The Communication Complexity of Efficient Allocation Problems*

Noam Nisan[†] Ilya Segal[‡]

First Draft: November 2001
This Draft: March 6, 2002

Abstract

We analyze the communication burden of surplus-maximizing allocations. We study both the continuous and discrete models of communication, measuring its burden with the dimensionality of the message space and the number of transmitted bits, respectively. In both cases, we offer a lower bound on the amount of communication. This bound is applied to the problem of allocating L heterogeneous objects among N agents, whose valuations are (i) unrestricted, (ii) submodular, or (iii) homogeneous in objects. In cases (i) and (ii), efficiency requires exponential communication in L. Furthermore, in case (i), polynomial communication in L cannot ensure a higher surplus than selling all objects as a bundle. On the other hand, in case (iii), exact efficiency requires the transmission of L numbers, but can be approximated arbitrarily closely using only $O(\log L)$ bits. When a Walrasian equilibrium with per-item prices exists, efficiency is achieved with deterministic communication that is polynomial in L.

*We are grateful to Paul Milgrom, Luis Rayo, Yoav Shoham, Steve Tadelis, Moshe Tenneholtz, and participants of seminars at Stanford and the DIMACS Workshop on Computational Issues in Game Theory and Mechanism Design for useful comments.

[†]Institute of Computer Science, Hebrew University, Jerusalem, Israel, email: noam@cs.huji.ac.il

[‡]Department of Economics, Stanford University, Stanford, CA 94305; email: ilya.segal@stanford.edu

1 Introduction

We have recently seen great interest in so-called *combinatorial auctions*, which allocate heterogeneous indivisible items among bidders whose preferences for combinations of items can exhibit complementarity or substitutability (see, e.g., Vohra and de Vries (2000) for an overview). The objective of an auction is to elicit enough information about bidders' preferences so as to realize an efficient or approximately efficient allocation. Recent important applications include auctions of spectrum licenses and online procurement.

The mechanism design literature has examined the bidders' incentives to reveal their valuations truthfully, using the Revelation Principle to focus on direct revelation mechanisms. For example, the Vickrey-Groves-Clarke direct revelation mechanism implements the efficient allocation and is incentive-compatible in the private-value environment. At the same time, it has been recognized that the full revelation of bidders' preferences may require a prohibitive amount of communication. Indeed, every bidder has a valuation for each subset of the items, and the number of such subsets is exponential in the number of objects. With 30 items, full revelation of such preferences would require the communication of more than one billion numbers, which is beyond the capabilities of any human or machine.[1]

Recognition of the communication problem has prompted researchers to examine the properties of simpler mechanisms, which do not fully reveal valuations. For example, auctions that quote only per-item prices were shown to be efficient in some restrictive settings (Gul and Stacchetti (2000), Milgrom (2000)). Also, ascending-bid auctions have been examined, in the hope of economizing on communication by asking bidders to announce their valuations only for the subsets revealed to be relevant in previous bidding (Ausubel and Milgrom (2001), Bikhchandani et al. (2001)). However, until now there has been no analysis of the *minimum* amount of communication required to achieve or

[1] Even if it is known a priori that the bidder is not interested in more than, say, six items, full revelation would require the communication of more than half a million numbers ($\binom{30}{6}$).

approximate efficiency. The present paper closes this gap, identifying the communication burden of the problem of implementing efficient or approximately efficient allocations.

We analyze the communication problem using techniques developed in parallel in economics and computer science.[2] In economics, these techniques were developed in the literature on the *dimensionality of message spaces*, originated by Hurwicz (1960) and Mount and Reiter (1974). This literature measures the communication burden of an allocation problem with the number of real variables that need to be announced to verify that a desired allocation is implemented. The key achievement of that literature is a formalization of Hayek's (1945) idea that in standard "convex" environments, the Walrasian mechanism is "informationally efficient," i.e., it realizes Pareto efficient allocations with the least amount of communication. The Walrasian mechanism involves only the announcement of prices, along with the allocation, which is much more economical than full revelation of agents' preferences. In nonconvex environments, however, a Walrasian equilibrium with linear prices need not exist, in which case much more extensive communication may be needed (see Calsamiglia (1977) for an early example). The problem of combinatorial allocation of indivisible items is nonconvex, and a Walrasian equilibrium with linear (per-item) prices in general does not exist (with some notable exceptions discussed below). The question examined in this paper can be formulated as how many real-valued "prices" need to be announced to verify efficiency in such discrete allocation problems.

A parallel approach to measuring communication has developed in the computer science field of *communication complexity*, pioneered by Yao (1979) and surveyed in Kushilevitz and Nisan (1997). The main difference from the economic approach is that it considers a discrete (finite) setting. All "inputs" (in our case, agents' valuations) are given with a finite precision, and the complexity of communication is measured with the number of bits transmitted.

Our main result is a lower bound on the amount of communication required to im-

[2] For an earlier discussion relating the economic and computer science approaches to communication complexity, see Marschak (1996).

plement efficiency in both the continuous and discrete cases. We obtain the bound by showing that all distinct states of the world in which the total surplus is constant across all outcomes must give rise to different communications. In the terminology of computer science, such states constitute a "fooling set." A lower bound on communication is given in the continuous case by the dimensionality of the fooling set, and in the discrete case by the binary log of its cardinality.

For the problem of allocating L heterogeneous objects to N agents whose valuations over subsets of objects are unrestricted, our lower bound implies that the required communication is exponential in the number L of items to be allocated. Even if the agents' valuations are known to be submodular (i.e., exhibit diminishing marginal utility of objects), exponential communication in L is still required.

These results are extended to the problem of finding *approximately efficient* allocations, i.e., those whose total surplus is within some predetermined factor from optimal. Since arbitrarily close approximation can be obtained with finite communication, here we focus on the discrete measure counting the communicated bits. Most of the approximation results follow from the simple observation that in the discrete case where the agents' valuations are described with a given precision, sufficiently close approximation is equivalent to exact efficiency. Hence, the lower bounds obtained for exact efficiency can be brought to bear on this case.

In the case of unrestricted valuations, we show that any mechanism guaranteed to achieve more than $1/N$ of the maximum available surplus involves the communication of an exponential number of bits in L. On the other hand, share $1/N$ of the surplus is attained by an auction selling all items as a bundle. Therefore, no "practical" mechanism can guarantee an improvement upon the simple bundled auction.[3]

We also identify a setting in which there is a drastic difference between the com-

[3] For the case of submodular valuations, we only establish a weaker result, ruling out so-called "Fully Polynomial Approximation Schemes," defined as approximating efficiency within ε using polynomial communication in the parameters and ε^{-1}. (An economic example of FPAS is an ascending-bid auction with L per-item prices and bid increment ε.)

munication burdens of exact and approximate efficiency. This setting has homogeneous valuations (i.e., agents only case about the number of objects we consume). Exact efficiency in this case requires the transmission of L real numbers, but arbitrarily close approximation can be achieved using only $O(\log L)$ bits. This setting is similar to that in Calsamiglia (1977), where the good infinitely divisible rather than discrete. By allocating the good in optimally chosen discrete units, an enormous savings in communication can be achieved in Calsamiglia's setting with only a slight sacrifice in economic efficiency.

The approximation results described above require uniform surplus approximation across states. However, we show that in the case of unrestricted valuations, the same lower bound applies for the weaker problem of approximating the *expected* surplus. Namely, using results of Nisan (2001), we establish that for a certain joint probability distribution over the agent's valuations, any mechanism achieving a higher expected surplus than $1/N$ of maximum possible must communicate, on expectation, an exponential number of bits. While the constructed distribution may not be realistic, the point is that obtaining non-trivial efficiency with sub-exponential communication is possible only for a restricted class of probability distributions over valuations.

Our lower bounds on communication are obtained under what is called the "verification" scenario in economics, and the "nondeterministic" scenario in computer science. This scenario supposes the existence of an omniscient oracle, who must prove to the agents that a given allocation should be implemented. (In economics, the oracle is known as the "Walrasian auctioneer", who can "guess" the equilibrium prices.) The problem of communication in the absence of such an oracle, called "deterministic communication," is more difficult. Deterministic communication is usually facilitated by having multiple rounds, where at each round an agent only transmits information that is revealed in the previous rounds to be relevant for the outcome. An economic example of such multi-round communication is given by "tatonnement" processes used to attain a Walrasian equilibrium. In particular, ascending-bid combinatorial auctions have recently been considered by Ausubel and Milgrom (2001), Bikhchandani et al. (2001), Parkes (1999), Parkes and Ungar (2001). The hope is that these auctions economize on communication, by asking the bidders in each round to send only their valuations for a small number of subsets that have been previously revealed to be "relevant." However, while such multi-round mechanisms may indeed reduce the amount of deterministic communication, they are not of any help when nondeterministic communication is allowed. Indeed, an omniscient oracle can announce the whole equilibrium communication sequence (in particular, the "relevant" subsets) up front, and needs only to verify that no agent objects to it. Therefore, our lower bounds apply to communication in any mechanism, be it nondeterministic or deterministic, single- or multi-round.

Our results imply that any practical combinatorial auction can only approximate efficiency for some restricted class of valuations (or probability distributions over valuations). For example, if the linear programming relaxation of the integer programming problem of efficient combinatorial allocation yields a feasible allocation, then a Walrasian equilibrium with per-item prices exists (Bikhchandani and Mamer (1997)), and it constitutes a polynomial nondeterministic protocol realizing efficient allocations. (One way to ensure this existence is with the "gross substitutes" property of Kelso and Crawford (1982) and Gul and Stacchetti (1999), but other conditions are known (Nisan (2000), Vohra and de Vries (2000)). For this case, we suggest a polynomial *deterministic* protocol that achieves efficiency.[4] This protocol mimics a separation-based linear programming algorithm, and it can be made incentive-compatible with little added communication. While this protocol is not very natural, it does show that whenever a Walrasian equilibrium with per-item prices is known to exist, efficiency can also be obtained with a *deterministic* incentive-compatible mechanism using polynomial communication.

