# Watch and Learn: Optimizing from Revealed Preferences Feedback

AARON ROTH
University of Pennsylvania
and
JONATHAN ULLMAN
Northeastern University
and
ZHIWEI STEVEN WU
University of Pennsylvania

A *Stackelberg game* is played between a *leader* and a *follower*. The leader first chooses an action, and then the follower plays his best response, and the goal of the leader is to pick the action that will maximize his payoff given the follower's best response. Stackelberg games capture, for example, the following interaction between a retailer and a buyer. The retailer chooses the *prices* of the goods he produces, and then the buyer chooses to buy a utility-maximizing bundle of goods. The goal of the retailer here is to set prices to maximize his profit—his revenue minus the production cost of the purchased bundle. It is quite natural that the retailer in this example would not know the buyer's utility function. However, he does have access to *revealed preference* feedback—he can set prices, and then observe the purchased bundle and his own profit. We give algorithms for efficiently solving, in terms of both computational and query complexity, a broad class of Stackelberg games in which the follower's utility function is unknown, using only "revealed preference" access to it. This class includes the profit maximization problem, as well as the optimal tolling problem in nonatomic congestion games, when the latency functions are unknown. Surprisingly, we are able to solve these problems even though the corresponding maximization problems are not concave in the leader's actions.

Categories and Subject Descriptors: G.1.6 [**Mathematics of Computing**]: Optimization

General Terms: Algorithms, Theory, Economics

## 1. INTRODUCTION

In a *revealed preferences* problem, there is an economic agent and a series of decisions made by that agent. Classically, the goal is to determine whether those decisions are *rationalizable*—i.e., to determine if they can be explained, ex-post, as the result of the agent maximizing some fixed utility function. The revealed preferences problem dates back to the work of Samuelson [Samuelson 1938] and there is a large literature on this topic [Mas-Colell et al. 1995; Rubinstein 2012; Varian 2006]. However, even when the agents' decisions can be rationalized in hindsight as maximizing some fixed utility function, that utility function may not generalize to predict future decisions. This deficiency was first observed by Beigman and Vohra [Beigman and Vohra 2006] and has led to a more recent interest in studying

Authors' addresses: aaroth@cis.upenn.edu, jullman@cs.columbia.edu, wuzhiwei@cis.upenn.edu

the problem of *learning from revealed preferences* [Zadimoghaddam and Roth 2012; Balcan et al. 2014; Amin et al. 2015].

In this note, we describe our work on this topic from [Roth et al. 2015]. We study a family of problems best exemplified by the problem of *maximizing profit using revealed preferences*:

> A retailer, who sells $d$ goods, repeatedly interacts with a buyer. In each interaction, the retailer decides how to price the $d$ goods by choosing $p \in \mathbb{R}^d_{\geq 0}$, and in response the buyer purchases the bundle $x \in \mathbb{R}^d_{\geq 0}$ that maximizes her utility $v(x) - \langle x, p \rangle$, where $v$ is an unknown concave valuation function. The retailer observes the bundle purchased, and therefore his profit, which is $\langle x, p \rangle - c(x)$, where $c$ is a convex cost function. The retailer would like to set prices that maximize his profit after only a small number of interactions with the buyer.

This problem fits naturally into the well studied model of *zeroth-order* or *bandit optimization*—the retailer is trying to maximize an *unknown* objective function given only query access to the function. That is, he can set prices and observe the value of his objective. Unfortunately, when this problem is posed as a bandit optimization problem, the objective function is not concave. This can be seen in the following very simple instance.

EXAMPLE 1. *Consider a setting with one good ($d = 1$). The buyer's valuation function is $v(x) = \sqrt{x}$, and the retailer's cost function is $c(x) = x$. The buyer's utility for buying $x$ units at price $p$ is $\sqrt{x} - xp$. Thus if the price is $p$, a utility-maximizing buyer will purchase $x^*(p) = 1/4p^2$ units. The profit of the retailer is then* $\mathrm{Profit}(p) = px^*(p) - c(x^*(p)) = 1/4p - 1/4p^2$. *Unfortunately, this profit function is not concave.*

Since the retailer's profit function is not concave in the price, it cannot be optimized efficiently using generic methods for concave maximization. Surprisingly, despite this lack of concavity we show that this problem and others can be solved efficiently subject to certain mild conditions.

The key idea is that *the retailer's objective is concave when written as a function of the follower's action,* which we can demonstrate by returning to the previous example.

EXAMPLE 2. *Recall that if the buyer's valuation function is $v(x) = \sqrt{x}$, then when she faces a price $p$, she will buy the bundle $x^*(p) = 1/4p^2$. In this simple case, we can see that setting a price of $p^*(x) = 1/2\sqrt{x}$ will induce the buyer to purchase $x$ units. In principle, we can now write the retailer's profit function as a function of the bundle $x$. In our example, the retailer's cost function is simply $c(x) = x$. So we have* $\mathrm{Profit}(x) = p^*(x) \cdot x - c(x) = \sqrt{x}/2 - x$, *which is concave.*

This example shows that when written in terms of $x$, the profit function is concave! As we show, this phenomenon continues in higher dimensions for arbitrary convex cost functions $c$, and for a wide class of economically meaningful concave valuation functions including the well studied families of CES and Cobb-Douglas utility functions.

