

Mechanism Design for Complexity-Constrained Bidders

Ravi Kumar¹, Mohammad Mahdian¹, and Amin Sayedi^{2*}

¹ Yahoo! Research, 701 Sunnyvale, CA 94089.

{ravikumar, mahdian}@yahoo-inc.com

² Carnegie Mellon University, Pittsburgh, PA 15213.

ssayedir@cmu.edu

Abstract. A well-known result due to Vickery gives a mechanism for selling a number of goods to interested buyers in a way that achieves the maximum social welfare. In practice, a problem with this mechanism is that it requires the buyers to specify a large number of values. In this paper we study the problem of designing optimal mechanisms subject to constraints on the complexity of the bidding language in a setting where buyers have additive valuations for a large set of goods. This setting is motivated by sponsored search auctions, where the valuations of the advertisers are more or less additive, and the number of keywords that are up for sale is huge. We give a complete solution for this problem when the valuations of the buyers are drawn from simple classes of prior distributions. For a more realistic class of priors, we show that a mechanism akin to the broad match mechanism currently in use provides a reasonable bicriteria approximation.

1 Introduction

Consider the following setting: there are m buyers who are interested in buying n goods, each with a unit supply. Each buyer has a value for each good, and her valuation for a bundle of goods is simply the sum of her valuations for each good in the bundle.

This is perhaps the simplest model for selling multiple non-identical goods, as the buyers' valuations are assumed to be additive and not combinatorial. From a mechanism design perspective, designing optimal (i.e., social welfare maximizing) auctions for this setting is trivial: simply run an independent second-price auction for each good. Each buyer will have the incentive to bid her true value for each good, and each good will be allocated to the buyer who has the maximum value for it.

The problem with this simple mechanism is that each bidder has to provide n values, one for each type of good, as her bid. This is especially problematic in applications such as sponsored search auctions, where the number of different types of goods is quite large or possibly infinite.³ This motivates the following problem, which is the main subject of

* Part of this work was done while the author was at Yahoo! Research.

³ The valuations of bidders (i.e., advertisers) in sponsored search auctions are quite close to being additive. The only non-additive valuations that sponsored search systems allow bidders to specify are those involving budget constraints. However, binding budget constraints seem to be rare. Furthermore, since in sponsored search auctions advertisers can bid on any user query, the number of goods that are available for sale is essentially infinite.

this paper: *what is the maximum social welfare that can be achieved with a mechanism that is restricted to ask each bidder for only a small amount of information?*

To make this more precise, we denote the value of a bidder i for a good j by v_{ij} , and assume a prior on the n -tuple $(v_{ij})_j$ of values, i.e., we assume these values are picked from a given joint distribution. This prior is supposed to capture the information on how the valuation of a bidder on different goods are correlated. The algorithm is allowed to query each bidder i for a fixed number k of v_{ij} 's, and based on the responses it gets, it allocates each good to one bidder. We evaluate the algorithm based on the social welfare it achieves, i.e., the sum of the valuations of the bidders for the goods they receive. Our objective is to design an allocation algorithm that achieves the maximum expected social welfare (where the expectation is over the draw of the valuations from the prior) among all algorithms that ask each bidder for at most k of her values.

Designing the optimal algorithm in the above model is often a complicated task, and depends on the type of prior we assume on the valuations. The main results of this paper include exact and approximate solutions of this problem for a few simple yet important classes of distributions, and a proof that for a more realistic class of distributions, clustering-based bidding languages — akin to the concept of *broad match* or *advanced match* that is currently in use — provide a reasonable bicriteria approximation for the optimal algorithm. Although the major technical contribution of the paper is on the problem of designing the optimal or approximately optimal allocation algorithms, we will also observe that these algorithms, combined with suitable payment schemes, result in mechanisms with good incentive properties.

