Market Equilibrium Models in Large-Scale Internet Markets

Christian Kroer and Nicolas E. Stier-Moses

Abstract Markets and their corresponding equilibrium concepts have traditionally been used as very powerful building blocks to find allocations and prices. This chapter provides examples of the use of Fisher markets in the technology industry. We focus on Internet advertising auctions, fair division problems, content recommendation systems, and robust abstractions of large-scale markets. After introducing these markets, we describe how these models fit the relevant application domains and what insights they can generate, exhibiting the most important theoretical and computational results from the recent literature on these topics.

1 Introduction

Firms in the technology industry often face situations where they must allocate goods to buyers, either literally or figuratively. Among them, internet companies have routinely employed mechanisms centered around auctions because they are robust and allow for changing market conditions and successful price discovery. Some of these mechanisms are 'static' in the sense that the whole market is cleared at once, while others are 'dynamic' meaning that decisions are made on a rolling horizon basis. In financial markets, to sell a newly released bond, potential buyers submit a supply function which specifies how many bonds they are willing to buy at what price. Then, the issuer computes a market-clearing price and use those functions to allocate bonds to buyers. In spectrum auctions, buyers and sellers submit combinatorial bids, and the market maker solves a large mixed integer program (MIP) to find the optimal allocation of spectrum to firms. In electricity markets, market equilibrium is used for pricing electricity in a way that incentivizes suppliers to generate the right amount of electricity. These prices are hard to compute due to non-convexities in the electricity production cost of a supplier (e.g., due to fixed costs of starting production), and integer programming is often used to compute these equilibria. In the technology industry, the volume of transactions and the dynamic nature of its markets make it hard to solve the whole allocation centrally and in one shot. Usually, firms resort to dynamic versions of the market that can be solved in a repeated way. For example, in the internet advertising use cases, an individual auction is run for the ad slots generated when a user interacts with the system. This may be triggered by a

Christian Kroer
Columbia University, 500 West 120th St., New York, New York, USA e-mail: christian.kroer@columbia.edu

Nicolas E Stier-Moses
Facebook Core Data Science, 1 Hacker Way, Menlo Park, California, USA e-mail: nicostier@yahoo.com

search query with keywords, by loading a news article, or by refreshing the feed of a social network. Similarly, there are applications to recommender systems used in the technology industry. In that setting, no explicit market and real money is exchanged. Nonetheless, the allocation of recommendation slots to different content creators can be modeled as a market with 'funny money,' where content creators use their funny-money budgets to optimize their allocation in recommendation slots to users. Another example of a market without money is robust content review problems, where different categories of sensitive content need to be reviewed, and the allocation of reviewing capacity towards the content categories can be modeled as a market allocation problem.

We will focus on the technology industry with its variety of use cases of market-equilibrium based allocations for divisible goods. These kind of market models can be used for a variety of purposes. The most immediate is to find a solution to these markets when the solution is needed and use it to perform the actual allocation. This would be comparable to the bond and spectrum markets mentioned earlier where the transactions are based on the solution to the market model. There are situations in which solving the problem in real-time is not feasible. In those cases, the solution may be computed offline and used as a benchmark. In an ex-post analysis, the firm can judge the merits of the allocation used in practice vis-a-vis the market approach and decide if the online solution approach should or should not be improved in terms of solution quality or computational efficiency. Yet another alternative is to use these models to compute features that can useful to forecast outcomes at the right level of granularity. If one would like to forecast relevant market metrics for next year—e.g., welfare, prices, revenues—running each ad auction individually, given that there might be millions of them per day, does not seem to be the right granularity. Instead, viewing the situation as recurrent realizations of a market for which we can predict the input parameters can provide a better handle to make the right forecast. Finally, another important use case is to evaluate counterfactuals. Having a market model that can deliver predictions allows us to change some of the underlying premises or interactions and find how the solution depends on those changes. An example of this could be to understand how a marketing promotion can provide incentives to advertisers and transform the resulting situation to another equilibrium.

An important factor in common in the use cases above is the need to *compute* solutions in those market models. It is not enough to know that a solution exist, one actually needs the solution itself to operate the system, to forecast it, or to make strategic decisions. To that end, we will discuss algorithmic approaches to solving these models, with a focus on large-scale methods.

To set the stage, we offer a few more details on these ideas by discussing how to use *fair recommender systems* on a job recommendations site. Such a site is a two-sided market. On one side are the users, who view job posts. On the other side, there are the employers creating job posts. Naively, a system might try to simply maximize the number of job posts that users click on, or apply to. This can lead to extremely imbalanced allocations, where a few job posts get a huge number of views and applicants, which is bad both for users and employers. Instead, the system may wish to fairly distribute user views across the many different job posts. To achieve a balance between fair distribution and market efficiency, market-equilibrium-based allocation can be used. In this setting the buyers are the job posts, and the goods are slots in the ranked list of job posts shown to job seekers.

In the next sections, we describe the various models that relate to the main use cases, including ad auctions, recommender systems, and fair division. Then we focus on algorithms and present several ideas that permit solving large-scale models as required by the use cases in the technology industry. This is an expository piece that exhibits existing theoretical research and computational studies done in the setting of internet-scale market applications.

2 Introduction to Fisher Markets and Market Equilibrium

This section introduces the *market equilibrium problem*, the basic modeling element of this chapter. We focus on a particular type of market, usually referred to as *Fisher market*, where there is a set of n buyers that are interested in buying goods from a distinct seller. Every buyer has a budget of $B_i > 0$ dollars. There is a set of m infinitely-divisible goods and each good j has a supply of $s_j > 0$ units that can be divided and sold arbitrarily. We refer to the full supply vector by s.

We will use $x \in \mathbb{R}_+^{n\times m}$ to denote an allocation of goods to buyers, where $x_{ij} \geq 0$ is the amount of good j that is allocated to buyer i. We also denote the bundle of goods given to buyer i as $x_i \in \mathbb{R}^m$ (the i'th row of x). Each buyer has a utility function $u_i(x_i) \mapsto \mathbb{R}_+$ that captures how much they like the bundle x_i. We make the following assumption to avoid degeneracy issues: there exists an allocation x such that $u_i(x_i) > 0$ for all buyers i. This means that it is possible to find an allocation such that all buyers get strictly-positive utilities.

When solving for *market equilibria*, the goal is to find a price $p \in \mathbb{R}_+^m$ for each of the m goods such that the market clears. Clearing the market means that there should exist a feasible allocation x such that every buyer is assigned an optimal allocation given their budgets and the prevailing prices. Formally, the *demand set* of a buyer i with budget B_i finds an optimal bundle under a set of prices by solving

$$D_i(p) = \text{argmax}_{x_i \geq 0} \{u_i(x_i) : \langle p, x_i \rangle \leq B_i\} .$$

A market equilibrium is an allocation-price pair (x, p) such that every buyer gets an optimal bundle and goods are not oversold. Mathematically, that corresponds to $x_i \in D_i(p)$ for all buyers i, and $\sum_i x_{ij} \leq s_j$ for all goods j, where the inequality has to be attained if $p_j > 0$.

Market equilibria have been thoroughly studied and found to have many attractive theoretical properties. One of the most celebrated properties is their *Pareto optimality*: a market equilibrium allocation x satisfies that, for *every* other allocation x', if a buyer is better off under x', then some other buyer is strictly worse off. In other words, x is such that no other allocation can simultaneously (weakly) improve all individuals' utilities. Either all utilities stay the same in other solutions or improving one buyer comes at the expense of another buyer. This is known as *the first fundamental theorem of welfare economics*.

Beyond Pareto optimality, there are several other interesting properties that are verified by market equilibria. They include envy-freeness, where every buyer prefers their own allocation to that of any other after correcting for budget sizes, and proportionality, where every buyer is at least as happy as if they were allocated a fraction of each good proportionally to their budget. These properties will be discussed in more detail when applying market equilibrium to fair division.

2.1 Convex Programming and Utility Functions

A very attractive feature of Fisher markets that make them particularly appealing for modeling purposes is that one can characterize equilibria in computationally efficient ways. Not only this implies that they are guaranteed to exist, but also that they are eminently *computable*, both in theory and in practice. Indeed, there is a nice convex program whose solutions satisfy the market equilibrium properties. Before writing the convex program, let us consider some properties that we would like an optimal allocation x to satisfy. As mentioned before, a feasible allocation necessitates the supply constraints to be respected: $\sum_i x_{ij} \leq s_j$ for all j.

Secondly, since a buyer's demand does not change even if we rescale their valuation by a constant, we require the optimal solution to the convex program to also remain unchanged. Similarly, splitting the budget of a buyer into two separate buyers with the same valuation function should leave the allocation unchanged. These conditions are satisfied by the budget-weighted geometric mean of the utilities:

$$\left(\prod_i u_i(x_i)^{B_i}\right)^{1/\sum_i B_i}.$$

Since taking roots and logs of the objective does not change the set of optimal solutions, we simplify the objective and include the supply constraints to get the so-called EG optimization problem:

$$\begin{aligned} \max_{x\geq 0} \quad & \sum_i B_i \log u_i(x_i) \\ s.t. \quad & \sum_i x_{ij} \leq s_j, \quad j = 1, \ldots, m, \end{aligned} \tag{EG}$$

We denote the dual variables corresponding to each of the supply constraints by p_j. If the utilities in EG are concave and non-negative then this yields a convex program, since composing a concave and nondecreasing function (the log) with a concave function (u_i) yields a concave function. Moreover, if the utilities are concave, continuous, non-negative, and homogeneous (CCNH) then an optimal solution x to EG satisfies the market equilibrium allocation conditions, and the dual variables p provide the equilibrium prices. Formulating the EG program was a seminal idea in the field of market equilibrium computation. It was originally done for linear utilities (which are CCNH) by Eisenberg and Gale [1959]. The general CCNH case was shown by Eisenberg [1961] a few years later. A more modern derivation for differentiable CCNH utilities can be found in Nisan et al. [2007]. For a derivation of the fully general CCNH statement with the more modern formulation of EG, see Gao and Kroer [2020].