Therefore, if the retailer had access to an oracle for the concave function $\text{Profit}(x)$, we could use an algorithm for bandit concave maximization to optimize the retailer's profit. Unfortunately, the retailer does not directly get to choose the bundle purchased by the buyer and observe the profit for that bundle. Instead, he can only set prices to observe the buyer's chosen bundle $x^*(p)$ at those prices and the resulting profit $\text{Profit}(x^*(p))$.

Nevertheless, we have reduced the retailer's problem to a possibly simpler one. In order to find the profit maximizing prices, it suffices to give an algorithm which simulates access to an oracle for $\text{Profit}(x)$ given only the retailer's query access to $x^*(p)$ and $\text{Profit}(x^*(p))$. Specifically, if for a given bundle $x$ the retailer could find prices $p$ such that the buyer's chosen bundle $x^*(p) = x$, then he could simulate access to $\text{Profit}(x)$ by setting prices $p$ and receiving $\text{Profit}(x^*(p)) = \text{Profit}(x)$.

The next key ingredient is a "tâtonnement-like" procedure that efficiently finds prices that approximately induce the buyer to purchase a target bundle $x$ given only access to the retailer's oracle $x^*(p)$. The procedure works as long as the buyer's valuation function is Lipschitz and strongly concave on the set of feasible bundles. Specifically, given a target bundle $x$, our procedure finds prices $p$ such that $|\text{Profit}(x^*(p)) - \text{Profit}(x)| \le \varepsilon$. Thus, we can use our procedure to simulate approximate access to the function $\text{Profit}(x)$. Our procedure requires only $\text{poly}(d, 1/\varepsilon)$ queries to $x^*(p)$. Fortunately, it turns out that recent algorithms for noise tolerant bandit optimization due to [Belloni et al. 2015] can maximize the retailer's profits efficiently even with only approximate access to $\text{Profit}(x)$.

Our tâtonnement procedure was inspired by an elegant recent paper of [Bhaskar et al. 2014], who gave an Ellipsoid-based procedure for the similar problem of finding tolls that induce a target flow in a routing game where the latency functions are unknown. Our procedure is rather general, and applies to a broad family of Stackelberg games in which the leader wishes to optimize his objective function, without knowing what the follower's utility function is, and has access only to observations of the follower's best responses. In particular, our techniques can also be applied to the flow problem studied by [Bhaskar et al. 2014], yielding incomparable guarantees. (Our results hold for a much broader class of latency functions, but our convergence time is slower).

We view our work as one of the first steps in a broader agenda of studying "revealed preferences problems" from a computational perspective. There are many interesting problems in this space, and we will now highlight one. In our profit maximization application, it would be very natural to consider a "Bayesian" version of our problem. At each round the producer sets prices, and then a new consumer, with valuation function drawn from an unknown prior, purchases her utility maximizing bundle. The producer's goal is to find the prices that maximize her *expected* profit, over draws from the unknown prior. Under what conditions can we solve this problem efficiently? The main challenge (and the reason why it likely requires new techniques) is that the *expected value* of the purchased bundle need not maximize any well behaved utility function, even if each individual consumer is maximizing a concave utility function.

## REFERENCES

AMIN, K., CUMMINGS, R., DWORKIN, L., KEARNS, M., AND ROTH, A. 2015. Online learning and profit maximization from revealed preferences. In *Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15)*.

BALCAN, M.-F., DANIELY, A., MEHTA, R., URNER, R., AND VAZIRANI, V. V. 2014. Learning economic parameters from revealed preferences. In *Web and Internet Economics*. Springer, 338–353.

BEIGMAN, E. AND VOHRA, R. 2006. Learning from revealed preference. In *Proceedings of the 7th ACM Conference on Electronic Commerce*. ACM, 36–42.

BELLONI, A., LIANG, T., NARAYANAN, H., AND RAKHLIN, A. 2015. Escaping the local minima via simulated annealing: Optimization of approximately convex functions. *CoRR abs/1501.07242*.

BHASKAR, U., LIGETT, K., SCHULMAN, L. J., AND SWAMY, C. 2014. Achieving target equilibria in network routing games without knowing the latency functions. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*. IEEE Computer Society, 31–40.

MAS-COLELL, A., WHINSTON, M. D., AND GREEN, J. R. 1995. *Microeconomic Theory*. Oxford University Press.

ROTH, A., ULLMAN, J., AND WU, Z. S. 2015. Watch and learn: Optimizing from revealed preferences feedback. *arXiv preprint arXiv:1504.01033*.

RUBINSTEIN, A. 2012. *Lecture notes in microeconomic theory: the economic agent*. Princeton University Press.

SAMUELSON, P. A. 1938. A note on the pure theory of consumers' behavior. *Economica 5,* 17, 61–71.

VARIAN, H. R. 2006. Revealed preference. In *Samuelsonian economics and the twenty-first century*, M. Szenberg, L. Ramrattan, and A. A. Gottesman, Eds. Oxford University Press, 99–115.

ZADIMOGHADDAM, M. AND ROTH, A. 2012. Efficiently learning from revealed preference. In *Internet and Network Economics - 8th International Workshop, WINE 2012, Liverpool, UK, December 10-12, 2012. Proceedings*, P. W. Goldberg, Ed. Lecture Notes in Computer Science, vol. 7695. Springer, 114–127.