Related work. Ronen [4] and Ronen and Saberi [5] studied the design of revenue-optimal mechanisms when the mechanism designer has communication complexity constraints in accessing the distribution of bidders' values. Mechanism design for single-item auctions under a constraint on the number of bits that each bidder can send to the auctioneer was studied by Blumrosen, Nisan, and Segal [1]. Earlier, Nisan and Segal [3] studied the communication complexity of maximizing social welfare in combinatorial auctions. The main difference between this line of work and ours is that in their models the constraint is on the *communication complexity* of the bidders, whereas our model focuses on *query complexity*, if each bidder is thought of as an oracle that can be queried for their value for each good. This is because in our motivating application (sponsored search auctions), the costly operation for an advertiser is to compute their value for a keyword, and not transmitting information to the auctioneer.

The setting of additive utilities for a large number of goods (motivated by sponsored search auctions) was studied by Mahdian and Wang [2]. Their focus is on a specific class of bidding languages called *clustering-based* bidding languages that are similar to the broad match scheme used in sponsored search systems. We address a more basic question: finding the most socially efficient algorithm without imposing any constraint other than the complexity constraint on the bidding language.

2 Model

Assume there are m buyers numbered 1 through m , and n different goods $1, \dots, n$ that are offered for sale ($n \gg m$). Without loss of generality, we can normalize the supply

of each good to one. Buyer i has a non-negative real value v_{ij} for good j . We assume that the valuations of buyers are additive, i.e., the value that buyer i has for a set S of goods is simply $\sum_{j \in S} v_{ij}$.

We assume a prior on the values v_{ij} . For simplicity, we assume that the valuation of different bidders are independently and identically distributed, i.e., there is a joint distribution \mathcal{D} over the set of all n -tuples of non-negative numbers, and for each buyer i , the tuple of valuations $(v_{ij})_j$ is drawn independently from \mathcal{D} .

Given the distribution \mathcal{D} and an integer k , the problem is to design an algorithm that asks each buyer for at most k of their values (i.e., asks buyer i for v_{ij} for at most k different values of j to be chosen by the mechanism), and among all such algorithms achieves the maximal social welfare. This problem can be studied both in an adaptive and in a non-adaptive framework. We focus on the *non-adaptive* variant, i.e., the algorithm asks all the questions at once and then receives the answers. The adaptive version (where the algorithm asks the questions one by one and can use the earlier answers to decide what question to ask next), although theoretically intriguing, is of less practical value in sponsored search systems.⁴

Perhaps the simplest possible case is when the distribution \mathcal{D} is a product distribution, i.e., all v_{ij} 's are drawn independently. Even in this case, finding the optimal algorithm is non-trivial. However, we show (Section 3) that a mechanism based on *spreading* the questions among different goods is optimal. A more interesting class of distributions correspond to the case where each buyer i has a one-dimensional *type* t_i drawn from a distribution \mathcal{D}^* , and the values of buyer i on various goods are independent *conditioned on the type* t_i (Section 4). Finally, we study a model where the set of goods are partitioned into a number of clusters, and the valuation of each buyer on each cluster followed the model of buyers with types described above, while the values are independent for distinct clusters (Section 5).

Incentive properties. Most of this paper is devoted to the problem of designing the optimal allocation algorithm subject to the number of queries this algorithm can make. Ideally, we would like to match such an algorithm with a suitable payment scheme to turn it into an incentive-compatible mechanism, i.e., a mechanism where bidders are better off answering the questions truthfully. However, in our setting, we need to be careful about the notion of incentive compatibility, for the following reason. If we assume that the bidder knows her value for all goods, then for any mechanism that infers something from the values of queried goods about the unknown values, there is some chance that the queried goods under- or over-represent the values of other goods. In such cases the bidder might have an incentive to bid untruthfully to “correct” the mistake of the allocation algorithm. This intuitive argument can be made precise to show that with the strict notion of incentive compatibility in dominant strategies, essentially no non-trivial learning can be done.

⁴ This is mainly because the algorithm cannot interleave the questions asked from two different advertisers in any way it wants, due to timing constraints. However, a hybrid between the adaptive and non-adaptive models, where the algorithm can use the answers provided by an advertiser to decide its next questions of the same (but not other) advertisers might be feasible in practice. We will comment on this in Section 6.