Let us review the implications of having the EG formulation. First, it gives us an immediate proof of market equilibrium existence for the CCNH Fisher market setting: the feasible set is clearly non-empty, and the max is guaranteed to be achieved. Second, it allows us to show Pareto optimality directly. A maximizer of EG is indeed Pareto optimal since another solution that simultaneously improves all utilities would be feasible and have a strictly higher objective, contradicting optimality. Third, the optimality of a solution to the EG formulation can also be used to show from first principles that the equilibrium utilities and prices must be unique. If there were more than one allocation at equilibrium, then by the strict concavity of the log function we would get that there is a strictly better solution, which is a contradiction. Thus, the set of equilibrium utilities must be unique. From there it can be seen that equilibrium prices are unique as well, which follows from the EG optimality conditions.

2.2 Classes of Utility Functions

In the previous section we saw that the EG formulation can be used to compute a market equilibrium as long as the utility functions belong to the fairly abstract class CCNH. To understand this class and to provide more context on what is used in practice, we present some concrete examples of its most common types of utilities. To get intuition on the generality of the class, one should primarily consider the homogeneity constraint. Imposing homogeneity disallows many potential utility functions but we will see that it is still a fairly rich class.

The easiest example of a utility function is a *linear utility* $u_i(x_i) = \langle v_i, x_i \rangle$ where $v_i \in \mathbb{R}^m_+$ is a vector of per-good utility rates. It is immediate that linear utilities are CCNH. They can be useful for modeling internet markets—in particular, both ad auctions and fair recommender systems—so they are of special interest to this chapter. More concretely, ad auction models rely on *quasilinear utilities*, a slight variation of linear utilities, where buyers subtract the price that they pay: $u_i(x_i, p) = \langle v_i - p, x_i \rangle$. Technically, this does not fall under the EG framework, since the utility now depends on the prices p. However, it was shown independently by Chen et al. [2007] and Cole et al. [2017] that a small modification to EG can handle quasilinear utilities.

Beyond linear utilities, the next list enumerates the most famous utility classes within CCNH. Let us consider i to be an arbitrary buyer and a_{ij} to be calibration parameters for every good j.

1. Leontief utilities: $u_i(x_i) = \min_j \frac{x_{ij}}{a_{ij}}$,
2. Cobb-Douglas utilities: $u_i(x_i) = \prod_j (x_{ij})^{a_{ij}/(\sum_j a_{ij})}$,
3. Constant elasticity of substitution (CES) utilities: $u_i(x_i) = \left(\sum_j a_{ij} x_{ij}^{\rho}\right)^{1/\rho}$, where ρ is another calibration parameter, with $-\infty < \rho \leq 1$ and $\rho \neq 0$.

CES utilities turn out to be the most general so far: Leontief utilities are obtained as ρ approaches $-\infty$, Cobb-Douglas utilities as ρ approaches 0, and linear utilities when $\rho = 1$. More generally, $\rho < 0$ implies that goods are complements, whereas $\rho > 0$ implies that goods are substitutes.

An interesting consequence of the existence of the EG formulation is that various natural iterative economic processes converge to a Fisher market equilibrium. This is because many such processes are formally equivalent to some form of iterative first-order optimization on the EG program. For example, various *tâtonnement* algorithms converge to a Fisher market equilibrium. A tâtonnement process is an iterative dynamic where a market operator repeatedly announces prices p_t at each time t, each buyer i reports their demand x_i^t under the given prices, and the market operator increases the price of over-demanded goods and decreases those of under-demanded ones. This can be reinterpreted as subgradient descent on the dual convex program of EG.

Other interesting dynamics based on the EG formulation also exist. Perhaps the most important one is the *proportional response* process, where buyers iteratively specify how much they wish to spend on each good, and the market operator sets the prices to the sum of these spends. This dynamic turns out to perform extremely well in practice, and we will review it in detail later. This was discovered by Wu and Zhang [2007] when analyzing bit-torrent sharing dynamics, and [Birnbaum et al., 2011] later gave a surprising convergence guarantee based on a mirror-descent equivalence.

3 Auction Markets and Budget Management Systems

Advertisers participate in internet ad markets to get impressions, clicks, or conversions of ads that are placed in content shown to users by the platform. To accomplish this, advertisers set up ad campaigns that indicate how much they are willing to bid in exchange for those events. Since the values per conversion are unknown to the platform, in the last decade, platforms turned to computing allocations and prices by running an auction every time a user shows up, and the competition in these auctions gave rise to ad auction markets. In addition to values, ad campaigns usually specify budget or ROI (return-on-investment) constraints. This allows them to control their total spend and to maximize the value they get out of the system while guaranteeing that they do not exceed the maximum amount of money that they are willing to spend. Bidding is typically performed by a *proxy bidder*, operated by the platform but acting on behalf of a given advertiser. This proxy bidder attempts to maximize the utility derived by the advertiser, while taking into account the specified constraints.

When designing the market mechanism and the corresponding proxy bidders, the platform needs to provide tools to allow advertisers to run ad campaigns that are effective. One of the issues that arises is that budgets and bids in a campaign may not necessarily be in agreement with each other. In light of that problem, a platform may offer ways to compute alternative campaign parameters that align budgets and bids. The two fundamental budget management systems that are dominant in practice and in the literature are (a) throttling, which uses a feedback loop to limit the number of auctions an ad participates in, and (b) pacing, which uses a feedback loop to shade bids. We provide more details about these mechanisms below.

The purpose of this section is to illustrate how market models and their equilibria can be used as a tool to understand tradeoffs in auction markets and budget management systems. We will focus on systems based on pacing mechanisms, since that is one of the dominant budget management methods used in practice, and these systems are particularly amenable to analysis via Fisher market models. As a simplification, we assume that each individual auction allocates a single good, usually referred to as a 'slot.' This is a simplification that allows us to model the repeated auctions as a quasi-linear Fisher market, and hence make available all the theory and results that apply to their equilibria. In practice, it is common for platforms to simultaneously auction several *impression opportunities* (slots to be filled with ads) in real time when they display a page or refresh a feed.

It is important to highlight that in the real-time operation of a platform, instead of relying on market equilibria as considered here, they typically rely on control algorithms which tune the parameters used by the proxy bidder to align the advertisers' campaign parameters. (The parameters may include the pacing multiplier which is relevant to our model, but in other implementations they may include participation probabilities for throttling campaigns.) The market equilibria that we describe in this section can be thought of as the desired steady state of the system. In practice, the control algorithms need to learn these parameters in an online fashion. Budget constraints and other pacing aspects invalidate traditional guarantees such as the strategyproofness of second price auctions.

We analyze the pacing equilibrium problem that results from the pacing system when the underlying allocation is produced by either a second or first price auction, in that order. By reinterpreting this problem as a game where players choose pacing parameters, we connect the equilibria of those games to solutions to suitable Fisher markets. After the static analysis, we also discuss the effects of adding temporal considerations to the model to get a dynamic auction market. This more closely parallels how campaigns are tuned in practice. We will see that the static game representation provides a good approximation that can be used as a starting point of dynamic procedures.

3.1 Market Models for Pacing Systems

We define an *auction market* similarly to a Fisher market (Section 2). We consider a set N of n buyers and a set M of m goods. Buyer i has value $v_{ij} \geq 0$ for good j, and each buyer has a budget $B_i > 0$. Each good j will be sold by itself in a sealed-bid auction, using either the first or second price as a payment rule. To disregard trivial cases, we assume that for all buyers i, there exists some good j such that $v_{ij} > 0$, and for all goods j there exists i such that $v_{ij} > 0$. Let $x \in \mathbb{R}_+^{n\times m}$ be an allocation of goods to buyers, with associated prices $p \in \mathbb{R}_+^m$. The utility that a buyer i derives from this allocation is

$$u_i(x_i, p) = \begin{cases} \langle v_i, x_i \rangle - \langle p, x_i \rangle & \text{if } \langle p, x_i \rangle \le B_i\,, \\ -\infty & \text{otherwise}\,. \end{cases} \tag{1}$$

We will use the abbreviations SP and FP for second and first price auctions markets, respectively.

Although auctions have several appealing properties when considered individually, budgets add a coupling constraint across auctions that influences those properties. For instance, it is well known that second price auctions in isolation are strategyproof, but the following example shows that second price auction markets with budgets are not: Consider an instance with two buyers and two goods, with valuations $v_1 = (100, 100)$, $v_2 = (1, 1)$ and budgets $B_1 = B_2 = 1$. If both buyers submit their true valuations then buyer 1 wins both goods, pays 2, and gets $-\infty$ utility. To fix this problem, each buyer needs to smooth out their spending across auctions to make sure that they remain within budget.

For large-scale internet auctions the smoothing is frequently achieved via *budget management systems* as mentioned at the beginning of this section. The following two mechanisms are widely used in practice by platforms. In both, each buyer i (or proxy bidder, as it may be) has to tune a parameter $\alpha_i \in [0, 1]$.

1. *Probabilistic throttling*: The parameter α_i encodes the probability that the buyer participates in each auction. For each auction j, an independent coin is flipped for buyer i. If it comes up heads (with probability α_i) then the buyer participates in the auction with a bid $b_{ij} = v_{ij}$. Otherwise the buyer is excluded from that particular auction.
2. *Multiplicative pacing*: The parameter α_i acts as a scalar multiplier on the reported bids from the advertiser. For each auction j, buyer i submits a bid $b_{ij} = \alpha_i v_{ij}$.

Figure 1 illustrates these options under second price auctions, in a simplified setting. For ease of presentation, the figures plot the opportunities in terms of time along the x-axis, even though these market abstractions are static. Time is inconsequential in this section, but we will revisit time in depth when we address dynamics in Section 3.2. We consider a focal buyer whose value is constant in all auctions and hence the bids are constant across them. The buyer participates in auctions as long as some budget remains, and then participation stops. Competition arising from other buyers present in the auctions cause resulting prices, plotted in the y-axis, to vary in different auctions. Since the figure represents second price auctions, the price in each auction is not necessarily the same, even though the focal buyer bids a single fixed amount. The circles represent the participation opportunities of the focal buyer and the shaded ones represent the auctions in which the focal buyer won.

