

# Non Parametric Learnability of Income-Lipschitz Demand Functions

[Extended Abstract]

Eyal Beigman  
CMS-EMS  
Kellogg School of Management  
Northwestern University  
Evanston IL 60208  
e-beigman@northwestern.edu

Rakesh Vohra  
MEDS  
Kellogg School of Management  
Northwestern University  
Evanston IL 60208  
r-vohra@northwestern.edu

## ABSTRACT

A sequence of prices and demands are *rationalizable* if there exists a concave, continuous and monotone utility function such that the demands are the maximizers of the utility function over the budget set corresponding to the price. Afriat [1] presented necessary and sufficient conditions for a finite sequence to be rationalizable. Varian [30] and later Blundell et al. [5, 6] continued this line of work studying nonparametric methods to forecasts demand. Their methods do not implement any probabilistic model and therefore fall short of giving a general degree of confidence in the forecast. The present paper complements this line of research by introducing a statistical model and a measure of complexity through which we are able to study the learnability of classes of demand functions and derive a degree of confidence in the forecasts.

In this paper we develop a framework to study the learnability of real vector valued demand functions through observations on prices and demand. Our results give lower and upper bounds on the sample complexity of PAC learnability and show that the sample complexity of learning a class of vector valued functions with finite fat shattering dimension increases by a linear factor of the dimension. We show that classes of income-Lipschitz demand functions with global bounds on the Lipschitz constant have finite fat shattering dimension.

## 1. INTRODUCTION

There are two traditions that run through through economics. One is normative and the other positivist. The first is defined by the belief that the goal of economic theorizing is to explain what *ought* to be, while the other with what *is*. Many of the recent applications of computer science to economics have been of the normative variety. In this paper

we provide an application of computer science to economics in the positivist tradition.

A central assumption of the positivist tradition has been that the preferences of economic agents can be represented by a monotone, concave, utility function. This in turn has led to a line of work devoted to verifying this assumption. The question is usually phrased this way: when is it that a finite collection of observations about the actions of an agent is consistent with the behavior of an agent endowed with a monotone, concave, utility function?

The particular class of observations we are interested in are price-demand pairs. Each member of the pair is a vector. The first a vector of prices (one price for each commodity) and the second a demand vector that lists the quantity demanded of each good at the relevant price. The set of price-demand pairs is said to be *rationalizable* if the observed demand choices at each price can be explained by supposing the agent chooses the demand at each price by maximizing a monotone, concave utility function subject to a budget constraint. Samuelson [28] first raised the question of when is observed demand rationalizable? Under what circumstances can we conclude that the data is consistent with the behavior of a utility maximizing agent equipped with a monotone concave utility function and subject to a budget constraint? When can we refute utility maximization? Samuelson gave a necessary but insufficient condition on the underlying preferences known as the *weak axiom of revealed preference*. Later work by Houthakker [12], Richter [25] and Varian [30] characterized the conditions for rationalizability. Conditions like these are important because they mean that theory of human behavior assumed by the positivist tradition is falsifiable and therefore a scientific theory.

Checking the consistency of a theory is not an end in itself. One would also like the theory to make predictions. Assuming the observations are rationalizable can we use them to predict the demand at as yet unobserved prices? If yes, this would be attractive, since it would provide a non-parametric way to estimate demand. This is the particular question that inspires this paper. It has a natural formulation in terms of PAC learnability. The main results are upper and lower bounds on the sample complexity of what are known as income-Lipschitz demand functions. These rely on an extension of the notion of fat shattering dimension to real vector valued functions that may be of independent interest.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

Previous work on learning in economic settings is highly diversified. Within the game theory community the most common approach is the Bayesian parametric approach. In these cases there is some prior assumption on the class of the demand function, like Cobb-Douglas or Leontief. The learning problem in these cases is relatively simple, and focused on estimating the parameters. The economic question is usually focused on the strategic aspects of eliciting information from the observed behavior. For more on parametric learning in economics and game theory see [3, 10].

Non parametric approaches to learning in economic settings were developed in parallel by econometricians, operation researchers and computer scientists, and this, to some extent, may explain the different assumptions on the information sets. In consumer theory an agent is faced with various budget sets and she must maximize her utility under this constraint. There are three types on observed information. The most informative are direct observations on the utility, namely, each observation would consist of a price, demand and some numeric measure of utility. In this case the learning problem is essentially a case of regression learning of real monotone concave functions. Kalai [13] studies learning algorithm for monotone utility functions describing medical data, however, the methods and efficiency results apply to utility functions as well. Note that direct access to utility approach is very hard to justify economically as utility functions are abstract entities which might be very difficult to quantify.

A second approach which was studied by Herbrich et al. [11] assumes that the information observed consists on pairs of price and demand and the relative preferences between the points. They showed that with this information the neural network can efficiently learn the indifference curves of the utility. The difficulty in this approach from an economical point of view is that the market mechanism does not naturally elicit preference information beyond what is implicit in the demand, and indeed the efficiency results of Herbrich et al. do not hold without the excess information.

The third approach known in the literature as *revealed preference* assumes only prices and demands are observable. As we see in section 3 the basic theorems addressed the issue of consistency vs. refutability, but in a sequence papers Varian [30, 31] showed how these relate to forecasting. Another recent work along these lines is was done by Blundell et al. [5, 6]. These papers introduce a model where an agent observes prices and Engel curves for these prices, this approach requires full information on a finite number of price trajectories.

This paper continues the line of research we began in [4], in that paper we studied the learnability of demand functions by a class of demand functions that are utility maximizers for piecewise linear monotone concave utility functions. Our main result in that paper showed that the class of all such demand functions had unbounded fat shattering dimension and therefore could not be used for learning.

## 2. DEMAND AND RATIONALITY

A *utility* function  $u : \mathbb{R}_+^n \rightarrow \mathbb{R}$  is a function relating bundles of goods to a cardinal in a manner reflecting the preferences over the bundles. A rational agent with a budget that w.l.g equals 1 facing a price vector  $p \in \mathbb{R}_+^n$  will choose from her *budget set*  $B(p) = \{x \in \mathbb{R}_+^n : p \cdot x \leq 1\}$  a bundle  $x \in \mathbb{R}_+^n$  that maximizes her private utility.

A standard assumption in economic theory is that the function is monotone increasing, namely, if  $x \geq y$ , in the sense that the inequality holds coordinatewise, then  $u(x) \geq u(y)$ . This reflects the assumption that agents will always prefer more of any one good. This, of course, does not necessarily hold in practice, as in many cases excess supply may lead to storage expenses or other externalities. However, in such cases the demand will be an interior point of the budget set and the less preferred bundles won't be observed. Another common assumption on the utility is that all the marginals (partial derivatives) are monotone decreasing. This is the *law of diminishing marginal utility* which assumes that the larger the excess of one good over the other the less we value each additional good of one kind over the other. These assumptions imply that the utility function is concave and monotone on the observations.

