(t) \leq z(t)$. By a similar argument, we have $w(t) \geq z(t)$. Thus $w(t) = z(t)$.

This shows that the task of assigning the pipe correctly at time 1 alone requires a message space of dimension at least T.

5.4 Possibility of Sequential Message Exchanges

The two cases we examined in this chapter exhibit quite a contrast; the multi-armed bandit problem admits spreading message exchanges over time so that only a scalar needs to be exchanged at each time, while the matching problem does not admit any sort of spreading at all. The difference is in the structures of optimal solutions of the respective *centralized* problems.

It would be nice to have a characterization of a problem structure which admits a sequential message exchanges, even if it were merely conceptual. The optimality principal of dynamic programming is at the heart of the following discussion.

Consider a centralized problem which requires making decisions over T time periods. Let E be a set of problem instances for the problem. Let $a(t)$ be a decision at time t, and $a := (a(1), \ldots, a(T))$. To simplify notation we write $a^{(t)} := (a(1), \ldots, a(t))$ for decisions made by time t. Let $\mathcal{A} := \mathcal{A}(1) \times \cdots \times \mathcal{A}(T)$ be an action space, where $\mathcal{A}(t)$ is a space of time t decision variables. Also, let $\mathcal{A}^{(t)} := \mathcal{A}(1) \times \cdots \times \mathcal{A}(t)$.[1] Let $F : E \to \mathcal{A}$ be the goal. For simplicity, we assume it is a function.

Level set of $a^{(t)}$ under F at time t is defined as the set of problem instances to which $a^{(t)}$ is a 'correct decision' during the first t time periods, and denoted by $F^{-1}(a^{(t)})$. More precisely, it is defined inductively (*backward in time*) as follows:

[1]The set of possible actions at time t could depend on actions taken by that time, say $\mathcal{A}(t, a^{(t-1)})$. $\mathcal{A}(t)$ might be thought of as $\mathcal{A}(t) := \bigcup_{a^{(t-1)} \in \mathcal{A}^{(t-1)}} \mathcal{A}(t, a^{(t-1)})$.

- for $t = T$, it is the level set of a under F, i.e.,

$$F^{-1}(a^{(T)}) := F^{-1}(a) = \{e \in E | F(e) = a\},$$

- for other t,

$$F^{-1}(a^{(t)}) := \bigcup_{b(t+1) \in \mathcal{A}(t+1)} F^{-1}(a^{(t)}, b(t+1)),\ t = T-1, T-2, \ldots, 1.$$

Note that the definitions have to be made backward in time in order not to be sorry for the current decision in future.

Now let us consider a decentralized mechanism which realizes F. Realizing the goal amounts to classifying a given problem instance (or environment) into the level set of the action to which it belongs. Thus for a mechanism to be goal-realizing, it is necessary to be able to classify a given instance into its proper level set at time t *by time* t.

We can consider the classification of problem instances into their proper level sets at time t as a goal of its own. The minimum size of message space needed to realize this goal gives a lower bound for amount of message exchanges needed by time t by a mechanism realizing F.

In the matching problem, it turns out that the classification into level sets at time 1 already requires at least as much message exchanges as the goal itself. Note this is inherent in the problem and there is nothing we can do in this regard.

The bandit problem has a particularly nice structure. The classification into level sets at time 2 can be done with 1-dimensional message exchanges, and within that level set, further classification into level sets at time 4 can be done with another 1-dimensional message exchanges, and so forth.

Chapter 6

Multi-armed Bandit Problem

6.1 Introduction

In the previous chapter, we had a glimpse of the multi-armed bandit problem. A stochastic version is studied in this chapter. As before, it is presented as an assignment problem of a digital pipe.

In a stochastic version, incremental rewards of pipe users are stochastic processes. The reward processes are (stochastically) independent of each other. As a user sends a packet, she learns more about her reward process. When she is not assigned to the pipe, her knowledge about her reward process remains unchanged (her knowledge or state is frozen). The task is to find an allocation which maximizes the expected aggregate reward.

This problem has received considerable attention with a good reason. The formulation covers a wide range of problems. For example, Gittins lists 'single machine scheduling', 'goldmining', 'industrial research', and other problems [8].

In 1972, Gittins and Jones[9] showed that, in a Markov control process framework, the optimal policy is obtained by

1. attaching an index to each user which is a function only of her own state, and

2. assigning the pipe to the user with the largest current index.

The index which Gittins termed *dynamic allocation index* is now, rightfully, called the *Gittins index* in the literature. The policy is generally referred as the index rule.

The index result is significant for our theme of decentralization of allocation processes. It means that, despite its apparent complexity, an optimal assignment can be sustained by exchanging one scalar at a time among participants. That is, our results in § 5.2 extend to the stochastic version.

In 1980, Whittle provided an elegant proof using dynamic programming, and what he called 'M-process'. Whittle's M-process will be related to the proof provided here in § 6.4.

Varaiya *et al.*[25] recognized that the following properties are essential for the optimality of the index rule:

1. the reward processes are independent,

2. states of those users who are not assigned to the pipe are frozen,

3. frozen users contribute no reward.

They extended the index result to non-Markovian cases. Their proof of the optimality of index rule is based on the interchange argument.

Mandelbaum[13] reformulated the problem as a control problem over a partially ordered set. He introduced the 'lower envelope' of the index processes. The lower envelope process is what we called a concave envelope of a reward process in § 5.2. It is a nonincreasing reward process 'equivalent' to the original reward process in the sense described there. It played a key role in his proof of optimality of the index rule for a continuous time multi-armed bandit problem[14].[1]

In this chapter, an alternative proof of the optimality of the index rule is obtained by utilizing properties of the equivalent reward processes. It adds a new insight to the problem. Weber[29] recently gave the very similar proof for a Markovian bandit.

In § 6.2 ideas of the proof are illustrated for a deterministic version of the problem in a discrete time setting. The same ideas are applied to show the asymptotic optimality of the index rule under the average reward criterion. They are also extended to a continuous time problem. The index is related to the Lagrange multiplier of the capacity constraint.

In § 6.3 stochastic counterparts of the problems in § 6.2 are studied. The difference is the addition of information constraints on admissible policies.

[1] He established the optimality of the index rule through the interchange argument for a discrete time bandit following Varaiya *et al.*[25]. It gave him an inequality involving the lower envelope processes which was central to his proof for the continuous time version.

In § 6.4 and § 6.5, superprocesses and an arm-acquiring bandit problem are discussed respectively. There we follow Varaiya *et al.*[25] closely.

6.2 Deterministic Bandit Problem

Ideas of the proof come from simple observations. They are best illustrated in a deterministic setting. A problem is an assignment of a digital pipe as in § 5.2. Users take turns to send their packets through a digital pipe. The pipe can send one packet per unit time. Utility of a user is represented as a series of incremental rewards. Unlike in § 5.2, users may have an infinite number of packets. The pipe is available over the infinite time horizon.

There are two criteria favored in the literature for an infinite horizon problem. One is the 'discounted reward' criterion, namely, to maximize the discounted total (aggregate) reward. The other is the 'average reward' criterion, namely, to maximize the time average of the reward earned over the infinite horizon. Both criteria are discussed below. Discrete time cases are studied in detail and the results are extended to continuous time problems.

In § 6.2.1, an assignment problem under the discounted reward criterion is examined in a discrete time setting. Key observations are made and steps of the proof are shown.

In § 6.2.2, a problem under the average reward criterion is examined. The asymptotic optimality of the index rule is derived.

In § 6.2.3, a multi-pipe case is considered. More than one digital pipe is available for the system. However, users are to send one packet at a time. They cannot occupy more than one pipe. The average reward criterion is considered. The result of § 6.2.2 is extended to this case. The derivations of asymptotic optimality are inspired by Weiss' work on parallel machines stochastic scheduling [30].

In § 6.2.4, the result of § 6.2.1 is extended to the continuous time counterpart. A hidden concavity of the problem, and a relation of the index and the Lagrange multiplier of the capacity constraint are revealed.

6.2.1 Discounted reward criterion: discrete time

We start our analysis with a problem under the discounted reward criterion in a discrete time setting.

User i is characterized by a **reward process** $Z^i := \{Z^i(l)\}_{l=1}^{\infty}$, where $Z^i(l)$ is the incremental reward of user i when her lth packet is sent.

Associated with clock time is a **discount factor** $\beta \in (0,1)$. A user's earning at time t is discounted by a factor of β^t. For example, when user i's lth packet is sent at time t, she receives incremental reward of $\beta^t Z^i(l)$ (in present value at time 0).

An **assignment** of the pipe is indicated by a variable $\Delta x^i(t)$. When user i is assigned to the pipe at time t, $\Delta x^i(t)$ takes value one, and zero otherwise. By a **load level** of user i at time t, we mean the cumulative number of packets of user i sent by time t, including the packet sent at t if user i is assigned to the pipe at t. It is denoted by $x^i(t)$. Thus $x^i(t) = \sum_{s=1}^{t} \Delta x^i(s)$. We set $x^i(0) \equiv 0$ and $Z^i(0) \equiv 0$ for all users. We also assume $\sum_{t=1}^{\infty} \beta^t |Z^i(t)| < \infty$ for all users.

Formally the problem is:

P1:

$$\begin{array}{lll} \max & \sum_{i=1}^{N} \sum_{t=1}^{\infty} \beta^t Z^i(x^i(t)) \Delta x^i(t) & \\ \text{sub. to} & \sum_{i=1}^{N} \Delta x^i(t) = 1, & t = 1,2,\ldots, \\ & \Delta x^i(t) \in \{0,1\}, & i = 1,2,\ldots,N,\ t = 1,2,\ldots. \end{array}$$

A **myopic policy**, assigning the pipe to the user with the largest current incremental reward at all times (breaking ties arbitrarily, if necessary), is optimal for the following special case.

Fact *When all reward processes are nonincreasing in their load levels, the myopic policy is optimal.*

This is intuitively clear and may be formally proved by the interchange argument. Note that when all reward processes are nonincreasing, the myopic policy is optimal regardless of the length of time horizon.

When reward processes are arbitrary sequences, future rewards have to be considered as well as the immediate reward. Note that the magnitude of discounting affects the trade-off between the immediate reward and future rewards.

The Gittins index, which may be viewed as the maximum attainable average reward rate (adjusted by the discount factor), resolves the dilemma. The idea is to assign the pipe to the user with the largest average reward rate at all times.

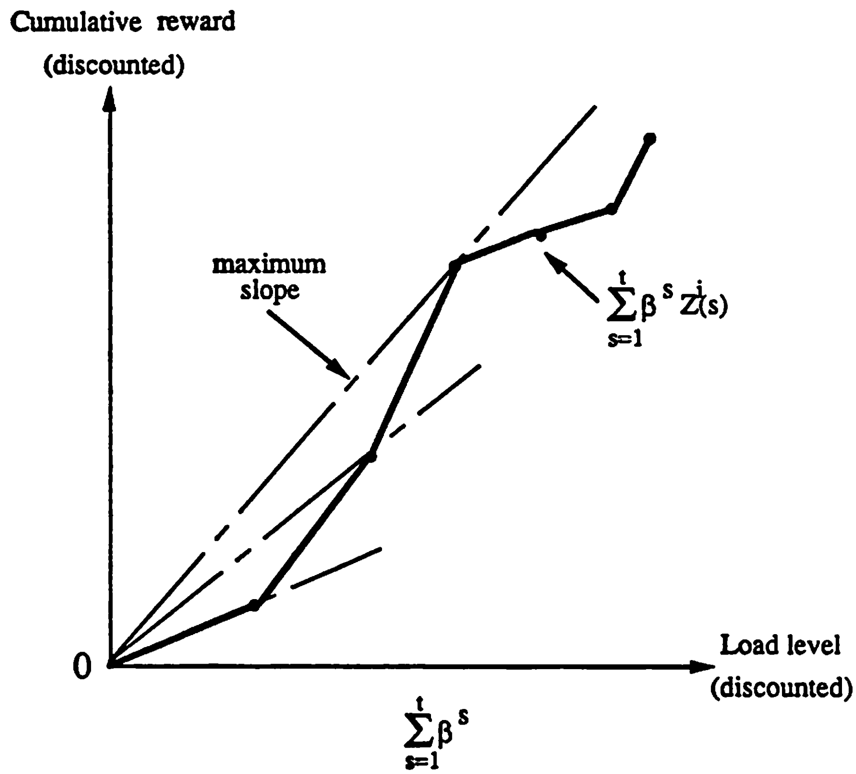

Figure 6.1. Index as the maximum slope

The index of user i at load level l is defined by

$$\nu^i(l+1) := \max_{\tau \geq l+1} \frac{\sum_{s=l+1}^{\tau} \beta^s Z^i(s)}{\sum_{s=l+1}^{\tau} \beta^s}, \quad l = 0, 1, \ldots. \tag{6.1}$$

The argument l appears in the definition in an awkward way, but it is meant to be consistent with the definition of the index for a stochastic version. The index at load level l, $\nu^i(l+1)$ is the largest attainable average reward rate from load level $l+1$ onward, computed after the lth packet is sent. Figure 6.1 illustrates how to find the index at load level 0.

Notice that the index is computed as if user i were the only user of the pipe, but the distinction between load level and clock time should be kept in mind.

The maximizing stopping time at load level l is the maximizer in the definition of the index (6.1) and denoted by $\tau^i(l+1)$. If there are two or more maximizers, the most immediate one is taken. The following fact about the maximizing stopping times can be proved:

$$\tau^i(l+1) = \inf\{s \geq l+1 | \nu^i(s+1) \leq \nu^i(l+1)\} \tag{6.2}$$

It says, the maximum average reward rate at any load level between l and $\tau^i(l+1)$ is greater than the rate at load level l. In other words, to stop when the future reward rate from that point onward is no greater than the rate we started with. A proof for a stochastic version is found in [25].

The **index rule** is the policy which assigns the pipe to the user with the largest index at the current load level at all times. A tie in the highest index can be broken arbitrarily without affecting the optimality of this policy.

The following process plays a key role in our proof. The **concave envelope** of the reward process of user i is defined by

$$\overline{Z}^i(l) := \min_{s \leq l} \nu^i(s), \; l = 1, 2, \ldots. \tag{6.3}$$

Figure 6.2 shows the picture of a reward process and its concave envelope. As seen there, the cumulative reward from this new process is the concave envelope to the cumulative reward of the original process. This process was first introduced by Mandelbaum [13] and called the lower envelope of the index process as the definition (6.3) suggests.

Properties of this process, which make our proof work, will be listed shortly in Lemma 6.2.1. But first we inductively define a sequence of stopping times $\{\tau_k^i\}_{k=0}^{\infty}$:

$$\tau_0^i \equiv 0 \tag{6.4}$$

$$\tau_{k+1}^i := \inf\{s > \tau_k^i | \nu^i(s+1) \leq \nu^i(\tau_k^i + 1)\}, \; k = 0, 1, \ldots. \tag{6.5}$$

This definition is made so that τ_{k+1}^i is the maximizing stopping time at load level τ_k^i, i.e., τ_{k+1}^i maximizes

$$\max_{\tau \geq \tau_k^i + 1} \frac{\sum_{s=\tau_k^i+1}^{\tau} \beta^s Z^i(s)}{\sum_{s=\tau_k^i+1}^{\tau} \beta^s}$$

over stopping times $\tau > \tau_k^i$. If for some k, $\tau_k^i = \infty$, we define $\tau_j^i := \infty$, for $j > k$. By the characterization of the maximizing stopping time (6.2) and the definition (6.3) of $\overline{Z}^i$,

$$\overline{Z}^i(s) = \nu^i(\tau_k^i + 1), \text{ for } \tau_k^i + 1 \leq s \leq \tau_{k+1}^i \tag{6.6}$$

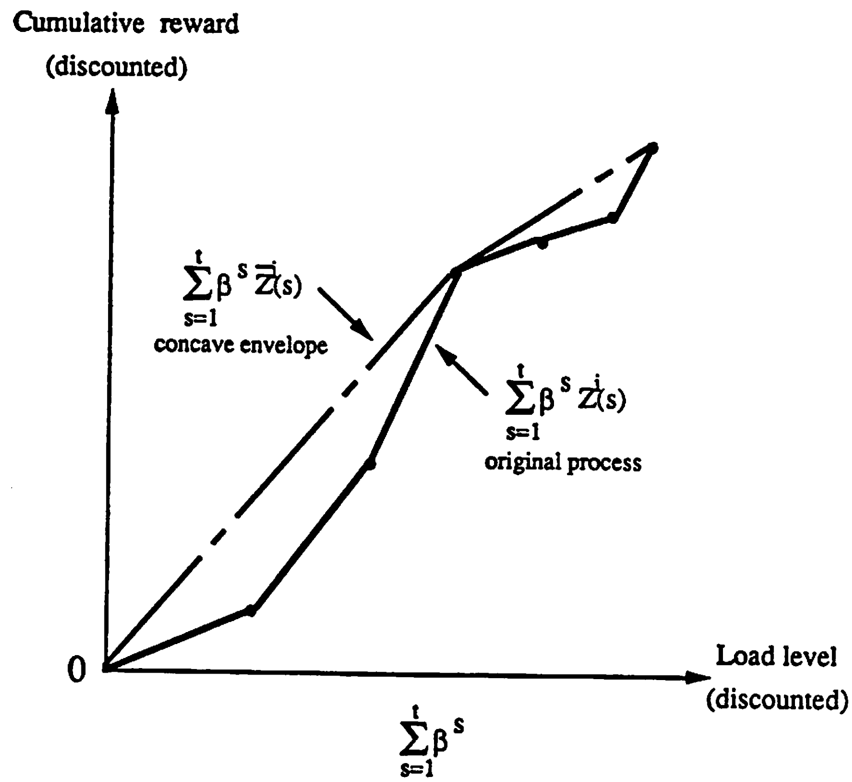

Figure 6.2. Original process and its concave envelope

Lemma 6.2.1 *The concave envelope has the following properties:*

1. $\overline{Z}^i(l)$ *is nonincreasing in* l.