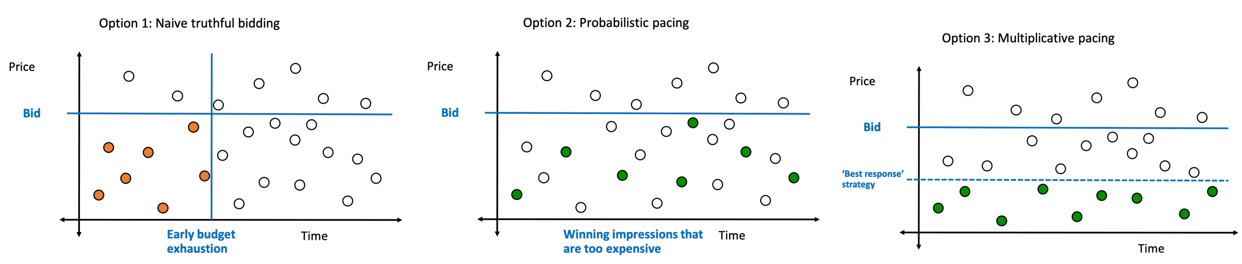

Fig. 1 Comparison of budget management systems. *Left*: no budget management, *middle*: probabilistic throttling, *right*: multiplicative pacing.

The left panel shows the outcome if the budgets are not managed and buyers bid naively: the focal buyer spends the budget too fast, and ends up running out of money prematurely. There are many low-price and high-value goods to the right of the budget exhaustion line that the buyer cannot get. This is a lost opportunity for the buyer. Furthermore, in practice, buyers tend to prefer to smoothly spend their budget throughout the day as opposed to running out of money long before the end of the planning window. The middle panel shows the effect of probabilistic throttling for an appropriately chosen parameter α_i. Buyers only participate in some auctions, allowing them to continue to have a remaining budget until the end of the planning horizon. As before, buyers end up winning some expensive auctions, while missing out on cheaper ones. From the buyer's perspective this is still sub-optimal in terms of utility, since all goods have the same value to the buyer. Finally, the right panel shows the effect of multiplicative pacing for an appropriate value of a pacing multiplier α_i. In this case, the buyer bids optimally in the many auctions, and is able to extract maximum value from their budget by buying the right set of goods. Note that the buyer ends up buying all goods over a certain bang-per-buck threshold (this holds in general for second price markets, if we allow the buyer to get a fraction of a good to reach their budget constraint exactly).

Other budget management systems discussed by Balseiro et al. [2017, see Table 1] include thresholding, reserve pricing, and multiplicative boosting. All these mechanisms work by modifying the participation, bidding or payment rules. For example, thresholding requires the buyer's bids to pass a given threshold to participate, thus forcing buyers to only bid on high-value goods. Reserve pricing is similar, except that the threshold is also used to compute the resulting winning price.

In this section we focus on static models of budget management systems, where the set of goods and values are known ahead of time. One advantage of this perspective is that we can model highly structured valuations across goods. On the other hand, it ignores the stochastic nature associated with impressions that arrive across a day. Several related papers consider goods that arrive stochastically and valuations are then drawn independently. For instance, Balseiro et al. [2015] show that when buyers get to select their bids in each individual auction, a multiplicative pacing equilibrium arises naturally via Lagrangian duality on the budget constraint, under a fluid-based mean-field market model. Balseiro et al. [2017] show the existence of pacing equilibrium for multiplicative pacing as well as the other pacing rules mentioned earlier in a stochastic model with independent valuations. They also give a very interesting comparison of revenue and social welfare properties of the various pacing mechanisms in the unique symmetric equilibrium of their setting. One of the main insights they provide is that multiplicative pacing achieves strong social welfare properties, while probabilistic throttling achieves higher revenue properties.

3.1.1 Second Price Auction Markets

We now explore the case of multiplicative pacing in a static market model with second price auctions. We follow the treatment in Conitzer et al. [2018], and direct the reader there for details, proofs and additional references. In the historical notes at the end of the section, we include additional references to papers that discuss probabilistic throttling.

In the right panel of Figure 1 we have seen that the focal buyer can optimize its utility by selecting a fixed shaded bid that depends on the total budget. The intuition that a buyer in a repeated auction setting should bid according to a single scalar times their valuations can be shown to hold more generally, even when goods have different values. Specifically, for a given set of bids by all the other buyers, a buyer can always specify a best response by choosing an optimal, constant pacing multiplier. The resulting bid for the buyer on a particular good would be the value of the good in that auction times the fixed pacing multiplier.

Theorem 1 *For arbitrary but fixed bids in each auction for buyers $k \neq i$, buyer i has a best response that consists of multiplicatively-paced bids. This assumes that if a buyer is tied for winning an auction, they can choose the fraction* *that they want to win. This holds even if the buyer (or proxy bidder) can dispose of some goods that they win, in order to avoid exceeding their budget.*

The previous result takes the perspective of a best response for an individual buyer. The main question we now wish to address is what happens at equilibrium when all buyers play best responses to each others' bids. We refer to such an outcome as a *pacing equilibrium*.

Definition 1 A second price pacing equilibrium (SPPE) is a vector of pacing multipliers $\alpha \in [0,1]^n$, a fractional allocation x_{ij}, and a price vector p that satisfies the following properties.

(Goods go to highest bidders) If $x_{ij} > 0$, then $\alpha_i v_{ij} = \max_{i' \in N} \alpha_{i'} v_{i'j}$ for each buyer $i \in N$ and good $j \in M$.
(Prices) The unit price of good $j \in M$ is $p_j = \max_{k \neq i} \alpha_k v_{kj}$ for any buyer $i \in N$ such that $x_{ij} > 0$.
(Budget-feasible) $\sum_{j \in M} x_{ij} p_j \leq B_i$ for each buyer $i \in N$.
(No unnecessary pacing) Additionally, if the budget inequality is strict then $\alpha_i = 1$.
(Demanded goods sold completely) $\sum_{i \in N} x_{ij} = 1$ for each good $j \in M$.

The conditions above enforce that winning bids get the goods and buyers pay the second price. The *no unnecessary pacing* condition comes from the practical consideration that buyer should only be paced if their budget constraint is binding. It is basically a complementarity condition that specifies that the mechanism does not want to pace buyers unless it has to. It follows (almost) immediately from Theorem 1 that in an SPPE every buyer is best responding.

Notice that the equilibrium not only includes the pacing multiplier but also the allocations. This is because there may be multiple winning bids for a given good j, and in that case the good may be split among the winning bids, such that each buyer hits their budget constraint exactly. This inclusion of the allocations as part of an SPPE makes it slightly different from a game-theoretic Nash equilibrium. More concretely, we can almost view the problem of finding an SPPE as a pure Nash equilibrium problem in terms of a pacing game that can be defined by the set of pacing multipliers. However, because we must specify the allocation as well, the resulting problem becomes more akin to a market equilibrium (in fact there are strong equivalences between SPPE and market equilibrium, as we shall see later). Nonetheless, it is also possible to formulate a static game with full information such that its pure Nash equilibria and the pacing multipliers α of an SPPE are in one-to-one correspondence. We refer the interested reader to Conitzer et al. [2018] for the details of this.

Importantly SPPE, as defined above, are always guaranteed to exist. This does not follow immediately from previous results such as the existence of Nash equilibria in a standard game. SPPE correspond to a specific type of pure-strategy Nash equilibria and the existence must be explicitly proved.

Theorem 2 *An SPPE of a pacing game is always guaranteed to exist.*

To illustrate this result, we include a quick sketch of the main elements of the proof. First, one constructs a smoothed pacing game, where the allocation is smoothly shared among all bids that are within ϵ of the maximum bid. This makes the allocation a deterministic function of the pacing multipliers α. Several other smooth approximations are also introduced to deal with other discontinuities. In the end, one gets a game in which each player simply has the interval $[0,1]$ of pacing multipliers as the action space and utilities are well-behaved continuous and quasi-concave functions. For this smoothed game, one can then appeal to a fixed-point theorem to guarantee the existence of a pure-strategy Nash equilibrium in the smoothed game. Finally, the limit point of smoothed games as the smoothing factor ϵ tends to zero yields an equilibrium in the original pacing equilibrium problem.

Unfortunately, while an SPPE is guaranteed to exist, there may be multiple solutions. Moreover, they can have large differences in revenue, social welfare, and other relevant statistics of interest. Figure 2 shows an example of this where the total platform revenue can be orders of magnitude different in two different SPPE. In practice this means that we might need to select the equilibrium that suits ours needs, instead of just solving for one. Although multiplicity of equilibria is a possibility, through simulations one can see that it is not a very common occurrence when looking at instances inspired by real-world ad markets.

Given the practical motivation of the use of market equilibria to understand, manage and forecast ad markets, one may want to actually compute SPPE for a given instance. For instance, the resulting pacing multipliers may be used to shade buyers' bids and drive the system to an operating point in which buyers do not have an incentive to adjust bids further. Although the computational complexity of finding an arbitrary SPPE is open, finding an extremal one (e.g., minimizing/maximizing revenue/social welfare) can be proved to be NP complete. Nevertheless, all the equilibrium conditions can be written as linear constraints with mixed-integer variables, leading to a mixed integer programming (MIP) formulation in which feasibility is equivalent to being at equilibrium. The formulation can be augmented with an objective function of interest to optimize among equilibria.

This formulation can be used to compute equilibria for modestly-sized instances, but as a method it is not very scalable. Instead, we can map SPPE to more general market equilibria to unlock more efficient methods.

Problem instance:

i	v_{i1}	v_{i2}	v_{i3}	v_{i4}	B_i
1	100	1	99	100	1
2	1	100	99		1
3				100	100

Equilibrium 1: Revenue = 102

α_i	b_{i1}	b_{i2}	b_{i3}	b_{i4}	spend
1	100	1	99	100	1
0.01	0.01	1	0.99		1
1				100	100

Equilibrium 2: Revenue = 3

α_i	b_{i1}	b_{i2}	b_{i3}	b_{i4}	spend
0.01	1	0.01	0.99	1	1
1	1	100	99		1
1				100	1

Fig. 2 Multiplicity of equilibria. *Left*: a problem instance. *Right*: two possible and very different SPPE.