However, it is possible (details omitted) to get around this problem by weakening the incentive requirement to one of the following:

(1) If the bidder does not know her value for goods about which she is not queried, then her best strategy is to answer the questions truthfully; this can be thought of as an *ex ante* notion of incentive compatibility.

(2) As the number of questions that can be asked of each bidder grows, the incentive to deviate from truthfulness quickly approaches zero.

3 Independent Valuations

In this section we consider the case where the prior distribution \mathcal{D} is a product distribution, i.e., each value v_{ij} is picked independently from a distribution. This is a simple case since there is no *learning* involved: the answers that the mechanism receives on one good cannot help with the allocation of other goods. However, the problem is still non-trivial as it involves optimally distributing the queries among different goods.

For simplicity of exposition, we further assume that all v_{ij} 's are independently and *identically* distributed according to a distribution with cdf F and pdf f . The assumption that the distributions for different goods are identical is not necessary, but will simplify the statement of our results. The following lemma gives the optimal allocation of the goods, given the questions that the mechanism asks and their answers.

Lemma 1. *Suppose that values of bidders for a good come from a distribution with cdf $F(\cdot)$ and pdf $f(\cdot)$ and expectation μ . If the auctioneer knows the values of $i < m$ bidders in a set S for this good, then the expected welfare $W(i)$ of allocating this good is maximized when it is given to the bidder in S with the maximum known value v if $v \geq \mu$ or to an arbitrary bidder not in S if $v < \mu$. Furthermore, $W(i) = M - \int_{\mu}^M F(x)^i dx$, where M is the upper bound for the valuation (i.e., $F(M) = 1$).*

Using this, we show that the expected welfare that the mechanism gets from a good is a concave function of the number of queries it makes about the value of that good.

Lemma 2. *The function $W(i) = M - \int_{\mu}^M F(x)^i dx$ is a concave function of i , i.e., $W(i) - W(i+1) \geq W(i+1) - W(i+2)$ for every i .*

Theorem 1. *If all values v_{ij} are drawn iid from a distribution F with expectation μ , then any mechanism of the following form is optimal: ask each bidder for their value for k goods in such a way that each good is asked about $\lfloor km/n \rfloor$ or $\lceil km/n \rceil$ times, and then allocate each good as prescribed by Lemma 1.*

If getting an approximation to the optimal social welfare is enough, we show that when $km \leq n$, a simple mechanism that allocates all the goods arbitrarily without asking any questions about the values extracts at least half of the social welfare of the optimal mechanism. In other words, when n is large, a mechanism with limited knowledge about values cannot do much better than a random mechanism.

Lemma 3. *If $n \geq mk$, and all the values v_{ij} are independent, a mechanism that allocates the good randomly to the bidders extracts welfare at least $\text{OPT}/2$, where OPT is the welfare of optimal mechanism subject to the same constraints. Furthermore, the bound is tight.*

The assumption $n \geq km$ in this lemma (also in Theorem 3) is indeed necessary.

4 Bidders with Types

In reality, buyers' valuations for different goods are not independent. For example, if a buyer has high valuation for a good, she is more likely to have high valuation for other related goods as well. To capture this, we consider the following class of distributions: each buyer i has a one-dimensional *type* t_i drawn from a distribution \mathcal{D}^* over non-negative numbers. Conditioned on the type t_i , the values of this buyer for the goods are independently and identically distributed according to some distribution $\mathcal{D}(t_i)$. This is a good model for settings where all the goods are related. In this case, the buyer i 's *base value* for the goods is determined by the type t_i , and her value for an individual good depends on her base value as well as another factor that is independently distributed. E.g., the value of the buyer for a good can be equal to her base value plus an iid noise. Another important example in the context of sponsored search is when the base value of the buyer (the advertiser) is her *value per conversion*, and her value for an individual good (keyword) is her value per conversion times a keyword-specific *conversion rate*.⁵

As in Section 3, we give an exact optimal allocation algorithm and a 2-approximation, in the case that the number of goods n is larger than mk . We start with a lemma that gives an upper bound on the social welfare of the optimal algorithm. Before stating the lemma, we need to define a few random variables $a_{i,j}$'s, μ_i 's, $\hat{\mu}_i$'s, and $\hat{\mu}$:

Definition 1. *Consider the following random experiment: for every $i = 1, \dots, m$, generate a random number t_i according to \mathcal{D}^* , and then for every $j = 1, \dots, k$, generate $a_{i,j}$ iid according to $\mathcal{D}(t_i)$. Let the random variable μ_i denote the expected value of a random variable distributed according to $\mathcal{D}(t)$ where t is drawn from \mathcal{D}^* , conditioned on k samples of this distribution (generated with the same t) being $a_{i,1}, \dots, a_{i,k}$. Finally, let $\hat{\mu} = \max_{j=1, \dots, m} \{\mu_j\}$ and $\hat{\mu}_i = \max_{j \neq i} \{\mu_j\}$. Note that the random variables μ_i 's, $\hat{\mu}_i$'s, and $\hat{\mu}$ are functions of the random variables $a_{i,j}$'s.*

Lemma 4. *Assume $n \geq mk$. Then the social welfare of any algorithm that asks at most k questions from each buyer is at most $\mathbb{E}[\sum_{i=1}^m \sum_{j=1}^k \max(a_{i,j}, \hat{\mu}_i) + (n - km)\hat{\mu}]$.*

Theorem 2. *Assume $n \geq mk$. The following algorithm is optimal: query each buyer for the values of k goods in such a way that no good is queried more than once, compute the values of μ_i 's as in Definition 1, and allocate each good j either to the bidder who queried about j , or to the one who has the maximum value of μ_i , whichever is larger.*

⁵ One might argue that the conversion rate of an advertiser for different keywords are not independent. This is in fact true, however, in sponsored search systems it is common to assume that conversion rates are separable, i.e., the conversion rate of an advertiser for a keyword is the product of an advertiser-specific conversion rate and a keyword-specific one. Such a system can be captured by our model by letting the type t_i be the value per conversion times the advertiser-specific conversion rate.

Since the algorithm given in the above theorem might be too complicated for implementation, or impractical as it treats different buyers asymmetrically, we also present a simple and natural algorithm that is a 2-approximation to the optimal welfare.

Theorem 3. *If $n \geq k(m + 1)$, the algorithm \mathcal{A} that asks all the buyers about their value for the first k goods, allocates each of these goods to the buyer with the highest value, and allocates all other goods to the buyer with the highest value of μ_i (as in Definition 1) gets a welfare that is at least half of the welfare of the optimal algorithm in expectation. Moreover, the bound is tight.*

Proof. Consider an optimal algorithm OPT. It is clear that the welfare that \mathcal{A} receives from the first k goods is at least as large as the welfare that OPT gets on those goods. For the remaining goods, it is easy to adapt the proof of Lemma 4 to show that the welfare that OPT gets from goods $k + 1, \dots, n$ is at most $E[\sum_{i=1}^m \sum_{j=1}^k \max(a_{i,j}, \hat{\mu}_i) + (n - k - km)\hat{\mu}]$, where $a_{i,j}$'s, μ_i 's, $\hat{\mu}_i$'s, and $\hat{\mu}$ are as in Definition 1. This is at most

$$E[\sum_{i=1}^m \sum_{j=1}^k (a_{i,j} + \hat{\mu}) + (n - k - km)\hat{\mu}]. \quad (1)$$

We claim that $E[a_{i,j}] = E[\mu_i]$. This can be proved using the following experiment: draw t from \mathcal{D}^* and then $k + 1$ numbers $a_{i,1}, \dots, a_{i,k+1}$ from $\mathcal{D}(t)$. Clearly, we have $E[a_{i,j}] = E[a_{i,k+1}]$. On the other hand, fixing the values of $a_{i,1}, \dots, a_{i,k}$, the value of $a_{i,k+1}$ has a distribution with mean μ_i . Therefore, taking the expectation over values of $a_{i,1}, \dots, a_{i,k}$, we have $E[a_{i,k+1}] = E[\mu_i]$. Therefore, $E[a_{i,j}] = E[\mu_i] \leq E[\hat{\mu}]$.