2. *At any time/load level, the concave envelope yields at least as much cumulative reward as the original reward process,*

$$\sum_{s=1}^{l} \beta^s \overline{Z}^i(s) \geq \sum_{s=1}^{l} \beta^s Z^i(s), \; l = 1, 2, \ldots$$

3. *At the stopping times defined by (6.4) and (6.5), the both processes yield the same cumulative reward,*

$$\sum_{s=1}^{l} \beta^s \overline{Z}^i(s) = \sum_{s=1}^{l} \beta^s Z^i(s), \; \textit{for} \;\; l = \tau_k^i, k = 1, 2, \ldots$$

Proof All properties are fairly direct consequences of the definitions of the index (6.1) and the concave envelope process (6.3). They represent the properties illustrated in Figure 6.2.

Property 1 is immediate from (6.3).

Property 2 is derived as follows:

Let $s \wedge t$ denote the minimum of s and t. We write

$$\sum_{s=1}^{l} \beta^s Z^i(s) = \sum_{k=0}^{\infty} \sum_{s=\tau_k^i+1}^{l\wedge\tau_{k+1}^i} \beta^s Z^i(s).$$

For each k,

$$\begin{aligned}\sum_{s=\tau_k^i+1}^{l\wedge\tau_{k+1}^i} \beta^s Z^i(s) &\le \nu^i(\tau_k^i+1) \sum_{s=\tau_k^i+1}^{l\wedge\tau_{k+1}^i} \beta^s \\ &= \overline{Z}^i(\tau_k^i+1) \sum_{s=\tau_k^i+1}^{l\wedge\tau_{k+1}^i} \beta^s \\ &= \sum_{s=\tau_k^i+1}^{l\wedge\tau_{k+1}^i} \beta^s \overline{Z}^i(s)\end{aligned}$$

In the above, the inequality is by the definition of the index, and the equalities are by (6.6). Summing over k, we obtain Property 2.

Property 3 is obtained by noting that when $l = \tau_k^i$ for some k, the inequality above becomes equality for each $j \le k$, again by the definition of the index. □

So far, we have converted the original reward processes to their concave envelopes. The next step is to splice the original reward processes, and to compare them with the processes obtained by splicing the concave envelopes in the same manner. Let Π be the set of feasible policies (i.e., the set of assignment patterns) of **P1**, and let π be a policy in it. Also let $V(Z^1, Z^2, \ldots, Z^N; \pi)$ be the total discounted reward earned by splicing processes $(Z^1, Z^2, \ldots, Z^N)$ according to policy π.

Our proof of the optimality of the index rule proceeds in the following steps:

Claim 1. *When the concave envelopes are seen as reward processes (rather than the original processes), the myopic policy is optimal, i.e., denoting the myopic policy by* π^*,

$$V(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N; \pi^*) = \max_{\pi\in\Pi} V(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N; \pi).$$

The myopic policy π^* for the envelope processes is precisely the index rule for the original problem.

Claim 2. *When the same policy is applied to two problem instances, one with original reward processes, the other with their concave envelopes, the latter yields at least as much total reward, i.e.,*

$$V(Z^1, Z^2, \ldots, Z^N; \pi) \leq V(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N; \pi), \ \forall \pi \in \Pi.$$

Therefore $V(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N; \pi^*)$ gives an upper bound to the optimal total reward from the original processes.

Claim 3. *When π^* is applied to the original processes, it yields the same total reward as it does from the envelope processes, i.e.,*

$$V(Z^1, Z^2, \ldots, Z^N; \pi^*) = V(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N; \pi^*).$$

Since the upper bound is attained, the index rule is optimal.

The first claim follows from monotonicity of the concave envelopes and the characterization of the maximizing stopping time (6.2).

We verify the second claim. It is a consequence of Property 2. The following lemma is useful.

Lemma 6.2.2 *Let $\{X(t)\}_{t=1}^{\infty}$ be a process such that*

$$\sup_{\tau > 0} \sum_{s=1}^{\tau} X(s) \leq 0,$$

and let $\{\gamma(t)\}_{t=1}^{\infty}$ be a nonincreasing sequence such that

$$1 \geq \gamma(1) \geq \gamma(2) \geq \ldots \geq 0.$$

Then,

$$\sum_{t=1}^{\infty} \gamma(t) X(t) \leq 0.$$

A proof for a stochastic version of this lemma is found in [25].

The next step is to convert our problem to the setting of the lemma above through a change of variables from clock time to load level.

Lemma 6.2.3 *Let $x := \{x(t)\}_{t=0}^{\infty}$ be an arbitrary load pattern of user i over the infinite horizon. ($x(0) \equiv 0$, and $\Delta x(t) = x(t+1) - x(t) \in \{0, 1\}$.)*

When x is applied to both the original process Z^i and its concave envelope $\overline{Z}^i$, $\overline{Z}^i$ yields at least as much total reward as Z^i does, i.e.,

$$\sum_{t=1}^{\infty} \beta^t (Z^i(x(t)) - \overline{Z}^i(x(t)))\Delta x(t) \leq 0.$$

Proof Let x^{-1} be the left inverse of the load pattern:

$$x^{-1}(l) := \inf\{t > 0 | x(t) \geq l\}, \quad l = 1, 2, \ldots.$$

That is, $x^{-1}(l)$ is the first time the load level reaches l. If l is never reached, we set $x^{-1}(l) := \infty$. Observe that $x^{-1}(l) - l$ is nonnegative and nondecreasing in l. Let us use the convention $\beta^\infty \equiv 0$.

$$\begin{aligned}
\sum_{t=1}^{\infty} \beta^t (Z_i(x(t)) - \overline{Z}_i(x(t)))\Delta x(t) &= \sum_{l=1}^{\infty} \beta^{x^{-1}(l)} (Z_i(l) - \overline{Z}_i(l)) \\
&= \sum_{l=1}^{\infty} \beta^{(x^{-1}(l)-l)} \beta^l (Z_i(l) - \overline{Z}_i(l))
\end{aligned}$$

Take $X(l) := \beta^l (Z_i(l) - \overline{Z}_i(l))$ and $\gamma(l) := \beta^{(x^{-1}(l)-l)}$ in Lemma 6.2.2.

$X(t)$ satisfies the condition of Lemma 6.2.2 because of Property 2 in Lemma 6.2.1, and so does $\gamma(l)$ by the observation made above. □

Since any feasible assignment is a combination of the type of assignments examined above, the second claim is verified.

The third claim is a direct consequence of Property 3 in Lemma 6.2.1.

Thus we have proved the optimality of the index rule:

Theorem 6.2.1 *The index rule achieves the optimal total reward for problem* **P1**.

Remark

We can imagine the following fictitious auction scheme to interpret a concave envelope and the index rule. An auction is held by a pipe owner. Auctioned is the right to use the pipe. A bid price is interpreted as a rent per unit time that a user is going to pay if he wins the bid. The highest bidder is allowed to use the pipe as long as he wishes, provided he keeps paying the rent. When it comes to the point that he incurs a loss if he keeps the pipe with the current rent, he simply returns the pipe to the owner. The owner then holds another auction and repeats the procedure. The concave envelope is then interpreted as the user's highest possible bids (and the subsequent rents) as his packets are sent out. Naturally his bids will decrease as more of his packets are sent out. The pipe owner's profit is maximized by taking the highest bid whenever the pipe becomes available.

Refer to Weber's interpretation of the concave envelope as a 'fair charge' of a bandit [29].

6.2.2 Average reward criterion

A bandit problem with a deadline is considered first. The pipe is available only for T ($< \infty$) time frames. There is no discounting.

The index rule is no longer optimal for a finite horizon problem. Concave envelopes are utilized to obtain a bound on the difference between the optimal total reward and the total reward earned under the index rule. When this bound is independent of the length of time horizon, it leads to the asymptotic optimality of the index rule under the average reward criterion.

To illustrate suboptimality of the index rule let us examine the following example:

> A pipe is available for $T = 4$ time frames.
> Users 1,2, and 3 are characterized by $Z^1 = Z^2 = (0,4,0,0)$, and $Z^3 = (0,0,7,0)$.

Each of user 1 and 2 has two packets to send and user 3 has 3 packets, and the rewards are earned only when all packets of respective users are sent.

The index rule suggests to serve user 3 and earn the reward of 7, but the optimal policy is to serve users 1 and 2 and earn the total reward of 8. In this example, the cause of suboptimality is the inefficient utilization of the pipe near the end of the service duration. It is conceivable that when each user has a relatively small number of packets compared to T, the index rule does not do too badly compared to an optimal policy. The difference in total reward will be bounded by the reward from the last job that the index rule started but did not finish.

The purpose of this subsection is to make those notions precise. The problem is

P2:

$$\begin{aligned} \max \quad & \sum_{i=1}^{N} \sum_{t=1}^{T} Z^i(x^i(t)) \Delta x^i(t) & \\ \text{sub. to} \quad & \sum_{i=1}^{N} \Delta x^i(t) \leq 1, & t = 1,2,\ldots,T \\ & \Delta x^i(t) \in \{0,1\}, & i = 1,2,\ldots,N,\ t = 1,2,\ldots,T. \end{aligned}$$

The index is defined as before:

$$\nu^i(l+1) := \max_{\tau \geq l+1} \frac{\sum_{s=l+1}^{\tau} Z^i(s)}{\tau - l}, l = 0, 1, \ldots. \qquad (6.7)$$

The concave envelopes are defined with a slight modification:

$$\overline{Z}^i(l) := \max[0, \min_{s \leq l} \nu^i(s)], \ l = 1, 2, \ldots.$$

When reward processes are positive, the capacity constraint in **P2** will be tight, and the above definition of the concave envelope reduces to the one given by (6.3). Here the pipe owner is allowed to stop renting the pipe. This may be thought of as having the reward process with 0 increments.

The maximizing stopping time is characterized as in (6.2).

Define a sequence of stopping times $\{\tau_k^i\}_{k=0}^{\infty}$ as in (6.4) and (6.5).

The *index rule* is the policy which assigns the pipe to the user with the largest index as long as there are users with positive indices. When there are no users with positive indices, it stops assigning the pipe.

Clearly, Properties 1 and 2 of the concave envelopes in Lemma 6.2.1 hold. Property 3 holds for the τ_k^i's before the index takes negative value. Hence, Claim 1 and 2 of the previous subsection remains valid. The following claim replaces Claim 3.

Claim 3′. *Let ΔV be the difference between the optimal total reward and the total reward earned under the index rule for a finite horizon problem. Then,*

$$\Delta V \leq V(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N; \pi^*) - V(Z^1, Z^2, \ldots, Z^N; \pi^*)$$

It is more convenient to have a bound in terms of τ_k^i and the index. Let κ be the user whose packet is sent at time T. Let l^i be the load level of user i at T (including the one delivered at T for user κ) under π^*. Also let τ_n^κ be such that $\tau_n^\kappa < l^\kappa \leq \tau_{n+1}^\kappa < \infty$. Then,

$$\begin{aligned}
\Delta V &\leq V(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N; \pi^*) - V(Z^1, Z^2, \ldots, Z^N; \pi^*) \\
&= \sum_{i=1}^{N} \sum_{s=1}^{l^i} \overline{Z}^i(s) - \sum_{i=1}^{N} \sum_{s=1}^{l^i} Z^i(s) \\
&\leq \sum_{i=1}^{N} \sum_{s=1}^{l^i} \overline{Z}^i(s) + \overline{Z}^\kappa(l^\kappa)(\tau_{n+1}^\kappa - l^\kappa) - \sum_{i=1}^{N} \sum_{s=1}^{l^i} Z^i(s) \\
&= \sum_{i=1}^{N} \sum_{s=1}^{l^i} Z^i(s) + \sum_{s=l^\kappa+1}^{\tau_{n+1}^\kappa} Z^\kappa(s) - \sum_{i=1}^{N} \sum_{s=1}^{l^i} Z^i(s)
\end{aligned}$$

$$= \sum_{s=l^\kappa+1}^{\tau^\kappa_{n+1}} Z^\kappa(s) \tag{6.8}$$

Let us impose uniform bounds on stopping times and indices.

Assumption 6.2.1 *We assume*

1. $\tau^i(l+1) - l \le D$, *for* $i = 1,2,\ldots,N$, $l = 1,2,\ldots$

2. $\nu^i(l) \le W$, *for* $i = 1,2,\ldots,N$, $l = 1,2,\ldots$

We can think of a situation that there are several groups of identical users. For such an instance, the assumption above is not an unreasonable one.

Under the assumption above, we have

$$\Delta V \le WD, \tag{6.9}$$

since

$$\sum_{s=l^\kappa+1}^{\tau^\kappa_{n+1}} Z^\kappa(s) \le \nu^\kappa(l^\kappa+1)(\tau^\kappa_{n+1} - l^\kappa) \le WD$$

by the definition of index at load level l^κ.

Since the bound on ΔV does not depend on the length of horizon,

$$\lim_{T\to\infty} \frac{\Delta V}{T} = 0.$$

Therefore, we have

Theorem 6.2.2 *Under Assumption 6.2.1, the index rule is asymptotically optimal under the average reward criterion.*

6.2.3 Multiple pipes

In the previous subsection, we saw that the suboptimality of the index rule was due to the inefficiency near the end of service duration. The same can be said when more than one pipe is available. We derive a bound on the difference between the optimal total reward and the total reward from an index rule as in the previous subsection. Let M (≥ 2) be the number of digital pipes available. The problem is

P3:

$$\begin{array}{lll} \max & \sum_{i=1}^{N} \sum_{t=1}^{T} Z^i(x^i(t))\Delta x^i(t) & \\ \text{sub. to} & \sum_{i=1}^{N} \Delta x^i(t) \le M, & t = 1,2,\ldots,T \\ & \Delta x^i(t) \in \{0,1\}, & i = 1,2,\ldots,N,\ t = 1,2,\ldots,T. \end{array}$$

Let us pose the following conditions on the reward processes.

Assumption 6.2.2 *1. Each user has a finite number of packets, σ^i, to send and*

$$\sigma^i \le D,\ i = 1,2,\ldots,N.$$

2. Indices are bounded, i.e.,

$$\nu^i(l) \le W,\ l = 1,2,\ldots,\sigma^i,\ i = 1,2,\ldots,N.$$

The first assumption here is much stronger than the first assumption in Assumption 6.2.1. It may be weakened in a specific application.

An *index rule* is a policy which assigns the pipes to the users with M highest indices as long as there are users with positive indices. Unlike the single-pipe case, a tie-breaking rule affects the total reward. Examine the following example:

Two pipes ($M = 2$) are available for $T = 2$ time frames.
Users 1,2, and 3 are characterized by $Z^1 = Z^2 = (1,0)$, and $Z^3 = (1,1)$.

In this example, all users have the same index 1 at the beginning, but sending packets of users 1 and 2 at time 1 results in a suboptimal assignment.

When all reward processes are nonincreasing, one may suspect that an optimal policy is found among index rules. But that is not the case as the following example illustrates.

Two pipes are available for $T = 2$ time periods.
Users 1,2, and 3 are characterized by $Z^1 = Z^2 = (2,0)$, and $Z^3 = (1,1)$.

Even though user 3 has smaller index at the beginning, it is optimal to send his packet at time 1 along with a packet of user 1 or 2.