To put SPPE in perspective, they can be seen as market equilibria, considering a market setting where each buyer has a quasi-linear demand function $D_i(p) = \text{argmax}_{0 \le x_i \le 1} u_i(x_i, p)$, where u_i was defined in (1). This characterization follows immediately by simply using the allocation x and prices p from the SPPE as a market equilibrium. Theorem 1 tells us that $x_i \in D_i(p)$, and the market clears by definition of SPPE. This implies that SPPE have several useful properties including no envy and Pareto optimality (if one considers the seller as a participant too). This yields the interesting guarantee that, in a budget-adjusted sense, no buyer prefers the allocation of any other buyer, given the prices.

3.1.2 First Price Auction Markets

We now switch to first price auctions in the context of pacing equilibria. Every other aspect of the definition of the market is the same as for SPPE. First price auctions were used initially in internet ad auctions in the 1990's, for example by Yahoo and others. But incentive and stability issues caused a shift to second price auctions. However, first price auctions have seen a recent resurgence of interest for these markets. Notably, several major ad exchanges switched to first price in recent years. For instance, Google Ad Manager switched in September 2019, while Twitter's MoPub exachange switched in August 2020. A major motivation cited by both exchanges is that a first price mechanism will increase transparency and reduce complexity. The incentive and stability issues observed in the 1990's are likely to be less of an issue in today's thicker and much larger-scale markets. We will see that, in the context of repeated auctions, a mechanism that relies on first price repeated auctions has several desirable properties. See also Paes Leme et al. [2020] for an interesting analysis in which firms endogenously choose first price. Our treatment here follows the work of Conitzer et al. [2019]; we refer the reader to that article for additional insights, results and proofs.

To build towards market equilibria, we start by defining *budget-feasible pacing multipliers*, which guarantee that buyers stay within budget for goods that are allocated according to first price auction rules.

Definition 2 A set of *budget-feasible first price pacing multipliers* (BFPM) is a vector of pacing multipliers $\alpha \in [0,1]^n$ and a fractional allocation $x_{ij} \in [0,1]$ that satisfies the following properties:

(Goods go to highest bidders) If $x_{ij} > 0$, then $\alpha_i v_{ij} = \max_{i' \in N} \alpha_{i'} v_{i'j}$ for each buyer $i \in N$ and good $j \in M$.
(Prices) The unit price of good j is $p_j = \max_{i \in N} \alpha_i v_{ij}$ for each good $j \in M$.
(Budget-feasible) $\sum_{j \in M} x_{ij} p_j \leq B_i$ for each buyer $i \in N$.
(Demanded goods sold completely) If $p_j > 0$, then $\sum_{i \in N} x_{ij} = 1$ for each good $j \in M$.
(No overselling) $\sum_{i \in N} x_{ij} \leq 1$ for each good $j \in M$.

To define a pacing equilibrium in the case of first price auctions, we take a BFPM and also impose a complementarity condition between the budget constraint and the pacing multiplier. This guarantees that buyers cannot be paced unless they spend their entire budget.

Definition 3 A *first price pacing equilibrium* (FPPE) is a BFPM (α, x) that also verifies the *no unnecessary pacing* condition, which means that if $\sum_{j \in M} x_{ij} p_j < B_i$, then $\alpha_i = 1$ for each buyer $i \in N$.

The only difference between an FPPE and an SPPE is the pricing condition, which now uses first price.

A very nice property of the first price setting is that BFPMs satisfy a monotonicity condition: if (α', x') and (α'', x'') are both BFPM, then the pacing vector $\alpha = \max(\alpha', \alpha'')$, where the max is taken componentwise, is also a BFPM. The associated allocation is that for each good j, we first identify whether the highest bid comes from α' or α'', and use the corresponding allocation of j (breaking ties towards α').

Intuitively, the reason that (α, x) is also BFPM is that for every buyer i, their bids are the same as in one of the two previous BFPMs (say (α', x') WLOG.), and so the prices they pay are the same as in (α', x'). Furthermore, since every other buyer is bidding at least as much as in (α', x'), they win weakly less of each good (using the tie-breaking scheme described above). Since (α', x') satisfied budgets, (α, x) must also satisfy budgets. The remaining conditions are easily checked.

In addition to componentwise maximality, there is also a *maximal* BFPM (α, x) (there could be multiple x compatible with α) such that $\alpha \geq \alpha'$ for all α' that are part of any BFPM. Consider $\alpha_i^* = \sup\{\alpha_i | \alpha \text{ is part of a BFPM}\}$. For any ϵ and i, we know that there must exist a BFPM such that $\alpha_i > \alpha_i^* - \epsilon$. For a fixed ϵ we can take componentwise maxima to conclude that there exists $(\alpha^\epsilon, x^\epsilon)$ that is a BFPM. This yields a sequence $\{(\alpha^\epsilon, x^\epsilon)\}$ as $\epsilon \to 0$. Since the space of both α and x is compact, the sequence has a limit point (α^*, x^*). By continuity (α^*, x^*) is a BFPM.

We can use this maximality to show existence and uniqueness (of multipliers) of FPPE:

Theorem 3 *An FPPE always exists and the set of pacing multipliers $\{\alpha\}$ that are part of an FPPE is a singleton.*

To prove this one can consider the component-wise maximal α and an associated allocation x such that they form a BFPM and show that it has no unnecessarily paced buyers. This follows from supposing that some buyer i is spending strictly less than B_i and $\alpha_i < 1$ and deriving a contradiction to the maximality of the pacing multipliers. Uniqueness follows from the component-wise maximality and the *no unnecessary pacing* condition.

Sensitivity Analysis

FPPE enjoy several nice monotonicity and sensitivity properties that SPPE do not. Several of these follow from the maximality property that we have seen earlier: the unique FPPE multipliers α are such that $\alpha \geq \alpha'$ for any other BFPM (α', x'). The following actions are all guaranteed to weakly increase the revenue at equilibrium.

Adding a buyer $n+1$. The original solution (α, x) together with $\alpha_{n+1} = 0, x_{n+1} = 0$ is a BFPM of the expanded instance. By the monotonicity property, prices must weakly increase.

Adding a good. The FPPE of the expanded instance α' satisfies $\alpha' \leq \alpha$. (To see this, consider the set of buyers whose multipliers increased, since they win more goods and prices are up, some buyer must strictly exceed their budget, a contradiction). The set of buyers $i \in N$ such that $\alpha'_i < \alpha_i$ must be spending their whole budget by the *no unnecessary pacing* condition. For buyers such that $\alpha'_i = \alpha_i$, they pay the same as before, and win weakly more goods.

Increasing a buyer i's budget. The original solution (α, x) is a BFPM in the updated instance. By the maximality of the FPPE solution, its multipliers must be larger.

It is also possible to show that revenue enjoys a Lipschitz property: increasing a single buyer's budget by Δ increases revenue by at most Δ. Similarly, social welfare can be bounded in terms of Δ, though multiplicatively, and it does not satisfy monotonicity.

Convex Program to Compute FPPE

As discussed earlier, besides theory, the motivation of formulating pacing systems as markets is to provide algorithms to compute them. Computing an FPPE turns out to be easier than an SPPE since we do not need to rely on an integer program. This is due to a direct relationship between pacing and market equilibria. FPPE are given exactly by the set of solutions to the *quasi-linear* variant of the Eisenberg-Gale convex program for computing a market equilibrium:

$$\begin{aligned} \max_{x \geq 0, \delta \geq 0, u} \quad & \sum_i B_i \log(u_i) - \delta_i \\ s.t. \quad & u_i \leq \sum_j x_{ij} v_{ij} + \delta_i, \quad i \in N \qquad (2) \\ & \sum_i x_{ij} \leq 1, \quad j \in M \qquad (3) \end{aligned}$$

$$\begin{aligned} \min_{p \geq 0, \beta \geq 0} \quad & \sum_j p_j - \sum_i B_i \log(\beta_i) \\ s.t. \quad & p_j \geq v_{ij} \beta_i, \quad i \in N, j \in M \\ & \beta_i \leq 1, \quad i \in N \end{aligned}$$

We show the primal convex program on the left and its corresponding dual convex program on the right. The variables

x_{ij} denote the fractional amount of good j that buyer i wins. The leftover budget is captured by δ_i, which is the primal variable corresponding to the dual constraint $\beta_i \leq 1$.

The dual variables β_i and p_j correspond to constraints (2) and (3), respectively. The variable p_j is the price of good j and $\beta_i = \min_{j:x_{ij}>0}\{p_j/v_{ij}\}$ can be interpreted as the inverse bang-per-buck for buyer i. With this definition of β_i, the constraint $\beta_i \leq 1$ is intuitively clear: a quasi-linear buyer only wishes to spend money if their price-per-utility is at most 1.

One can show via Karush-Kuhn-Tucker conditions that FPPE and EG are equivalent. Informally, the correspondence between them follows because β_i specifies a single price-per-utility rate per buyer which exactly yields the pacing multiplier $\alpha_i = \beta_i$. Complementary slackness then guarantees that if $p_j > v_{ij}\beta_i$ then $x_{ij} = 0$, so any good allocated to i has rate β_i exactly. Similarly, complementary slackness on $\beta_i \leq 1$ and the associated primal variable δ_i guarantee that buyer i is only paced if they spend their whole budget.

Theorem 4 *An optimal solution to the quasi-linear Eisenberg-Gale convex program corresponds to an FPPE with pacing multiplier $\alpha_i = \beta_i$ and allocation x_{ij}, and vice versa.*

It follows that an FPPE can be computed in polynomial time, and that we can apply various first-order methods to compute large-scale FPPE. Such first-order methods will be discussed in Section 5.2.

3.1.3 Comparison between SPPE and FPPE

The SPPE and FPPE properties have interesting differences, which we summarize in Table 1. For additional details, see the literature in the historical notes at the end of the section. FPPE are unique (this can be shown from the convex program, or directly from the monotonicity property of BFPM) while SPPE are not. In practice SPPE instances admitting multiple equilibria seem rare. FPPE can be computed in polynomial time. While the complexity of SPPE is unknown, it is NP-hard to maximize revenue or social welfare. FPPE are robust to perturbations (e.g., revenue increases smoothly as budgets are increased). Both equilibrium concepts correspond to (different) market equilibria but SPPE requires buyer demands to be "supply aware." SPPE correspond to a pure-strategy Nash equilibria, and thus buyers are best responding to each other. Neither FPPE nor SPPE are strategyproof, but the market equilibrium connection can be used to show strategyproofness in an appropriate "large market" sense.