The remainder of the paper is organized as follows. Section 2 describes the general setup. Section 3 discusses the different measures of communication complexity. Section 4 obtains a general lower bound on communication using the "fooling set" technique.

[4]This protocol works for more general valuations than those suggested by Kelso and Crawford (1982) and Ausubel (2000). Also, the mechanism of Kelso and Crawford (1982) and Ausubel (2000) only achieve FPAS - i.e., obtain ε-approximation in time proportional to ε^{-1}, while our mechanism takes so-called "true polynomial time," i.e., polynomial in $\log \varepsilon^{-1}$, the representation length of ε.

In Section 5, this bound is applied to the problem of efficient combinatorial allocation for different classes of agents' valuations. In Section 6, the same technique is applied to the problem of approximate efficiency. Section 7 suggests a polynomial deterministic mechanism that implements efficiency whenever a Walrasian equilibrium with per-item prices exists. In conclusion, we discuss the relation of our results to the computational complexity and mechanism design literatures.

2 Setup

Let N be the finite set of agents, and K be the set of outcomes. (With a slight abuse of notation, we will use the same letter to denote a set and its cardinality when this causes no confusion. At this point, the set K need not be finite, though it will be in most applications.) An agent's valuation assigns real values to all outcomes, and is therefore represented with a vector in $\mathbb{R}^K$. The class of possible valuations of agent $i \in N$ is denoted by $V^i \subset \mathbb{R}^K$. Agent i's valuation $v^i \in V^i$ is assumed to be his privately observed "type." A *state* is a profile of valuations: $\left(v^1, ..., v^N\right) \in V \equiv V^1 \times \ldots \times V^N \subset \mathbb{R}^{NK}$.

A leading application considered in this paper is the problem of allocating L items among the agents. The outcome set in this problem is $K = N^L$.

The efficient choice correspondence $K^* : V \twoheadrightarrow K$ is defined by[5]

$$K^*(v) = \arg\max_{k \in K} \sum_{i \in N} v_k^i \ \forall v \in V.$$

Observe that valuations $(v_1^i, \ldots, v_K^i)$ and $(v_1^i + a, \ldots, v_K^i + a)$ describe the same preferences for any $a \in \mathbb{R}$. It will be convenient to rule out such situations, by assuming that each agent i's class of valuations V^i is *normalized*, meaning that it does not contain the two above valuations for any $a \neq 0$. A simple way to normalize valuations is by assigning the value of zero to one of the outcomes.

[5] This notion of efficiency is based on the implicit assumption that the agents' utilities are quasilinear in monetary transfers, and that such transfers can be used to compensate agents.

3 The Communication Problem

3.1 Nondeterministic Communication

We begin with a formal definition:

Definition 1 *A* (nondeterministic) protocol *is a triple* $\langle M, \mu, h \rangle$*, where* M *is the message set,* $\mu : V \twoheadrightarrow M$ *is the message correspondence, and* $h : M \to K$ *is the outcome function, and the message correspondence* μ *has the following two properties:*

(a) $\mu(v) \neq \emptyset \ \forall v \in V$*,*

(b) $\mu(v) = \cap_{i \in N} \mu^i(v^i) \ \forall v \in V$*, where* $\mu^i : V^i \twoheadrightarrow M \ \forall i \in N$*.*

The protocol $\langle M, \mu, h \rangle$ realizes *choice correspondence* $F : V \twoheadrightarrow K$ *if* $h(\mu(v)) \subset F(v)$ $\forall v \in V$*.*

This definition corresponds to a communication scenario called "verification" in economics and "nondeterministic" in computer science. In the scenario, an omniscient oracle knows the state of the world v and consequently the set $F(v)$ of "desirable" outcomes, but needs to prove to an outsider that an outcome $k \in F(v)$ is desirable. He does this by publicly announcing a message $m \in M$. Each agent i either accepts or rejects the message, doing this on the basis of his own type v^i. The message correspondence $\mu(v)$ describes the set of messages acceptable to all agents in state v. If message m is accepted by all agents, then the oracle implements outcome $h(m)$.

In this interpretation, condition (a) requires that there exist an acceptable message in each state. Condition (b) follows from the fact that each agent does not observe other agents' types when making his acceptance decision, thus the set of messages he accepts is a function $\mu^i(v^i)$ of his own type v^i only. This condition is known as "privacy preservation" in the economic literature.

Conditions (a) and (b) can also be given a geometric interpretation. Condition (a) can be interpreted as saying that the collection of sets $\{\mu^{-1}(m)\}_{m \in M}$ is a covering of the state space V. In this interpretation, condition (b) says that each element of the covering is a product ("rectangular") set (see Figure 1 for an illustration). This is why condition (b) is known as the "rectangle property" in computer science.

Finally, the definition of realization requires that acceptable messages give rise only to desirable outcomes, which means that outcome $h(m)$ is "proven" to be desirable whenever message m is accepted by all agents.

A famous nondeterministic communication protocol in economics is the Walrasian equilibrium. The role of the oracle is played by the "Walrasian auctioneer," who guesses the equilibrium prices and quantities.[6] Each agent accepts the message if and only if the announced quantities constitute his optimal choice from the budget set given by the announced prices. This protocol realizes the Pareto efficient allocation correspondence.

3.2 Measures of Complexity

The economic literature considers problems with continuous type spaces V^i, in which efficiency usually requires protocols with continuous message spaces M. The amount of communication is measured with the dimensionality of the message space M. A potential problem with this approach is that the message correspondence could be chosen to be a one-to-one function $\mu : \mathbb{R}^{NK} \to \mathbb{R}$ (e.g., the inverse Peano function), which "smuggles" any amount of information in a one-dimensional message space. To rule out such smuggling, it is customary to impose a regularity restriction on the message correspondence. We follow Walker (1977) and Sato (1981) in requiring that the message space M be a Hausdorff topological space, and that the message correspondence μ be *locally threaded* - which means that it has a continuous selection on an open neighborhood of every $v \in V$.

As suggested by Walker (1977), we define the "dimensionality" of the message space by comparing it with Euclidean spaces using the *Frechet ordering.* According to this

[6] The choice of quantities can often be delegated to the agents, but not always - for example, when production exhibits constant returns to scale, the oracle needs to assign market shares.

ordering, $M \geq_F \mathbb{R}^S$ if $\mathbb{R}^S$ can be homeomorphically embedded in M (in this case, we will also write $\dim M \geq S$).

The computer science literature, in contrast, considers finite state spaces, in which case a finite message space M obviously suffices. The relevant complexity measure is then the number of bits needed to encode a message from M, which equals $\log_2 |M|$.

When the agents' valuation classes V^i are continuous, efficiency will typically require an infinite (continuous) message space, and the continuous complexity measure will be used. On the other hand, when valuations given with a finite precision, we will use the discrete complexity measure. We also consider the more interesting case in which only approximate efficiency is required, and it can be achieved by announcing valuations that are rounded-off with a finite precision. The precision can then be chosen endogenously to achieve a given degree of approximation.

3.3 Relation to Deterministic Communication

The notion of communication defined above does not describe a realistic process by which a state v results in a message m. For example, in the case of the Walrasian protocol, it does not explain how the Walrasian prices are obtained. A more realistic concept, called deterministic communication, defines communication as a sequence of messages sent by the agents (see Kushilevitz and Nisan 1997). An economic example of deterministic communication is a "tatonnement" process that converges to the Walrasian outcome.

In the language of game theory, a deterministic protocol is an extensive-form message game along with the agents' strategies (complete plans contingent on their types and on observed history) specified in the game.[7] Since incentives are not considered, without loss of generality the game can taken to have perfectly observed histories.

The computer science literature considers finite deterministic communication, and

[7] More precisely, it is a "game form" because the agents' payoffs are not specified, and are in fact irrelevant since the agents' behavior is not required to be incentive-compatible. Instead, the agents are assumed to follow blindly the strategies assigned to them. See the Conclusion for a discussion of the effect of imposing incentive-compatibility constraints.

restricts attention to games with two actions at each node (interpreted as sending a bit of information). The ("worst-case") deterministic communication complexity is then defined as the maximum depth of the game tree, i.e., the maximum number of bits transmitted over all states.[8]

It is well known that deterministic communication is "harder" than nondeterministic. To see this, note that any deterministic protocol can be converted into nondeterministic without an increase in communication. The converted protocol works as follows: the oracle announces the equilibrium path of the game in the given state. Each agent i accepts a path if and only if at each node on it at which it is his turn to move, the strategy assigned to his type v^i is to move along the path. This protocol yields the same outcome as the original deterministic protocol. Its complexity equals the binary log of the number of terminal nodes in the game tree, which cannot exceed the maximum depth of the tree, which is the complexity of the deterministic protocol. Therefore, nondeterministic complexity offers a lower bound on deterministic complexity. (A similar argument can be made for continuous communication, where real-valued messages can be sent at each node.)