The *demand function* of a rational agent is the correspondence  $f_u : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n$  satisfying

$$f(p) = \operatorname{argmax}\{u(x) : p \cdot x \leq I\}$$

In general this correspondence is not necessarily single valued, but it is single valued for income Lipschitz demand functions and demand functions maximizing strictly concave utility.

Since large quantities of any good are likely to create utility decreasing externalities, we assume the prices are limited to a compact set. W.l.g. we assume  $u$  has marginal utility zero outside  $[0, 1]^d$ . Any budget set that is not a subset of the support is maximized on any point on the boundary of the support and it is therefore unpredictable for these prices. We are thus interested in forecasts for budget sets that are below the simplex  $\Delta_d = \operatorname{conv}\{(0, \dots, 1, \dots, 0)\}$ , these budget sets are determined by price vectors  $p = (p_1, \dots, p_n)$  such that  $p_i > 1$  for  $i = 1, \dots, n$ . This set of prices, denoted  $\Gamma$  is metrizable with:

$$d_P(p, p') = \max\{|\frac{1}{p_i} - \frac{1}{p'_i}| : i = 1, \dots, d\}$$

Note that with this metric  $\Gamma$  is compact.

## 3. REVEALED PREFERENCE

### 3.1 Rationalizing Finite Observations

Addressing these issues Houthakker [12] noted that an observer can see only finite quantities of data. He asks when can it be determined that a finite set of observations is consistent with utility maximization without making parametric assumptions? He shows that rationalizability of a finite set of observations is equivalent to the *strong axiom of revealed preference*. Richter [25] showed that strong axiom of revealed preference is equivalent to rationalizability by a strictly concave monotone utility function. Afriat [1] gives another set of rationalizability conditions the observations must satisfy. Varian [30] introduces the *generalized axiom of revealed preference* (GARP), an equivalent form of Afriat's consistency condition that is easier to verify computationally. It is interesting to note that these necessary and sufficient conditions for rationalizability are essentially versions of the well known Farkas lemma [8] (see also [32]).

Afriat [1] proved his theorem by an explicit construction of a utility function witnessing consistency. Varian [30] took this one step further progressing from consistency to forecasting. Varian's forecasting algorithm basically rules out

bundles that are revealed inferior to observed bundles and finds a bundle from the remaining set that together with the observations is consistent with GARP. Furthermore, he introduces Samuelson's "money metric" as a canonical utility function and gives upper and lower envelope utility functions for the money metric. Knoblauch [18] shows these envelopes can be computed efficiently. Varian [31] provides an up to date survey on this line of research.

### 3.2 Piecewise Linear Utility Functions

A sequence of prices and demands  $(p_1, x_1), \dots, (p_n, x_n)$  is *rationalizable* if there exists a utility function  $u$  such that  $x_i = f_u(p_i)$  for  $i = 1, \dots, n$ . We begin with a trivial observation, if  $p_i \cdot x_j \leq p_j \cdot x_i$  and  $x_i = f(p_i)$  then  $x_i$  is preferred over  $x_j$  since the latter is in the budget set when the former was chosen. It is therefore revealed that  $u(x_j) \leq u(x_i)$  implying  $p_j \cdot x_j \leq p_j \cdot x_i$ .

Suppose there is a sequence  $(p_{i_1}, x_{i_1}), \dots, (p_{i_k}, x_{i_k})$  such that  $p_{i_j} \cdot (x_{i_j} - x_{i_{j+1}}) \leq 0$  for  $j = 1 \dots k-1$  and  $p_{i_k} \cdot (x_{i_k} - x_{i_1}) \leq 0$ . Then the same reasoning shows that  $u(x_{i_1}) = u(x_{i_2}) = \dots = u(x_{i_k})$  implying  $p_{i_1} \cdot (x_{i_1} - x_{i_2}) = p_{i_2} \cdot (x_{i_2} - x_{i_3}) = \dots = p_{i_{k-1}} \cdot (x_{i_{k-1}} - x_{i_k}) = 0$ . We call the latter condition the *Afriat condition* (AC). This argument shows that AC is necessary for rationalizability; the surprising result in Afriat's theorem is that this condition is also sufficient.

Let  $A$  be an  $n \times n$  matrix with entries  $a_{ij} = p_i \cdot (x_j - x_i)$  ( $a_{ij}$  and  $a_{ji}$  are independent),  $a_{ii} = 0$  and let  $D(A)$  be the weighted digraph associated with  $A$ . The matrix satisfies AC if every cycle with negative total weight includes at least one edge with positive weight.

**THEOREM 1.** *There exists  $y = (y_1, \dots, y_n) \in \mathbb{R}^n$  and  $s = (s_1, \dots, s_n) \in \mathbb{R}_+^n$  satisfying the set of inequalities  $L(A)$ ,*

$$y_j \leq y_i + s_i a_{ij} \quad i \neq j \quad 1 \leq i, j \leq n$$

*iff  $D(A)$  satisfies AC.*

*Proof:* If  $L(A)$  is feasible then it is easy to see that

$$u(x) = \min_i \{y_i + s_i p_i (x - x_i)\}$$

is a concave utility function that is consistent with the observations, and from our previous remark it follows that  $D(A)$  satisfies AC.

In the other direction it is shown by explicit construction that Afriat's condition for  $D(A)$  implies  $L(A)$  is feasible. The construction provides a utility function that is consistent with the observations. Teo and Vohra [27] give a strongly polynomial time algorithm for this construction which will be the heart of our learning algorithm.

The construction is executed in two steps. First, the algorithm finds  $s \in \mathbb{R}_+^n$  such that the weighted digraph  $D(A, s)$  defined by the matrix  $\tilde{a}_{ij} = s_i a_{ij}$  has no cycle with negative total weight if  $D(A)$  satisfies AC and returns a negative cycle otherwise.

The dual of a *shortest path* problem is given by the constraints:

$$y_j - y_i \leq s_i a_{ij} \quad i \neq j$$

It is a standard result (see [?] p 109) that the system is feasible iff  $D(A, s)$  has no negative cycles. Thus, in the second step, if  $D(A)$  satisfies AC, the algorithm calls a **SHORT-EST-PATH** algorithm to find  $y \in \mathbb{R}^n$  satisfying the constraints.