As we will discuss in Section 3.3, FPPE and SPPE have also been studied experimentally, using instances generated from real ad auction data. To complete the comparison, we report the most interesting takeaways from those experiments here:

- Manipulation is hard in both SPPE and FPPE if you can only lie about your value-per-click.
- FPPE dominates SPPE on revenue.
- Social welfare can be higher in either FPPE or SPPE. Experimentally, there was not a clear winner of which of the two solution concepts provides a higher social welfare.

3.2 Dynamic Budget Management Systems

The previous section explored a market with repeated auctions viewed as a static game between advertisers that set pacing multipliers. Since that view ignores time, this section presents a dynamic view, where a buyer or a proxy bidder

	SPPE	FPPE
Exists?	Yes	Yes
Is unique?	No	Yes, up to buyer utilities
Is efficiently computable?	NP-hard	Convex program
Is welfare monotone?	No	Yes, in goods
Is revenue monotone?	No	Yes, in goods/buyers/budgets
Is shill proof?	No	Yes
Pacing eq. is best response?	Yes	No
Simulated revenue	SPPE ≤ FPPE	
Simulated welfare	Ambiguous	

Table 1 A comparison of FPPE and SPPE.

has to sequentially tune its pacing multiplier to manage their bids over time. The goal is to hit the 'right' pacing multiplier as before and match the spend and the budget, but each buyer has to learn that multiplier as the market plays out. We will see how to approach this problem using ideas from regret minimization. The exposition closely follows the work of Balseiro and Gur [2019].

3.2.1 Dynamic Auctions Markets

In this section we consider second-price mechanisms with n buyers who participate in auctions sequentially at times $t = 1, \dots, T$. At time t an auction takes place and each buyer samples a valuation v_{it} independently from a cumulative distribution (CDF) function F_i supported in $[0, \bar{v}_i]$ which is assumed to be absolutely continuous and with bounded density f_i. We use the vector notation v_i to denote the sequence of realized valuations across all auctions. As earlier, we assume that each buyer has some budget B_i to be spent across all auctions. We define by $\rho_i = B_i/T$ the per-period target expenditure, which we assume to be bounded by $\bar{v}_i$. Each buyer is characterized by a type $\theta_i = (F_i, \rho_i)$.

After realizing the valuation v_{it}, buyer i submits a bid b_{it}. We let $d_{it} = \max_{k \neq i} b_{kt}$ denote the highest bid other than that of i, and we use the vector notation d_i to refer to the sequence across all auctions. The buyers' utilities continue to be quasi-linear: they receive a utility of $u_{it} = \mathbb{1}\{d_{it} \leq b_{it}\}(v_{it} - d_{it})$, where the first term is an indicator function that equals one if buyer i wins auction t, and pay $z_{it} = \mathbb{1}\{d_{it} \leq b_{it}\}d_{it}$.[1]

We assume that each buyer has no information on the valuation distributions, including their own. Instead, they just know their own target spend rate ρ_i (i.e., spend per time period) and the total number of time periods T. Buyers also do not know how many other buyers are in the market. At time t, buyer i knows the *history* $(v_{i\tau}, b_{i\tau}, z_{i\tau}, u_{i\tau})_{\tau=1}^{t-1}$ of their own values, bids, payments, and utilities. Furthermore, they know their current value v_{it}. Based on this history, they choose a bid b_{it}. We will say that a bidding strategy for buyer i is a sequence of mappings $\beta = \beta_1, \dots$ where β_t maps the current history to a bid (potentially in randomized fashion). The strategy β is budget feasible if the bids b_{it}^{β} generated by β are such that $\sum_{t=1}^{T} \mathbb{1}\{d_{it} \leq b_{it}^{\beta}\}d_{it} \leq B_i$ under any vector of highest competitor bids d_i. For given d_i and valuation vectors v_i, we denote the expected value of a strategy β as

$$\pi_i^{\beta}(v_i, d_i) = \mathbb{E}_{\beta}\left[\sum_{t=1}^{T} \mathbb{1}\{d_{it} \leq b_{it}^{\beta}\}(v_{it} - d_{it})\right],$$

[1] In this case with continuous distributions, the probability of ties is zero.

where the expectation is taken with respect to randomness in β.

We would like to compare this outcome to the *hindsight optimal* strategy. We denote the expected value of that strategy as

$$\begin{aligned} \pi_i^H(v_i, d_i) := \max_{x_i \in \{0,1\}^T} \; & \sum_{t=1}^{T} x_{it}(v_{it} - d_{it}) \\ s.t. \; & \sum_{t=1}^{T} x_{it} d_{it} \le B_i \,. \end{aligned} \tag{4}$$

The hindsight optimal strategy has a simple structure: a buyer simply chooses the optimal subset of goods to win while satisfying the budget constraint. In the case where the budget constraint is binding, this is a knapsack problem.

Ideally we would like to consider a strategy π_i^β that approaches π_i^H. However, this turns out not to be possible. We will use the idea of asymptotic γ-competitiveness to see this. Formally, β is asymptotically γ-competitive if

$$\limsup_{\substack{T\to\infty,\\ B_i=\rho_i T}} \sup_{\substack{v_i \in [0,\bar{v}_i]^T,\\ d_i \in \mathbb{R}_+^T}} \frac{1}{T}\left(\pi_i^H(v_i, d_i) - \gamma \pi_i^\beta(v_i, d_i)\right) \le 0\,.$$

Intuitively, the condition says that asymptotically, β should achieve at least $1/\gamma$ of the hindsight-optimal expected value.

For any $\gamma < \bar{v}_i/\rho_i$, asymptotic γ-competitiveness turns out to be impossible to achieve. Thus, if our target expenditure ρ_i is much smaller than our maximum possible valuation, we cannot expect to perform anywhere near as well as the hindsight optimal strategy. The general proof of this fact is quite involved, but the high-level idea is not too complicated. We show the construction for $\bar{v}_i = 1, \rho_i = 1/2$, and thus the claim is that $\gamma < \bar{v}_i/\rho_i = 2$ is unachievable. The impossibility is via a worst-case instance. In this instance, the highest other bid comes from one of the two following sequences:

$$\begin{aligned} d^1 &= \left(d_{high}, \ldots, d_{high}, \bar{v}_i, \;\ldots\ldots\;, \bar{v}_i\right) \\ d^2 &= \left(d_{high}, \ldots, d_{high}, d_{low}, \ldots, d_{low}\right), \end{aligned}$$

for $\bar{v}_i \ge d_{high} > d_{low} > 0$. The general idea behind this construction is that in the sequence d^1, buyer i must buy many of the expensive goods to maximize their utility, since they receive zero utility for winning goods with price $\bar{v}_i$. However, in the sequence d^2, buyer i must save money so that they can buy the cheaper goods priced at d_{low}.

For the case we consider here, there are $T/2$ of each type of highest other bid (assuming that T is even for convenience). Now, we may set $d_{high} = 2\rho_i - \epsilon$ and $d_{low} = 2\rho_i - k\epsilon$, where ϵ and k are constants that can be tuned. For sufficiently small ϵ, buyer i can only afford to buy a total of $T/2$ goods, no matter the combination they get. Furthermore, buying a good at price d_{low} yields k times as much utility as buying a good at d_{high}.

To achieve at least half of the optimal utility under d^1, buyer i must purchase at least $T/4$ of the goods priced at d_{high}. Since they do not know whether d^1 or d^2 occurred until after deciding whether to buy at least $T/4$ of the d_{high} goods, this must also occur under d^2. But then buyer i can at most afford to buy $T/4$ of the goods priced at d_{low} when they find themselves in the d^2 case. Finally, for any $\gamma < 2$, we can pick k and ϵ such that achieving $\gamma\pi_i^H$ requires buying at least $T/4 + 1$ of the d_{low} goods.

It follows that we cannot hope to design an online algorithm that competes with $\gamma\pi_i^H$ for $\gamma < \bar{v}_i/\rho_i$. However, it turns out that a subgradient descent algorithm can achieve exactly $\gamma = \bar{v}_i/\rho_i$

3.2.2 An Adaptive Pacing Strategy

In this section, we present a pacing strategy that optimizes the pacing multipliers by adjusting them over time. We consider a focal buyer $i \in N$ for whom we set $\alpha_i = \frac{1}{1+\mu}$ and iteratively tune it by running a subgradient descent scheme on the value for μ, which will allow the buyer to smoothly spend the budget across the T time periods.

The algorithm takes as input a step size $\epsilon_i > 0$ and some initial value $\mu_1 \in [0, \bar{\mu}_i]$ where $\bar{\mu}_i$ is an upper bound on μ. We use $P_{[0,\bar{\mu}_i]}$ to denote projection onto the interval $[0, \bar{\mu}_i]$. The algorithm APS, proposed by [Balseiro and Gur, 2019] and motivated by Lagrangian duality, proceeds as follows:

Algorithm 1: APS [Balseiro and Gur, 2019]

Initialize the pacing parameter μ_1 and the remaining budget $\tilde{B}_{i1} = B_i$.
for *every time period* $t = 1, \ldots, T$ **do**
    Observe v_{it}, construct a paced bid $b_{it} = \min(\frac{v_{it}}{1+\mu_t}, \tilde{B}_{it})$.
    Observe spend z_{it}, and refine the pacing multiplier using the update rule $\mu_{t+1} = P_{[0,\bar{\mu}_i]}(\mu_t - \epsilon_i(\rho_i - z_{it}))$.
    Update remaining budget $\tilde{B}_{i,t+1} = \tilde{B}_{it} - z_{it}$.