3.4 Relation to "Black-Box" Computational Complexity

One can consider the problem of *computing* a selection from the efficient choice correspondence K^*. However, in our settings of interest, such as that of combinatorial allocation of a large number L of objects, the agents' valuations v^i that serve as inputs to computation may hold a huge amount of information. In such cases, one must consider how a computational algorithm will access this information. The most general model will allow the algorithm to ask a "black box" an arbitrary question about a valuation v^i and receive an answer for any such question. More restricted models will define which queries the black box for v^i will answer - e.g., the "valuation oracle" model will only answer queries

[8]One may instead be interested in the expected number of bits transmitted given a certain probability distribution over states, rather than the worst-case number. Such *distributional* complexity is considered in subsection 6.6.

of the form "what is v_k^i".

A trivial, yet important, observation is that the general black box model is equivalent to the deterministic communication complexity model. A stage in a deterministic communication protocol where agent i has observed the message history $\langle m_1...m_t\rangle$ and responds with his own message m_{t+1} is equivalent to a black box that is asked the query $\langle m_1...m_t\rangle$ and responds with the answer m_{t+1}. In both cases m_{t+1} depends only on v^i and $\langle m_1...m_t\rangle$. Thus, our lower bounds on communication also imply lower bounds on computational complexity in the general black box model.

3.5 Relation to Walrasian Equilibria: Messages as Prices

A Walrasian Equilibrium is the classical economic example of continuous nondeterministic communication. The efficiency of this protocol is ensured by the First Welfare Theorem. However, the existence property (a) is traditionally established under convexity assumptions that are not satisfied in the combinatorial allocation setting.

One way to ensure existence is by introducing *Lindahl prices*, i.e., personalized prices that depend on the whole outcome. The *Lindahl-Walras protocol* can be defined as follows: the designer announces a Lindahl price vector $(p_k^i)_{i\in N,k\in K}$ and an outcome k, and each agent accepts if and only if $k \in \arg\max_{j\in K} \left[v_j^i - p_j^i\right]$ - i.e., k is his preferred outcome given the price vector. We require that $\sum_i p_k^i = 0$ for all $k \in K$, reflecting the fact that each outcome k is costless to the designer. Also, we can normalize $\sum_k p_k^i = 0$ for all $i \in N$, since adding a constant to prices for all outcomes does not changes agents' choices. Thus, the dimensionality of the message space in the Lindahl-Walras protocol is $(N-1)(K-1)$.

By the First Welfare Theorem argument, a Lindahl-Walras equilibrium is efficient. Also, such an equilibrium always exists. For example, for each k we can take $p_k^i = v_k^i - \frac{1}{K}\sum_j v_j^i$ for $i = 1, ..., N-1$ - that is, let personalized prices to all agents but agent N mirror their valuations, and let agent N's prices be $p_k^N = -\sum_{i=1}^{N-1} p_k^i$. Thus, $(N-1)(K-1)$ gives an upper bound on the dimensionality of message space required to realize efficiency. This upper bound can also be realized with *deterministic* communication, in which all agents but agent N reveal their valuations, and then agent N announces an efficient decision.

The problem with the Lindahl-Walras protocol is that it requires a large amount of communication – equivalent to full revelation of valuations by all but one agent. This raises the question whether we could restrict attention to a lower-dimensional subset of the Lindahl price space while preserving the existence property. For example, per-item Walrasian prices can be interpreted as Lindahl prices that are restricted to be linear in objects and anonymous, which drastically reduces the dimensionality of the price space, but in general violates the existence property. In this terminology, our search for the message space of minimum dimensionality can be interpreted as looking for the minimum-dimensional subspace of the Lindahl price space that ensures the existence of a Walrasian equilibrium.

4 The Basic Lower Bound

Our lower bound on communication is obtained by constructing a set that is called a "fooling set" in computer science and a "set with the uniqueness property" in economics. The simplest illustration of this construction obtains in the case where a single indivisible object is to be allocated between two agents. Letting v^i denote agent i's valuation for this object, a state of the world is described by a pair (v^1, v^2). Efficiency requires giving the object to the agent with the higher valuation (see Figure 2).

The basic idea is that an efficient protocol cannot use the same message in two different "diagonal" states of the world, (v, v) and (w, w), where, say, $w > v$. Suppose in negation that there exists a single message that occurs in both states. Then each agent 1 and 2 accepts this message when his valuation is v or w. But then the "privacy preservation"/"rectangle" property of communication implies that both agents accept the same message in states (w, v) and (v, w), and therefore the same outcome will obtain in the two states. But this contradicts efficiency, which requires giving the object to agent

1 in state (w, v) and to agent 2 in state (v, w). Therefore, all distinct diagonal states in the example must give rise to distinct messages. A set of states with this property is called a "fooling set." Communication of the fooling set bounds from below the number of messages required to realize efficiency. In particular, in the case where each agent's valuation can be in $[0, 1]$, there is a continuum of diagonal states, and therefore at least a continuum of possible messages is needed to realize efficiency.[9]

Extending this argument to arbitrary many agents and outcomes yields

Lemma 1 *Suppose each V^i is normalized, and $v, w \in V^1 \times \ldots \times V^N$ such that $v \neq w$ and $K^*(v) = K^*(w) = K$. Then in any efficient protocol $\langle M, \mu, h \rangle$, $\mu(v) \cap \mu(w) = \emptyset$.*

Proof. Suppose in negation that $m \in \mu(v) \cap \mu(w)$. Then by the rectangle property, we must also have for all $i \in N$, $m \in \mu(v^i, w^{-i})$ and $m \in \mu(w^i, v^{-i})$. Hence, for outcome $k^* = h(m) \in K$ we must have

$$k^* \in \arg\max_{k \in K} \left[v_k^i + \sum_{j \neq i} w_k^j \right] \text{ and } k^* \in \arg\max_{k \in K} \left[w_k^i + \sum_{j \neq i} v_k^j \right].$$

But by assumption, $\sum_j v_k^j$ and $\sum_j w_k^j$ do not depend on k, which allows us to rewrite the above display as

$$k^* \in \arg\max_{k \in K} \left[v_k^i - w_k^i \right] \text{ and } k^* \in \arg\max_{k \in K} \left[w_k^i - v_k^i \right] = \arg\max_{k \in K} \left[v_k^i - w_k^i \right].$$

This is only possible when $v_k^i - w_k^i$ does not depend on $k \in K$, and since V^i is normalized, we must have $v^i = w^i$. Since this argument applies for each $i \in N$, we must have $v = w$, contradicting our assumption. ■

In words, the Lemma says that any two distinct states in which all outcomes are indifferent from the social viewpoint must give rise to distinct messages. Formally, the message correspondence in any efficient protocol is injective on the set

$$\mathfrak{F} = \left\{ v \in V^1 \times \ldots \times V^N : K^*(v) = K \right\},$$

[9] A one-dimensional continuum indeed suffices here - e.g., consider the Walrasian equilibrium where the auctioneer announces price $\left(v^1 + v^2\right)/2$.

i.e., the set can serve as a "fooling set." The communication burden of efficiency is at least that of communicating a point in the fooling set. For the discrete and continuous communication measures, this implies

Proposition 1 *Suppose each V^i is normalized, and $\langle M, \mu, h \rangle$ is an efficient protocol. Then $\log_2 |M| \geq \log_2 |\mathfrak{F}|$ and $\dim M \geq \dim \mathfrak{F}$.*

Proof. The first inequality follows immediately from Lemma 1. The second inequality follows from Lemma 1 and our topological assumptions on the protocol using Lemma 1 in Sato (1981). ■

5 Application to Combinatorial Allocation Problems

In this section we apply Proposition 1 to the problem of allocating L items among N agents. The outcome set in this problem is $K = N^L$, where $k(l)$ denotes the agent holding object $l \in L$ in outcome $k \in K$. We will maintain the following assumptions on agents' valuations:

- **No Externalities (NE):** For each agent i, $v_k^i = u^i(k^{-1}(i))$, where $u^i : 2^L \to \mathbb{R}$.

In words, each agent i cares only about the subset of items $k^{-1}(i)$ allocated to him. Let $U^i \subset \mathbb{R}^{2^L}$ denote the class of agent i's valuations for his own consumption.

- **Monotonicity (M):** For each $u^i \in U^i$, $u^i(S)$ is nondecreasing in $S \subset L$.

- **Normalization (N):** $u^i(\emptyset) = 0$ for all $i \in N$.

- **Boundedness (B):** $u^i(S) \leq 1$ for all $i \in N$, $S \subset N$.

All these restrictions could only reduce the communication burden, and thus only strengthen our lower bounds. Restriction (N) is in fact without loss of generality, as argued in Section 2. Restrictions (M) and (B) are not needed in this section, but will play a role in the next section's analysis of approximation.

We will examine in turn the communication requirements for three classes of valuations:

(i) **Unrestricted valuations:** Each U^i is the set of valuations satisfying (NE), (M), (N), and (B).

(ii) **Submodular valuations:** Each U^i is the set of valuations $u^i : 2^L \to \mathbb{R}$ satisfying (NE), (M), (N), (B), and in addition *submodularity*:

$$u^i(S \cup T) + u^i(S \cap T) \leq u^i(S) + u^i(T) \text{ for all } S, T \subset L.$$

It is well known that submodularity can be alternatively defined by requiring that the marginal benefit of an item $l \in L$, $u^i(S \cup l) - u^i(S)$ be nonincreasing in $S \subset L$.

(iii) **Homogeneous valuations:** Each U^i is the set of valuations $u^i : 2^L \to \mathbb{R}$ satisfying (NE), (M), (N), (B), and in addition taking the form $u^i(S) = \phi(|S|)$ for all $S \subset N$, where $\phi : \{0...L\} \to \mathbb{R}$. That is, agents care only about the number of items they receive.

We will also formulate results for the *discrete case*, which in addition satisfies

- **Discreteness (D):** $u^i(S) \in \{0, 1/R, 2/R, ..., 1\}$ for all $i \in N$, $S \subset N$.