As we saw through these examples, Claim 1 in § 6.2.1 does not make sense anymore. But the first half of Claim 2 is still valid, i.e.,

$$V(Z^1, Z^2, \ldots, Z^N; \pi) \le V(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N; \pi),$$

where π is any feasible policy for **P3**, and V denotes the total reward as before. For simplicity, assume there are enough packets and users (and positive rewards) so that none of the pipes become idle at time T. Let $\kappa_1, \kappa_2, \ldots, \kappa_M$ be the users whose packets are sent at time T under an index policy. Let l^i be the load level of user i after the delivery at T. By definition of an index policy, none of the users who have unsent packets and are not assigned to the pipes have greater current indices than those who are assigned to the pipe at T. This observation leads to the following lemma:

Lemma 6.2.4 *Let π be an index policy and κ_m's and l^i's be defined as above. If π were allowed to finish sending packets of users κ_1 through κ_M and collect rewards from them accordingly, it would yield at least as much total reward as an optimal policy, say π^*, would, i.e.,*

$$V(Z^1, Z^2, \ldots, Z^N; \pi^*) \leq V(Z^1, Z^2, \ldots, Z^N; \pi) + \sum_{m=1}^{M} \sum_{s=l^{\kappa_m}+1}^{\sigma^{\kappa_m}} Z^{\kappa_m}(s). \tag{6.10}$$

Proof By property 2 of the concave envelope in Lemma 6.2.1,

$$V(Z^1, Z^2, \ldots, Z^N; \pi^*) \leq V(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N; \pi^*).$$

We will argue that

$$V(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N; \pi^*) \leq V(\overline{Z}^1, \overline{Z}^2, \ldots, \overline{Z}^N; \pi) + \sum_{m=1}^{M} \sum_{s=l^{\kappa_m}+1}^{\sigma^{\kappa_m}} \overline{Z}^{\kappa_m}(s). \tag{6.11}$$

Since the right hand side of (6.11) is equal to the right hand side of (6.10) by property 3 in Lemma 6.2.1, this proves the lemma.

Let us regard the $\overline{Z}^i$'s as reward processes (nonincreasing reward sequences). Note that the optimal total reward from a single-pipe case with the deadline MT is at least as large as the optimal total reward from M pipes with the deadline T.

Let us assume that κ_1 is the user with the smallest immediate reward among the κ_m's, i.e.,

$$\overline{Z}^{\kappa_1}(l^{\kappa_1}) = \min_{1 \leq m \leq M} \overline{Z}^{\kappa_m}(l^{\kappa_m}).$$

Let τ^{κ_m} be such that

$$\overline{Z}^{\kappa_m}(\tau^{\kappa_m} + 1) \leq \overline{Z}^{\kappa_1}(l^{\kappa_1}), \text{ for } m = 2, \ldots, M.$$

Let $l^{\kappa_m} \vee \tau^{\kappa_m}$ be the larger of the two. If π were allowed to continue until $l^{\kappa_m} \vee \tau^{\kappa_m}$ were reached for each user $(\kappa_2, \ldots, \kappa_M)$ and would collect the associated rewards, then π would yield as much total reward as an optimal policy for a single-pipe problem with the deadline $MT + \sum_{m=2}^{M}(l^{\kappa_m} \vee \tau^{\kappa_m} - l^{\kappa_m})$ would yield. Because all the rewards greater than $\overline{Z}^{\kappa_1}(l^{\kappa_1})$ would be collected, and no reward less than $\overline{Z}^{\kappa_1}(l^{\kappa_1})$ would be included under π in that total pipetime. Thus,

$$\begin{aligned} V(\overline{Z}^1,\overline{Z}^2,\ldots,\overline{Z}^N;\pi^*) &\le V(\overline{Z}^1,\overline{Z}^2,\ldots,\overline{Z}^N;\pi) + \sum_{s=2}^{M}\sum_{s=l^{\kappa_m}+1}^{\tau^{\kappa_m}} \overline{Z}^{\kappa_m}(s) \\ &\le V(\overline{Z}^1,\overline{Z}^2,\ldots,\overline{Z}^N;\pi) + \sum_{m=1}^{M}\sum_{s=l^{\kappa_m}+1}^{\sigma^{\kappa_m}} \overline{Z}^{\kappa_m}(s), \end{aligned}$$

as desired. □

Assumption 6.2.2 and the lemma above give the following bound on the error of an index policy:

$$\begin{aligned} \Delta V^{\pi} &:= V(Z^1,Z^2,\ldots,Z^N;\pi^*) - V(\overline{Z}^1,\overline{Z}^2,\ldots,\overline{Z}^N;\pi) \\ &\le \sum_{m=1}^{M}\sum_{s=l^{\kappa_m}+1}^{\sigma^{\kappa_m}} Z^{\kappa_m}(s) \\ &\le \sum_{m=1}^{M} \nu^{\kappa_m}(l^{\kappa_m}+1)(\sigma^{\kappa_m} - l^{\kappa_m}) \\ &\le MWD \end{aligned}$$

The second inequality is by the definition of the indices.

The asymptotic optimality of an index rule under the average reward criterion follows.

Let us now assume that the number of the available pipes changes over time, say, $M(t), t = 1,\ldots,T$. And let $M := \max_t M(t)$. When an index rule is applied, there are at most M users such that $\tau^i_{n^i} < l^i \le \tau^i_{n^i+1}$, where $\{\tau^i_{n^i}\}$ is the sequence of the maximizing stopping times defined by (6.4) and (6.5). Note those users have greater indices at T than the others. An argument similar to the one above leads to $\Delta V^{\pi} \le MWD$.

6.2.4 Discounted reward criterion: continuous time

When a pipe owner has many pipes and the ability to switch among users in relatively short time, the problem will be well approximated by treating both load level and time as continuous variables. The problem instance is stated below and the the optimality of the index rule is proved through similar steps as in § 6.2.1. At the end of this subsection, the index is related to the Lagrange multiplier to the capacity constraint.

A reward process is defined by $\{Z^i(l), l \geq 0\}$, where l is the load level as before except that it is continuous rather than discrete. When the load level of user i is $l(t)$ at time t, the prediscounted reward in a small time interval $[t, t+dt]$ is approximated by $Z^i(l(t))\dot{l}(t)dt$. Associated with the clock time is a **discount rate** $\alpha > 0$. A user's earning at time t is discounted at the rate $e^{-\alpha t}$.

An *allocation* of the pipe capacity is indicated by a variable $\dot{x}^i(t)$. By the *load level* of user i at time t, we mean the cumulative amount of capacity allocated to user i by time t. It is denoted by $x^i(t)$. Thus $x^i(t) = \int_0^t \dot{x}^i(t)dt$. Let us set $x^i(0) \equiv 0$ and $Z^i(0) \equiv 0$ for all users. Let M be a fixed capacity (rate) limit. At any time instance, the sum of load rates is not allowed to exceed this limit.

We assume that the $Z^i(\cdot)$ are continuous and $\int_0^\infty e^{-\alpha t}|Z^i(t)|dt < \infty$.

The problem is

P4:

$$\begin{aligned} \max \quad & \int_0^\infty e^{-\alpha t} \sum_{i=1}^N Z^i(x^i(t))\dot{x}^i(t)dt \\ \text{sub. to} \quad & \sum_{i=1}^N \dot{x}^i(t) \leq M, \ t \geq 0, \\ & \dot{x}^i(t) \geq 0, \ i = 1, 2, \ldots, N, \ t \geq 0. \end{aligned}$$

Since $\dot{x}^i(t) \leq M$, and $x^i(\cdot)$ is nondecreasing, $x^i(\cdot)$ is absolutely continuous and $\dot{x}^i(\cdot)$ is defined almost everywhere.

The **index of user i at load level** l is defined by

$$\nu^i(l) := \sup_{\tau > l} \frac{\int_l^\tau e^{-\alpha s/M} Z^i(s)ds}{\int_l^\tau e^{-\alpha s/M}ds} \tag{6.12}$$

where l, τ and s are load levels.

The concave envelope of reward process of user i is defined as before:

$$\overline{Z}^i(l) := \max[0, \inf_{0 \leq s \leq l} \nu^i(s)], \ l \geq 0. \tag{6.13}$$

The concave envelope has similar properties as its discrete counterpart.

Properties of an envelope process

1. $\overline{Z}^i(l)$ is nonincreasing in l.
2. $\int_0^l e^{-\alpha s/M} \overline{Z}^i(s)ds \geq \int_0^l e^{-\alpha s/M} Z^i(s)ds, \ l \geq 0$.
3. $\int_0^l e^{-\alpha s/M} \overline{Z}^i(s)ds = \int_0^l e^{-\alpha s/M} Z^i(s)ds$, for l such that $\overline{Z}(l)$ is strictly decreasing.

Property 1 is immediate from the definition of the envelope processes. Properties 2 and 3 may be obtained by appropriately discretizing and applying the results in § 6.2.1, and taking limits.

An *index rule* is the policy which allocates the capacity to the users with the highest indices as long as there are users with positive indices.

Optimality of the index rule is proved through the same three steps as in § 6.2.1. Let us prove the second claim for the continuous-time version. The continuous-time version of Lemma 6.2.2 is useful.

Lemma 6.2.5 *Let $\{X(t), t \geq 0\}$ be a process such that*

$$\sup_{T>0} \int_0^T X(t)dt \leq 0,$$

and let $\{\gamma(t), t \geq 0\}$ be a nonincreasing process such that

$$1 \geq \gamma(t) \geq \gamma(s) \geq 0, t \geq s.$$

Then,

$$\int_0^\infty \gamma(t)X(t)dt \leq 0.$$

The next step is to prove the continuous-time version of Lemma 6.2.3 by the change of variables.

Lemma 6.2.6 *Let $\{x(t), t \geq 0\}$ be an arbitrary load pattern of user i over time. ($x(0) \equiv 0$, and $0 \leq \dot{x}(t) \leq M$.)*

When $x(\cdot)$ is applied to both the original process Z^i and its concave envelope $\overline{Z}^i$, $\overline{Z}^i$ yields at least as much total reward as Z^i does, i.e.,

$$\int_{t=0}^\infty e^{-\alpha s/M}(Z^i(s) - \overline{Z}^i(s))ds \leq 0.$$

Proof Let the left inverse of the load pattern be x^{-1}:

$$x^{-1}(l) := \inf\{t > 0 | x(t) \geq l\}, \quad l \geq 0.$$

If l is never reached, set $x^{-1}(l) = \infty$. Observe that $x^{-1}(l) - l/M$ is nonnegative and increasing in l. Let us use the convention $e^{-\infty} \equiv 0$.

$$\begin{aligned}
\int_{t=0}^\infty e^{-\alpha s/M}(Z^i(s) - \overline{Z}^i(s))ds &= \int_{t=0}^\infty e^{-\alpha x^{-1}(l)}(Z^i(l) - \overline{Z}^i(l))dl \\
&= \int_{t=0}^\infty e^{-\alpha(x^{-1}(l)-l/M)}e^{-\alpha l/M}(Z^i(l) - \overline{Z}^i(l))dl
\end{aligned}$$

Take $X(t) := e^{-\alpha l/M}(Z^i(l) - \overline{Z}^i(l))$ and $\gamma(l) := e^{-\alpha(x^{-1}(l) - l/M)}$ in the previous lemma. □

To complete the proof of the optimality of the index rule, we need to show that the index rule does maintain a tie in the largest index while allocating the capacity to those users with the largest index. Refer to Mandelbaum[14] for details.[2]

Theorem 6.2.3 *The index rule achieves the optimal total reward for problem* **P4**.

We observe a hidden concavity property of **P4**, and relate the concave envelopes to the Lagrange multiplier of the capacity constraint through the *dual* problem.

Let $R^i(Z^i, x)$ be the total reward of user i with reward process $\{Z^i(l), l \geq 0\}$ under an allocation $x(\cdot)$, i.e.,

$$R^i(Z^i, x) := \int_0^\infty e^{-\alpha t} Z^i(x(t))\dot{x}(t)dt$$

Let $\mathcal{X}$ be defined by $\{x | x(0) \equiv 0 \text{ and } 0 \leq \dot{x}(t) \leq M, t \geq 0\}$.

Lemma 6.2.7 *$R^i(Z^i, x)$ is concave in $x(\cdot) \in \mathcal{X}$ if and only if $Z^i(\cdot)$ is nonincreasing in load level.*

Proof First we assume $Z^i(l)$ is nonincreasing in l. let $x_1(\cdot), x_2(\cdot) \in \mathcal{X}$ and $\theta \in [0,1]$. Since allocations are nondecreasing in time, we have

$$Z^i(\theta x_1(t) + (1-\theta)x_2(t)) \leq Z^i(\theta x_1(t)) \leq Z^i(x_1(t)), t \geq 0.$$

Similarly, $Z^i(\theta x_1(t) + (1-\theta)x_2(t)) \leq Z^i(x_2(t))$. Therefore,

$$\begin{aligned}
R^i(Z^i, \theta x_1 + (1-\theta)x_2) &= \int_0^\infty e^{-\alpha t} Z^i(\theta x_1(t) + (1-\theta)x_2(t))(\theta\dot{x}_1(t) + (1-\theta)\dot{x}_2(t))dt \\
&= \int_0^\infty e^{-\alpha t} Z^i(\theta x_1(t) + (1-\theta)x_2(t))\theta\dot{x}_1(t)dt + \\
&\quad \int_0^\infty e^{-\alpha t} Z^i(\theta x_1(t) + (1-\theta)x_2(t))(1-\theta)\dot{x}_2(t))dt \\
&\leq \theta \int_0^\infty e^{-\alpha t} Z^i(x_1(t))\dot{x}_1(t)dt + (1-\theta)\int_0^\infty e^{-\alpha t} Z^i(x_2(t))\dot{x}_2(t)dt \\
&= \theta R^i(Z^i, x_1) + (1-\theta)R^i(Z^i, x_2)
\end{aligned}$$

Thus R is concave in load pattern.

[2] Maintaining a tie involves knowing how fast indices are changing. Hence the additional information needs to be communicated.

Conversely, assume R is concave in load pattern. Let $0 \le s_1 < s_2$, and $x_1(\cdot)$ and $x_2(\cdot)$ be such that

$$\dot{x}_1(t) := \begin{cases} 1, & \text{for } 0 \le t \le s_1 \\ 0, & \text{for } t > s_1 \end{cases}$$

$$\dot{x}_2(t) := \begin{cases} 1, & \text{for } 0 \le t \le s_2 \\ 0, & \text{for } t > s_2 \end{cases}$$

Then

$$R^i(Z^i, \theta x_1 + (1-\theta)x_2) = \int_0^{s_1} e^{-\alpha t} Z^i(t)dt + (1-\theta)\int_{s_1}^{s_2} Z^i((1-\theta)t)dt$$

and

$$\theta R^i(Z^i, x_1) + (1-\theta)R^i(Z^i, x_2) = \theta\int_0^{s_1} e^{-\alpha t} Z^i(t)dt + (1-\theta)\int_0^{s_2} e^{-\alpha t} Z^i(t)dt$$

Therefore, concavity implies

$$\int_{s_1}^{s_2} e^{-\alpha t}\{Z^i((1-\theta)t) - Z^i(t)\}dt \ge 0.$$

Since θ, s_1, and s_2 are arbitrary, Z^i is nonincreasing in load level. □

Remark

An (incremental) reward process $Z^i(l)$ can be thought of a derivative of $U(l) := \int_0^l Z^i(s)ds$. When $Z^i(l)$ is nondecreasing, $U(l)$ is a concave function of l. The lemma above is a consequence of this fact.

From the lemma above and the definition of $\overline{Z}^i$, we see that when the original processes are replaced by their concave envelopes in problem **P4**, we have a concave program. We saw that the optimal solution to the concave envelope version is the optimal solution to the original problem, and yields the same total reward. The recognition of this hidden concavity leads to the dual formulation. Let us define the Lagrangian by

$$\begin{aligned} L(\dot{x}(\cdot), \lambda(\cdot)) \;&:=\; \int_0^\infty e^{-\alpha t}\sum_{i=1}^N Z^i(x^i(t))\dot{x}^i(t)dt - \int_0^\infty e^{-\alpha t}\lambda(t)(\sum_{i=1}^N \dot{x}^i(t) - M)dt \\ &=\; \int_0^\infty e^{-\alpha t}[\sum_{i=1}^N \dot{x}^i(t)\{Z^i(x^i(t)) - \lambda(t)\} + M\lambda(t)]dt, \end{aligned} \tag{6.14}$$

where $\dot{x}(\cdot) := (\dot{x}^1(\cdot), \ldots, \dot{x}^N(\cdot))$. Define the functional $G(\lambda)$ by

$$G(\lambda) := \sup_{\dot{x} \ge 0} L(\dot{x}, \lambda(\cdot)) \tag{6.15}$$

Above, the supremum is taken over $\dot{x}^i \geq 0$ but free of the constraint $\sum_{i=1}^N \dot{x}^i \leq M$.