The problem $\max_{x \in \{0,1\}^T} \sum_{t=1}^{T} [x_{it}(v_{it} - (1-\mu)d_{it}) + \mu\rho_i]$ is the Lagrangian relaxation of the hindsight optimal optimization problem (4). The optimal solution for the relaxed problem is easy to characterize: we set $x_{it} = 1$ for all t such that $v_{it} \geq (1-\mu)d_{it}$. Importantly, this is achieved by the bid $b_{it} = \frac{v_{it}}{1+\mu}$ that we use in APS.

The Lagrangian dual is the minimization problem

$$\inf_{\mu \geq 0} \sum_{t=1}^{T} \left[(v_{it} - (1-\mu)d_{it})^+ + \mu\rho_i \right], \tag{5}$$

where $(\cdot)^+$ denotes thresholding at 0. This dual problem upper bounds π_i^H (but we do not necessarily have strong duality since we did not even start out with a convex primal program). The minimizer of the dual problem yields the strongest possible upper bound on ϕ_i^H. However, solving this requires us to know the entire sequences of v_i and d_i. APS approximates this optimal μ by taking a subgradient step on the t'th term of the dual:

$$\partial_\mu \left[(v_{it} - (1-\mu)d_{it})^+ + \mu\rho_i \right] \ni \rho_i - d_{it}\mathbb{1}\{b_{it} \geq d_{it}\} = \rho_i - z_{it}.$$

Thus APS is taking subgradient steps based on the subdifferential of the t'th term of the Lagrangian dual of the hindsight optimal optimization problem.

The APS algorithm achieves exactly the lower bound we derived earlier, and is thus asymptotically optimal.

Theorem 5 APS *with step size $\epsilon_i = O(T^{-1/2})$ is asymptotically $\frac{v_{it}}{\rho_i}$-competitive, and converges at a rate of $O(T^{-1/2})$.*

This result holds under adversarial conditions: for example, the sequence of highest other bids may be as d^1, d^2 in the lower bound. However, in practice we do not necessarily expect the world to be quite this adversarial. In a large-scale auction market, we would typically expect the sequences v_i, d_i to be more stochastic in nature. In a fully stochastic setting with independence, APS turns out to achieve π_i^H asymptotically:

Theorem 6 *Suppose* $(v_{it,d_{it}})$ *are sampled independently from stationary, absolutely continuous CDFs with differentiable and bounded densities. Then the expected payoff from* APS *with step size* $\epsilon_i = O(T^{-1/2})$ *approaches* π_i^H *asymptotically at a rate of* $T^{-1/2}$.

Theorem 6 shows that if the environment is well-behaved then we can expect much better performance from APS. It can also be shown that when all buyers use APS with appropriate step sizes, then each buyer converges to a solution that achieves the optimal dual value (5) (note that since we do not have strong duality this does not imply that π_i^H is achieved).

3.3 Numerical Experiments

In previous sections, we have illustrated how pairing auction markets and market equilibria can allow us to derive theoretical properties and can give us a tool to effectively compute equilibria in auction markets. Recalling the focus on equilibrium computation, in this section we present an empirical study, drawing from the material in Conitzer et al. [2018, 2019]. Relying on SPPE and FPPE computed for a set of stylized and realistic instances, we discuss how these equilibria compare in terms of revenue and welfare, provide evidence that incentives in FPPE arising from the first price auction are not problematic, and show how the static FPPE can be used to effectively seed the dynamic pacing algorithm.

The experiments are mainly based on realistic instances derived from the real-world auction markets at Facebook and Instagram. The instances were constructed in two steps, as explained in Conitzer et al. [2019]. The first step is to take bidding data for a region during a period and use it to create n buyers and m goods. The buyers are the top n advertisers that participate in the most auctions in that period in that region. The set of goods is constructed by applying a k-means algorithm to the auctions in which the advertisers participated. The features used for this are the n-dimensional vector of bids of each advertiser in each auction. The valuation of a buyer to a good is set to the average valuation of auctions in the cluster. The budgets are set equal to the expected value that the buyer would receive in a uniform random allocation of goods to buyers, i.e., $B_i = \frac{1}{n}\sum_j v_{ij}$. The motivation for this is that it leads to a good mixture of budget-constrained and unconstrained buyers, since in aggregate this constrains the sum of prices to be the sum of average valuations, whereas it would be the sum of maximum valuations if every buyer were unconstrained. The set of constructed instances combines different days, platforms, number of buyers ($n \in \{6, 8, 10, 12, 14\}$), and number of goods ($m \in \{10, 20, 30\}$) for a total of 210 instances for FPPE. Instances for SPPE require to be slightly smaller to be able to solve the MIP. The numerical study includes a set of instances with $\{3, \dots, 8\}$ buyers and $\{4, \dots, 8\}$ goods, for a total of $2 \times 7 \times 6 \times 5 = 420$ instances.

3.3.1 Computational Comparison between SPPE and FPPE

This section compares revenue and social welfare under FPPE and SPPE, as shown in Figure 3. The left panel shows the CDF of the ratio of FPPE revenue to that of SPPE. The right panel is similar but with social welfare. We see that FPPE revenue is always higher than SPPE revenue, though both coincide for about 30% of instances, and almost never more than 4.5 times as high. For social welfare, perhaps surprisingly, neither solution concept is dominating, with most instances having relatively similar social welfare under either solution concept, though FPPE does slightly better. There

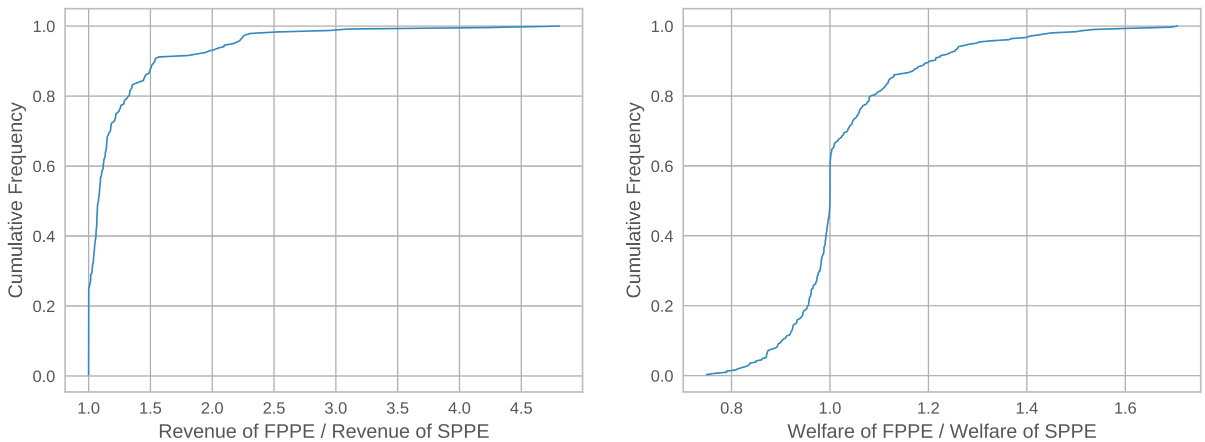

Fig. 3 CDF of the FPPE / SPPE ratio of revenue (left) and social welfare (right).

are two caveats to keep in mind for these results: (a) the numerical study did not compute the social-welfare-maximizing SPPE so it is possible that there is a better equilibrium (although this is highly unlikely given that most instances admit a single equilibrium); (b) many buyers are budget constrained in the FPPE of our setting, and so these insights might not translate to cases where many buyers are not budget constrained. These experiments show that an FPPE is not necessarily worse than an SPPE with respect to social welfare (at least with nonstrategic buyers), while potentially having a significantly higher revenue.

3.3.2 Incentives in FPPE

This section summarizes why incentives for advertisers are less of a problem than expected when using first price auctions. The incentive to deviate is quantified through the ex-post regret of buyers at FPPE, which capture what can be achieved when they unilaterally deviate to a different pacing multiplier while keeping the FPPE multipliers fixed for all other buyers. Figure 4 displays those ex-post relative regrets as the fraction of utility that is lost if the buyer uses the best-response pacing multiplier instead of reporting truthfully. The median regrets are very close to zero for instances of all sizes and the third quartile is below 0.02. The conclusion is that ex-post incentives to shade bids for individual advertisers when they can report a lower value per conversion or budget is very small in almost all cases. Furthermore, the incentive for misreporting the value per conversion or budget as inputs to the mechanism is vanishingly small. In the unrealistic case when advertisers have the power to shade bids at the auction level, the level of ex-post regret depends on the market thickness. Even in this extreme case, the average relative regret is never above 0.2.

As a hypothesis, this conclusion has to do with the coarseness at which manipulations can be performed when buyers do not have the ability to shade bids in individual auctions. Even if there is a large gap between the first and second price in a given auction, the winning buyer may not be able to exploit this, because once they start lowering their value per conversion, they might start losing some other auction much closer to their first price bid. Thus, a buyer need not face a “thick” market in every auction as one would conclude with first price auctions. It is enough for the incentive to deviate to be small if just a fraction of the auctions targeted by each advertiser is competitive.

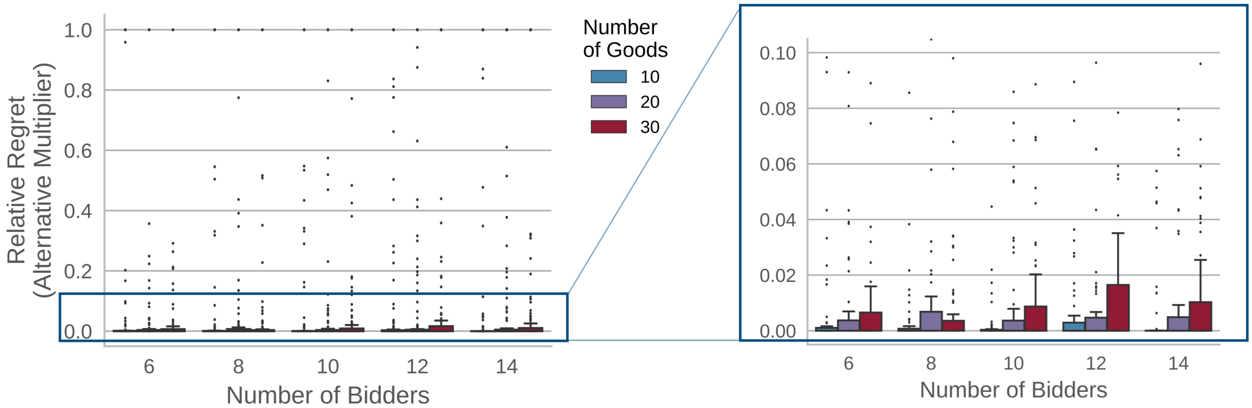

Fig. 4 Summary statistics of relative ex-post regret at an FPPE (ratio of best-response utility keeping competitors' bids fixed to utility under the FPPE). There is a data point for each buyer at the FPPE of each instance. The lower and upper edges of the boxes represent the first and third quartiles; the lines extending from the box show outliers within 1.5 times the inter-quartile range; and the dots represent individual outliers outside that range. The plot on the right is a zoomed-in version of the plot on the left.