In the discrete case, valuations for each set are given with finite precision R, and can be encoded with $\log_2 (R + 1)$ bits. For now we treat R as given, but in the next section the precision with which agents' valuations are revealed will be chosen endogenously to achieve the desired approximation. The *continuous case* is one where (D) is not imposed, and the range of valuations is $[0, 1]$.

Our lower bounds are obtained by focusing on the case where $N = 2$. (The bounds extend to $N > 2$, by considering the case where all agents but two have valuations that are identically zero.) In most cases, we apply Proposition 1 in the following way: for each valuation $u \in U^1$, define a *dual* valuation u^* by

$$u^*(S) = u(L) - u(L \backslash S).$$

The idea is that if agent 1 has valuation u and agent 2 has the dual valuation u^*, the total surplus for any allocation of objects between the agents is $u(L)$. Dual valuations are illustrated in Figure 3 (where the horizontal axis represents one "cut" of the allocation space, corresponding to reallocating items from agent 1 to agent 2 in a particular order). By Proposition 1, the set of states $\mathfrak{F} = \{(u, u^*) : (u, u^*) \in U^1 \times U^2\}$ constitutes a fooling set, and the dimensionality of the message space in the continuous case is at least $\dim \mathfrak{F}$, while the number of bits in the discrete case is at least $\log_2 |\mathfrak{F}|$.

5.1 Unrestricted Valuations

One may easily verify that the dual of any valuation in the set U of unrestricted valuations is also in U, thus $\{(u, u^*) : u \in U\}$ is a fooling set. Communicating the fooling set is equivalent to communicating a valuation from U. Thus, by Proposition 1, the lower bound is $\dim U$ and $\log_2 |U|$ in the continuous and discrete case respectively. Note that this lower bound is essentially tight for $N = 2$, since it is achieved by the protocol in which agent 1 announces his valuation and then agent 1 announces an efficient allocation.

In the continuous case, $\dim U = 2^L - 1$. In the discrete case, $|U|$ is the number of monotone functions on subsets of L with range $\{0, 1/R, ..., 1\}$. This number can be bounded below by $(R+1)^{\binom{L}{L/2}}$, by considering valuations with $u(S) = 0$ for $|S| < L/2$ and $u(S) = 1$ for $|S| > L/2$. Thus, we have

Corollary 1 *With continuous unrestricted valuations, the dimensionality of the message space in an efficient protocol is at least* $2^L - 1$. *With discrete unrestricted valuations, the number of bits communicated by an efficient protocol is at least* $\binom{L}{L/2} \log_2 (R+1)$.

Therefore, the communication burden is exponential in the number of objects, both in the continuous and discrete case.[10] It an interesting open problem how the required communication grows with the number N of agents.[11]

[10] More precisely, for the discrete case, Stirling's formula implies that
$\binom{L}{L/2} \sim \sqrt{2/(\pi L)} \cdot 2^L$.

[11] That such an increase might be required is suggested by the finding of Bikhchandani and Ostroy

5.2 Submodular Valuations

The approach of the previous subsection cannot be applied to this case directly, since the dual of a submodular valuation is not submodular (unless both are identically constant). We modify the approach by defining duality in such a way that the social surplus is constant only on the outcomes involving even division of objects. Namely, let

$$\widetilde{K} = \left\{k \in K : \left|k^{-1}(1)\right| = \left|k^{-1}(2)\right| = L/2\right\}.$$

Consider the set $\widetilde{U}$ of valuations $u \in \mathbb{R}^{2^L}$ such that

- $u(S) = 4|S|d$ for $|S| < L/2$,
- $u(S) = 2Ld$ for $|S| > L/2$,
- $u(S) \in [(2L-1)d, 2Ld]$ for $|S| = L/2$,
- $u(\{1, ..., L/2\}) + u(\{L/2+1, ..., L\}) = (4L-1)\,d$,

where in the continuous case $d = 1/\,(2L)$, and in the discrete case it is rounded down to a multiple of $1/R$, i.e., $d = [R/\,(2L)]\,/R$, where $[\cdot]$ stands for the integer part. One can easily verify that in both cases $\widetilde{U}$ is a subset of the class U of submodular valuations.

Note that in any state $(u^1, u^2) \in \widetilde{U} \times \widetilde{U}$, all efficient outcomes lie in $\tilde{K}$. (Indeed, any allocation from $\tilde{K}$ brings a total surplus of at least $(4L-2)\,d$, while any other allocation brings a total surplus of at most $(4L-4)\,d$.) This allows us to apply Proposition 1 with the two agents' valuations restricted to $\widetilde{U}$ and the outcomes restricted to $\widetilde{K}$ (the last bullet above ensures that thus constructed valuation classes on $\widetilde{K}$ are normalized). For this purpose, for each $u \in \widetilde{U}$, define its "quasi-dual" $\widehat{u} \in \widetilde{U}$ as follows:

- $\widehat{u}(S) = (4L-1)\,d - u(L \backslash S)$ for $|S| = L/2$,
- $\widehat{u}(S) = 4|S|d$ for $|S| < L/2$.

(2001) that Lindahl prices, i.e., personalized prices for subsets of objects, are needed to ensure the existence of a Walrasian equilibrium.

- $\widehat{u}(S) = 2Ld$ for $|S| > L/2$,

By construction, the set of efficient outcomes in any state $(u, \widehat{u}) \in \widetilde{U} \times \widetilde{U}$ is exactly $\widetilde{K}$, hence by Proposition 1 these outcomes constitute a fooling set. In the continuous case, the dimensionality of this set is $\dim \widetilde{U} = \left|\widetilde{K}\right| - 1$, while in the discrete case, its cardinality is $\left|\widetilde{U}\right| = (dR+1)^{|\widetilde{K}|-1}$ (where $\left|\widetilde{K}\right| = \binom{L}{L/2}$). Thus, Proposition 1 implies

Corollary 2 *With continuous submodular valuations, the dimensionality of the message space in an efficient protocol is at least* $\binom{L}{L/2} - 1$*. With discrete submodular valuations, the number of bits communicated by an efficient protocol is at least* $\left[\binom{L}{L/2} - 1\right] \log_2 \left(\left[\frac{R}{2L}\right] + 1\right)$*.*

In the continuous case, this Corollary implies that exponential communication is still needed for efficiency. In the discrete case, however, note that the lower bound is non-trivial only when $R \geq 2L$. Thus, for a fixed precision R, the discrete bound is not very useful. However, the bound will have implications in the next section, where input precision will be chosen endogenously to achieve progressively better and better approximation.

5.3 Homogeneous Valuations

Since the dual of a homogeneous valuation is homogeneous, the fooling set can be constructed here in the same way as with unrestricted valuations: $\{(u, u^*) : u \in U\}$. Communicating the fooling set is equivalent to communicating a valuation from U. Thus, by Proposition 1, the lower bound is $\dim U$ and $\log_2 |U|$ in the continuous and discrete case respectively.

In the continuous case, $\dim U = L$. In the discrete case, $|U|$ is the number of monotone functions with domain $\{1..L\}$ and range $\{0, 1/R, ..., 1\}$, which is exactly the number of ways that R indistinguishable balls (corresponding to the function's jumps) can be partitioned into $L+1$ bins: $\binom{R+L}{R}$. This yields

Corollary 3 *With continuous homogeneous valuations, the dimensionality of the message space in an efficient protocol is at least L. With discrete homogeneous valuations,* *the number of bits communicated by an efficient protocol is at least* $\log_2 \binom{R+L}{R}$.

Just as with unrestricted valuations, this lower bound is essentially tight for $N = 2$, since it is achieved by the protocol in which agent 1 announces his valuation and then agent 1 announces an efficient allocation. The problem of how the required communication grows with N is again open.

The result for the continuous case can be linked to a result obtained by Calsamiglia (1977). In his model, instead of L indivisible goods there exists one unit of an infinitely divisible good. In this case, U is the space of nondecreasing functions $[0, 1] \to [0, 1]$, which is infinitely-dimensional under a reasonable topology. The same argument as that before Corollary 3 establishes that the dimensionality of the message space in an efficient mechanism must be at least $\dim U = \infty$. Therefore, finite-dimensional communication cannot realize efficiency, which is exactly the result obtained by Calsamiglia (1977).[12]

These results should be contrasted with the case where both agents' valuations are known to be concave. In this case, a Walrasian equilibrium with a single price exists, and thus efficiency is achieved with a one-dimensional message space (regardless of whether the objects are divisible or not).

Note also that the communication burden stated in the Corollary for the discrete case with a given input precision R is only proportional to $\log L$. This foreshadows the finding in Section 6 that any given approximation of efficiency can be achieved using communication that is $O(\log L)$.

[12] Calsamiglia (1977) restricts the valuation of agent 1 to be concave and that of agent 2 to be convex. Since the dual of a concave valuation is convex, the analysis goes through without modification. Similarly, the agents' valuations can be restricted to be arbitrarily smooth (even analytical), without changing the argument.

6 Approximation

6.1 The measure of approximation

This section considers the problem of approximating the maximum total surplus, rather than achieving it exactly. The standard measure of approximation used in computer science is the "approximation factor," defined as the inverse of the fraction of the maximum possible value that is guaranteed by the algorithm (in our case, protocol).[13] However, this approximation measure has an important shortcoming in our case when the agents' valuations are continuous: no finite approximation factor can be achieved with finite communication. The simplest illustration of this is in the problem of allocating a single indivisible object between two agents, whose valuations for the object are in $[0, 1]$. Pick some $a > 1$, and consider the restricted problem in which both agents' valuations lie in the set $\{a^{-r}\}_{r=0}^{\infty} \subset [0, 1]$. In this simpler problem, achieving an approximation factor less than a is equivalent to exact efficiency. The fooling set constructed in Lemma 1 for the exact efficiency problem consists of all possible valuations for an agent, hence it is countable. Since any $a > 1$ could be chosen, this implies that no finite approximation factor can be achieved with finite communication. On the other hand, it can be seen that any approximation factor can be realized with countable communication, so the continuous communication measure used in economics is not useful, either.