The dual of **P4** is defined by

D4:

$$\inf_{\lambda \geq 0} G(\lambda) \tag{6.16}$$

The weak duality relation follows from the definition: let $\dot{y}$ and η be feasible solutions to **P4** and **D4** respectively, then

$$\begin{aligned} G(\eta) &= \sup_{\dot{x}} L(\dot{x}, \eta) \\ &\geq L(\dot{y}, \eta) \\ &= \int_0^\infty e^{-\alpha t} \sum_{i=1}^N Z^i(y^i(t)) \dot{y}^i(t) dt - \int_0^\infty e^{-\alpha t} \lambda(t) (\sum_{i=1}^N \dot{y}^i(t) - M) dt, \end{aligned}$$

since $\dot{y}$ and η are feasible, $\lambda(t)(\sum_{i=1}^N \dot{y}^i(t) - M) \geq 0$ for all t, and

$$L(\dot{y}, \eta) \geq \int_0^\infty e^{-\alpha t} \sum_{i=1}^N Z^i(y^i(t)) \dot{y}^i(t) dt.$$

Moreover, the strong duality relation holds. To see that replace Z^i by $\overline{Z}^i$. Let x^* be an allocation under an index rule, and set the corresponding multiplier by

$$\lambda^*(t) := \max_{i \in \{1, \ldots, N\}} \overline{Z}^i(x^{*i}(t)).$$

Note

$$G(\lambda^*) = \sup_{\dot{x}} \int_0^\infty e^{-\alpha t} [\sum_{i=1}^N \dot{x}^i(t) \{\overline{Z}^i(x^i(t)) - \lambda^*(t)\} + M \lambda^*(t)] dt$$

is maximized by x^*. Thus,

$$\int_0^\infty e^{-\alpha t} \sum_{i=1}^N \overline{Z}^i(x^{*i}(t)) \dot{x}^{*i}(t) dt = L(x^*, \lambda^*) = G(\lambda^*) = M \int_0^\infty e^{-\alpha t} \lambda^*(t) dt.$$

Remarks

1. We may formulate a *conjugate dual* of **P4** and obtain the concave envelopes as the optimal dual variables.

2. Note the similarity of the bandit problem to a linear program. If the incremental rewards were associated with the clock time rather than the load levels, it is a linear program with timewise separability. As in LP, integer constraints $\Delta x^i(t) \in \{0, 1\}$ can be relaxed to $\Delta x^i(t) \in [0, 1]$ without affecting the optimality (provided a fractional assignment is properly interpreted).

6.3 Stochastic Bandit Problem

The results obtained through the analysis of deterministic problems are extended to their stochastic counterparts. The main difference is an additional information requirement on feasible policies: outcomes of future events should not be used in current decisions.

Since learning from past history becomes an essential part of the problem, the dependency of reward processes across pipe users becomes an important issue. Unfortunately, the optimality of the index rule holds only under the restrictive condition of the independence among reward processes.

In a stochastic setting, the index is defined as a forecast (or conditional expectation) of the maximum attainable average reward rate (adjusted by the discount rate) based on the past history of a reward process. The index rule is defined accordingly using this index.

6.3.1 Discounted reward criterion: discrete time

User i is characterized by a *reward process* $Z^i := \{Z^i(l), \mathcal{F}^i(l-1)\}_{l=1}^{\infty}$. $Z^i(l)$ is a (stochastic) prediscounted incremental reward from sending her lth packet. $\mathcal{F}^i(l-1)$ is the σ-field representing the information available after her load level reaches $l-1$. $Z^i(l)$ is in general not $\mathcal{F}^i(l-1)$-measurable, but we assume that it is part of the information contained in $\mathcal{F}^i(l)$. Let

$$\mathcal{F}^i(\infty) := \bigvee_{l=0}^{\infty} \mathcal{F}^i(l).$$

To ease the notational burden, $\mathcal{F}^i(0)$ is taken to be trivial for all users. Associated with clock time is a fixed *discount factor* $\beta \in (0,1)$.

Assumptions on reward processes are listed below.

Assumption 6.3.1 *We assume*

1. *independence of reward processes, i.e.,* $\mathcal{F}^i(\infty)$*'s are independent,*

2. *information is never forgotten,*

$$\mathcal{F}^i(l) \subseteq \mathcal{F}^i(l+1), \; i = 1,2,\ldots,N, \; l = 0,1,2,\ldots,$$

3. *expected total discounted reward is finite,*

$$E\sum_{t=1}^{\infty} \beta^t |Z^i(t)| < \infty, \; i = 1,2,\ldots,N.$$

An assignment is denoted by $\Delta x^i(t)$ as before. Let $\Delta x(t) := (\Delta x^1(t), \Delta x^2(t), \ldots, \Delta x^N(t))$. The problem is:

SP1:

$$\begin{array}{lll} \max & E\sum_{i=1}^{N}\sum_{t=1}^{\infty}\beta^t Z^i(x^i(t))\Delta x^i(t) & \\ \text{sub. to} & \sum_{i=1}^{N}\Delta x^i(t) = 1, & t = 1,2,\ldots, \\ & \Delta x^i(t) \in \{0,1\}, & i = 1,2,\ldots,N,\ t = 1,2,\ldots. \\ & \Delta x(t) \text{ is } \bigvee_{i=1}^{N}\mathcal{F}^i(x^i(t-1))\text{-measurable}, & t = 1,2,\ldots \end{array}$$

The last constraint is referred as the *information constraint*: an assignment at time t must be based on information available from the actions taken by time $t-1$. All constraints are meant to be satisfied almost surely.

Mandelbaum [13] formulates the problem as an optimal control over partially ordered sets and characterizes an admissible policy as an optional path of multiparameter processes. Let π be an admissible policy. One technical point involved in the proof by the interchange argument is to show that the 'filtration associated with π', say $\{\mathcal{F}_\pi(t)\}_{t=1}^{\infty}$, is well-defined as a single parameter (clock time) filtration, and what is expected, i.e., $\mathcal{F}_\pi(t) = \bigvee_{i=1}^{N}\mathcal{F}^i(x_\pi^i(t))$, where $x_\pi^i(t)$ is the load level of user i at time t under policy π. Also a '$\mathcal{F}_\pi$-stopping time' need to be well-defined. The multiparameter process framework provides a way to justify these intuitive notions. We take this for granted.

The **index of user i at load level l** is defined by

$$\nu^i(l+1) := \operatorname{esssup}_{\tau \geq l+1} \frac{E[\sum_{s=l+1}^{\tau}\beta^s Z^i(s)|\mathcal{F}^i(l)]}{E[\sum_{s=l+1}^{\tau}\beta^s|\mathcal{F}^i(l)]}, \tag{6.17}$$

where the essential supremum is taken over all $\mathcal{F}^i$-stopping times.[3] Infinite is allowed as a value of a stopping time. Varaiya *et al.* [25] and Mandelbaum [13] showed that the essential supremum in the definition is actually attained by a $\mathcal{F}^i$-stopping time under our third assumption,[4]

$$\tau^i(l+1) = \inf\{s \geq l+1 | \nu^i(s+1) \leq \nu^i(l+1)\} \tag{6.18}$$

Let us call this stopping time as the *maximizing stopping time* at load level l as before.

[3]The term 'stopping time' is used as customary in probability theory. However, the parameter is not a clock time but a load level.

[4]Thus it may be justified to use 'max' instead of somewhat cumbersome 'esssup'. We use 'max' in most of the rest of this chapter without further justification.

The **index rule** is the policy which assigns the pipe to a user with the largest index at the current load level at all times (almost surely).

We make an observation which we use later in the proof of the optimality of the index rule. Let $\mathcal{F}^i(l;\infty)$ be the information field representing the user i's information about her own reward process up to load level l and all the information about the other users' reward processes, i.e.,

$$\mathcal{F}^i(l;\infty) := \mathcal{F}^i(l) \bigvee (\bigvee_{j\neq i} \mathcal{F}^j(\infty)).$$

Because of the independence of the reward processes, the essential supremum of

$$\frac{E[\sum_{s=l+1}^{\tau} \beta^s Z^i(s) | \mathcal{F}^i(l;\infty)]}{E[\sum_{s=l+1}^{\tau} \beta^s | \mathcal{F}^i(l;\infty)]}$$

over $\mathcal{F}^i(\cdot;\infty)$-stopping times $\tau \geq l+1$ will again be obtained by the maximizing stopping time of (6.17), yielding the $\nu^i(l+1)$ as the maximum value.

The **concave envelope** of reward process of user i is defined pathwise by,

$$\overline{Z}^i(l) := \inf_{s\leq l} \nu^i(s), \; l = 1,2,\ldots.$$

A sequence of stochastic maximizing stopping times $\{\tau_k^i\}_{k=0}^{\infty}$ is defined as in (6.4) and (6.5).

The properties of the concave envelope are listed below.

Lemma 6.3.1 *The concave envelope has the following properties:*

1. $\overline{Z}^i(l)$ *is* $\mathcal{F}^i(l-1)$*-measurable for all* l*, and pathwise nonincreasing in* l*.*

2. *Stopped at an arbitrary* $\mathcal{F}^i$*-stopping time (or* $\mathcal{F}^i(\cdot;\infty)$*-stopping time), the concave envelope yields at least as much expected cumulative reward as the original reward process,*

$$E\sum_{s=1}^{\tau} \beta^s \overline{Z}^i(s) \geq E\sum_{s=1}^{\tau} \beta^s Z^i(s), \text{ for all } \mathcal{F}^i\text{- and } \mathcal{F}^i(\cdot;\infty)\text{-stopping time } \tau.$$

3. $E\sum_{s=1}^{\tau} \beta^s \overline{Z}^i(s) = E\sum_{s=1}^{\tau} \beta^s Z^i(s)$*, for* $\tau = \tau_k^i, k = 1,2,\ldots$

Proof Property 1 is immediate from the definition of the concave envelope. Let us derive Property 2. Let τ be an arbitrary $\mathcal{F}^i$-stopping time. Also let $\Omega_k := \{\tau_k^i < \tau\}$ for $k = 0,1,\ldots$ and 1_{Ω_k} their indicators. We write

$$E\sum_{s=1}^{\tau} \beta^s Z^i(s) = E\sum_{k=0}^{\infty} \sum_{s=\tau_k^i+1}^{\tau\wedge\tau_{k+1}^i} \beta^s Z^i(s) 1_{\Omega_k}.$$

For each k,

$$\begin{aligned}
E\sum_{s=\tau_k^i+1}^{\tau\wedge\tau_{k+1}^i} \beta^s Z^i(s)1_{\Omega_k} &= E[1_{\Omega_k}E[\sum_{s=\tau_k^i+1}^{\tau\wedge\tau_{k+1}^i} \beta^s Z^i(s)|\mathcal{F}^i(\tau_k^i)]] \\
&\leq E[1_{\Omega_k}E[\nu^i(\tau_k^i+1)\sum_{s=\tau_k^i+1}^{\tau\wedge\tau_{k+1}^i} \beta^s|\mathcal{F}^i(\tau_k^i)]] \\
&= E[1_{\Omega_k}\overline{Z}^i(\tau_k^i+1)E[\sum_{s=\tau_k^i+1}^{\tau\wedge\tau_{k+1}^i} \beta^s|\mathcal{F}^i(\tau_k^i)]] \\
&= E[1_{\Omega_k}E[\sum_{s=\tau_k^i+1}^{\tau\wedge\tau_{k+1}^i} \beta^s \overline{Z}^i(s)|\mathcal{F}^i(\tau_k^i)]] \\
&= E\sum_{s=\tau_k^i+1}^{\tau\wedge\tau_{k+1}^i} \beta^s \overline{Z}^i(s)1_{\Omega_k}
\end{aligned}$$

Summing over k, we obtain Property 2. When τ is a $\mathcal{F}^i(\cdot;\infty)$ stopping time, condition on $\mathcal{F}^i(\tau_k^i;\infty)$.

Property 3 is obtained by noting that when $\tau = \tau_k^i$ for some k, the inequality above becomes equality for each $j \leq k$. □

The claims made in § 6.2.1 is valid for the stochastic version provided we interpret the total reward as the expected total reward. Claim 1 is valid because the myopic policy for concave envelopes are informationally feasible and optimal by Property 1. Claim 2 is verified by utilizing stochastic counterpart of Lemma 6.2.2 cited below from Varaiya *et al.*[25].

Lemma 6.3.2 *Let $\{X(t)\}_{t=1}^{\infty}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{G}, \mathcal{P})$. Let $\{\mathcal{F}(t)\}_{t=1}^{\infty}$ be an increasing family of sub-σ-field of $\mathcal{G}$, and suppose that $E\sum_{t=1}^{\infty}|X(t)| < \infty$. If*

$$\max_{\tau\geq 1} E[\sum_{t=1}^{\tau} X(t)|\mathcal{F}(1)] \leq 0 \tag{6.19}$$

where the supremum is taken over all $\mathcal{F}$-stopping times, then,

$$E[\sum_{t=1}^{\infty}\alpha(t)X(t)|\mathcal{F}(1)] \leq 0 \tag{6.20}$$

for all $\mathcal{F}$-adapted random sequences $\{\alpha(t)\}_{t=1}^{\infty}$ such that

$$1 \geq \alpha(t) \geq \alpha(t+1) \geq 0,\ t = 1, 2, \ldots$$

The following lemma is the stochastic counterpart of Lemma 6.2.3.

Lemma 6.3.3 *Let* $\{x(t)\}_{t=0}^{\infty}$ *be an arbitrary load pattern of user* i *such that* $x(0) \equiv 0$, $\Delta x(t) \in \{0,1\}$, *and* $\Delta x(t)$ *is* $\mathcal{F}^i(x(t-1);\infty)$*-measurable. Then,*

$$E\sum_{t=1}^{\infty}\beta^t(Z^i(x(t)) - \overline{Z}^i(x(t)))\Delta x(t) \le 0.$$

Proof As before let x^{-1} be the left inverse of the load pattern.

$$E\sum_{t=1}^{\infty}\beta^t(Z^i(x(t)) - \overline{Z}^i(x(t)))\Delta x(t) = E\sum_{t=1}^{\infty}\beta^{(x^{-1}(l)-l)}\beta^l(Z^i(l) - \overline{Z}^i(l))$$

Since $x^{-1}(l) - l$ is pathwise nonnegative and nondecreasing in l, it remains to show that it is $\mathcal{F}^i(\cdot;\infty)$-adapted, i.e., $x^{-1}(l)$ is $\mathcal{F}^i(l-1;\infty)$-measurable. But this follows from the load pattern's measurability. □

Since any informationally feasible assignment satisfies the condition on a load pattern in the lemma above for each user, the second claim is verified.

The third claim follows from Property 3. Thus the optimality of the index rule is proved.

Theorem 6.3.1 *The index rule achieves the optimal reward for problem* **SP1**.

6.3.2 Average reward criterion

The result of § 6.2.2, the asymptotic optimality of the index rule, is extended to the stochastic counterpart of the problem.

We replace the third assumption of Assumption 6.3.1 by the following.

Assumption 6.3.2 *We assume*

3′. *The essential supremum of the following is attained by a* $\mathcal{F}^i$*-stopping time.*

$$\text{esssup}_{\tau\ge l+1}\frac{E[\sum_{s=l+1}^{\tau} Z^i(s)|\mathcal{F}^i(l)]}{E[\tau - l|\mathcal{F}^i(l)]},\ i = 1,2,\ldots,N,\ l = 0,1,\ldots$$

As before let us start with an assignment of a single pipe with the deadline T.

SP2:

$$\begin{array}{lll} \max & E\sum_{i=1}^{N}\sum_{t=1}^{T} Z^i(x^i(t))\Delta x^i(t) & \\ \text{sub. to} & \sum_{i=1}^{N}\Delta x^i(t) \le 1, & t = 1,2,\ldots,T \\ & \Delta x^i(t) \in \{0,1\}, & i = 1,2,\ldots,N,\ t = 1,2,\ldots,T \\ & \text{information constraint.} & \end{array}$$

The information constraint is same as in **SP1**.

The index at load level l is defined by

$$\nu^i(l+1) := \max_{\tau \geq l+1} \frac{E[\sum_{s=l+1}^{\tau} Z^i(s)|\mathcal{F}^i(l)]}{E[\tau - l|\mathcal{F}^i(l)]}, \; l = 0, 1, \ldots$$

The concave envelope is defined by

$$\overline{Z}^i(l) := \max[0, \inf_{s \leq l} \nu^i(s)], \; l = 1, 2, \ldots.$$

The index rule is defined accordingly. Claims 1,2, and 3′ made in deterministic case are valid provided the total reward is interpreted as the expected total reward. A bound in terms of τ_k^i and the index is desirable. Let π be an index rule. We employ the same notation as in § 6.2.2. Now κ is the user to whom π assigns the pipe at time T, and $l^i := x_\pi^i(T)$. τ_n^κ is random. Let $E\Delta V$ be the difference between the optimal expected total reward and the expected total reward earned under π. The stochastic counterpart of (6.8) is

$$E\Delta V \leq E \sum_{s=l^\kappa+1}^{\tau_{n+1}^\kappa} Z^\kappa(s).$$

(Note that (6.8) does *not* hold pathwise.)

We pose the following uniform bounds on conditional expectations of maximizing stopping times and indices.