3.3.3 Seeding Dynamics with SPPE

As we described in Section 3.2, real-world pacing heuristics rely on tractable adaptive algorithms that update buyers' pacing multipliers over time. This section looks at the rate at which the APS algorithm converges since the longer it takes to converge, the worse it is at optimizing the buyer's utility. In the evaluation, the algorithm is seeded with the solutions to the static SPPE and the resulting regret compared to other starting solutions such as constant or random pacing multipliers. The seed is computed from a static instance that abstracts away the dynamics but captures the market structure. The MIP mentioned earlier is used to solve the problem.

For each set of initial pacing multipliers, the runs are done with parameters ϵ and $\alpha^{\mathtt{min}}$, determined through grid search by choosing those that minimize the *average ex-post relative regret* (i.e., the average amount that a buyer could have improved its utility by playing a single best-response multiplier, given the other bids are fixed). As shown on the left of Figure 5, running APS on stylized random instances with MIP-based initial multipliers produces a lower regret than with other choices of initial multipliers. The performance of the MIP-based solution degrades as the noise parameter σ on the input data grows, but even at the highest levels we considered, this solution outperforms the others. For the fixed initial multipliers, the resulting regret is highly sensitive to choices in the step size: low initial multipliers would often not reach the MIP's equilibrium multipliers by the time the algorithm terminated. For realistic instances, the right plot of Figure 5 shows that the regret experienced by buyers when starting from the MIP-based initial multipliers was lower than in the other cases, for every learning rate ϵ that was considered. The worst learning rate for the MIP was better than the best learning rate for any other set of starting points. These findings are robust to different number of clusters m when producing the realistic instances.

In conclusion, using an SPPE of a static representation of an instance to warm-start an adaptive algorithm on the dynamic instance resulted in better convergence, and these improvements were robust to noise in the input data. This robustness provides evidence that the MIP does not need the exact valuation distribution or exact market structure to be useful. In Section 5.3, we discuss how to compress a large instance to create a smaller, approximate representative instance that could be tractably solved by the MIP.

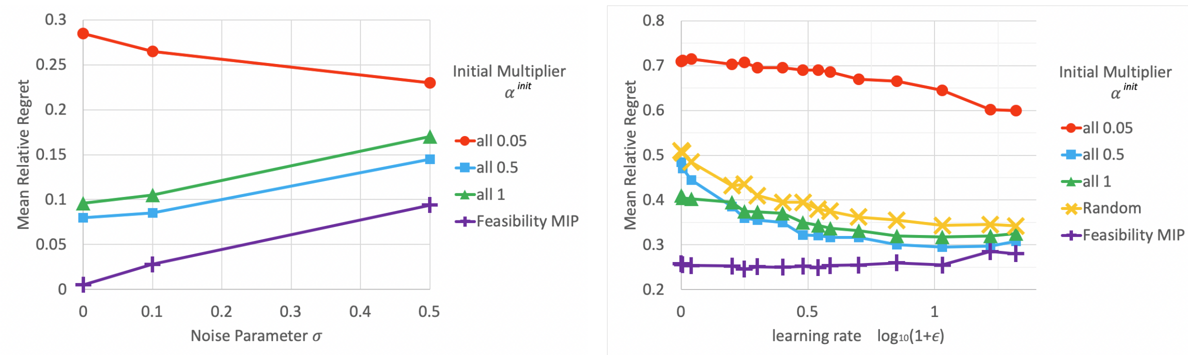

Fig. 5 Mean relative regret from running APS. Each curve plots different initial pacing multipliers α_i^{init}. *Left*: Stylized instances with random perturbations. Regret as a function of the noise parameter σ. *Right*: Realistic instances with 8 clusters (no noise). Regret as a function of the learning rate ϵ (shown in log scale as $\log 10(1+\epsilon)$).

3.4 Historical Notes

Borgs et al. [2007] study a dynamic bid optimization scheme based on first and second price auctions with perturbed allocation rules. While they do not discuss pacing as an equilibrium, their perturbation scheme in the first-price case leads to FPPE. Balseiro et al. [2015] started the study of pacing equilibria and showed that when buyers get to select their bids individually, multiplicative pacing equilibrium arises naturally via Lagrangian duality on the budget constraint, under a fluid-based mean-field market model. The literature has generally studied models where goods arrive stochastically and valuations are then drawn independently. Balseiro et al. [2017] show existence of pacing equilibrium for multiplicative pacing as well as several other pacing rules for such a setting; they also give a very interesting comparison of revenue and social welfare properties of the various pacing options in the unique symmetric equilibrium of their setting. Most notably, multiplicative pacing achieves strong social welfare properties, while probabilistic throttling achieves higher revenue properties. The static multiplicative pacing equilibrium results that we presented in this chapter were developed by Conitzer et al. [2018] for second price auction markets, and by Conitzer et al. [2019] for first price ones. The fixed-point theorem that is invoked to guarantee existence of a pure-strategy Nash equilibrium in the smoothed game is by Debreu [1952], Glicksberg [1952], and Fan [1952].

The quasi-linear variant of Eisenberg-Gale was given by Chen et al. [2007] and later rediscovered by Cole et al. [2017]. For discussion on strong duality and optimality conditions for these problems, see Bertsekas et al. [2003, Proposition 6.4.4]. The KKT conditions can be significantly generalized beyond convex programming.

The dynamic model of budget management was developed by Balseiro and Gur [2019]. Beyond auction markets, the idea of using paced bids based on the Lagrange multiplier μ has been studied in the revenue management literature, see e.g., Talluri and Van Ryzin [1998], where it is shown that this scheme is asymptotically optimal as T tends to infinity. There is also recent work on the adaptive bidding problem using multi-armed bandits [Flajolet and Jaillet, 2017].

The numerical study that we presented was performed by Conitzer et al. [2018] for second price auction markets and the comparison of dynamic pacing under the various starting points, and by Conitzer et al. [2019] for first price auction markets and the comparison between them.

4 Fair Division Problems and Applications At Scale

Market equilibrium is also intimately related to the problem of fairly dividing goods among agents. In *fair division problems* the setup is completely analogous to the Fisher market setting: we have m divisible goods to allocate to n individuals. The preferences of individuals are captured by utility functions $u_i(x_i)$. The goal is to find a "good" assignment x of goods to buyers. However, what is considered "good" turns out to be complicated in the setting of fair division, as there are many possible desiderata we may wish to satisfy.

First, we would like the allocation to be efficient, meaning that it should lead to high utilities for the individuals. One option would be to try to maximize the *social welfare* $\sum_i u_i(x_i)$. However, this turns out to be incompatible with the fairness notions that we will introduce later. An easy criticism of social welfare in the context of fair division is that it favors *utility monsters*: individuals with much greater capacity for utility are given more goods. Since social welfare maximization is typically incompatible with fairness, fair division mechanisms typically opt for the less stringent notion of *Pareto optimality* of the resulting allocation x. That requires that for *every* other allocation x', if one individual i is better off under x' than under x, then some other individual i' is strictly worse off. In other words, x should be such that no other allocation weakly improves all individuals' utilities, unless all utilities stay the same.

In addition to Pareto optimality, fair division mechanisms typically strive for allocations that satisfy various fairness conditions. We will be concerned with the following two desiderata:

Envy free: An allocation x has no envy if $u_i(x_i) \geq u_i(x_{i'})$ for every pair of individuals i and i'. In other words, every individual prefers their own bundle at least as much as that of anyone else.

Proportionality: An allocation x satisfies proportionality if $u_i(x_i) \geq u_i\left(\frac{1}{n} \cdot s\right)$. In other words, every individual prefers their own bundle at least as much as receiving a bundle composed of an nth of every good.

An allocation that satisfies Pareto optimality, no envy, and proportionality turns out to be achievable using the so-called *competitive equilibrium from equal incomes* (CEEI), a classic economic solution concept based on market equilibrium. In CEEI, a fair allocation is obtained as follows. First, we give each individual a unit budget of *funny money* that represents a fake currency to operate the content recommendation system. Second, we compute a market equilibrium for the Fisher market consisting of the individuals and their utility functions, along with the unit budgets of funny money. Finally, we take the corresponding market equilibrium allocation x, call it our fair division, and forget about the funny money.

CEEI is an appealing solution with respect to the previous desiderata. It is Pareto optimal since every market equilibrium satisfies it as discussed in Section 2. It has no envy since each buyer has the same budget. That means that each buyer can afford the bundle of any other buyer, and every buyer buys an optimal allocation given the prices and budgets. Finally, proportionality is satisfied since each buyer can afford the bundle where they get s_j/n of each good. This is easily shown by noting that the sum of prices must equal the sum of budgets.

In the divisible setting, CEEI is guaranteed to exist, and it is computable both using convex programming (via the Eisenberg-Gale convex program), and at scale via first-order methods as we shall see in Section 5.2. In contrast, for the indivisible setting, CEEI is not necessarily guaranteed to exist. In that case, one could rely on approximate-CEEI, a relaxed solution concept where buyers get slightly unequal budgets [Budish, 2011].