Intuitively, the problem here is that the desired precision depends on the maximum total surplus available, which in turn depends on the unknown scale of the agents' valuations, and transmission of this scale requires infinite communication.[14] From the economic viewpoint, the problem is not significant, because it is not important to achieve close approximation in states in which the maximum total surplus is very low.

[13] This is a "worst-case" definition, since it requires uniform approximation across all states. The weaker requirement of "average-case" approximation is considered in subsection 6.6 below.

[14] The problem does not arise if only the discrete case is considered, as in computer science. Indeed, if valuations are given with a fixed precision R, the maximum achievable surplus is bounded below by $1/R$ (except in the trivial state where all valuations are identically zero).

To avoid the problem, we consider an approximation requirement bounding the *absolute*, rather than relative, loss of surplus. However, for our results to be comparable with the existing computer science literature, we express the bound as $(1-1/c)\,N$, where $c>1$ is called an "approximation factor." Formally, we consider the realization of the relaxed choice correspondence

$$\begin{aligned} K_c^*(v) &= \left\{ k \in K : \sum_{i\in N} v_k^i \geq S(v) - \left(1-\frac{1}{c}\right) N \right\}, \\ \text{where } S(v) &= \max_{j\in K} \sum_{i\in N} v_j^i. \end{aligned}$$

Note that replacing N (which is the maximum potential surplus across all states) with $S(v) \leq N$ (the maximum surplus in the given state) would yield the relative approximation requirement used in computer science:

$$\sum_{i\in N} v_k^i \geq \frac{1}{c} S(v).$$

Thus, our absolute notion of approximation within factor c is slightly weaker than the computer science notion.

Given our definition of approximation, arbitrary approximation factor $c>1$ can be realized with a finite communication in which agents report their valuations rounded off with a sufficiently fine precision. Therefore, in this section we measure communication with the number of bits transmitted, and examine how it depends on the desired approximation factor c and parameters L, N.

Most of our lower bounds on approximation of the continuous case are obtained by replacing the problem with the simpler problem of approximating the discrete case. In the discrete case with input precision R, any inefficiency loses at least surplus $1/R$, hence realizing the approximation factor $c < \left(1-\frac{1}{NR}\right)^{-1} = 1 + \frac{1}{NR-1}$ is equivalent to realizing exact efficiency. Therefore, the obtained lower bounds on exactly efficient protocols in the discrete case can be used to derive lower bounds on approximating protocols.

We will be able to compare our lower bounds with the approximation factors achieved by several protocols suggested in computer science, which realize the stronger relative approximation requirement in the discrete case. Namely, suppose we have a protocol realizing relative (and therefore absolute) approximation factor c in the discrete case with input precision R, and it is polynomial in $\log R$ (e.g., it involves the full revelation of a certain number of valuations). Then in the continuous case, we can let agents discretize their valuations in multiples of ε and run the protocol on the discretized valuations with $R = \varepsilon^{-1}$, which yields the absolute approximation factor $c + \varepsilon$ using polynomial communication in $\log \varepsilon^{-1}$ (often called "truly polynomial").

For a simple example, consider the auction in which all objects are sold together as a bundle to the agent who announces the highest valuation for it. The social surplus realized by this auction is

$$\max_i u^i(L) \geq \frac{1}{N} \sum_i u^i(L) \geq \frac{1}{N} \max_k \sum_i v_k^i = \frac{1}{N} S(v),$$

hence the bundled auction realizes approximation factor N (relative, and hence absolute). In the discrete case, the auction communicates $N \log_2 (R+1)$ bits. In the continuous case, absolute approximation factor $N + \varepsilon$ is achieved by letting agents reveal their valuations for the bundle rounded off to multiples of ε, requiring the communication of $N \log_2 (\varepsilon^{-1} + 1)$ bits. In this and similar cases, we will say for brevity that the protocol has agents "reveal their valuations" for the bundle and "realize factor N-approximation," which is shorthand for saying that agents "reveal their valuations rounded off to multiples ε" and "realize approximation factor $N + \varepsilon$."

6.2 Unrestricted Valuations

Using Corollary 1 for the case $N = 2$ and $R = 1$, we see that approximation within a factor $c < 1 + \frac{1}{NR-1} = 2$ requires the communication of at least $\binom{L}{L/2}$ bits. This implies

Corollary 4 *With unrestricted valuations (continuous or discrete), approximation within a factor $c < 2$ requires the communication of at least $\binom{L}{L/2}$ bits.*

This result is generalized in Nisan (2001) as follows[15]:

Theorem 1 *(Nisan (2001)) With unrestricted valuations, approximation within a factor* $c < N$ *requires communication that is exponential in* L/c^2*.*

Since factor N-approximation is achieved by the bundled auction, the Theorem implies that for any given N, improvement upon the bundled auction requires exponential communication in L.

This result should be contrasted with the findings of Lehmann et al. (1999) and Holzman et al. (2001), who suggest "simple" protocols improving upon the bundled auction. The improvements are achieved when the number of objects L either stays fixed or goes to infinity along with, and not much faster than, the number N of agents. For example, Holzman et al. (2001) note that auctioning L off in two equal-sized bundles achieves approximation factor $L/2$ for any N, thus improving over the single-bundle auction for $L < 2N$ (splitting L into more bundles achieves further improvement). Lehmann et al. (1999) suggest a polynomial protocol in L that realizes approximation factor $c(L) = \sqrt{L}$, which is better than the bundled auction when $N > \sqrt{L}$.[16] Note that this does not contradict the above theorem, which is vacuous for $c(L) = \sqrt{L}$. Intuitively, the theorem implies that in large problems in which the number N of agents is "substantially smaller" than the number L of items (e.g., smaller than $L^{1/2-\varepsilon}$), simple protocols (i.e., polynomial in L) cannot improve over bundled auctions. When N is either comparable with or larger than L, simple protocols *can* improve over bundled auctions, but both bundled

[15] Nisan (2001) establishes this lower bound by considering the following *set packing problem*: each of the N agents holds a collection of subsets of L, and the aim is to approximate the maximum number of subsets in the union of their collections that can be packed together, i.e., that are pairwise disjoint. This problem is a special case of the combinatorial allocation problem (where the range of valuations is $\{0, 1\}$ and $u^i(S) = 1$ if the subset S is in the collection of agent i).

[16] At each stage of the protocol, each agent i who is not yet allocated any items requests a subset S_i of yet unallocated items that maximizes the ratio $u^i(S_i)/\sqrt{|S_i|}$, along with the ratio itself. The agent who announces the highest ratio receives the requested subset.

auctions and all other simple protocols realize a vanishing share of the available surplus as $N, L \to \infty$.

6.3 Submodular Valuations

Using Corollary 2 for the case $N = 2$ and $R = 2L$, we see that approximation within a factor $c < 1 + \frac{1}{NR-1} = 1 + \frac{1}{4L-1}$requires the communication of at least $\binom{L}{L/2} - 1$ bits. This implies

Corollary 5 *With continuous submodular valuations, approximation within a factor* $c < 1 + \frac{1}{4L-1}$ *requires the communication of at least* $\binom{L}{L/2} - 1$ *bits.*

This result is substantially weaker than that for unrestricted valuations. For example, it does not rule out the possibility that any given approximation factor greater than 1 can be realized with polynomial communication. Nevertheless, Corollary 5 does rule out the possibility of a so-called "Fully Polynomial Approximation Scheme" - i.e., $1 + \varepsilon$ approximation with communication that is polynomial in L, N, ε^{-1}. (Indeed, the corollary implies that realizing $\varepsilon(L) = 1/(4L)$ requires exponential communication in L.) To have an economic example, consider an ascending-bid auction with per-item bids and bid increment ε. The number of bits transmitted by agents in this auction (whether they bid or abstain on a given object at a given price) is at most $NL\varepsilon^{-1}$. Corollary 5 implies that such an auction cannot yield an efficient allocation up to the bid increment ε, for otherwise it would be a FPAS.

The only upper bound known for the submodular case is the 2-approximation of Lehmann et al. (2001), which is achieved by allocating the objects sequentially to the agents with the highest current marginal benefit for them.

6.4 Homogeneous Valuations

Consider the protocol in which all agents reveal their valuations rounded down to a multiple of ε. This revelation can be done by having each agent submit the minimum number of items for which he is willing to pay $r\varepsilon$, for each integer $r \in [0, \varepsilon^{-1}]$. Since each agent sends ε^{-1} numbers from $\{0..L\}$, the total communication involves $N\varepsilon^{-1}\log_2(L+1)$ bits. The protocol then chooses the allocation to maximize the sum of agents' payments, which approximates the maximum surplus within $N\varepsilon$.[17] Thus, we achieve factor $(1+\varepsilon)$ -approximation using a protocol that is polynomial in $\log L, N, \varepsilon^{-1}$ – i.e., a Fully Polynomial Approximation Scheme in parameters $\log L$ and N. In particular, this implies that efficiency can be approximated arbitrarily closely using $O(\log L)$ bits, in contrast to the finding of Corollary 3 that exact efficiency in the continuous case requires the communication of L real variables – an exponentially larger number.