Assumption 6.3.3 *We assume*

1. $E[\tau^i(l+1) - l|\mathcal{F}^i(l)] \leq D$, *for* $i = 1, 2, \ldots, N$, $l = 1, 2, \ldots$
2. $\nu^i(l) \leq W$, *for* $i = 1, 2, \ldots, N$, $l = 1, 2, \ldots$

These are very strong conditions. However, when there are specific structures on reward processes, there will be better and obvious bounds.

Proposition 6.3.1 *Under Assumption 6.3.3, we have*

$$E\Delta V \leq WD.$$

Proof Let $\{\mathcal{F}_\pi(t)\}_{t=1}^\infty$ be the filtration associated with π. $\mathcal{F}_\pi(t)$ represents the information available at time t (after the delivery at the time).

$$E \sum_{s=l^\kappa+1}^{\tau_{n+1}^\kappa} Z^\kappa(s) \;=\; E[E[\sum_{s=l^\kappa+1}^{\tau_{n+1}^\kappa} Z^\kappa(s)|\mathcal{F}_\pi(T)]]$$

$$
\begin{aligned}
&= E[E[\sum_{s=l^\kappa+1}^{\tau_{n+1}^\kappa} Z^\kappa(s)|\mathcal{F}^\kappa(l^\kappa)]] \\
&\leq E[\nu^\kappa(l^\kappa+1)E[\tau_{n+1}^\kappa - l^\kappa|\mathcal{F}^\kappa(l^\kappa)]] \\
&\leq E[WE[\tau_{n+1}^\kappa - l^\kappa|\mathcal{F}^\kappa(l^\kappa)]] \\
&\leq WD
\end{aligned}
$$

as desired. □

Since the bound on $E\Delta V$ does not depend on the length of horizon, we have

Theorem 6.3.2 *Under Assumption 6.3.3, the index policy is asymptotically optimal under the average expected reward criterion.*

6.3.3 Multiple pipes

We move on to an assignment of multiple pipes. A finite horizon problem is

SP3:

$$
\begin{aligned}
\max \quad & E\sum_{i=1}^N \sum_{t=1}^T Z^i(x^i(t))\Delta x^i(t) & \\
\text{sub. to} \quad & \sum_{i=1}^N \Delta x^i(t) \leq M, & t = 1,2,\ldots,T \\
& \Delta x^i(t) \in \{0,1\}, & i = 1,2,\ldots,N,\ t = 1,2,\ldots,T. \\
& \text{information constraint.} &
\end{aligned}
$$

Again, the information constraint is as in **SP1**.

Let us pose the following conditions on the reward processes.

Assumption 6.3.4 *1. The expected number of remaining packets is uniformly bounded,*

$$E[\sigma^i - l|\mathcal{F}^i(l)] \leq D,\ i = 1,2,\ldots,N,\ l = 1,2,\ldots$$

2. The index is uniformly bounded,

$$\nu^i(l) \leq W,\ i = 1,2,\ldots,N,\ l = 1,2,\ldots$$

By arguing as in § 6.2.3 and § 6.3.2, we have

Proposition 6.3.2 *Under the assumption above, the difference between the optimal expected total reward and the expected total reward from an arbitrary index policy is bounded by MWD.*

When the number of the available pipes changes independently of reward processes and of assignments, and satisfies $E\max_t M(t) \leq M$, we have an error bound of MWD.

6.3.4 Discounted reward criterion: continuous time

The stochastic version of the continuous time bandit is briefly discussed.

Mandelbaum[14] showed the optimality of the index rule. Details are left to [14].

A reward process is defined by $\{Z^i(l), \mathcal{F}^i(l), l \geq 0\}$. Interpretation of $Z^i(l)$ is as in § 6.2.4, and $\mathcal{F}^i(l)$ is the σ-field representing the information available after user i's load level reaches l. The capacity of the pipe is fixed at M over the infinite horizon. Assumptions on reward processes include

1. independence of reward processes, i.e., $\mathcal{F}^i(\infty)$'s are independent,

2. information is never forgotten,

$$\mathcal{F}^i(l) \subseteq \mathcal{F}^i(l'),\ i = 1, 2, \ldots, N,\ 0 \leq l \leq l',$$

3. $\{\mathcal{F}^i(\cdot)\}$ is right continuous, $i = 1, 2, \ldots, N$,

4. $Z^i(\cdot)$ is pathwise continuous, $i = 1, 2, \ldots, N$,

5. expected total discounted reward is finite,

$$E \int_0^\infty e^{-\alpha t} |Z_i(t)| dt < \infty$$

The problem is
SP4:

$$\begin{array}{rl} \max & E \int_0^\infty e^{-\alpha t} \sum_{i=1}^N Z^i(x^i(t)) \dot{x}^i(t) dt \\ \text{sub. to} & \sum_{i=1}^N \dot{x}^i(t) \leq M,\ t \geq 0, \\ & \dot{x}^i(t) \geq 0,\ i = 1, 2, \ldots, N,\ t \geq 0. \\ & \text{information constraint} \end{array}$$

The index of user i at load level l is defined by

$$\nu^i(l) := \max_{\tau > l} \frac{E[\int_l^\tau e^{-\alpha s/M} Z^i(s) ds | \mathcal{F}^i(l)]}{E[\int_l^\tau e^{-\alpha s/M} ds | \mathcal{F}^i(l)]}, \tag{6.21}$$

where l, τ and s correspond to load levels, and the maximum is taken over $\mathcal{F}^i$-stopping times.

The concave envelope of the reward process of user i is defined pathwise by

$$\overline{Z}^i(l) := \max[0, \inf_{0 \le s \le l} \nu^i(s)], \; l \ge 0. \qquad (6.22)$$

Properties of an envelope process

1. $\overline{Z}^i(l)$ is pathwise nonincreasing in l.
2. $E\int_0^\tau e^{-\alpha s/M}\overline{Z}^i(s)ds \ge E\int_0^\tau e^{-\alpha s/M}Z^i(s)ds$, for all $\mathcal{F}^i$- and $\mathcal{F}^i(\cdot;\infty)$-stopping time τ.
3. $E\int_0^\tau e^{-\alpha s/M}\overline{Z}^i(s)ds = E\int_0^\tau e^{-\alpha s/M}Z^i(s)ds$, for τ such that $\overline{Z}(\tau)$ is strictly decreasing.

A proof of the above involves successive discretization of reward processes and consideration of stopping times which take values at the grid points created by the discretization. The optimality of the index rule may be shown by following the steps in the proof of the discrete time version.

The Lagrangian is defined by

$$L(\omega, \dot{x}(\cdot), \lambda(\cdot)) = \int_0^\infty e^{-\alpha t}\sum_{i=1}^N Z^i(x^i(t))\dot{x}^i(t)dt - \int_0^\infty e^{-\alpha t}\lambda(t)(\sum_{i=1}^N \dot{x}^i(t) - M)dt \qquad (6.23)$$

where $\dot{x} := (\dot{x}^1, \ldots, \dot{x}^N)$.

A functional $G(\omega, \lambda)$ is defined by

$$G(\omega, \lambda) := \sup_{\dot{x} \ge 0} L(\omega, \dot{x}, \lambda).$$

The dual of **SP4** is defined by

SD4:

$$\text{essinf}_{\lambda \ge 0} EG(\omega, \lambda).$$

The weak duality result follows from the definitions. Let x^* be the allocation under the index policy. By setting

$$\lambda^*(t) := \max_{i \in \{1,\ldots,N\}} \overline{Z}^i(x^{*i}(t)),$$

the strong duality, too, can be obtained.

6.4 Superprocess

In this and the following sections, we reexamine the results about superprocesses and an arm-acquiring bandit problem in § C and § D of Varaiya *et al.*[25].

The definition of *domination* is cited and expressed in terms of concave envelopes. The expression clarifies the relation of the domination to Whittle's M-process.

6.4.1 Problem instance

The problem instance of superprocesses in [25] is repeated here.

A superprocess is a collection of reward processes. We assume there are N superprocesses. Superprocess I is denoted by $\mathbf{X}^I := \{X^i, i \in I\}$, where by abuse of notation, I represents some index set. We assume that superprocesses are independent, i.e., given any collection of N reward processes, one from each superprocess, $\{X^i \in \mathbf{X}^I, I = 1, \ldots, N\}$, $(\mathcal{F}^{X^1}(\infty), \ldots, \mathcal{F}^{X^N}(\infty))$ are independent. Note that independence of the reward processes *within* a superprocess is *not* assumed.

For each selection $\{X^i \in \mathbf{X}^I, I = 1, 2, \ldots, N\}$, let $V^*(X^1, X^2, \ldots, X^N)$ be the maximum expected reward of the 'standard' bandit in § 6.3.1. The bandit problem associated with the N superprocesses is to find $\{X^i \in \mathbf{X}^I, I = 1, 2, \ldots, N\}$ to maximize

$$\max_{\{X^i \in \mathbf{X}^I\}} V^*(X^1, X^2, \ldots, X^N).$$

The selection of the optimal collection $(X^1, X^2, \ldots, X^N)$ will usually have to be jointly determined. However, when there is a dominant process (the definition will be given in the next subsection) within each superprocess, then the selection can be made independently of each other.

6.4.2 Domination among Reward Processes

The definition of domination is cited from [25]. Let $X := \{X(l), \mathcal{F}^X(l-1)\}_{l=1}^{\infty}$ and $Y := \{Y(l), \mathcal{F}^Y(l-1)\}_{l=1}^{\infty}$ be two reward processes.

Process X *dominates* process Y if

$$\forall a \in R, \; \max_{\tau^X > 0} E \sum_{l=1}^{\tau^X} \beta^l (X(l) - a) \geq \max_{\tau^Y > 0} E \sum_{l=1}^{\tau^Y} \beta^l (Y(l) - a), \tag{6.24}$$

where τ^X ranges over $\mathcal{F}^X$-stopping times and τ^Y over $\mathcal{F}^Y$-stopping times.

The stopping times in (6.24) are easily found when we consider the concave envelopes $\overline{X}$ and $\overline{Y}$.

Lemma 6.4.1 *The original reward process and its concave envelope dominate each other,*

$$\max_{\tau>0} E\sum_{l=1}^{\tau}\beta^l(X(l)-a) = \max_{\tau>0} E\sum_{l=1}^{\tau}\beta^l(\overline{X}(l)-a), \ \forall a \in R.$$

Moreover, there is a common maximizing stopping time for both processes for each value of a.

Proof $LHS \leq RHS$ follows from Property 2 of the concave envelope in Lemma 6.3.1. We will find an optimal stopping time for RHS and show that the same stopping time applied to LHS yields equality.

Since $\overline{X}(l)$ is nonincreasing in l, the optimizing stopping time is easily found. Let

$$\tau_a := \inf\{l > 0 | \overline{X}(l+1) \leq a\}. \tag{6.25}$$

Since $\overline{X}(l+1)$ is $\mathcal{F}^X(l)$-measurable, τ_a is indeed $\mathcal{F}^X$-stopping time. Thus by Property 3 of the concave envelope in Lemma 6.3.1.

$$E\sum_{l=1}^{\tau_a}\beta^l X(l) = E\sum_{l=1}^{\tau_a}\beta^l\overline{X}(l),$$

as desired. □

From definition (6.25), we see that when $a < b$, $\tau_a \leq \tau_b$. Also notice that X dominates Y if and only if $\overline{X}$ dominates $\overline{Y}$. The domination expressed in terms of the concave envelopes has an intuitive interpretation which relates to Whittle's M-processes. Let $X(l) \vee a := \max(X(l), a)$. Since

$$\begin{aligned}
\max_{\tau>0} E\sum_{l=1}^{\tau}\beta^l(X(l)-a) &= E\sum_{l=1}^{\tau_a}\beta^l(\overline{X}(l)-a) \\
&= E\sum_{l=1}^{\tau_a}\beta^l(\overline{X}(l)\vee a-a) \\
&= E\sum_{l=1}^{\infty}\beta^l(\overline{X}(l)\vee a) - \frac{a\beta}{1-\beta},
\end{aligned}$$

X dominates Y if and only if

$$E\sum_{l=1}^{\infty}\beta^l(\overline{X}(l)\vee a) \geq E\sum_{l=1}^{\infty}\beta^l(\overline{Y}(l)\vee a), \ \forall a \in R. \tag{6.26}$$

The process $\{\overline{X}(l) \vee a\}_{l=1}^{\infty}$ may be interpreted as the reward process obtained under the optimal retirement plan. Thus X dominates Y, if for all values of retirement pension a, X is favored over Y.

Given a reward process X and a retirement pension a, let $V(a)$ be the maximum expected reward under the optimal retirement plan, i.e.,

$$V(a) := E\sum_{l=1}^{\infty}\beta^l(\overline{X}(l) \vee a).$$

Lemma 6.4.2 *$V(a)$ is nondecreasing and convex in a.*

Proof Nondecreasing part is clear from the definition.

Let $v(a)$ be the pathwise total reward under the optimal retirement plan when the pension is a,

$$\begin{aligned} v(a) &:= \sum_{l=1}^{\infty}\beta^l(\overline{X}(l) \vee a) \\ &= \sum_{l=1}^{\tau_a}\beta^l\overline{X}(l) + a\beta^{\tau_a}\frac{\beta}{1-\beta}. \end{aligned}$$

We will show that $v(a)$ is convex in a. Let $a < b$, $0 \le \lambda \le 1$, and $c := \lambda a + (1-\lambda)b$. We have $\tau_b \le \tau_c \le \tau_a$. Consider retirement plans which run the process up to τ_c and retire with pension b and a respectively. We note that $v(c)$ is a convex combination of the total rewards from these plans. Clearly, $v(b)$ and $v(a)$ are at least as large as the total rewards from these plans respectively. Thus the claimed convexity. In equations,

$$\begin{aligned} v(b) &= \sum_{l=1}^{\tau_b}\beta^l\overline{X}(l) + \sum_{l=\tau_b+1}^{\tau_c}\beta^l b + b\beta^{\tau_c}\frac{\beta}{1-\beta} \\ &\ge \sum_{l=1}^{\tau_b}\beta^l\overline{X}(l) + \sum_{l=\tau_b+1}^{\tau_c}\beta^l\overline{X}(l) + b\beta^{\tau_c}\frac{\beta}{1-\beta} \\ &= \sum_{l=1}^{\tau_c}\beta^l\overline{X}(l) + b\beta^{\tau_c}\frac{\beta}{1-\beta} \\ v(a) &= \sum_{l=1}^{\tau_c}\beta^l\overline{X}(l) + \sum_{l=\tau_c+1}^{\tau_a}\beta^l\overline{X}(l) + a\beta^{\tau_a}\frac{\beta}{1-\beta} \\ &\ge \sum_{l=1}^{\tau_c}\beta^l\overline{X}(l) + \sum_{l=\tau_c+1}^{\tau_a}\beta^l a + a\beta^{\tau_a}\frac{\beta}{1-\beta} \\ &= \sum_{l=1}^{\tau_c}\beta^l\overline{X}(l) + a\beta^{\tau_c}\frac{\beta}{1-\beta} \end{aligned}$$

$$v(c) \quad = \quad \sum_{l=1}^{\tau_c} \beta^l \overline{X}(l) + c\beta^{\tau_c}\frac{\beta}{1-\beta}$$

Therefore, $\lambda v(a) + (1-\lambda)v(b) \geq v(c)$. To complete the proof, take expectation. □

6.4.3 Domination and optimal selection

Our goal is to prove Lemma 3.2 of [25] cited below.

Lemma 6.4.3 *Let X, Y, Z be three reward processes such that $\mathcal{F}^X(\infty)$ and $\mathcal{F}^Z(\infty)$ are independent, and $\mathcal{F}^Y(\infty)$ and $\mathcal{F}^Z(\infty)$ are independent. If X dominates Y, then $V^*(X,Z) \geq V^*(Y,Z)$.*

Let X, Y, Z be reward processes satisfying the conditions of the lemma above. For simplicity we assume they are nonnegative. The pair $\{X, Y\}$ is viewed as a superprocess here. From this lemma, it is straightforward to derive

Theorem 6.4.1 *Suppose $X^i \in \mathbf{X}^I$ dominates every other $Y^i \in \mathbf{X}^I$. Then*

$$V^*(X^1, X^2, \ldots, X^N) = \max_{\{Y^i \in \mathbf{X}^I\}} V^*(Y^1, Y^2, \ldots, Y^N).$$

Refer to [25] for the derivation of the theorem.

Given two independent reward processes X and Z, let the reward sequence obtained by applying the index rule to this pair be denoted by $\{\nu^{[X,Z]}(t), \mathcal{F}^{[X,Z]}(t-1)\}_{t=1}^{\infty}$. We will show that $\nu^{[X,Z]}$ dominates $\nu^{[Y,Z]}$, that is, the optimal sequencing of the pair $[X, Z]$ with the optimal retirement plan yields at least as large expected reward as the pair $[Y, Z]$ does for every retirement pension. This implies Lemma 6.4.3.