Using this inequality, the expression in (1) is at most $(n - k - km + 2km)E[\hat{\mu}] \leq 2(n - k)E[\hat{\mu}]$, where the latter inequality follows from the assumption $n \geq k(m + 1)$.

On the other hand, we analyze the expected welfare that algorithm \mathcal{A} receives from goods $k + 1, \dots, n$ as follows: the answers that \mathcal{A} gets on queries that it makes on the first k goods are distributed as the $a_{i,j}$'s of Definition 1. Therefore, fixing the values of $a_{i,j}$'s, the expected welfare that \mathcal{A} gets on any of the goods $k + 1, \dots, n$ is precisely $\hat{\mu}$. Thus, the total expected welfare of \mathcal{A} on these goods is $(n - k)E[\hat{\mu}]$.

To sum things up, the total welfare of \mathcal{A} on the first k goods is at least that of OPT on these goods, and its total welfare on the other $n - k$ goods is at least half that of OPT. Therefore, \mathcal{A} is a 2-approximation to OPT.

For the tightness, suppose there is only one type. Value of a bidder for a good is 0 with probability $1 - \epsilon$ and is 1 with probability ϵ . Also, let $n = k(m + 1)$. Since all bidders are of the same type, \mathcal{A} allocates all goods $k + 1, \dots, n$ to an arbitrary bidder for expected welfare of $(n - k)\epsilon = m k \epsilon$. Therefore, the total welfare of this algorithm is at most $k + m k \epsilon$. On the other hand, OPT asks one question about each of the first $m k$ goods, and gets an expected welfare of $\epsilon + (1 - \epsilon)\epsilon$ on each such good. Therefore, the total welfare of OPT is at least $2m k \epsilon - m k \epsilon^2$. Taking $\epsilon = 1/\sqrt{m}$ and letting m grow, the welfare of \mathcal{A} will tend to half that of OPT. \square

The proof of the above theorem also implies the following, which will be useful in Section 5.

Lemma 5. *Assume $n \geq km$. An algorithm that knows the types of all buyers and allocates all goods to the buyer that has the maximum expected value (given the type) without asking any question gets a welfare of at least half of the optimal algorithm.*

Non-uniform supplies. We can extend these results to a more general model in which every good j has some supply s_j . In other words, value of buyer i for good j is $s_j v_{ij}$. By slightly modifying algorithm \mathcal{A} in Theorem 3 and replacing the condition $n \geq (m+1)k$ by $\sum_{j=1}^n s_j \geq (m+1) \sum_{j=1}^k s_j$, we can show that algorithm \mathcal{A} can be adopted for this more general model. More formally, assume without loss of generality that $s_1 \geq \dots \geq s_n$. If algorithm \mathcal{A} in Theorem 3 asks all the buyers about the k goods with largest amount of supply, namely s_1, \dots, s_k , it gets a welfare that is at least half of the welfare of the optimal algorithm in expectation.

5 A Cluster Model

The model studied in Section 4 captures situations where all goods are related, and therefore buyers' valuations for these goods are correlated by their *types*. Our final model captures the more realistic setting where goods can be partitioned into a number c of *clusters*, with each cluster containing a number of related goods. Different clusters are unrelated, and each buyer has a type for each cluster. The valuations of a buyer for the goods within a cluster follows the model from Section 4 with the buyer's type for the cluster. This is a good model for applications such as sponsored search, where the goods (keywords) can be partitioned based on their *topic*, with keywords within a topic being related and keywords from different topics being essentially independent.

Formally, there are c disjoint clusters, with the j th cluster containing n_j goods. Buyer i has a type t_{ij} for cluster j , drawn iid according to a distribution \mathcal{D}^* .⁶ Given these types, the value of this buyer for each good in cluster j is picked iid according to a distribution $\mathcal{D}(t_{ij})$. We denote the expected value of this distribution by μ_{ij} . Note that μ_{ij} is a function of t_{ij} and is therefore a random variable.