The lesson here is that by insisting on exact rather than approximate efficiency, the economic literature on the dimensionality of message spaces can enormously complicate communication. This also happens in the model of Calsamiglia (1977), which only differs from the one considered here in that the homogeneous good to be allocated is infinitely divisible. Indeed, suppose that the agents' valuation functions over consumption $x \in [0,1]$ satisfy the following condition:

$$\left|u^i(x+\Delta x) - u^i(x)\right| \leq [-\log|\Delta x|]^{-A}, \text{ for some } A > 0.$$

(This is a very mild strengthening of continuity – for example, it is weaker than Lipschitz continuity of any degree $k > 0$, since $|\Delta x|^k (\log|\Delta x|)^A \to 0$ as $|\Delta x| \to 0$.) Then restricting the agents to demand the good in $L = 2^{\varepsilon^{-1/A}}$ identical discrete units can reduce the maximum surplus by at most $N(\log L)^{-A} = N\varepsilon$. Running the protocol described above with agents submitting their discretized demands will approximate the maximum surplus within $2N\varepsilon$. Thus, we realize the approximation factor $1+2\varepsilon$ while communicating only $\varepsilon^{-1}\log_2(L+1) \approx \varepsilon^{-1}N\varepsilon^{-1/A}$ bits – i.e., we have a Fully Polynomial Approximation Scheme, even though exact efficiency here requires infinite-dimensional communication.

[17] This protocol is clearly not incentive-compatible, but it can be made into such by having each agent pay the Vickrey-Groves-Clarke transfer based on the announced bids, rather than paying his own bid.

6.5 The Procurement Problem

The procurement problem is the combinatorial allocation problem in which the "items" are obligations to supply objects, thus the valuations are the negative costs of producing combinations of objects: $u^i(S) \equiv -c^i(S)$. Now it is the cost functions, rather than the valuations, that satisfy "free disposal." The analysis of exact efficiency in the procurement problem is exactly the same as in the "selling" problem. However, the analysis of approximation of cost-minimization in the procurement problem differs from the approximation of value-maximization in the "selling" problem.

To illustrate the difference, consider the setting with two agents whose costs take only 0/1 values. Here *any* finite approximation requires realizing the total cost of zero whenever this is possible. However, by the same argument as in Corollary 1, this requires exponential communication.

Better approximation can be achieved when costs are known to be subadditive, i.e., satisfying $c^i(S \cup T) \leq c^i(S) + c^i(T)$. Nevertheless, even for this case, Nisan (2001) obtains the following lower bound:[18]

Theorem 2 *(Nisan (2001)) For procurement auctions with subadditive costs, the realization of a $c \log L$-approximation for any constant $c < 1/2$ requires communication that is exponential in L.*

Nisan (2001) observes that a matching upper bound follows from classic algorithms of Lov'asz (1975) for approximate set covering using the following iterative procedure: repeatedly procure the set of items S with the minimal average cost until all items are procured.

[18]This is done by considering the following *set covering problem*: each of two agents holds a collection of subsets of L, and the aim is to approximate the minimal number of sets in the combined collection whose union is L. This is a special case of the procurement auction with subadditive costs, if agent i's cost of producing set S is defined to be the minimal number of sets in his own collection that together cover S (or infinity if such covering is impossible).

6.6 Average-case approximation

One may relax the notion of approximation by requiring only that the *expected* surplus be close to optimal. Furthermore, we can count the expected rather than worst-case number of bits transmitted. In the terminology of communication complexity, this concept is called "distributional complexity," since the results clearly depend on the distribution of the agents' valuations.

It turns out that the results of Nisan (2001) also imply lower bounds on distributional complexity:

Proposition 2 *In the combinatorial allocation problem with unrestricted valuations, there exists a probability distribution over states* $\left(v^1, \ldots, v^N\right)$ *such that any protocol realizing fraction* $1/N + \varepsilon$ *of the maximum expected total surplus (for any fixed* $\varepsilon > 0$*) requires communication of an expected number of bits that is exponential in* L/N^2*.*

Proof. The lower bound in Nisan (2001) applies even for the restricted problem of distinguishing between the states with maximum total surplus $S(v) = N$ and maximum total surplus $S(v) = 1$. Furthermore, the bound applies also for randomized protocols with any bounded error. Using the well-known equivalence of randomized complexity and distributional complexity (see Kushilevitz and Nisan (1997), Section 3.4), it follows that for some distribution over states, any protocol that distinguishes states with $S(v) = N$ from those with $S(v) = 1$ correctly with probability at least $1/2 + \varepsilon/4$ requires exponential communication. In particular, we can choose the distribution so that $\Pr\{S(v) \in \{1, N\}\} = 1$. Note in particular that we must have $\Pr\{S(v) = N\} \leq 1/2 + \varepsilon/4$, for otherwise always declaring that $S(v) = N$ would be correct with a probability higher than $1/2 + \varepsilon/4$.

Now take any protocol, and let δ be the probability that the protocol realizes an allocation with surplus exceeding 1 conditional on $S(v) = N$, given the distribution constructed above. Then by declaring that $S(v) = N$ if and only if the surplus at the realized allocation exceeds 1, we err with probability

$$\Pr\{S(v) = N\} \cdot (1 - \delta) \leq (1/2 + \varepsilon/4) \cdot (1 - \delta).$$

On the other hand, if the protocol is subexponential, then the probability of error must be at least $1/2 - \varepsilon/4$, hence we must have

$$1 - \delta \geq \frac{1/2 - \varepsilon/4}{1/2 + \varepsilon/4} \Rightarrow \delta \leq \frac{\varepsilon/2}{1/2 + \varepsilon/4} < \varepsilon.$$

Now consider the conditional distribution on states with $S(v) = N$, and assign probability zero to states with $S(v) = 1$. The expected surplus achieved on this conditional distribution is $1 + \delta(N - 1)$, which is less than fraction $1/N + \varepsilon$ of the total expected surplus, N. ■

In the distribution constructed in the above proposition, the valuations are not necessarily independently distributed. We can obtain a (weaker) lower bound on approximation even for independent valuations using the distributional lower bounds of Babai, Frankl, and Simon (1986):

Proposition 3 *There exist distributions D_1, D_2 on unrestricted valuations such that any protocol for the combinatorial allocation problem with $N = 2$ realizing fraction c of the maximum expected surplus (for some fixed $c < 1$) when the agents' valuations are distributed independently according to D_1, D_2, respectively, requires communication of an expected number of bits that is exponential in L.*

Proof. We will use a reduction to the "disjointness problem" from communication complexity theory (See Kushilevitz and Nisan 1997). In this problem, two agents are each given a subset of a set of size m, and the objective is to decide whether the sets are disjoint. Babai, Frankl, and Simon (1986) prove a lower bound on the distributional complexity of disjointness for product distributions:

Theorem 3 *(Babai, Frankl, and Simon (1986)) There exists a distribution D on subsets of M with $|M| = m$ and a fixed $d > 0$ such if the two agents' sets are drawn according to D, then any protocol that communicates in expectation at most $d\sqrt{m}$ bits must err with at least 1% probability when attempting to solve the disjointness problem.*

We will now show that any protocol for combinatorial allocation that achieves 99.5% expected efficiency when agents' valuations are drawn according to D_1, D_2 (to be defined below), can be used to obtain a protocol for disjointness for $m = \binom{L}{L/2}$ that errs on at most 1% of inputs (drawn according to D). Thus the lower bound of $d\sqrt{m} = d\sqrt{\binom{L}{L/2}}$ (which is exponential in L) communication applies to the combinatorial allocation problem.

Here is the definition of the distributions D_1, D_2 on valuations: Let M be the collection of subsets of L of size exactly $L/2$, hence $|M| = m$. The valuation v is chosen by first choosing a random subset X of M according the distribution D of Babai, Frankl, and Simon (1986). In both D_1 and D_2, we define $v(S) = 0$ for $|S| < L/2$; $v(S) = 1$ for $|S| > L/2$. In D_2 we define for $|S| = L/2$, $v(S) = 1$ if $S \in X$ and $v(S) = 0$ otherwise. In D_2 we define for $|S| = L/2$, $v(S) = 1$ if $N \backslash S \in X$ and $v(S) = 0$ otherwise. Now in order to solve the disjointness problem on X_1 and X_2, the two parties can each create a valuation according to the rule specified above and then solve the combinatorial allocation problem. Finding an allocation with surplus 2 means finding a partition of L into two sets $(S, N \backslash S)$ of size $L/2$ each such that $S \in X_1$ and $S \in X_2$, thus proving X_1 and X_2 are not disjoint. Any inefficient allocation has at most surplus 1. Now, if the allocation protocol loses at most 0.5% of expected total surplus, then the probability that it produces an inefficient allocation in a state with maximum surplus 2 is at most 1%. Thus, if we declare that X_1 and X_2 are disjoint whenever the obtained allocation has value 1, we err with probability of at most 1%. ■

The proof is done for $c = 99.5\%$, which is derived from a constant quoted in Babai, Frankl, and Simon (1986). No optimization of the constant was attempted and it seems likely that a substantial improvement is possible.

7 Deterministic Protocols using Linear Programming

The communication lower bounds obtained in this paper leave room for simple protocols that work well on restricted classes of valuations. In this section we sketch a possibility result along these lines.

It is well known that the allocation problem of combinatorial auctions may be phrased as an integer programming problem (see Vohra and de Vries (2000)). This integer programming problem is commonly relaxed to a linear programming problem, and in some cases it is known that this linear program will indeed return integer allocations, solving the original problem as well. In particular it is known that this is the case if all valuations satisfy the "gross substitutes" property (Kelso and Crawford 1982, Gul and Stacchetti 1999). What we wish to point here is that in all such cases, efficiency may be realized with a protocol that only requires polynomial communication.