Note that the concave envelope of $\nu^{[X,Z]}$ is identical to the concave envelope of $\nu^{[\overline{X},Z]}$. Also note that process $\nu^{[\overline{X},Z]}(t)$ is nonincreasing, since $\overline{X}$ is nonincreasing. Therefore the concave envelope of $\nu^{[\overline{X},Z]}$ is itself. So it is enough to show that $\nu^{[\overline{X},Z]}$ dominates $\nu^{[\overline{Y},Z]}$.

The following observation is the key to prove the dominance relation.

Lemma 6.4.4 *Let Z be a deterministic reward process defined by*

$$Z(l) := \begin{cases} b, & l = 1, 2, \ldots, s, \\ 0, & l > s, \end{cases}$$

where $b \geq 0$. Let a $(< b)$ be a retirement pension. Then

$$\sum_{t=1}^{\infty} \beta^t (\nu^{[\overline{X},Z]}(t) \vee a) = (1-\beta^s)\sum_{t=1}^{\infty} \beta^t(\overline{X}(t) \vee b) + \beta^s \sum_{t=1}^{\infty} \beta^t(\overline{X}(t) \vee a). \quad (6.27)$$

Proof Note that

$$\begin{aligned}\sum_{t=1}^{\infty}\beta^t(\overline{X}(t)\vee b) &= \sum_{t=1}^{\tau_b}\beta^t\overline{X}(t)+b\beta^{\tau_b}\frac{\beta}{1-\beta}\\ \sum_{t=1}^{\infty}\beta^t(\overline{X}(t)\vee a) &= \sum_{t=1}^{\tau_b}\beta^t\overline{X}(t)+\beta^{\tau_b}\sum_{t=1}^{\infty}\beta^t(\overline{X}(\tau_b+t)\vee a),\end{aligned}$$

and

$$\begin{aligned}\sum_{t=1}^{\infty}\beta^t(\nu^{[\overline{X},Z]}(t)\vee a) &= \sum_{t=1}^{\tau_b}\beta^t\overline{X}(t)+b\beta^{\tau_b}\sum_{t=1}^{s}\beta^t+\beta^{\tau_b+s}\sum_{t=1}^{\infty}\beta^t(\overline{X}(\tau_b+t)\vee a)\\ &= \sum_{t=1}^{\tau_b}\beta^t\overline{X}(t)+b\beta^{\tau_b}\frac{\beta}{1-\beta}(1-\beta^s)+\beta^s\beta^{\tau_b}\sum_{t=1}^{\infty}(\overline{X}(\tau_b+t)\vee a).\end{aligned}$$

The claim of the lemma follows. □

Note that the coefficients of the convex combination in (6.27) depends only on Z. So we have the same coefficient of the convex combination for Y process. Thus taking expectation of (6.27) and its counterpart for Y, and using the domination of X over Y, we have $E\sum_{t=1}^{\infty}\beta^t(\nu^{[\overline{X},Z]}(t)\vee a)\geq E\sum_{t=1}^{\infty}\beta^t(\nu^{[\overline{Y},Z]}(t)\vee a)$ for $a<b$. When $a\geq b$, $\nu^{[\overline{X},Z]}(t)\vee a=\overline{X}(t)\vee a$ and $E\sum_{t=1}^{\infty}\beta^t(\nu^{[\overline{X},Z]}(t)\vee a)\geq E\sum_{t=1}^{\infty}\beta^t(\nu^{[\overline{Y},Z]}(t)\vee a)$ follows immediately from the domination of X over Y. Thus for this Z, $\nu^{[\overline{X},Z]}$ dominates $\nu^{[\overline{Y},Z]}$.

Now let Z be a deterministic and nonincreasing reward process defined by

$$Z(l):=b_n,\ s_{n-1}<l\leq s_n,\ n=1,2,\ldots, \tag{6.28}$$

where $s_0\equiv 0$ and $b_1\geq b_2\geq\cdots\geq 0$. We will show the dominance of $\nu^{[\overline{X},Z]}$ over $\nu^{[\overline{Y},Z]}$ for this case. Let $Z_n(l)$ be the reward sequence curtailed at time s_n, i.e.,

$$Z_n(l):=\begin{cases} Z(l), & 1\leq l\leq s_n,\\ 0, & l>s.\end{cases}$$

Let $w_n^X(a):=\sum_{t=1}^{\infty}\beta^t(\nu^{[\overline{X},Z_n]}(t)\vee a)$ be the optimal (pathwise) reward from pair $[\overline{X},Z_n]$ when the retirement pension is a. Set $b_0\equiv\infty$, and $w_0^X(a):=\sum_{t=1}^{\infty}\beta^t(\overline{X}(t)\vee a)$. Then arguing as in the proof of Lemma 6.4.4, we find for $n\geq 1$,

$$w_n^X(a)=\begin{cases} w_{n-1}^X(a), & a\geq b_n,\\ (1-\beta^{s_n-s_{n-1}})w_{n-1}^X(b_n)+\beta^{s_n-s_{n-1}}w_{n-1}^X(a), & a<b_n.\end{cases} \tag{6.29}$$

Thus we can inductively prove $Ew_n^X(a)\geq Ew_n^Y(a)$ for all a. Since $\sum_{t=1}^{\infty}\beta^t(\nu^{[\overline{X},Z]}(t)\vee a)=w_n^X(a)$ for $b_n>a\geq b_{n+1}$, the desired dominance follows.

Finally, let Z be random but independent of both X and Y. We argue through their concave envelopes.

Lemma 6.4.5 *Let X, Y, Z be as in Lemma 6.4.3. Then $\nu^{[X,Z]}$ dominates $\nu^{[Y,Z]}$.*

Proof Let $\{\sigma_k\}_{k=0}^{\infty}$ be the sequence of the maximizing stopping times defined as in § 6.3.1, i.e.,

$$\begin{aligned} \sigma_0 &\equiv 0, \\ \sigma_{k+1} &:= \inf\{l > \sigma_k | \nu^Z(l+1) \leq \nu^Z(\sigma_k+1)\},\ k = 0, 1, \ldots \end{aligned}$$

Let $\overline{Z}_n$ be the $\overline{Z}$ curtailed at σ_n, $w_n^{\overline{X}}(a) := \sum_{t=1}^{\infty} \beta^t(\nu^{[\overline{X},\overline{Z}_n]}(t) \vee a)$ for $n \geq 1$. As before $w_0^{\overline{X}}(a) := \sum_{t=1}^{\infty} \beta^t(\overline{X}(t) \vee a)$. Then as in (6.29), for $n \geq 1$,

$$w_n^X(a) = \begin{cases} w_{n-1}^X(a), & \text{on } \{a \geq \overline{Z}(\sigma_n)\}, \\ (1-\beta^{\sigma_n - \sigma_{n-1}}) w_{n-1}^X(\overline{Z}(\sigma_n)) + \beta^{\sigma_n - \sigma_{n-1}} w_{n-1}^X(a), & \text{on } \{a < \overline{Z}(\sigma_n)\}. \end{cases}$$

We claim

$$E[w_n^X(a) | \mathcal{F}^Z(\sigma_n)] \geq E[w_n^Y(a) | \mathcal{F}^Z(\sigma_n)],\ a \geq 0,\ n = 0, 1, \ldots$$

We prove the claim by induction. Recall that $\overline{Z}(l) = \overline{Z}(\sigma_k+1) = \overline{Z}(\sigma_{k+1})$ for $\sigma_k + 1 \leq l \leq \sigma_{k+1}$, and also that $\overline{Z}(l+1)$ is $\mathcal{F}^Z(l)$-measurable. Hence $\overline{Z}(\sigma_{k+1})$ is $\mathcal{F}^Z(\sigma_k)$-measurable.

For $n = 0$, the claim is true by the domination of X over Y.

Suppose that the claim is true for $n-1$. Fix a. On $\{a < \overline{Z}(\sigma_n)\}$,

$$E[w_n^X(a)|\mathcal{F}^Z(\sigma_n)] = (1-\beta^{\sigma_n-\sigma_{n-1}}) E[w_{n-1}^X(\overline{Z}(\sigma_n))|\mathcal{F}^Z(\sigma_n)] + \beta^{\sigma_n - \sigma_{n-1}} E[w_{n-1}^X(a)|\mathcal{F}^Z(\sigma_n)].$$

Since $\overline{Z}(\sigma_n)$ is $\mathcal{F}^Z(\sigma_{n-1})$-measurable and by its definition $w_{n-1}^X(a)$ does not include any incremental reward from $\overline{Z}$-process beyond $\overline{Z}(\sigma_{n-1})$,

$$E[w_{n-1}^X(\overline{Z}(\sigma_n))|\mathcal{F}^Z(\sigma_n)] = E[w_{n-1}^X(\overline{Z}(\sigma_n))|\mathcal{F}^Z(\sigma_{n-1})].$$

Also the independence of X and Z implies

$$E[w_{n-1}^X(a)|\mathcal{F}^Z(\sigma_n)] = E[w_{n-1}^X(a)|\mathcal{F}^Z(\sigma_{n-1})].$$

Thus the claim for $n-1$ implies that for n on $\{a < \overline{Z}(\sigma_n)\}$. A similar argument can be made on $\{a \geq \overline{Z}(\sigma_n)\}$.

Since $\sum_{t=1}^{\infty}\beta^t(\nu^{[\overline{X},\overline{Z}]}(t)\vee a) = w_n^X(a)$ on $\Omega_n := \{\overline{Z}(\sigma_n) > a \geq \overline{Z}(\sigma_{n+1})\} \in \mathcal{F}^Z(\sigma_n)$, we may write

$$\sum_{t=1}^{\infty}\beta^t(\nu^{[\overline{X},\overline{Z}]}(t)\vee a) = \sum_{n=0}^{\infty} 1_{\Omega_n} w_n^X(a),$$

where $\Omega_0 := \{a \geq \overline{Z}(1)\}$ is either the empty set or the entire space. Thus the lemma follows from the claim above. □

Remark

Consider a problem involving several controlled machines. States of each machine are independent of states and control actions of other machines. We may regard each machine with all its possible 'local' feedback laws as a superprocess.

As Varaiya *et al.*[25] shows, when each superprocess contains a dominant machine, the optimal strategy for the controlled multi-armed bandit problem is to operate these dominating machines according to the index rule. Note this result does not immediately follow from Theorem 6.4.1, since an admissible control policy can use the information gathered through past activities. See [25] for details.

6.5 Arm-acquiring Bandit

We move on to the case in which the system admits new arrivals of pipe users (and hence reward processes). Along with Assumption 6.3.1, we assume:

Assumption 6.5.1 *The future arrivals are independent of the past and the present control actions.*

We introduce notation for the arrival processes here. Throughout this section, $A(t)$ is the set of pipe users that arrived at time t *after* the assignment of the pipe at the time (hence they become eligible for an assignment at time $t+1$, but not at time t). $A(0)$ is the set of pipe users initially present in the system. $\mathbf{A}(t)$ is the set of users present at the end of time t, i.e., $\mathbf{A}(t) := \bigcup_{s=0}^{t} A(s)$.

In § 6.5.1 it is shown that when the unassigned user processes are frozen as assumed in the case of the ordinary bandit problem, an optimal policy has a *greedy* nature in the sense that it maximizes the expected average reward rate (adjusted by the discount factor) at every state as in the ordinary bandit problem.

In § 6.5.2 the nested structure of the optimal policy for an arm-acquiring bandit with *i.i.d.* arrival processes is shown.

6.5.1 Domination by an optimal policy

The index rule for the ordinary bandit problem in § 6.3.1 is a greedy policy in the sense mentioned above. We assume the existence of an optimal policy and show that under our assumptions the greedy policy is optimal for the arm-acquiring bandit problem.

We approach the task through the examination of optimal policies for problems with retirement pensions. In the following we assume the existence of an optimal policy for every problem with a retirement pension. Also all the relevant essential supremums are assumed to be attained by well-defined stopping times.

Our immediate goal is to prove Lemma 3.5 in [25] restated below as Proposition 6.5.1. The proof given in [25] is incomplete, and we aim to fill the gap.

Let π be an arbitrary admissible policy. Let $X^{\pi} := \{X^{\pi}(t), \mathcal{F}^{\pi}(t-1)\}_{t=1}^{\infty}$ be the reward sequence realized under π. The concave envelope of X^{π} can be formed as usual. It is denoted by $\overline{X^{\pi}}$. Let $V(a;\pi)$ be the expected reward from policy π with the optimal retirement when the pension is a,

$$V(a;\pi) := E\sum_{t=1}^{\infty} \beta^t(\overline{X^{\pi}}(t) \vee a).$$

Let $\tau^{\pi}(a)$ be the time of the earliest optimal retirement,

$$\tau^{\pi}(a) := \inf\{t > 0 | \overline{X^{\pi}}(t+1) \leq a\}. \tag{6.30}$$

Let $\pi(b)$ be an optimal policy for the problem with pension b, i.e., $V^*(b) := \max_{\pi} V(b;\pi) = V(b;\pi(b))$. In Proposition 6.5.1, it will be shown that for a smaller pension $a < b$, there is an optimal policy $\pi(a)$ which is a continuation of $\pi(b)$ from $\tau^{\pi(b)}(b)$ on. Therefore, in effect, the optimal policy for the assignment problem without any retirement pension is an optimal policy for all problems with retirement pensions, and hence it maximizes the expected average reward rate in any state. In view of Lemma 6.4.2, $V^*(\cdot)$ is a nondecreasing convex function as Whittle[32] showed.

We make a few observations on admissible policies.

Let a reward sequence from an admissible policy π be denoted by

$$\underbrace{X(1),\ldots,X(\zeta_1)}_{\text{block 1}},\underbrace{X(\zeta_1+1),\ldots,X(\zeta_2)}_{\text{block 2}},\underbrace{X(\zeta_2+1),\ldots,X(\zeta_3)}_{\text{block 3}},\underbrace{X(\zeta_3+1),\ldots,X(\zeta_4)}_{\text{block 4}},\ldots,$$

where the ζ_n are $\mathcal{F}^\pi$-stopping times.

Observation 1: Assume that the sequence obtained by exchanging block 2 and block 4,

$$\underbrace{X(1),\ldots,X(\zeta_1)}_{\text{block 1}},\underbrace{X(\zeta_3+1),\ldots,X(\zeta_4)}_{\text{block 4}},\underbrace{X(\zeta_2+1),\ldots,X(\zeta_3)}_{\text{block 3}},\underbrace{X(\zeta_1+1),\ldots,X(\zeta_2)}_{\text{block 2}},\ldots$$

constitutes a reward sequence from another admissible policy. Then, policy π has chosen the reward sequence of block 4 independent of the events in block 2 and 3 (including arrivals in these periods and the associated information) given the events in block 1. Thus for example,

$$E[\sum_{t=\zeta_3+1}^{\zeta_4} \beta^t X(t)|\mathcal{F}^\pi(\zeta_3)] = E[\sum_{t=\zeta_3+1}^{\zeta_4} \beta^t X(t)|\mathcal{F}^\pi(\zeta_1)].$$

(Here and in the following, reward sequences and their filtrations are indexed by the time appeared in their original sequences.) Also policy π has chosen the reward sequence of block 3 independent of the events in block 2 given the events in block 1. Therefore, any sequence obtained by placing blocks 2, 3, and 4 in an arbitrary order can be the result from an admissible policy. Note that this makes sense only because the unassinged processes are frozen and the arrival processes are independent of the past control actions.

Observation 2: Assume that the sequence obtained by removing a block, say block 2,

$$\underbrace{X(1),\ldots,X(\zeta_1)}_{\text{block 1}},\underbrace{X(\zeta_2+1),\ldots,X(\zeta_3)}_{\text{block 3}},\underbrace{X(\zeta_3+1),\ldots,X(\zeta_4)}_{\text{block 4}},\ldots.$$

constitutes a sequence from another admissible policy. Then, policy π has chosen the reward sequence from the time ζ_2 onward independent of the events in block 2 given the events in block 1.

Lemma 6.5.1 *Let $\pi(b)$ be an optimal policy when the retirement pension is b, and let $\tau^{\pi(b)}(b) =: \tau$ be as in (6.30). Let*

$$\underbrace{Z(1),\ldots,Z(\zeta_1)}_{\textit{block 1}},\underbrace{Z(\zeta_1+1),\ldots,Z(\zeta_2)}_{\textit{block 2}},\underbrace{Z(\zeta_2+1),\ldots,Z(\zeta_3)}_{\textit{block 3}},$$
$$\underbrace{Z(\zeta_3+1),\ldots,Z(\zeta_4)}_{\textit{block 4}},\underbrace{Z(\zeta_4+1),\ldots,Z(\tau)}_{\textit{block 5}},b,b,b,\ldots.$$

be the associated reward sequence.