The problem of designing the optimal algorithm boils down to deciding how the queries of each buyer should be allocated across different clusters. In one extreme, we might want to ask many queries about the same cluster to get a better estimate of the values of the buyer for that cluster, and forgo other clusters. In the other extreme, we might want to spread the queries evenly across different clusters, to get a rough estimate of the values of the buyer for all clusters. Interestingly, there are cases where each of these strategies outperforms the other by an arbitrary factor. This suggests that finding the optimal algorithm in the cluster model is a difficult problem, as the optimal allocation of the queries can be highly dependent on the nature of underlying distribution.

On the positive side, we can show that if we are allowed to ask more queries than OPT, enough to ask a logarithmic number of queries for each cluster, then we can get a good estimate of μ_{ij} 's for each cluster and therefore the machinery from Section 4 gives us a good approximation to the optimal welfare. This result (Theorem 4) is essentially a bicriteria approximation of the optimal algorithm. To prove this result, we start with the following lemma (proof omitted), which shows that an algorithm that knows all μ_{ij} 's can guarantee welfare of at least half of the optimal algorithm, if all clusters are large.

⁶ It is not hard to see that our result holds in the more general case where the distribution of types for different clusters are different (but still independent).

Lemma 6. *Suppose we are given all expected values μ_{ij} , and for every j , $n_j \geq mk$. An algorithm \mathcal{A} that allocates all goods of each cluster to the buyer with highest expected value gets welfare of at least $\text{OPT}/2$ where OPT is the optimal expected welfare.*

Let L denote the maximum, over the choice of the type t , of the ratio between the maximum value from the distribution $\mathcal{D}(t)$ and the expected value of $\mathcal{D}(t)$; this parameter captures how difficult it is to estimate the mean of the distribution by sampling.

Theorem 4. *Assume $n_j \geq mk, \forall j$. There is an algorithm \mathcal{A} that asks $O(cL^2\epsilon^{-2} \log m)$ queries, and achieves at least a $\frac{0.9}{2(1+\epsilon)^2}$ fraction of the welfare of the optimal algorithm.*

6 Conclusions

We studied the problem of designing the optimal allocation algorithms for an auction setting where buyers have additive values, but the number of different types of goods is large and we are limited in the number of queries we can make to each buyer. We believe this is a promising direction for research, as in many realistic situations, we are faced with valuations that lie in a high-dimensional space, even though the distribution that these valuations come from can allow for “learning” the valuation given only a small-dimensional sample. The problem we studied is essentially about the interaction between this learning aspect, and the optimization aspect of queries. In other words, in our optimization, we must take into account not only the additional information that a query gives us, but also the amount of additional welfare it can lead to.

There are many open problems left for future research. For example, our results for the model of typed buyers (Section 4) cannot handle the case with non-identical distributions, except for the case of non-uniform supplies. Extending the results to such a setting seems difficult, as it requires ways to capture the information contents of each distribution. Also, it would be nice to get rid of the $n \geq mk$ assumption in our results. In particular, in Section 4, it might be possible to prove that the algorithm that spreads the questions almost evenly across the goods is optimal. Finally, it would be interesting to solve the adaptive variant of the problem, or at least the more practical “hybrid” between adaptive and non-adaptive variants where the questions the mechanism asks can depend on the previous answers of the same buyer but not the other buyers.

References

1. L. Blumrosen, N. Nisan, and I. Segal. Auctions with severely bounded communication. *JAIR*, 28:233–266, 2007.
2. M. Mahdian and G. Wang. Clustering-based bidding languages for sponsored search. In *Proc. 17th ESA*, 2009.
3. N. Nisan and I. Segal. The communication requirements of efficient allocations and supporting prices. *JET*, 129(1):192–224, 2006.
4. A. Ronen. On approximating optimal auctions. In *Proceedings of the 3rd ACM conference on Electronic Commerce (EC’01)*, pages 11–17, 2001.
5. A. Ronen and A. Saberi. On the hardness of optimal auctions. In *Proceedings of Annual IEEE Symposium on Foundations of Computer Science (FOCS’02)*, page 396, 2002.