Now we describe how to choose the  $s_i$ 's. Define  $S = \{(i, j) : a_{ij} < 0\}$ ,  $E = \{(i, j) : a_{ij} = 0\}$  and  $T = \{(i, j) : a_{ij} > 0\}$  and let  $G = ([n], S \cup E)$  be a digraph with weights  $w_{ij} = -1$  if  $(i, j) \in S$  and  $w_{ij} = 0$  otherwise.  $D(A)$  has no negative cycles, hence  $G$  is acyclic and breadth first search can assign potentials  $\phi_i$  such that  $\phi_j \leq \phi_i + w_{ij}$  for  $(i, j) \in S \cup E$ . We relabel the vertices so that  $\phi_1 \geq \phi_2 \geq \dots \geq \phi_n$ . Let

$$\delta_i = (n-1) \frac{\max_{(i,j) \in S} (-a_{ij})}{\min_{(i,j) \in T} a_{ij}}$$

if  $\phi_i < \phi_{i-1}$  and  $\delta_i = 1$  otherwise, and define

$$s_i = \prod_{j=2}^i \delta_j = \delta_i \cdot s_{i-1}$$

We show that for this choice of  $s$ ,  $D(A, s)$  contains no negative weight cycle. Suppose  $C = (i_1, \dots, i_k)$  is a cycle in  $D(A, s)$ . If  $\phi$  is constant on  $C$  then  $a_{i_j i_{j+1}} = 0$  for  $j = 1, \dots, k$  and we are done. Otherwise let  $i_v \in C$  be the vertex with smallest potential satisfying w.l.o.g.  $\phi(i_v) < \phi(i_{v+1})$ .

For any cycle  $C$  in the digraph  $D(A, s)$ , let  $(v, u)$  be an edge in  $C$  such that (i)  $v$  has the smallest potential among all vertices in  $C$ , and (ii)  $\phi_u > \phi_v$ . Such an edge exists, otherwise  $\phi_i$  is identical for all vertices  $i$  in  $C$ . In this case, all edges in  $C$  have non-negative edge weight in  $D(A, s)$ .

If  $(i_v, i_{v+1}) \in S \cup E$ , then we have

$$\phi(i_{v+1}) \leq \phi(i_v) + w_{i_v, i_{v+1}} \leq \phi(i_v)$$

a contradiction. Hence  $(i_v, i_{v+1}) \in T$ . Now, note that all vertices  $q$  in  $C$  with the same potential as  $i_v$  must be incident to an edge  $(q, t)$  in  $C$  such that  $\phi(t) \geq \phi(q)$ . Hence the edge  $(q, t)$  must have non-negative weight. i.e.,  $a_{q,t} \geq 0$ . Let  $p$  denote a vertex in  $C$  with the second smallest potential. Now,  $C$  has weight

$$s_v a_{vu} + \sum_{(k,l) \in C \setminus (v,u)} s_k a_{k,l} \geq s_v a_{v,u} + s_p (n-1) \max_{(i,j) \in S} \{a_{ij}\} \geq 0,$$

i.e.,  $C$  has non-negative weight  $\square$

Algorithm 1 returns in polynomial time a hypothesis that is a piecewise linear function and agrees with the labeling of the observation namely sample error zero. To use this function to forecast demand for unobserved prices we need algorithm 2 which maximizes the function on a given budget set. Since  $u(x) = \min_i \{y_i + s_i p_i (x - x_i)\}$  this is a linear program and can be solved in time polynomial in  $d$ ,  $n$  as well as the size of the largest number in the input.

## 4. SUPERVISED LEARNING

In a supervised learning problem, a learning algorithm is given a finite sample of labeled observations as input and is required to return a model of the relationship underlying the labeling. This model, which we refer to as a *hypothesis*, is usually a computable function that is used to forecast the labels of future observations. The labels are often binary values indicating the membership of the observed points in the set that is being learned. However, we are not limited to binary values and, indeed, in the demand functions we are studying the labels are real vectors.

The learning problem has three major components: *estimation*, *approximation* and *complexity*. The estimation problem is concerned with the tradeoff between the size of

---

**Algorithm 1** Utility Algorithm

---

**Input**  $(x_1, p_1), \dots, (x_n, p_n)$   
 $S \leftarrow \{(i, j) : a_{ij} < 0\}$   
 $E \leftarrow \{(i, j) : a_{ij} = 0\}$   
**for all**  $(i, j) \in S$  **do**  
     $w_{ij} \leftarrow -1$   
**end for**  
**for all**  $(i, j) \in E$  **do**  
     $w_{ij} \leftarrow 0$   
**end for**  
**while** there exist unvisited vertices **do**  
    visit new vertex  $j$   
    assign potential to  $\phi_j$   
**end while**  
reorder indices so  $\phi_1 \leq \phi_2 \dots \leq \phi_n$   
**for all**  $1 \leq i \leq n$  **do**  
     $\delta_i \leftarrow (n-1) \frac{\max_{(i,j) \in S} (-a_{ij})}{\min_{(i,j) \in T} a_{ij}}$   
     $s_i \leftarrow \prod_{j=2}^i \delta_j$   
**end for**  
SHORTEST\_PATH( $y_j - y_i \leq s_i a_{ij}$ )  
**Return**  $y_1, \dots, y_n \in \mathbb{R}^d$  and  $s_1, \dots, s_n \in \mathbb{R}_+$

---

---

**Algorithm 2** Evaluation

---

**Input**  $y_1, \dots, y_n \in \mathbb{R}^d$  and  $s_1, \dots, s_n \in \mathbb{R}_+$   
     $\max z$   
     $z \leq y_i + s_i p_i (x - x_i)$  for  $i = 1, \dots, n$   
     $px \leq 1$   
**Return**  $x$  for which  $z$  is maximized

---

the sample given to the algorithm and the degree of confidence we have in the forecast it produces. The approximation problem is concerned with the ability of hypotheses from a certain class to approximate target functions from a possibly different class. The complexity problem is concerned with the computational complexity of finding a hypothesis that approximates the target function.

In the *probably approximately correct* (PAC) paradigm, the learning of a target function is done by a class of hypothesis functions, that does or does not include the target function itself; it does not necessitate any parametric assumptions on this class. It is also assumed that the observations are generated independently by some distribution on the domain of the relation and that this distribution is fixed. If the class of target functions has finite 'dimensionality' then a function in the class is characterized by its values on a finite number of points. The basic idea is to observe the labeling of a finite number of points and find a function from a class of hypotheses which "tends to agree" with this labeling. The theory tells us that if the sample is large enough then any function that "tends to agree" with the labeling will, with high probability, be a good approximation of the target function for future observations. The prime objective of PAC theory is to develop the relevant notion of dimensionality and to formalize the tradeoff between dimensionality, sample size and the level of confidence in the forecast.

In the revealed preference setting, our objective is to use a set of observations of prices and demand to forecast demand for unobserved prices. It is assumed the observations are the demand response of a rational agent for different prices.

Thus the relationship being modeled is between prices and consumption bundles. The hypothesis class is the class of piecewise linear demand functions, it is a proper subset of the class of all rational demand functions, but as we have seen in section ?? for any rational demand function and any set of prices there exists a piecewise linear demand that approximates the demand with zero error on the sample. It is not difficult to show that this implies the class of piecewise linear demands is dense in the class of rational demands with the supremum norm. This shows any rational demand can be approximated with any degree of accuracy by this class hence, there is no approximation issue for this class.