Assume that the sequence obtained by removing block 2 constitutes a sequence from another admissible policy. Then the expected average reward rate in block 2 given the events *in block 1 (i.e., conditioned on $\mathcal{F}^{\pi(b)}(\zeta_1)$) is no less than that in blocks 3, 4, and 5, and also the expected average reward rate in blocks 3, 4, and 5 given the events in block 1 is in turn no less than b, i.e.,*

$$\frac{E[\sum_{t=\zeta_1+1}^{\zeta_2} \beta^t Z(t)|\mathcal{F}^{\pi(b)}(\zeta_1)]}{E[\sum_{t=\zeta_1+1}^{\zeta_2} \beta^t|\mathcal{F}^{\pi(b)}(\zeta_1)]} \geq \frac{E[\sum_{t=\zeta_2+1}^{\tau} \beta^t Z(t)|\mathcal{F}^{\pi(b)}(\zeta_1)]}{E\sum_{t=\zeta_2+1}^{\tau} \beta^t|\mathcal{F}^{\pi(b)}(\zeta_1)]} \geq b, \quad a.s. \tag{6.31}$$

Assume that the sequence obtained by exchanging block 2 and block 4 constitutes a sequence from another admissible policy. Then, the expected average reward rate in block 2 given the events in block 1 is no less than that in block 3, and which is in turn no less than that in block 4. Also the expected average reward rate in block 5 given the events in block 1 is no less than b.

Assume that a block of rewards $Y(1), \ldots, Y(\rho)$ is inserted after a $\mathcal{F}^{\pi(b)}$-stopping time ζ ($< \tau$) and the resulting sequence constitutes a reward process from another admissible policy. Then, the expected average reward rate from the inserted block given the events by time ζ (i.e., conditioned on $\mathcal{F}^{\pi(b)}(\zeta)$) is no greater than the expected average reward rate during $\zeta < t \leq \tau$ nor b.

Proof We prove the first inequality of (6.31). Other claims can be proved in a similar manner. Note that the sequence obtained from exchanging block 2 and the rest of reward sequence up to time τ (block 1, block 3, block 4, block 5, block 2, $b, b, b, \ldots$) constitutes a reward sequence from some admissible policy, say $\tilde{\pi}$. To ease the notation, let us write $\delta := \zeta_2 - \zeta_1$, and $\eta := \tau - \zeta_2$. Let Δ be the difference in the (pathwise) rewards from $\pi(b)$ and $\tilde{\pi}$, i.e.,

$$\begin{aligned}
\Delta &:= \beta^{\zeta_1}\Big[[\sum_{t=1}^{\delta} \beta^t Z(\zeta_1 + t) + \beta^{\delta} \sum_{t=1}^{\eta} \beta^t Z(\zeta_2 + t)] - [\sum_{t=1}^{\eta} \beta^t Z(\zeta_2 + t) + \beta^{\eta} \sum_{t=1}^{\delta} \beta^t Z(\zeta_1 + t)]\Big] \\
&= \beta^{\zeta_1}\Big[(1 - \beta^{\eta}) \sum_{t=1}^{\delta} \beta^t Z(\zeta_1 + t) - (1 - \beta^{\delta}) \sum_{t=1}^{\eta} \beta^t Z(\zeta_2 + t)\Big].
\end{aligned}$$

Let $\Omega_0 := \{E[\Delta|\mathcal{F}^{\pi(b)}(\zeta_1)] < 0\}$. On $\{\zeta_1 = \infty\}$, we use the convention $\Delta \equiv 0$. We assume $\text{Prob}(\{\zeta_1 = \infty\}) < 1$. Note that δ and $\sum_{t=1}^{\eta} \beta^t Z(\zeta_2 + t)$ are conditionally independent given $\mathcal{F}^{\pi(b)}(\zeta_1)$, and similarly for η and $\sum_{t=1}^{\delta} \beta^t Z(\zeta_1 + t)$. Thus

$$\begin{aligned}
E[\Delta|\mathcal{F}^{\pi(b)}(\zeta_1)] &= \beta^{\zeta_1}\Big[E[1 - \beta^{\eta}|\mathcal{F}^{\pi(b)}(\zeta_1)]E[\sum_{t=1}^{\delta} \beta^t Z(\zeta_1 + t)|\mathcal{F}^{\pi(b)}(\zeta_1)] \\
&\quad -E[1 - \beta^{\delta}|\mathcal{F}^{\pi(b)}(\zeta_1)]E[\sum_{t=1}^{\eta} \beta^t Z(\zeta_2 + t)|\mathcal{F}^{\pi(b)}(\zeta_1)]\Big]
\end{aligned}$$

We will show that $\text{Prob}(\Omega_0) = 0$, which leads to the desired result.

Consider the policy which follows $\pi(b)$ except for events in Ω_0, for which it follows $\bar{\pi}$ after ζ_1. The policy is admissible since both $\pi(b)$ and $\bar{\pi}$ are admissible and $\Omega_0 \in \mathcal{F}^{\pi(b)}(\zeta_1)$. Thus if $\text{Prob}(\Omega_0) > 0$, this policy yields strictly larger expected reward than $\pi(b)$ does. But this contradicts the optimality of $\pi(b)$. □

Remarks

1. Inequalities involving b, for example the second inequality in (6.31), may be strengthened to strict inequalities by using the definition (6.30) of $\tau = \tau^{\pi(b)}(b)$ as the earliest optimal retirement time.

2. Roughly speaking the lemma above says that an optimal policy arranges exchangeable blocks, if there are any, in the decreasing order of the expected average reward rates. We call this the 'greedy' property of an optimal policy.

Proposition 6.5.1 *Let $\pi(b)$ be an optimal policy for the problem with the retirement pension b. Let τ be an optimal retirement time under $\pi(b)$ (not necessarily the earliest possible one). Let a be another retirement pension strictly less than b. Then there is an optimal policy for the problem with the retirement pension a which follows $\pi(b)$ up to τ.*

Proof Let $\{Z(1), Z(2), \ldots, Z(\tau)\}$ be the reward sequence under $\pi(b)$ up to τ. Let $\pi(a)$ be an optimal policy for the problem with the pension a. Let the corresponding (pathwise) reward sequence under $\pi(a)$ be

$$Y(1), \ldots, Y(v_1), Z(\zeta_1 + 1), \ldots, Z(\zeta_1 + \delta_1), \ldots, Y(v_1 + 1), \ldots, Y(v_2),$$
$$Z(\zeta_2 + 1), \ldots, Z(\zeta_2 + \delta_2), \ldots, Y(v_\kappa + 1), \ldots, Y(\tau^{\pi(a)}(a)), a, a, a, \ldots.$$

where the $Z(\cdot)$ in this sequence represent the incremental rewards appearing as a part of $\{Z(1), Z(2), \ldots, Z(\tau)\}$, but not necessarily in the same order. By the observations we made about admissible policies, we see that the sequence

$$Z(1), \ldots, Z(\tau), Y(1), \ldots, Y(v_1), Y(v_1 + 1), \ldots, Y(v_2), \ldots,$$
$$Y(v_\kappa + 1), \ldots, Y(\tau^{\pi(a)}(a)), a, a, a, \ldots$$

constitutes a reward sequence from an admissible policy. Lemma 6.5.1 may be applied to show that the second sequence can be obtained from the first one by inserting blocks and exchanging blocks pairwise in a manner it does not decrease the expected reward while maintaining admissibility. □

Remark

From the optimality of $\pi(a)$ and Lemma 6.5.1, we see that $\pi(a)$ retired at $\tau^{\pi(a)}(b)$, too, is optimal to the problem with the retirement pension b. Also note that when $\tau := \tau^{\pi(b)}(b)$ is the earliest optimal retirement time, there is no room to insert another block in $\{Z(1), Z(2), \ldots, Z(\tau)\}$. It is the optimal reward sequence up to the interchange within itself.

Theorem 6.5.1 *A policy for the assignment problem without a retirement option is optimal if and only if it is optimal for all problems with retirement pensions. In other words, the reward sequence from an optimal policy necessarily dominates reward sequences from all the admissible policies.*

Proof Immediate from Proposition 6.5.1. □

6.5.2 Nested structure of an optimal policy for *i.i.d.* arrival processes

In this subsection, we assume that arrival processes $\{A(t)\}_{t=1}^{\infty}$ are *i.i.d.*. With this additional restriction, an optimal policy is shown to have a nested structure which allows decentralization of decision making.

The nested structure of an optimal policy may be explained through a fictitious auction. The pipe owner holds the initial auction at time 1. The participants of this auction are the users present in the system initially (at time 0). Auctioned is the right to use the pipe oneself or to sublet the pipe to another pipe user. A bid is interpreted as a rent in each period while a user occupies the pipe as its user or subletter. A user is free to terminate the lease. In general, each user's bid price will depend on the information about reward processes of the other users present at the time as well as the information about her own reward process and the arrival processes. When the arrival processes are *i.i.d.*, the matter simplifies. It will be shown that in this case, a primary renter sublets to only those who arrive after she wins the bid, if she ever sublets. Therefore, each participant decides her bid price based on the information about her own reward process and the arrival processes alone. Once the primary renter sublets (to the highest bidder of auction held by her, if the highest bid is higher than her rent), the secondary renter behaves just as the primary renter, either he uses the pipe or sublets. When he is through (with on going rent to the primary renter), the pipe is returned to the primary renter, upon which she uses it herself or sublets it or returns to the owner. We call this process of auctions and sublets the *nested (auction) structure* of the optimal policy. The expected total reward from an optimal policy is the amount the pipe owner collects from the primary renters.

We utilize the following properties of a reward process and maximizing stopping times in showing the nested structure.

Lemma 6.5.2 *Let $\{Z(l), \mathcal{F}(l-1)\}_{l=1}^{\infty}$ be a reward sequence. Let indices $\{\nu(l+1)\}_{l=0}^{\infty}$ be defined as usual. Let $\tau^* := \tau(1) = \inf\{s \geq 1 | \nu(s+1) \leq \nu(1)\}$ be the earliest stopping time at load level 0 which maximizes the expected average reward rate. Let σ be an arbitrary $\mathcal{F}$-stopping time. Then on $\{\sigma < \tau^*\}$,*

$$\frac{E[\sum_{t=\sigma+1}^{\tau^*} \beta^t Z(t) | \mathcal{F}(\sigma)]}{E[\sum_{t=\sigma+1}^{\tau^*} \beta^t | \mathcal{F}(\sigma)]} > \nu(1), \ a.s.$$

Proof Let

$$\Omega_0 := \{\sigma < \tau^*\} \cap \left\{ \frac{E[\sum_{t=\sigma+1}^{\tau^*} \beta^t Z(t) | \mathcal{F}(\sigma)]}{E[\sum_{t=\sigma+1}^{\tau^*} \beta^t | \mathcal{F}(\sigma)]} \leq \nu(1) \right\}.$$

We will show that $\text{Prob}(\Omega_0) = 0$. Consider the stopping time ζ which takes value σ on Ω_0, and τ^* otherwise.

$$\begin{aligned} E\sum_{t=1}^{\zeta} \beta^t Z(t) &= \nu(1) E\sum_{t=1}^{\tau^*} \beta^t Z(t) - E[1_{\Omega_0} \sum_{t=\sigma+1}^{\tau^*} \beta^t Z(t)] \\ &= \nu(1) E\sum_{t=1}^{\tau^*} \beta^t Z(t) - E[1_{\Omega_0} E[\sum_{t=\sigma+1}^{\tau^*} \beta^t Z(t) | \mathcal{F}(\sigma)]] \\ &\geq \nu(1) E\sum_{t=1}^{\tau^*} \beta^t Z(t) - \nu(1) E[1_{\Omega_0} E[\sum_{t=\sigma+1}^{\tau^*} \beta^t | \mathcal{F}(\sigma)]] \\ &= \nu(1) E\sum_{t=1}^{\zeta} \beta^t \end{aligned}$$

From the above we have

$$\frac{E\sum_{t=1}^{\zeta} \beta^t Z(t)}{E\sum_{t=1}^{\zeta} \beta^t} \geq \nu(1).$$

By the definition of $\nu(1)$, the inequality above is satisfied with equality. But this contradicts the assumption that τ^* is the earliest stopping time maximizing the expected average reward rate. □

The result extends to the case involving the maximizing stopping time at any load level. Let $\tau(l+1) = \inf\{s \geq l+1 | \nu(s+1) \leq \nu(l+1)\}$, then on $\{\sigma < \tau(l+1)\}$,

$$\frac{E[\sum_{t=\sigma+1}^{\tau(l+1)} \beta^t Z(t) | \mathcal{F}(\sigma)]}{E[\sum_{t=\sigma+1}^{\tau(l+1)} \beta^t | \mathcal{F}(\sigma)]} > \nu(l+1), \quad \text{a.s.}$$

The next property is immediate from the definition of $\tau(l+1)$.

Lemma 6.5.3 *Fix l, then for $l < s < \tau(l+1)$, $\nu(s+1) > \nu(l+1)$ and $\tau(s+1) \leq \tau(l+1)$.*

Let us go back to the arm-acquiring bandit with *i.i.d.* arrival processes.

Let π^* be an optimal policy and $\{Z(t), \mathcal{F}^*(t-1)\}_{t=1}^{\infty}$ be the reward process under π^*, where $\mathcal{F}^*(t-1)$ represents information available at time t gathered through control actions by time $t-1$.

Let $\nu(\cdot)$ and $\tau(\cdot)$ be defined as usual, i.e.,

$$\begin{aligned}
\nu(t+1) &:= \max_{\tau \geq t+1} \frac{E[\sum_{s=t+1}^{\tau} \beta^s Z(s) | \mathcal{F}^*(t)]}{E[\sum_{s=t+1}^{\tau} \beta^s | \mathcal{F}^*(t)]}, \; t = 0, 1, 2, \ldots, \\
\tau(t+1) &:= \inf\{s \geq t+1 | \nu(s+1) \leq \nu(t+1)\}, \; t = 0, 1, 2, \ldots.
\end{aligned}$$

Theorem 6.5.2 *Let k be the user to whom the optimal policy π^* assigns the pipe at time 1. Then for the time interval $1 < s \leq \tau(1)$, π^* does not assign the pipe to the other users present at time 0.*

In general, let $\kappa(t)$ be the user to whom π^* assigns the pipe at time t. Then for the time interval $t < s \leq \tau(t)$, π^* does not assign the pipe to the other users present at the end of time $t-1$ (users in $\mathbf{A}(t-1) \setminus \{\kappa(t)\}$).

Proof We prove the theorem for the initial assignment. The general case can be proved in a similar manner.

Let $\tau^* := \tau(1)$. Let $\sigma_1 + 1$ be the first time π^* assigns the pipe to some user other than k who are initially present in the system. Since the assignment of the pipe at time $s+1$ is made based on the information gathered by time s, σ is a $\mathcal{F}^*$-stopping time. Let $\Omega_0 := \{\sigma_1 < \tau^*\}$. We will show that on Ω_0,

$$\frac{E[\sum_{s=\sigma_1+1}^{\tau^*} \beta^s Z(s) | \mathcal{F}^*(\sigma_1)]}{E[\sum_{s=\sigma_1+1}^{\tau^*} \beta^s | \mathcal{F}^*(\sigma_1)]} \leq \nu(1), \quad \text{a.s.}$$

In view of Lemma 6.5.2, it follows $\text{Prob}(\Omega_0) = 0$, which is the desired result.

Let $\tilde{A}(\sigma_0) := A(0) \setminus \{k\}$ and $\tilde{A}(\sigma_1) := A(\sigma_1) \setminus \tilde{A}(\sigma_0)$. $\tilde{A}(\sigma_0)$ is the set of users who do not receive the initial assignment. $\tilde{A}(\sigma_1)$ is the set consisting of users who arrive by the end of time σ_1 and user k. Thus $\tilde{A}(\sigma_0)$ is the set of users who do not receive an assignment in the time interval $1 \le s \le \sigma_1$, while $\tilde{A}(\sigma_1)$ is the set of users who may receive assignments in the same interval. $\sigma_1 + 1$ starts the time interval during which the users in $\tilde{A}(\sigma_1)$ do not receive an assignment of the pipe.