The basic correspondence between many auction protocols and primal-dual methods of linear programming was pointed out in Bikhchandani et al. (2001). However, to prove a theorem about communication we will need to use separation-based LP algorithms. A separation-based linear programming algorithm may be used in cases of linear programs that have an exponential number of constraints (inequalities) that are given implicitly. Specifically, such an algorithm does not receive the inequalities as an explicit input, but rather is provided with a "separation oracle" as its input: whenever an infeasible solution is presented to this oracle, it must be able to produce a violated inequality. Algorithms of this type can solve linear programs in polynomial time, given just this type of an oracle. The reader is referred to any textbook on linear programming (e.g., Karloff 1991) for more information.

The allocation problem itself is usually phrased as having an exponential number of variables (x^i_S specifying the allocation of the bundle S of goods to bidder i), but only a polynomial number of significant inequalities (that each item is sold only once, and that each bidder is allocated only one set):

Maximize: $\sum_{i,S} x^i_S u^i(S)$
Subject to:

- For all items j: $\sum_{i,S|j\in S} x^i_S \leq 1$.

- For all bidders i: $\sum_S x^i_S \leq 1$.

- For all i, S: $x_S^i \geq 0$.

In order to use a separation-based linear programming algorithm, we move to the dual that has just a polynomial number of variables:

Minimize: $\sum_j p_j + \sum_i w_i$
Subject to:

- For all i, S: $w_i + \sum_{j \in S} p_j \geq u^i(S)$.
- For all j: $p_j \geq 0$.
- For all i: $w_i \geq 0$.

It is important to note that in the dual, each inequality specifies a condition that depends on the valuation of a single agent.

Now consider the protocol running the separation-based LP algorithm, but instead of making an oracle query, asking all agents to suggest a violated inequality at the current solution, if one exists.[19] If a violated equality exists is suggested, and the protocol continues. The communication that takes place at each such stage is polynomial, and since the algorithm is known to terminate within a polynomial number of steps, the whole protocol requires polynomial communication.

Truthfulness on the part of all the bidders can be ensured by adding a final stage in which all agents announce their final utilities, and each agent receives a payment equal to the sum of everyone else's announcements. The agents will have no incentives to lie in the final stage, since an agent's announcement does not affect his own payment. Furthermore, since each agent's final payoff equals to the total surplus, he will have no incentive to deviate from a surplus-maximizing protocol. The only shortcoming of this mechanism is its large budget deficit, but it can be covered by charging each agent an extra amount

[19] A violated inequality can be described as a subset S of items that would give player i at least utility w_i under the prices p_j. Finding this set may be a computationally difficult problem depending on the representation of individual valuations u^i, but this does not concern us here.

that depends only on the others' messages. In particular, the designer can implement the Vickrey-Groves-Clarke payments by running n extra linear programs, each excluding a single agent, and its value determining the extra charge to the excluded agent. The agents will have no incentive to lie in the extra linear programs, since an agent's message will not affect his own payment. Thus, truthtelling constitutes an ex post Nash equilibrium of the proposed mechanism.[20] While not being very elegant, the mechanism illustrates that for some restricted classes of valuations communication may not be a bottleneck, and that incentive-compatibility can in such cases be ensured without a large increase in communication (the latter point is also made by Reichelstein (1984)).

8 Conclusion

The communication problem examined here is different from the often considered problem of *computing* the optimal allocation once valuations are known. The computational complexity of a problem is defined relative to its input size, but in our cases of interest the input size itself is enormous. In some cases the problem of computing the optimal combinatorial allocation is known to be NP-complete even when its "input size" is small, and so the communication problem is easy. (A simple example is when each agent is known to be interested in a single bundle of items.) Nevertheless, it seems that in practice the communication bottleneck may be more severe than the computational one. First, the NP-completeness results only indicate the asymptotic complexity of the problem as the number of items goes to infinity. In practice, computational complexity can be handled for up to hundreds of items (and thousands of bids) optimally (Vohra and de Vries 2000, Sandholm et al. 2001) and thousands of items (with tens of thousand of bids)

[20]In general it will not be a dominant-strategy equilibrium. Indeed, if agent j believes that agent i will send the message corresponding to type $v^{i\prime}$ after one of j's announcements and to type $v^{i\prime\prime}$ after another, j's incentives to be truthful may be destroyed. Thus, dominant-strategy incentive-compatibility is destroyed when agents send their messages sequentially rather than simultaneously, but it is such sequentiality that is needed to reduce deterministic communication complexity.

near-optimally (Zurel and Nisan 2001). In contrast, our lower bounds on communication are exact, and they "kick in" already with a few dozen items as long as valuations are general. Second, unlike the communication burden, the computational burden may be sidestepped by transferring it to the bidders themselves, e.g., by asking the bidders to suggest allocations or matches to their package (Banks et al. 1989, Nisan and Ronen 2000).

Through most of the paper we have also assumed that the agents follow the strategies suggested by the designer, rather than behaving in their self-interest. We found that the communication requirement *by itself* often constitutes a "bottleneck" preventing exact or approximate efficiency. If agents behave in their self-interest, this imposes further incentive-compatibility constraints that the designer must honor. In the deterministic communication model, these constraints require that the agents' strategies constitute an equilibrium of the extensive-form message game. (Incentive-compatibility cannot be directly imposed on nondeterministic communication, since it does not specify what an agent would obtain by "rejecting" the oracle's message.) Given an equilibrium concept such as Bayesian Nash or ex post Nash, a version of the revelation principle still holds – that is, incentive-compatibility constraints can be imposed directly on the choice rule. As shown in Reichelstein (1984) and in the previous subsection, in cases in which the communication problem is easy, the imposition of incentive constraints does not require a drastic increase in communication. However, it would be interesting to explore the interaction of communication and incentive constraints in a more general setting.

References

[1] Ausubel, L. (2000). "An efficient dynamic auction for heterogeneous commodities," mimeo, University of Maryland.

[2] Ausubel, L., and P. Milgrom (2001). "Ascending Auctions with Package Bidding." Available from http://www.ausubel.com/auction-papers/package-bidding.pdf

[3] Babai, L., P. Frankl, and J. Simon. "Complexity Classes in Communication Complexity Theory," *Proceedings of FOCS*, 337–347, 1986.

[4] Banks, J., J. Ledyard, and D. Porter (1989). "Allocating uncertain and unresponsive resources: An experimental approach," *Rand Journal of Economics*, 20:1 – 25.

[5] Bikhchandani, S., and J.W. Mamer (1997). "Competitive equilibrium in an exchange economy with indivisibilities," *Journal of Economic Theory*, 74:385 – 413.

[6] Bikhchandani, S., and J. Ostroy (2001). "The Package Assignment Model," Available from http://www.anderson.ucla.edu/faculty/sushil.bikhchandani/bikh_wp.htm

[7] Bikhchandani, S., R. Vohra, S. de Vries, and J. Schummer (2001). "Linear programming and Vickrey auctions," mimeo, 2001. Availailabe from http://www.kellogg.nwu.edu/faculty/vohra/htm/res.htm.

[8] Calsamiglia, X. (1977). "Decentralized Resource Allocation and Increasing Returns," *Journal of Economic Theory* 14: 262-283.

[9] DeMartini, C., A. Kwasnica, J. Ledyard, and D. Porter (1998). "A new and improved design for multi-object iterative auctions," Available from http://www.hss.caltech.edu/~jledyard/rad.pdf

[10] Gul, F., and E. Stacchetti (1999). "Walrasian equilibrium with gross substitutes," *Journal of Economic Theory*, 87:95 – 124.

[11] Gul, F., and E. Stacchetti (2000). "The English Auction with Differentiated Commodities," *Journal of Economic Theory.*

[12] Hayek, F. (1945). "The Use of Knowledge in Society," *American Economic Review* 35: 519-30.

[13] Holzman, R., N. Kfir-Dahav, D. Monderer, and M. Tennenholtz (2001), "On Bundling Equilibrium in Combinatorial Auctions" mimeo, Technion University, http://iew3.technion.ac.il/~moshet/rndm11.ps

[14] Hurwicz, L. (1960), "Optimality and Informational Efficiency in Resource Allocation Processes," in K.J. Arrow, S. Karlin, and P. Suppes, eds., *Mathematical Methods in the Social Sciences*, 27-46, Stanford: Stanford University Press.

[15] Hurwicz, L. (1977). "On the Dimensional Requirements of Informationally Decentralized Pareto-Satisfactory Processes," in K.J. Arrow and L. Hurwicz, eds., *Studies in Resource Allocation Processes*, 413-424, New York: Cambridge University Press.

[16] Karloff, H. (1991). *Linear Programming.* Birkhäuser Verlag.

[17] Kelso, A.S. Jr., and V.P. Crawford (1982). "Job matching, coalition formation, and gross substitutes." *Econometrica*, 50:1483-1504.

[18] Kushilevitz, E., and N. Nisan (1997). *Communication Complexity.* Cambridge University Press.

[19] Lang, S. (1987). *Linear Algebra*, New-York: Springer-Verlag, 3rd edition.

[20] Lehmann, B., D. Lehmann, and N. Nisan (2001). "Combinatorial auctions with decreasing marginal utilities." In *ACM conference on electronic commerce.*

[21] Lehmann, D., L. Ita O'Callaghan, and Y. Shoham (1999). "Truth revelation in rapid, approximately efficient combinatorial auctions." In *1st ACM conference on electronic commerce*, 1999.

[22] Lov'asz, L. (1975) "The ratio of optimal integral and fractional covers." *Discrete Mathematics*, 13.