For comparison, a parametric paradigm assumes that the underlying functional relationship comes from a well defined family, for instance a class of Cobb-Douglas production functions; the system must learn the parameters characterizing this family. A parametric learning algorithm observing a finite set of production data assumes, *a priori*, it comes from a Cobb-Douglas production function and returns a hypothesis that is usually from the same class but can also come from different class like the class of polynomials with bounded degrees. The estimation problem in the latter case would be to assess the sample size needed to obtain a good estimate of the coefficients. The approximation problem would be to assess the error sustained from approximating a rational function by a polynomial and the complexity problem would be the assessment of the time required to compute the polynomial coefficients.

The theory of PAC learning for real valued functions is concerned predominantly with learning functions with scalar values, in this paper we extend the fundamental notions of PAC learning to vector valued functions and use them to prove lower and upper bounds for sample complexity of learning demand functions. Before we can proceed with the formal definition, we must clarify what we mean by *forecast* and *tend to agree*. In the case of discrete learning, we would like to obtain a function  $f$  that with high probability agrees with the observations. We would then take the probability  $\mathbb{P}(f(x) = y)$  as the measure of the quality of the estimation. Demand functions are real vector functions and we therefore do not expect  $f$  to exactly agree with the observation, we are content with having small mean square errors on all coordinates. Thus, our measure of *estimation error* is given by:

$$er(f) = er_{\mathbb{P}}(f) = \mathbb{E}_{(p,x)}(\|f(p) - x\|_{\infty}^2).$$

For given observations  $S = \{(p_1, x_1), \dots, (p_n, x_n)\}$  we measure the agreement by the *sample error*

$$\hat{er}_S(f) = \sum_j \|x_j - f(p_j)\|_{\infty}^2.$$

A *sample error minimization* (SEM) algorithm is an algorithm that finds a hypothesis minimizing  $er_S(f)$ . In the usual setting such an algorithm is used to minimize the estimation error. We have seen that in the revealed preference setting algorithms, like algorithm 1, return hypothesis with zero sample error.

**DEFINITION 1.** *A set of demand functions  $\mathcal{C}$  is probably approximately correct (PAC) learnable if for any  $\varepsilon, \delta > 0$ ,  $f \in \mathcal{C}$  and distribution  $\mathbb{P}$  on the prices and demands, there exists an algorithm  $A$  that for a set of observations of length  $m_A = m_A(\varepsilon, \delta) = \text{Poly}(\frac{1}{\varepsilon}, \frac{1}{\delta})$  finds a function  $f \in \mathcal{C}$  such that  $er(f) < \varepsilon$  with probability  $1 - \delta$ .*

There may be several learning algorithms for  $\mathcal{C}$  with different sample complexities. The minimal  $m_A$  is called the *sample complexity* of  $\mathcal{C}$ .

Note that in the definition there is no mention of the time complexity of finding  $f$  in  $\mathcal{C}$  or evaluating  $f(p)$ . A class  $\mathcal{C}$  is *efficiently* PAC-learnable if there is a  $\text{Poly}(\frac{1}{\delta}, \frac{1}{\epsilon})$  time algorithm for finding  $f$  and evaluating  $f(p)$ . Algorithm 1 shows that the class of piecewise linear demands is efficiently PAC-learnable if it is PAC-Learnable.

For discrete function sets, sample complexity bounds may be derived from the VC-dimension of the set (see [29, 17]). An analog to this notion of dimension for real functions introduced by Kearns [16] is the fat shattering dimension. We use an adaptation of this notion to real vector valued function sets. Let  $\Gamma \subset \mathbb{R}_+^d$  and let  $\mathcal{C}$  be a set of real functions from  $\Gamma$  to  $\mathbb{R}_+^d$ .

**DEFINITION 2.** For  $\gamma > 0$ , a set of points  $p_1, \dots, p_n \in \Gamma$  is  $\gamma$ -shattered by a class of real functions  $\mathcal{C}$  if there exists  $x_1, \dots, x_n \in \mathbb{R}_+^d$  and pairs of parallel affine hyperplanes  $(H_{0,1}, H_{1,1}) \dots, (H_{0,n}, H_{1,n})$  such that  $H_{0,i}, H_{1,i} \subset \mathbb{R}^d$ ,  $0 \in H_{0,i}^- \cap H_{1,i}^+$ ,  $\text{dist}(H_{0,i}, H_{1,i}) > \gamma$  for  $i = 1 \dots n$  and for each  $b = (b_1, \dots, b_n) \in \{0, 1\}^n$  there exists a function  $f_b \in \mathcal{C}$  such that  $f_b(p_i) \in x_i + H_{0,i}^+$  if  $b_i = 0$  and  $f_b(p_i) \in x_i + H_{1,i}^-$  if  $b_i = 1$ .

We say that  $x_1, \dots, x_n$  and  $(H_{0,1}, H_{1,1}) \dots, (H_{0,n}, H_{1,n})$  witness the shattering of  $p_1, \dots, p_n$ . Abusing notation somewhat, we say that the functions  $\{f_b : b \in \{0, 1\}^n\}$  also witness the shattering. We define the  $\gamma$ -fat shattering dimension of  $\mathcal{C}$ , denoted  $\text{fat}_{\mathcal{C}}(\gamma)$  as the maximal size of a  $\gamma$ -shattered set in  $\Gamma$ . If this size is unbounded then the dimension is infinite.

## 5. LOWER BOUND

We use the fat shattering dimension to derive a lower bound on the sample complexity of a class of demand functions.

**LEMMA 2.** Suppose the functions  $\{f_b : b \in \{0, 1\}^n\}$  witness the shattering of  $\{p_1, \dots, p_n\}$ . Then, for any  $x \in \mathbb{R}_+^d$  and labels  $b, b' \in \{0, 1\}^n$  such that  $b_i \neq b'_i$  for  $1 \leq i \leq n$  either  $\|f_b(p_i) - x\|_{\infty} > \frac{\gamma}{2\sqrt{d}}$  or  $\|f_{b'}(p_i) - x\|_{\infty} > \frac{\gamma}{2\sqrt{d}}$ .