We claim that on Ω_0, π^* assigns the pipe to users in $\tilde{A}(\sigma_1)$ at some time in the interval $\sigma_1 + 1 < s \le \tau^*$ (a.s.). Let $\tilde{\Omega}_0 \subseteq \Omega_0$ be the set on which this claim fails. By Lemma 6.5.2 and the definition of $\tau(\sigma_1 + 1)$, on $\tilde{\Omega}_0$

$$\begin{aligned}
\nu(1) &< \frac{E[\sum_{s=\sigma_1+1}^{\tau^*} \beta^s Z(s) | \mathcal{F}^*(\sigma_1)]}{E[\sum_{s=\sigma_1+1}^{\tau^*} \beta^s | \mathcal{F}^*(\sigma_1)]}, \\
&\le \max_\tau \frac{E[\sum_{s=\sigma_1+1}^{\tau} \beta^s Z(s) | \mathcal{F}^*(\sigma_1)]}{E[\sum_{s=\sigma_1+1}^{\tau} \beta^s | \mathcal{F}^*(\sigma_1)]} \\
&= \frac{E[\sum_{s=\sigma_1+1}^{\tau(\sigma_1+1)} \beta^s Z(s) | \mathcal{F}^*(\sigma_1)]}{E[\sum_{s=\sigma_1+1}^{\tau(\sigma_1+1)} \beta^s | \mathcal{F}^*(\sigma_1)]}, \quad \text{a.s.}
\end{aligned}$$

Note $\tau(\sigma_1 + 1) \le \tau^*$ by Lemma 6.5.3 on $\tilde{\Omega}_0$. Since the maximum expected reward rate is attained without assigning the pipe to the users in $\tilde{A}(\sigma_1)$ (the users arriving in $1 \le s \le \sigma_1$ and k, who contribute to $\mathcal{F}^*(\sigma_1)$) and the reward processes are independent, the last term is equal to

$$\max_\tau \frac{E \sum_{s=\sigma_1+1}^{\tau} \beta^s Z(s)}{E \sum_{s=\sigma_1+1}^{\tau} \beta^s}$$

(i.e., $\mathcal{F}^*(\sigma_1)$ may randomize the optimal stopping time but the randomization does not increase the expected reward rate). Now consider a slightly different problem in that the users initially present in the system do not include k, the one who receives the first assignment in the original problem, but otherwise the same as the original problem. Let π be an optimal policy for this new problem and $\{X(t), \mathcal{F}^\pi(t-1)\}$ be the reward sequence under π. Then

$$\max_\tau \frac{E \sum_{s=\sigma_1+1}^{\tau} \beta^s Z(s)}{E \sum_{s=\sigma_1+1}^{\tau} \beta^s} = \max_\tau \frac{E \sum_{s=1}^{\tau} \beta^s X(s)}{E \sum_{s=1}^{\tau} \beta^s},$$

where the maximum is taken over $\mathcal{F}^*$-stopping time in the left hand side and over $\mathcal{F}^\pi$-stopping time in the right hand side. By the greedy property of optimal policies for the original problem, the right hand side is not greater than $\nu(1)$. This is possible only when $\text{Prob}(\tilde{\Omega}_0) = 0$. Thus the claim is verified.

Let $\sigma_2 + 1$ be the first time after $\sigma_1 + 1$ at which π^* assigns the pipe to a user in $\tilde{A}(\sigma_1)$. Note σ_2 is a $\mathcal{F}^*$-stopping time. Arguing in a similar manner, we obtain

$$\frac{E[\sum_{s=\sigma_1+1}^{\sigma_2} \beta^s Z(s)|\mathcal{F}^*(\sigma_1)]}{E[\sum_{s=\sigma_1+1}^{\sigma_2} \beta^s|\mathcal{F}^*(\sigma_1)]} \leq \nu(1).$$

Also note that $\text{Prob}(\{\sigma_2 + 1 = \tau^*\}) = 0$ by the greedy property of an optimal policy.

Let $\tilde{A}(\sigma_2) := A(\sigma_2) \setminus \tilde{A}(\sigma_1)$. $\tilde{A}(\sigma_2)$ is the set consisting of users who arrive in the interval $\sigma_1 + 1 \leq s \leq \sigma_2$ and user k. Note that $\tilde{A}(\sigma_1)$ is the set of users who do not receive an assignment in the interval $\sigma_1 + 1 \leq s \leq \sigma_2$, while $\tilde{A}(\sigma_2)$ is the the set of users who may receive assignments in the same interval. $\sigma_2 + 1$ starts the time interval during which users in $\tilde{A}(\sigma_2)$ do not receive an assignment of the pipe.

We claim that on Ω_0, π^* assigns the pipe to users in $\tilde{A}(\sigma_2)$ at some time in the interval $\sigma_2 + 1 < s \leq \tau^*$ (a.s.). This claim can be verified by following the steps in the verification of the previous claim. Reset $\tilde{\Omega}_0$ to the set on which the new claim fails. By Lemma 6.5.2 and the definition of $\tau(\sigma_2 + 1)$, on $\tilde{\Omega}_0$,

$$\nu(1) < \frac{E[\sum_{s=\sigma_2+1}^{\tau(\sigma_2+1)} \beta^s Z(s)|\mathcal{F}^*(\sigma_2)]}{E[\sum_{s=\sigma_2+1}^{\tau(\sigma_2+1)} \beta^s|\mathcal{F}^*(\sigma_2)]}, \quad \text{a.s.}$$

where $\sigma_2 + 1 \leq \tau^*$ by Lemma 6.5.3. Noting the absence of an assignment to the users in $\tilde{A}(\sigma_2)$ in maximizing the expected reward rate, we see that the right hand side is equal to

$$\max_{st} \frac{E[\sum_{s=\sigma_2+1}^{\tau} \beta^s Z(s)|\mathcal{F}^*(\sigma_1)]}{E[\sum_{s=\sigma_2+1}^{\tau} \beta^s|\mathcal{F}^*(\sigma_1)]}.$$

Now we may consider a problem which starts at time σ_1 with the same state as the original problem except the users in $\tilde{A}(\sigma_2)$ are absent. Again write an optimal policy for this problem as π and the reward sequence under π as X. Then we have

$$\frac{E[\sum_{s=\sigma_1+1}^{\sigma_2} \beta^s Z(s)|\mathcal{F}^*(\sigma_1)]}{E[\sum_{s=\sigma_1+1}^{\sigma_2} \beta^s|\mathcal{F}^*(\sigma_1)]} \leq \nu(1) < \max_{\tau} \frac{E[\sum_{s=\sigma_1+1}^{\tau} \beta^s X(s)|\mathcal{F}^*(\sigma_1)]}{E[\sum_{s=\sigma_1+1}^{\tau} \beta^s|\mathcal{F}^*(\sigma_1)]}.$$

Note that given $\mathcal{F}^*(\sigma_1)$, the first term and the third term can be viewed as the expected reward rates from exchangeable (with respect to π^*) blocks. But this contradicts the greedy property of π^*. Thus $\text{Prob}(\tilde{\Omega}_0) = 0$.

Let $\sigma_3 + 1$ be the first time after $\sigma_2 + 1$ at which π^* assigns the pipe to a user in $\tilde{A}(\sigma_2)$. Then by arguing in a similar manner (by the greedy property of π^*),

$$\nu(1) \geq \frac{E[\sum_{s=\sigma_1+1}^{\sigma_2} \beta^s Z(s)|\mathcal{F}^*(\sigma_1)]}{E[\sum_{s=\sigma_1+1}^{\sigma_2} \beta^s|\mathcal{F}^*(\sigma_1)]} \geq \frac{E[\sum_{s=\sigma_2+1}^{\sigma_3} \beta^s Z(s)|\mathcal{F}^*(\sigma_2)]}{E[\sum_{s=\sigma_2+1}^{\sigma_3} \beta^s|\mathcal{F}^*(\sigma_2)]}.$$

The steps above may be carried out inductively to show that on Ω_0, $\tau^* = \infty$ a.s. and

$$\nu(1) \geq \frac{E[\sum_{\sigma_1+1}^{\infty} \beta^s Z(s) | \mathcal{F}^*(\sigma_1)]}{E[\sum_{\sigma_1+1}^{\infty} \beta^s | \mathcal{F}^*(\sigma_1)]},$$

which completes the proof. □

Consider an arm-acquiring bandit problem with only user i with load level l is initially present in the system. Let π be an admissible policy for this problem and X^π the associated reward sequence. Let

$$\nu^i(l+1) := \max_{\pi} \max_{\tau \geq l+1} \frac{E[\sum_{s=l+1}^{\tau} \beta^s X^\pi(s) | \mathcal{F}^i(l)]}{E[\sum_{s=l+1}^{\tau} \beta^s | \mathcal{F}^i(l)]}. \tag{6.32}$$

We cite Theorem 3.5 of [25].

Theorem 6.5.3 *For the bandit problem with i.i.d. arrivals, it is optimal to assign the pipe at each time to the user with the largest current index defined by (6.32).*

Proof It follows from Proposition 6.5.1 and Theorem 6.5.2. □.

Remark

When each reward process is a finite state Markov chain, Theorem 6.5.3 reduces to the result in Whittle[32] where the index is associated to each state of each class of arms.

Chapter 7

Conclusions

The main results are summarized, and topics for future work are suggested.

7.1 Summary

The most important message in this thesis is the most obvious one: we must recognize the need for message exchanges in achieving a system goal when information is distributed among participants of the system.

Throughout this thesis, the goals of the system were the efficient allocation of resources. Participants' valuations of resources were private knowledge.

Though the need for message exchange may appear obvious, it can be overlooked. We saw an example in Chapter 4 where we examined pricing schemes of electric power. A 'centralized' pricing scheme proposed in a literature turned out to be an informationally infeasible procedure.

Minimality results about message space size obtained in this thesis are in accord to what we have known intuitively.

By introducing a 'message space of a level set' (§ 2.2), we succeeded to extract the dimension of a space of 'prices' from that of a message space previously studied by economists. In § 3.1, under the assumption of agentwise separability the number of resources—the dimension of price vector, was shown to be the minimum size of a message space of a level set. Then in § 3.3, we saw that the presence of a joint cost or an 'externality' increases size of a necessary message space. Through the analysis of a two-stage recourse model for the electric power pricing in § 4.4, we saw that uncertainty and intertemporality in users' valuations necessitates the 'price forecast'. Again this is in accord to what the Arrow-Debreu economy model suggests.

Two assignment problems of a digital pipe examined in Chapter 5 offered two extreme cases in regard to the possibility of sequential message exchanges. A multi-armed bandit problem admits sequential message exchanges and a matching problem does not. The difference was traced in the problem structures of the respective (centralized) problems. In general, intertemporality in valuation forces the exchange of messages at the beginning of planning periods.

In Chapter 6, a multi-armed bandit problem was further examined. In a stochastic version of the problem, pipe users were assumed statistically independent, which preserved the agentwise separability in stochastic setting.

An alternative proof of the optimality of the index rule was derived for a discrete time problem under the discounted reward criterion. The key idea was to convert the reward process to its concave envelope, which is pathwise nonincreasing and both dominated by and dominating the original process. The same idea was applied to prove the asymptotic optimality of the index rule under the average reward criterion. The analysis of the continuous time version of the problem revealed the hidden concavity of the problem and relation between the index and the Lagrange multiplier.

7.2 Future Work

Computation

In mechanism theory, equilibrium messages and the size of message space size have received the most attention. It does not address how fast the equilibrium is reached. However, the time is an important factor in design of procedures.

The theory of parallel and distributed computation in computer science may provide 'complexity measures' to quantify the amount of computational effort in the decentralized system. Each participant with his private valuation of resources may be thought of as a processor with the initial relevant data in its local memory. Then the design of a dynamical procedure involves networking the processors and devising an algorithm which runs on the networks.

Two of the relevant complexity measures in the area of the distributed computation are[2]

- time complexity: the time until the algorithm terminates.
- communication complexity: the number of messages transmitted in the course of the algorithm.

It would be nice if we could incorporate the minimum requirement on information-carrying capacity obtained from the 'static' analysis of the mechanism theory into the design of the networks and algorithms.

Limitation on communication capability

The results in this thesis indicate that inter-participant and intertemporal factors, and uncertainty in the system increase the necessary information-carrying capacity in order to sustain an efficient allocation in a decentralized manner.

However, the real-world systems, most notably our economic systems, have a physical limitation on communication capability.

Since it is impractical to implement the required information-carrying capacity in many cases, it is important to identify the capacity requirements of existing or proposed procedures and see 'how far off' they can be for typical cases and/or the worst case.

Stochastic system control

In a decentralized stochastic system, the underlying stochastic process may be partially and privately observed by participants. In such an instance 'learning about the system' in the sense of centralized system control will not be complete unless the observations are communicated.

Communication of private information about the stochastic process from the viewpoint of decentralized control is an unexplored but important area.

Bibliography

[1] M. A. Armstrong. *Basic Topology.* McGraw Hill, 1979. (Reprinted by Springer-Verlag in 1983).

[2] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation.* Prentice-Hall, 1989.

[3] G. Brown and M. B. Johnson. Public utility pricing and output under risk. *American Economic Review*, 59:119–128, Mar. 1969.

[4] M. C. Caramanis, B. R. E., and F. C. Schweppe. Optimal spot pricing: Practice and theory. *IEEE Transactions on Power Apparatus and Systems*, PAS-101:3234–3245, Sept. 82.

[5] H.-P. Chao and R. B. Wilson. Priority service: Pricing, investment, and market organization. *American Economic Review*, 77(5):899–916, Dec. 87.

[6] M. C. Crew and P. R. Kleindorfer. *The Economics of Public Utility Regulation.* MIT Press, Cambridge, Massachusetts, 1987. MIT Press Series on the Regulation of Economic Activity.

[7] T. W. Gedra and P. P. Varaiya. Markets and pricing for interruptible electric power. In *IEEE PES Winter 1992 Meeting*, 1992. Paper No.92 WM 169-3-PWRS. (To appear in IEEE Trans. Power Systems).

[8] J. C. Gittins. *Multi-armed Bandit Allocation Indices.* Wiley, 1989.

[9] J. C. Gittins and D. Jones. A dynamic allocation index for the sequential design of experiments. In J. Gani, K. Sarkadi, and I. Vince, editors, *Progress in Statistics, Euro. Meet. Statis., 1972, vol. 1*, pages 241–266. North Holland, 1974.

[10] L. Hurwicz. On informational decentralization and efficiency in resource allocation mechanisms. In S. Reiter, editor, *Studies in Mathematical Economics*, pages 238–350. The Mathematical Association of America, 1986.

[11] R. J. Kaye, F. F. Wu, and P. Varaiya. Pricing for system security. *IEEE Transactions on Power Apparatus and Systems*. To appear.

[12] T. C. Koopmans and M. Beckmann. Assignment problems and the location of economic activities. *Econometrica*, 25(1):53–76, Jan. 1957.

[13] A. Mandelbaum. Discrete multi-armed bandits and multi-parameter processes. *Probability Theory and Related Fields*, 71:129–147, 1986.

[14] A. Mandelbaum. Continuous multi-armed bandits and multiparameter processes. *The Annals of Probability*, 15(4):1527–1556, 1987.

[15] K. Mount and S. Reiter. The information size of message spaces. *Journal of Economic Theory*, 8:161–192, 1974.

[16] K. Mount and S. Reiter. A model of computing with human agents. Discussion Paper 890, North Western University, June 1990.

[17] S. S. Oren, S. A. Smith, R. B. Wilson, and H.-P. Chao. Priority service: Unbundling the quality attributes of electric power. EPRI Interim Report, Electric Power Research Institute, Palo Alto, California, 1986.

[18] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, Englewood Cliffs, New Jersey, 1982.

[19] S. Reiter. Informational incentive and performance in the *new* welfare economics. In S. Reiter, editor, *Studies in Mathematical Economics*, pages 219–237. The Mathematical Association of America, 1986.

[20] R. T. Rockafellar and R. J.-B. Wets. The optimal recourse problem in discrete time: L1-multipliers for inequality constraints. *SIAM J. Control and Optimization*, 16(1), Jan. 1978.

[21] J. N. R. Sandell, P. Varaiya, M. Athans, and M. G. Safonov. Survey of decentralized control methods for large scale systems. *IEEE Transactions on Automatic Control*, AC-23(2):108–128, Apr. 1978.

[22] F. C. Schweppe, M. C. Caramanis, R. E. Tabors, and R. E. Bohn. *Spot Pricing of Electricity*. Kluwer Academic Publishers, 1988.

[23] C.-W. Tan and P. Varaiya. A model for pricing interruptible electric power service. In G. B. Di Masai, A. Gombani, and A. B. Kurzhansky, editors, *Modeling, Estimation and Control of Systems with Uncertainty*, pages 423–443. Birkhauser, 1991. Volume 10 in Progress in Systems and Control Theory.

[24] R. Turvey. Public utility pricing and output under risk: Comment. *American Economic Review*, 60:485–486, June 1970.

[25] P. Varaiya, J. Warland, and C. Buyukkoc. Extensions of the multi-armed bandit problem: The discounted case. *IEEE Trans. Automatic Control*, AC-30(5), May 1985.

[26] H. R. Varian. *Microeconomic Analysis*. Norton, second edition, 1984.

[27] W. Vickrey. Responsive pricing of public utility services. *Bell Journal of Management Science*, Spring:337–346, 1971.

[28] M. Walker. On the informational size of message spaces. *Journal of Economic Theory*, 15:366–375, 1977.

[29] R. Weber. On the optimality of the Gittins index policy. *The Annals of Applied Probability*. To appear.

[30] G. Weiss. Turnpike optimality of Smith's rule in parallel machines stochastic scheduling. Report Series J-90-04, Georgia Institute of Technology, Industrial and Systems Engineering, 1990.

[31] P. Whittle. Multi-armed bandits and the Gittins index. *J. Roy. Statis. Soc.*, 42:143–149, 1980.

[32] P. Whittle. Arm-acquiring bandits. *The Annals of Probability*, 9:284–292, 1981.