[23] Marschak, T. (1996) "On Economies of Scope in Communication," *Economic Design* 2(1): 1-30.

[24] Milgrom, P. (2000). "Putting Auction Theory to Work: The Simultaneous Ascending Auction," *Journal of Political Economy*.

[25] Mount, K., and S. Reiter (1974). "The Information Size of Message Spaces," *Journal of Economic Theory* 28:1-18.

[26] Nisan, N. (2000). "Bidding and allocation in combinatorial auctions." In *ACM Conference on Electronic Commerce*.

[27] Nisan, N. (2001). "The communication complexity of approximate set packing and covering." Preliminary version available from http://www.cs.huji.ac.il/˜noam/mkts.html.

[28] Nisan, N., and A. Ronen (2000). Computationally feasible vcg-based mechanisms. In *ACM Conference on Electronic Commerce*.

[29] Parkes, D.C. (1999). "ibundle: An efficient ascending price bundle auction," mimeo, Harvard University.

[30] Parkes, D.C., and Lyle H. Ungar (2000). Iterative combinatorial auctions: Theory and practice. In *AAAI/IAAI*, pages 74–81.

[31] Reichelstein, S. (1984). "Incentive Compatibility and Informational Requirements," *Journal of Economic Theory* 34(1), 32-51.

[32] Rothkhof, M.H., A. Pekeč, and R.M. Harstad (1998). "Computationally manageable combinatorial auctions." *Management Science*, 44(8):1131–1147.

[33] Sato, F. (1981). "On the Informational Size of Message Spaces for Resource Allocation Processes in Economies with Public Goods," *Journal of Economic Theory* 24: 48-69.

[34] Sandholm, T., S. Suri, A. Gilpin, and D. Levine "Cabob: A fast optimal algorithm for combinatorial auctions." In *IJCAI*, 2001.

[35] Vohra, R., and S. de Vries (2000). "Combinatorial auctions: A survey," Availailabe from http://www.kellogg.nwu.edu/faculty/vohra/htm/res.htm.

[36] Walker, M. (1977). "On the Informational Size of Message Spaces," *Journal of Economic Theory* 15, 366-375.

[37] Yao, A. C.-C. (1979), "Some complexity questions related to distributive computing," In *ACM Symposium on Theory of Computing*, pages 209–213.

[38] Zurel, E., and N. Nisan (2001). "An efficient approximate allocation algorithm for combinatorial auctions," *ACM conference on electronic commerce.*

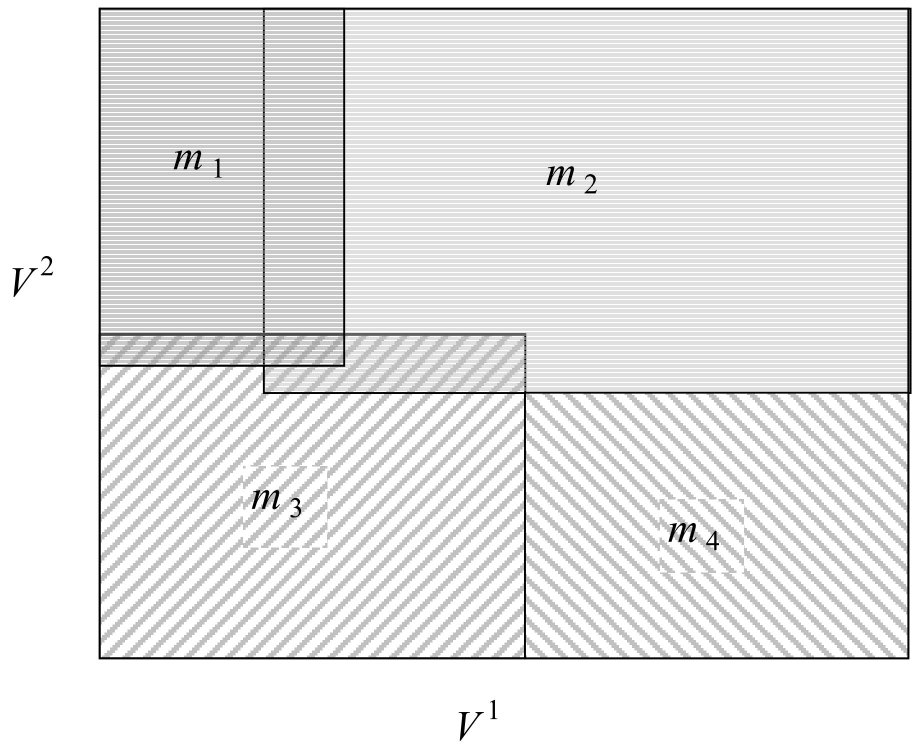

Figure 1: Message Correspondence

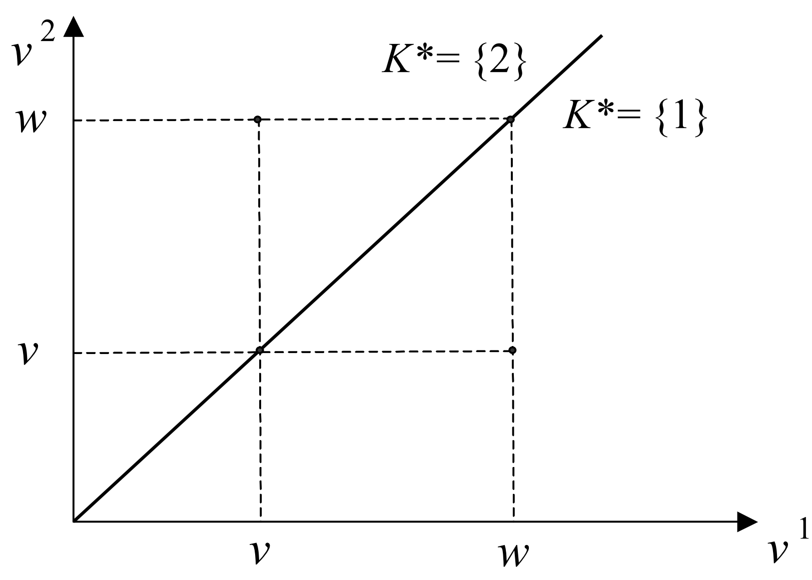

Figure 2: The Fooling Set

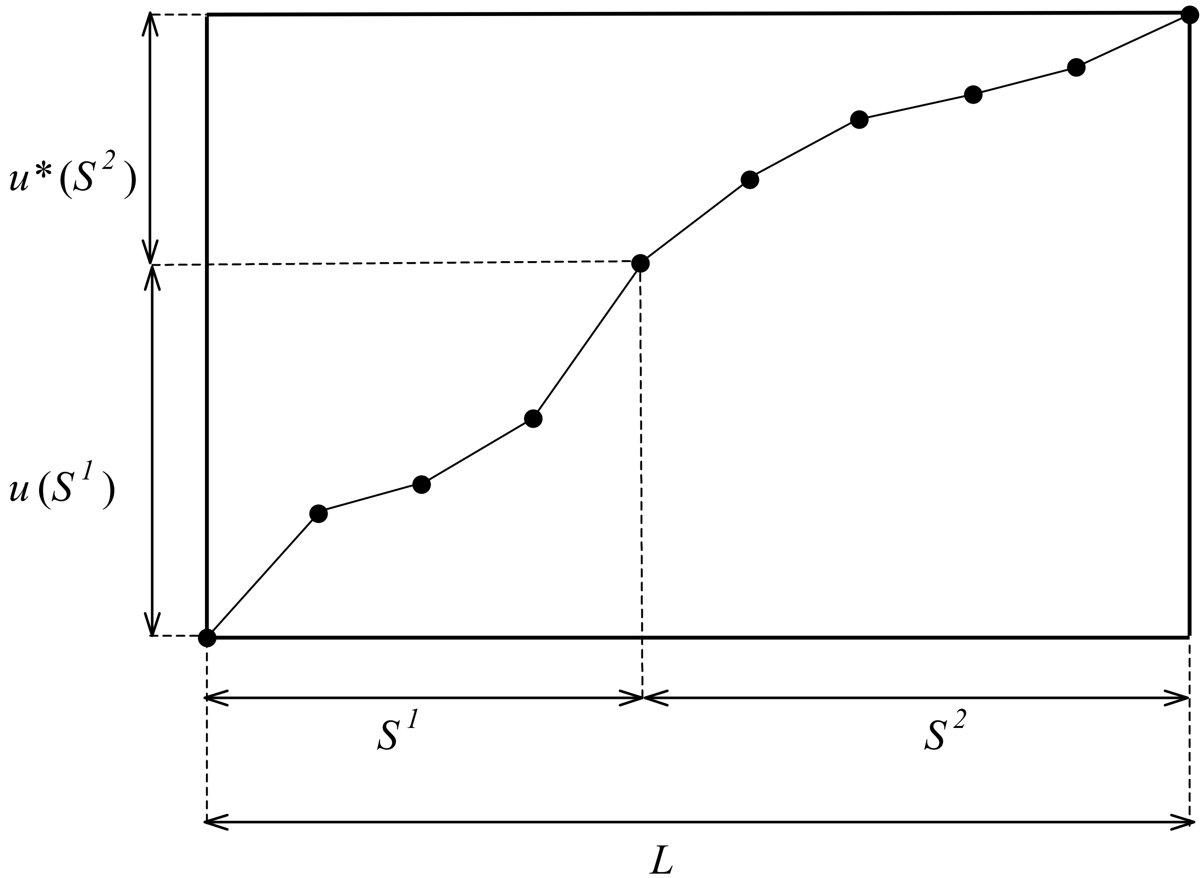

Figure 3: Dual Valuations