*Proof:* Since the max exceeds the mean, it follows that if  $f_b$  and  $f_{b'}$  correspond to labels such that  $b_i \neq b'_i$  then

$$\|f_b(p_i) - f_{b'}(p_i)\|_{\infty} \geq \frac{1}{\sqrt{d}} \|f_b(p_i) - f_{b'}(p_i)\|_2 > \frac{\gamma}{\sqrt{d}}.$$

This implies that for any  $x \in \mathbb{R}_+^d$  either  $\|f_b(p_i) - x\|_{\infty} > \frac{\gamma}{2\sqrt{d}}$  or  $\|f_{b'}(p_i) - x\|_{\infty} > \frac{\gamma}{2\sqrt{d}}$   $\square$

**THEOREM 3.** Suppose that  $\mathcal{C}$  is a class of piecewise linear demand functions. Then any learning algorithm  $A$  for  $\mathcal{C}$  has sample complexity

$$m_A(\epsilon, \delta) \geq \frac{1}{2} \text{fat}_{\mathcal{C}}(8\delta\epsilon)$$

*Proof:* Suppose  $n = \frac{1}{2} \text{fat}_{\mathcal{C}}(8\delta\epsilon)$  then there exists a set  $P = \{p_1, \dots, p_{2n}\}$  that is shattered by  $\mathcal{C}$  and witnessed by  $\mathcal{C}_P = \{f_b : b \in \{0, 1\}^{2n}\}$ . To show the sample complexity  $m_A(\epsilon, \delta) > n$  it suffices to show that any learning algorithm  $\mathcal{C}$  will, with probability bounded away from zero, return a function with error bounded away from zero for  $\mathbb{P}$ , the uniform distribution on  $P$ .

We do the following randomization, first be choose uniformly  $(i_1, \dots, i_n) \subset [2n]$  and  $b \in \{0, 1\}^{2n}$ . Suppose for the random sample

$$S_{b, i_1, \dots, i_n} = \{(p_{i_1}, f_b(p_{i_1})) \dots, (p_{i_{2n}}, f_b(p_{i_{2n}}))\}$$

a learning algorithm returns a function  $f_{S_{b, i_1, \dots, i_n}}$ . Let

$$X_{b, i_1, \dots, i_n} = \text{er}(f_{S_{b, i_1, \dots, i_n}})$$

where the error randomization is over the uniform distribution on  $\{(p_1, f_b(p_1)) \dots, (p_{2n}, f_b(p_{2n}))\}$ . Obviously, for  $j = 1, \dots, n$  if  $b_{i_j} = b'_{i_j}$  then  $f_{S_{b, i_1, \dots, i_n}} = f_{S_{b', i_1, \dots, i_n}}$ . It follows from lemma 2 (with  $\gamma = 4\sqrt{d}\epsilon$ ) that the probability that

$$\|f_{S_{b, i_1, \dots, i_n}}(p) - f_b(p)\|_{\infty} > 4\epsilon$$

for  $p \in P - \{p_{i_1}, \dots, p_{i_n}\}$  is at least as high as getting heads on a fair coin toss, therefore

$$\mathbb{E}_b \left( \mathbb{E}_{i_1, \dots, i_n} (X_{S_{b, i_1, \dots, i_n}} | b) \right) = \mathbb{E}(X_{S_{b, i_1, \dots, i_n}}) > 2\epsilon$$

Hence  $\mathbb{E}_{i_1, \dots, i_n} (X_{S_{b_0, i_1, \dots, i_n}}) > 2\epsilon$  for some  $b_0 \in \{0, 1\}^{2n}$ . Since  $X_{S_{b, i_1, \dots, i_n}}$  is bounded by some  $M$  it follows that

$$\mathbb{P}_{i_1, \dots, i_n} \{X_{S_{b_0, i_1, \dots, i_n}} > \epsilon\} > \frac{\epsilon}{M}.$$

This shows  $\text{er}(f_{b_0}) > \epsilon$  with probability greater than  $\epsilon M$  for some  $f_{b_0}$ , hence,  $f_{b_0}$  is not PAC-learnable with a sample of size  $n$   $\square$

**THEOREM 4.** Let  $\mathcal{C}$  be the set of piecewise linear demand functions from  $\Gamma$  to  $\mathbb{R}_+^d$  then

$$\text{fat}_{\mathcal{C}}(\gamma) = \infty$$

The theorem was proved in [4]. This implies that the class of all piecewise linear demands is too general to use as a learning class.

## 6. UPPER BOUNDS

In this section we prove an upper bound on the sample complexity of a class of demand functions with finite fat shattering dimension. We sketch a proof along the classical lines of proof for real function classes. We redefine the appropriate notions for vector valued functions and prove the key lemma and some of the corollaries for this case with the new bound. We quote without proof essential results which do not require any new ideas beyond those used for the real value case, the full proofs can be found in Anthony and Bartlett [2].

### 6.1 Covering and Packing Numbers

A set of  $n$ -samples  $W_n = \{(x_1, \dots, x_n) : x_i \in \mathbb{R}_+^d \ i = 1 \dots n\} \subset \mathbb{R}_+^{nd}$  is *totally covered* if for every  $\epsilon > 0$  there exists a finite set  $C_{n, \epsilon} \subset W_n$  such that for every  $(x_1, \dots, x_n) \in W_n$  there exists  $(y_1, \dots, y_n) \in C_{n, \epsilon}$  satisfying

$$\max_{i=1 \dots n} \{\|x_i - y_i\|_{\infty}\} < \epsilon$$

$C_{n, \epsilon}$  is then called a  $\epsilon$ -cover of  $W_n$ . The  $\epsilon$ -covering number of  $W_n$  is the minimal cardinality of an  $\epsilon$ -cover, denoted  $\mathcal{N}(\epsilon, W_n)$ .

$B_{n, \epsilon} \subset W_n$  is an  $\epsilon$ -packing (or a  $\epsilon$ -separated subset) of  $W_n$  if

$$\max_{i=1 \dots n} \{\|x_i - y_i\|_{\infty}\} > \epsilon$$

for every  $(x_1, \dots, x_n), (y_1, \dots, y_n) \in B_{n, \varepsilon}$ . The  $\varepsilon$ -packing number of  $W_n$  is the maximal cardinality of an  $\varepsilon$ -packing, denoted  $\mathcal{M}(\varepsilon, W_n)$ .

Let  $\mathcal{C}$  be a class of functions from  $\Gamma$  to  $\mathbb{R}_+^d$ . For a sequence of prices  $p = (p_1, \dots, p_n) \in \Gamma^n$  let

$$W_p = \{(f(p_1), \dots, f(p_n)) : f \in \mathcal{C}\}.$$

The covering number of the class  $\mathcal{C}$  is given by

$$\mathcal{N}_\infty(\varepsilon, \mathcal{C}, n) = \max\{\mathcal{N}(\varepsilon, W_p) : p \in \Gamma^n\}$$

and infinity if there is no maximum. The packing number of  $\mathcal{C}$  is given by

$$\mathcal{M}_\infty(\varepsilon, \mathcal{C}, n) = \max\{\mathcal{M}(\varepsilon, W_p) : p \in \Gamma^n\}$$

and infinity if there is no maximum.

In the learning context, it is sometimes easier to use packing numbers and othertimes packing numbers, the following lemma, a standard result in functional analysis, shows the relation between these terms.

LEMMA 5.

$$\mathcal{M}_\infty(2\varepsilon, \mathcal{C}, n) \leq \mathcal{N}_\infty(\varepsilon, \mathcal{C}, n) \leq \mathcal{M}_\infty(\varepsilon, \mathcal{C}, n)$$

## 6.2 Quantization

In this section we reduce the problem from a continuous setting to a discrete setting, and show why the later gives a bound to the former. For  $r \in \mathbb{R}_+$  denote by  $[r]$  the greatest integer that is less than  $r$ . For  $r = (r_1, \dots, r_d) \in \mathbb{R}_+^d$  denote  $[r] = ([r_1], \dots, [r_d])$ . The quantization of  $r = (r_1, \dots, r_d) \in \mathbb{R}_+^d$  by quantization width  $\alpha > 0$  is given by:

$$Q_\alpha(r) = \alpha \cdot \left( \left\lfloor \frac{r_1}{\alpha} \right\rfloor, \dots, \left\lfloor \frac{r_d}{\alpha} \right\rfloor \right)$$

with scalar multiplication. In accord, the quantization of a demand function is

$$Q_\alpha f(x) = Q_\alpha(f(x))$$

If the demand function  $f$  maps to  $[0, 1]^d$  then  $Q_\alpha f$  maps to the finite lattice  $\{0, \alpha, 2\alpha, \dots, \lfloor \frac{1}{\alpha} \rfloor \alpha\}^d$ . Finally, we denote  $Q_\alpha \mathcal{C} = \{Q_\alpha f : f \in \mathcal{C}\}$  the class of lattice valued functions.

The following lemma shows the effect of quantization on the packing number and the fat shattering dimension. Idea of the proof is identical to the real value case.

LEMMA 6. For  $\varepsilon > 0$  and  $n \in \mathbb{N}$

$$1. \mathcal{M}_\infty(\varepsilon, \mathcal{C}, n) \leq \mathcal{M}_\infty(\varepsilon, Q_{\frac{\varepsilon}{2}}(\mathcal{C}), n)$$

$$2. \text{fat}_{Q_{\frac{\varepsilon}{2}}(\mathcal{C})}(\frac{\varepsilon}{2}) \leq \text{fat}_{\mathcal{C}}(\frac{\varepsilon}{4})$$

## 6.3 A Combinatorial Lemma

This lemma is the key to the upper bound on the sample complexity, the proof follows the line of proof of [2] with the adjustments necessary for vector valued functions.

LEMMA 7. Let  $X = \{0, \dots, b\}^d$  be a  $d$ -dimensional integer lattice,  $|\Gamma| = n$  and  $H \subset X^\Gamma$ . If  $s = \text{fat}_H(1)$  then

$$\mathcal{M}(2, H) \leq 2(b^{2d})^{\lceil \log_2 y \rceil}$$

where  $y = \sum_{i=1}^s \binom{m}{i} b^{id}$ .

*Proof:* For  $k \geq 2$  we define:

$$\zeta(\Gamma, G) = \{(P, x) : G \text{ 1-shatters } \emptyset \neq P \subset \Gamma \text{ witnessed by } x_1, \dots, x_i \text{ where } i = |P|\}$$

$$t(k, n) = \min\{|\zeta(\Gamma, G)| : |\Gamma| = n, G \subset X^\Gamma, |G| = k, \text{ and } G \text{ is 2-separated}\}$$

and  $t(k, n) = \infty$  if there is no 2-separated  $G$  as above.

If  $|P| < i$  then there can be no more than  $b^{id}$  tuples  $x_1, \dots, x_i$  witnessing the 1-shattering of  $P$ , hence

$$|\{(P, x) \in \zeta(\Gamma, G) : 0 < |P| \leq s\}| \leq y.$$

This implies that if  $t(k, n) \geq y$  then every 2-separated set  $G$  with cardinality  $k$  1-shatters some set  $P$  with  $|P| > s$ . Since  $s = \text{fat}_H(1)$  it implies that  $H|_P$  for any  $p = (p_1, \dots, p_n)$  is not 2-separated hence  $\mathcal{M}(2, H) < k$ .

We compute  $t(k, n)$  for  $k = 2(b^{2d})^{\lceil \log_2 y \rceil}$  by recursion. We take an arbitrary partition of  $G$  into two sets. The 2-separation implies for every pair  $(g, g')$  there exists associated  $p_i \in \Gamma$  such that  $\|g(p_i) - g'(p_i)\|_\infty > 2$ . The pigeonhole principle shows there exists  $p_0 \in p$  associated with at least  $\frac{k}{2n}$  pairs.

There are  $\binom{b^d}{2}$  possible pairs of values for  $(g(p_0), g'(p_0))$ , hence again by the pigeonhole principle there are at least  $\frac{k}{nb^{2d}}$  pairs such that  $\{g(p_0), g'(p_0)\}$  are identical. Thus, there exists a partition  $G_0$  and  $G_1$  of  $G$  with  $|G_0| = |G_1|$  such that  $g(p_0) = z_0$  for  $g \in G_0$ ,  $g(p_0) = z_1$  for  $g \in G_1$  and  $\|z_0 - z_1\|_2 \geq \|z_0 - z_1\|_\infty > 2$ . By definition if  $x_1, \dots, x_i$  is a witness to the shattering of  $P$  (with appropriate pairs of hyperplanes) by  $G_0$  or  $G_1$  then the same is true for  $G$ . Obviously,  $p_0 \notin P$  in this case. In the case that  $x_1, \dots, x_i$  is a witness to the shattering of  $P$  by both  $G_0$  and  $G_1$  then  $x_1, \dots, x_i, x_{i+1}$  witnesses the shattering of  $P \cup \{p_0\}$  where  $H_{1, i+1}$  is a lattice point between two hyperplanes  $H_{0, i+1}$  and  $H_{1, i+1}$  passing through  $z_0$  and  $z_1$ .

This gives the relation  $t(k, n) \geq 2t(\lfloor \frac{k}{mb^{2d}} \rfloor, n-1)$  implying

$$t(2(b^{2d})^{\lceil \log_2 y \rceil}, n) \geq 2^{\lceil \log_2 y \rceil} t(2, n - \lceil \log_2 y \rceil) \geq y$$

since by definition  $t(2, n) = 1$  for every  $n$ .

COROLLARY 8. Let  $\mathcal{C}$  from  $\Gamma$  to  $[0, 1]^d$ . Then for  $\varepsilon > 0$

$$\mathcal{M}_\infty(\varepsilon, \mathcal{C}, n) < 2(nb^{2d})^{\lceil \log_2 y' \rceil}$$

where  $b = \frac{2}{\varepsilon}$ ,  $s = \text{fat}_{\mathcal{C}}(\frac{\varepsilon}{4})$  and

$$y' = \sum_{i=1}^s \binom{n}{i} b^{id}$$

*Proof:* Let  $p \in \Gamma^n$  be fixed. Denote  $H = Q_{\frac{\varepsilon}{2}}(\mathcal{C})|_p$  and let  $s = \text{fat}_{Q_{\frac{\varepsilon}{2}}(\mathcal{C})}(\frac{\varepsilon}{2})$ . Lemma 7 with an appropriate rescaling and  $b = \lfloor \frac{2}{\varepsilon} \rfloor$  implies  $\mathcal{M}(\varepsilon, H) \leq 2(\frac{4n}{\varepsilon^{2d}})^{\lceil \log_2 y \rceil}$  and since  $p$  was arbitrary we get

$$\mathcal{M}_\infty(\varepsilon, \mathcal{C}, n) \leq \mathcal{M}_\infty(\varepsilon, Q_{\frac{\varepsilon}{2}}(\mathcal{C}), n) < 2(nb^2)^{\lceil \log_2 y \rceil} < 2(nb^2)^{\lceil \log_2 y' \rceil}$$

the first inequality follows from lemma 6(1) and the last one, with  $y$  as in lemma 7, from lemma 6(2).

COROLLARY 9.

$$\mathcal{N}_\infty(\varepsilon, \mathcal{C}, n) < 2 \left( n \cdot \frac{4^d}{\varepsilon^{2d}} \right)^{s(\log_2(\frac{4n}{\varepsilon^2})) + d \log(\frac{2}{\varepsilon})}$$

for  $n \geq s$ .

*Proof:*

$$\sum_{i=1}^s \binom{n}{i} b^{id} < b^{sd} \sum_{i=1}^s \binom{n}{i} < \left(\frac{enb^d}{s}\right)^s$$

Hence corollary 8 implies

$$\mathcal{N}_\infty(\varepsilon, \mathcal{C}, n) < 2 \left(n \cdot \frac{4^d}{\varepsilon^{2d}}\right)^{s(\log_2(\frac{en}{s}) + d \log(\frac{2}{\varepsilon}))} \quad \square$$

## 6.4 Upper Bound on the Sample Complexity

The following theorem gives a bound on the uniform convergence.

**THEOREM 10.** *Let  $\mathcal{C}$  be a class of demand functions,  $\mathbb{P}$  be a distribution on  $\Gamma$ ,  $0 < \varepsilon < 1$  and  $n \in \mathbb{N}$ . Then*

$$\mathbb{P}(\exists f \in \mathcal{C} \text{ er}_p(f) > \varepsilon) \leq 4d \mathcal{N}_\infty(\varepsilon, \mathcal{C}, 2n) e^{-\frac{\varepsilon^2 n}{32}}$$

*Proof:* The proof of the theorem is almost identical to the proof in Anthony and Bartlett [2]. The main idea on the proof is that the probability for a bad event where there is a wide gap between the estimation error and the sample error for some function in  $\mathcal{C}$  is bounded by the probability that there is a large difference between the sample error of two independent samples for some function in  $\mathcal{C}$ . That in turn is bounded by the probability that there is a wide gap between the sample error of two independent samples for a function in some covering set of  $\mathcal{C}$ . Next they bound this probability with the size of the covering set and the probability of a some sort of permuted event. The latter reduces to a sum of independent random variables for which the Chernoff bound applies. The only difference with vector values is the covering number is different and there are  $d$  different sums of random variables, for which we add factor of  $d$   $\square$

We are now able to give an upper bound on the sample complexity of algorithm 1.

**THEOREM 11.** *Let  $\mathcal{C}$  be a class of demand functions with finite fat shattering dimension. Then any algorithm that finds function in  $\mathcal{C}$  that agrees with the sample, and in particular algorithm 1 with respect to piecewise linear classes, is a learning algorithm with sample complexity*

$$m_A(\varepsilon, \delta) \leq \frac{c_1}{\varepsilon^2} \left(d \cdot \text{fat}_{\mathcal{C}}(c_2 \varepsilon) \log^2\left(\frac{c_3}{\varepsilon}\right) + \log\left(\frac{c_4 d}{\delta}\right)\right)$$

for constant  $c_1, c_2, c_3, c_4 > 0$  and every  $\varepsilon, \delta > 0$ .

*Proof:* For  $\varepsilon, \delta > 0$  we need to find  $N$  such that

$$\mathbb{P}(\exists f \in \mathcal{C} \text{ er}_p(f) > \varepsilon) < \delta$$

for  $n > N$ . Theorem 10 shows that it suffices to satisfy

$$\begin{aligned} & 4d \mathcal{N}_\infty(\varepsilon, \mathcal{C}, 2n) e^{-\frac{\varepsilon^2 n}{32}} \\ & < 2 \left(n \cdot \frac{4^d}{\varepsilon^{2d}}\right)^{s(\log_2(\frac{en}{s}) + d \log(\frac{2}{\varepsilon}))} e^{-\frac{\varepsilon^2 n}{32}} \\ & < \delta \end{aligned}$$

We will go over the rather tedious computations we used to derive the final formula, though we used several approximations on which it may be possible to improve  $\square$

## 7. LEARNING FROM REVEALED PREFERENCE

Uzawa [26] and Mas-Colell [19, 20] introduced a notion of income-Lipschitz and showed that demand functions with this property are rationalizable. These properties do not require any parametric assumptions and are technically refutable, but they do assume knowledge of the entire demand function and rely heavily on the differential properties of demand functions. Hence, an infinite amount of information is needed to refute the theory.

**DEFINITION 3.** *A demand function is  $L$ -income-Lipschitz, for  $L \in \mathbb{R}_+$ , if*

$$\frac{\|f(p) - f(p')\|_\infty}{d_P(p, p')} \leq L$$

for any  $p, p' \in \Gamma$ .

This property reflects an assumption that preferences and demands have some sort of stability. It rules out dramatically different demands for similar prices, and in particular implies strong concavity. We assume from here on that demand functions are single valued.

We show that the class of utility functions that have marginal utilities with compact support and for which the relevant demand functions are income-Lipschitzian has finite fat shattering dimension.

**THEOREM 12.** *Let  $\mathcal{C}$  be a set of  $L$ -income-Lipschitz demand functions from  $\Gamma$  to  $\mathbb{R}_+^d$  for some global constant  $L \in \mathbb{R}$ . Then*

$$\text{fat}_{\mathcal{C}}(\gamma) \leq \left(\frac{L}{\gamma}\right)^d$$

*Proof:* Let  $p_1, \dots, p_n \in \Gamma$  be a shattered set with witnesses  $x_1, \dots, x_n \in \mathbb{R}_+^d$ . W.l.g.  $x_i + H_0^+ \cap x_j + H_0^- = \emptyset$  implying  $x_i + H_1^- \cap x_j + H_1^+ = \emptyset$ , for a labeling  $b = (b_1, \dots, b_n) \in \{0, 1\}^n$  such that  $b_i = 0$  and  $b_j = 1$ ,  $\|f_b(p_i) - f_b(p_j)\|_\infty > \gamma$  hence  $\|p_i - p_j\|_\infty > \frac{\gamma}{L}$ . A standard packing argument implies  $n \leq \left(\frac{L}{\gamma}\right)^d$   $\square$

The implication of this theorem is that for a sequence of observations on the demand of a utility maximizing agent with a utility function in this class, algorithm 1 will return a piecewise linear utility function that is income-Lipschitz with the same bounds on the Lipschitz coefficient. This implies that, with high probability, after a polynomial number of observations we can construct a piecewise linear demand functions that gives reliable.

It is useful to contrast Theorem 12 with the main result of Mas-Colell [??]. In that paper he supposes the preferences of an agent are monotone, convex and continuous. From observation of their entire demand curve can one recover their preferences? Mas-Colell presents an example of two distinct preferences orderings that yield the same demand curve. So, no, one cannot pin down exactly the underlying preference relation by observing the demand at all possible prices. However, if the underlying preferences are Lipschitzian, then, Mas-Colell shows how to recover the underlying preferences. In this paper we show that it is not necessary to observe the entire demand curve. Rather, a sufficiently large sample of the demand curve is enough to pin down reasonably accurately, the underlying preferences. In addition we provide a bound on on just how large a sample is needed.

## 8. FUTURE RESEARCH

The first question is, of course, whether the bounds can be improved. In the proof of the combinatorial lemma as well as the other proofs, we did not use the assumption that demand functions are derived from monotone concave utility functions. Can we use these fact to improve the upper bounds?

We are focused on the class of piecewise linear demand functions as a learning class. We chose this class because we have an efficient learning algorithm and because most related applications in operations research use piecewise linear functions as first approximations. However, piecewise linear demand functions are by no means the only possible class of learning functions. We conjecture that for any learning class of demand functions with zero estimation error the fat shattering dimension can only go down.

CONJECTURE 1. *If  $\mathcal{L}$  is a class of demand functions such that for any demand function  $f$  and sample  $P$  there exists  $h \in \mathcal{L}$  such that  $\hat{e}_P(f, h) = 0$  then  $\text{fat}_{\mathcal{L}}(\gamma) \leq \text{fat}_{\mathcal{C}}(\gamma)$ .*

## 9. ACKNOWLEDGMENTS

The authors would like to thank Rosa Matzkin, Adam Galambos, Michael Kearns,

## 10. REFERENCES

- [1] Afriat S. N. (1967) The Construction of a Utility Function from Expenditure Data *International Economic Review* **8**, 67–77.
- [2] Anthony M. and Bartlett P. L. (1999) *Neural Network Learning: Theoretical Foundations* Cambridge University Press.
- [3] Aumann R. and Maschler M. (1995) *Repeated Games of Incomplete Information* The MIT Press Cambridge MA.
- [4] Beigman E. and Vohra R. (2006) Learning From Revealed Preference. To appear in *Proceedings of The Seventh ACM Conference on Electronic Commerce (EC'06)*.
- [5] Blundell R., Browning M. and Crawford I. (2003) Nonparametric Engel curves and revealed preference. *Econometrica*, **71**(1):205–240.
- [6] Blundell R. (2005 ) How revealing is revealed preference? *European Economic Journal* **3**, 211 – 235.
- [7] Diewert E. (1973) Afriat and Revealed Preference Theory *Review of Economic Studies* **40**, 419 – 426.
- [8] Farkas J. (1902) Über die Theorie der Einfachen Ungleichungen *Journal für die Reine und Angewandte Mathematik* **124** 1–27
- [9] Fostel A., Scarf H. and Todd M. (2004) Two New Proofs of Afriats Theorem *Economic Theory* **24**, 211-219.
- [10] Fudenberg D. and Levine D. (1998) *The Theory of Learning in Games* The MIT Press Cambridge MA.
- [11] Herbrich R., Keilbach M., Graepel T., Bollmann-Sdorra P. and Obermayer K. (1999) Neural Networks in Economics: Background, Applications, and New Developments *Advances in Computational Economics* **11** 169–196.
- [12] Houthakker H. (1950) Revealed Preference and the Utility Function *Economica* **17**, 159 – 174.
- [13] Kalai A. (2004) Learning Monotonic Linear Functions *Proceedings of the 17th Annual Conference on Learning Theory (COLT'04)* 487–501.
- [14] Kalai, G. (2003) Learnability and rationality of choice *Journal of Economic Theory* **113**, 104 – 117.
- [15] Kalai, G., Rubenstein A. and Spiegel R. (2002) Rationalizing Choice functions by multiple rationals *Econometrica* **70**, 2481-2488.
- [16] Kearns M. J., Schapire R. E. and Sellie L. M. (1994) Toward efficient agnostic learning *Machine Learning* **17**(2/3) 115–142.
- [17] Kearns M. and Vazirani U. (1994) *An Introduction to Computational Learning Theory* The MIT Press Cambridge MA.
- [18] Knoblauch V. (1992) A Tight Upper Bound on the Money Metric Utility Function. *The American Economic Review*, **82**(3):660–663.
- [19] Mas-Colell A. (1977) The Recoverability of Consumers' Preferences from Market Demand. *Econometrica*, **45**(6):1409–1430.
- [20] Mas-Colell A. (1978) On Revealed Preference Analysis. *The Review of Economic Studies*, **45**(1):121–131.
- [21] Mas-Colell A., Whinston M. and Green J. R. (1995) *Microeconomic Theory* Oxford University Press.
- [22] Matzkin R. and Richter M. (1991) Testing Strictly Concave Rationality. *Journal of Economic Theory*, **53**:287–303.
- [23] Papadimitriou C. H. (2001) Algorithms, Games, and the Internet *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC'01)*.
- [24] Papadimitriou C. H. and Steiglitz K. (1982) *Combinatorial Optimization* Dover Publications inc.
- [25] Richter M. (1966) Revealed Preference Theory. *Econometrica*, **34**(3):635–645.
- [26] Uzawa H. (1960 ) Preference and rational choice in the theory of consumption. In K. J. Arrow, S. Karlin, and P. Suppes, editors, *Mathematical Models in Social Science* Stanford University Press, Stanford, CA.
- [27] Teo C. P. and Vohra R. V. (2003) Afriat's Theorem and Negative Cycles *Working Paper*
- [28] Samuelson P. A. (1948) Consumption Theory in Terms of Revealed Preference *Economica* **15**, 243 – 253.
- [29] Vapnik V. N. (1998) *Statistical Learning Theory* John Wiley & Sons Inc.
- [30] Varian H. R. (1982) The Non-Parametric Approach to Demand Analysis *Econometrica* **50**, 945 – 974.
- [31] Varian H. R. (2005) Revealed Preference, In Michael Szenberg editor, *Samuelson Economics and the 21st Century*.
- [32] Ziegler G. M. (1994) *Lectures on Polytopes* Springer.