Let us now discuss the main applications of CEEI. The most obvious fair division settings that come to mind as practical examples involve indivisible objects, e.g., housing assignment, school choice, fair estate division, and so on. There have been several interesting applications of market-equilibrium ideas to the indivisible case. In spite of the possible non-existence of market equilibria, Budish et al. [2016] apply an approximate market equilibrium notion to the problem of fairly assigning course seats to MBA students. While market equilibria are not guaranteed to exist in that setting, Budish [2011] shows that approximate equilibria exist, and that they have appealing fairness and incentive properties. This approach is currently applied at several business schools. Another interesting application is that of fairly dividing goods such as items in an estate. A publicly available implementation of this can be found at `www.spliddit.org`. That webpage allows users to set up fair division instances of moderate size, and a fair allocation is offered by computing the discrete allocation that maximizes the geometric mean of utilities. This is a direct extension of EG to the discrete setting. A market equilibrium approximation is not guaranteed, but approximations to discrete variants of envy freeness and proportionality are guaranteed, as shown by Caragiannis et al. [2019].

The indivisible approaches mentioned above are, however, far from scalable to the size of allocation problems faced in auction markets and content ranking applications. Indivisible fair division problems can be converted into divisible ones by allowing randomized allocation. But this may not lead to particularly fair or acceptable solutions (for example, flipping a coin to decide who inherits a house would not be acceptable to many people). Although sometimes this randomization can be resolved successfully [e.g., Budish et al., 2013], generally speaking the randomization can lead to large ex-post regret. However, internet firms can usually circumvent this issue because of the scale of their markets. For instance, if they allocate content to users attempting to provide an balanced distribution from creators, then a randomized allocation may be enough. Content gets shown a large number of times, which smooths out randomization issues that can crop up in settings where each individual is assigned only a few goods. In the following sections we will describe applications of the divisible CEEI solution concept to such internet-scale problems.

4.1 Fair Recommender Systems and Diversity in Ranking

As a concrete application 'at scale,' in this section we go deeper to describe how Fisher markets and their equilibria can be used to model large-scale content recommendation systems. The main aspect of these recommendations is that we want to explicitly consider fairness and diversity goals. The problem can be summarized as follows: we have a set of n pieces of *content* (e.g., songs we could recommend to users, or job posts that we could show to users). We have m opportunities to show content (e.g., each job seeker is shown a ranked list of five job posts, which would generate five opportunities). The goal is to allocate the content to the different opportunities in a way that maximizes a relevant efficiency metric (e.g., likes of recommended songs, or actual job applications). At the same time, we also wish to treat each piece of content fairly. For example, in the jobs setting, we may wish to avoid showing a small set of job posts over and over, even if most job seekers were likely to apply to them. Not only this is not conducive to improving the ecosystem in this two sided market but also it is unlikely that all applications can be accepted by the company offering the job.

For the content recommendation problem, the analogy to market equilibrium is that the content creators are the buyers, and the content recommendation slots are the goods (e.g., if there are five job posts shown per page then there would be five goods to 'sell' in the market). By reducing the content recommendation problem to a market equilibrium problem, we can guarantee Pareto-efficient content recommendation, where every content creator will have budget-adjusted envy-freeness, and receive at least their budget-proportional utility. The budget of 'funny money' given to each content creator can be chosen according to what fraction of the recommendations we want each content creator to receive. As an example, consider the music recommendation setting and two musicians where one is up-and-coming while the other is a superstar. While the goal is to fairly give some exposure to the up-and-coming musician, we likely would not want to give equal budgets to the two musicians. Instead, the superstar might get a larger budget of 'funny money', while the up-and-coming musician might get a smaller budget that is large enough to ensure exposure.

Similarly to the discussion about proxy bidders in Section 3, this vision of a recommender system may be implemented by the platform in its entirety. A proxy bidder submits the value of each opportunity acting on behalf of the buyer. In this setup the funny money and budgets are merely abstractions used in the market, and are entirely controlled by the platform. There is no actual money changing hands in this mechanism. The whole point is to make the recommendations fair, and not that anybody can buy their way into more appealing slots.

More concretely, suppose that we have n content creators (say job posts), and m possible recommendation slots to fill (e.g., every time a job seeker shows up we show them a single job post). Furthermore, suppose that we measure the quality of recommending job post j to job seeker i by the probability that the job seeker clicks on (or applies to) the job post. This probability v_{ij} will represent the value in the market model. We wish for a given job post i to get allocated roughly to some percentage of all content slots. We can set the budget B_i for job post i equal to that percentage. Having defined buyers, goods, values and budgets, now we have a Fisher market model that admits an equilibrium. Allocating the job posts according to that equilibrium guarantees several desirable properties. For any job post i, proportionality guarantees that the expected number of job seekers that will click on the job post is at least as high as if that post were shown to each job seeker with probability B_i. Secondly, the no-envy property guarantees that every job post has at least as high of an expected number of clicks as if they receive the allocation for any other job post k, after adjusting the allocation to k by the factor $\frac{B_i}{B_k}$.

An interesting extension of the CEEI application to fair allocation of content would be to more directly guarantee fairness on both sides of the market. In particular, CEEI gives a one-sided fairness guarantee: it only requires that content creators receive a fair share of the set of recommendations. Ideally, we would also like users being shown recommendations to also get explicit fairness guarantees. Mehrotra et al. [2018] describe this problem in the context of *music recommendations* motivated by Spotify. A naive content recommendation approach based on relevance prediction will typically allocate the majority of recommendations to a small set of superstar musicians. Instead, the platform would like to recommend music in a way that more equitably recommends songs by less famous artists. At the same time, there is a clear tradeoff in that user satisfaction is extremely important. Mehrotra et al. [2018] study a group fairness notion where musicians are each assigned one out of K "popularity bins," and they measure fairness as $\sum_{k=1}^{K} \sqrt{|A_k|}$ where $|A_k|$ is the number of artists from bin k assigned a recommendation. This yields a form of regularization towards fair treatment of each artist bin, though it does not yield the kind of per-artist fairness that market equilibrium guarantees.

Let us briefly mention that a market-equilibrium-based allocation is not the only possible approach towards achieving more diverse recommendations. A very related approach is to use a linear programming (LP) approach that maximizes total social welfare, but ensures that each content creator is allocated at least some minimum amount of utility. Such an approach can be adapted to an online setting by adding a Lagrange multiplier on each utility lower bound. This multiplier can be tuned over time using a control algorithm, based on whether the creator is receiving the right amount of utility. In a certain sense, these two approaches can be viewed as equivalent: by the second welfare theorem we know that any Pareto-efficient allocation can be implemented in market equilibrium by an appropriate redistribution of budgets. However, in practice the two approaches are calibrated very differently. The LP approach requires us to specify an exact utility lower bound for each content creator. On the one hand this gives us very concrete utility guarantees, but on the other hand it may be hard to pick appropriate utility lower bounds, especially if we do not know the exact market composition ahead of time. Some choices of utility lower bounds may even lead to infeasibility. By using a market-equilibrium approach one can instead specify the budgets of funny money for each content creator. This is always guaranteed to be feasible, and is akin to instead specifying a fractional share of the market that we would like to allocate to each content creator.

4.2 More Connections to Large-Scale Internet Applications

Besides fair content recommendation systems, allocations based on market equilibria apply to several other problems that are related to large-scale internet applications where no real money is involved. In the *robust content review problem*, we are faced with the task of filtering several types of harmful social media content (e.g., fake news, impersonation, hate speech, . . .). Let us consider n categories of harmful content, each with some forecasted amount of content to review in each time period. We also have *m review groups*, which are groups of reviewers (typically in different geographic locations) that have been trained to handle a certain subset of harmful content categories. Each review group has some total amount of reviewing capacity that can be provided during each time period. The goal is to allocate review time to the content categories in a way that satisfies all forecasted review amounts, and then allocate the excess reviewing capacity across the content types to be robust to variations from the forecast. Allouah et al. [2021] show that this problem can be formulated as a variation of the Fisher market equilibrium problem.

Another fair allocation problem that is a component of large-scale internet applications is recommending donation opportunities to people who previously registered as being interested in donating blood [McElfresh et al., 2020]. Opportunities arise from requests by blood centers, temporary events such as blood drives, and emergency situations where blood is needed to save lives. The opportunities are submitted to a social network or a donation-specific app, and the recommendation system has to decide what set of users to offer the donation suggestion to. The goal is to optimally allocate suggestions in a way that maximizes the total amount of blood donated, while also treating each donation opportunity equitably.

4.3 Historical Notes

The CEEI solution concept was introduced by Varian et al. [1974]. Assigning course seats to students fairly via market equilibrium was studied by Budish [2011]. Goldman and Procaccia [2015] created an online service called `spliddit.org` which has a user-friendly interface for fairly dividing many things such as estates, rent, fares, and others. The motivating example of fair recommender systems, in which we fairly divide impressions among content creators via CEEI was suggested in Kroer et al. [2019], Kroer and Peysakhovich [2019], Murray et al. [2020a].[2] The robust content review problem was introduced by Allouah et al. [2021], where they also show that extensions of CEEI lead to desirable properties. The blood donations problem was introduced by McElfresh et al. [2020]. A comprehensive overview of recent fair division work was given by Freeman and Shah [2020]. There are also interesting fair division problems related to large-scale internet settings where CEEI is not the preferred solution method. An interesting example of this is the sharing a large set of compute resources such as cloud computing infrastructure or a large compute cluster [Parkes et al., 2015, Ghodsi et al., 2018].

[2] See also Murray et al. [2020b].

5 Computing Large-Scale Market Equilibria

In Section 2 we pointed out that one can compute equilibria of Fisher markets solving the Eisenberg-Gale convex program (EG). In this section we discuss the practical considerations of solving this problem for large instances.

This EG formulation can be solved with off-the-shelf software, though it requires the ability to solve convex programs involving exponential cones. For small instances these can be handled with open-source solvers such as SCS [O'Donoghue et al., 2016], and are readily expressed using the CVXPY interface [Diamond and Boyd, 2016]. However, in our experience, open-source solvers quickly run into numerical issues for solving EG (around 120 buyers and goods). For moderate-to-large problems, we can rely on the commercial conic solver *Mosek* [Mosek, 2010]. If the model is such that the interior-point method in Mosek is able to perform iterations then this is typically the best approach. Mosek is very fast, and has industrial-grade capacity for dealing with numerical issues.

However, for extremely large-scale problems such as fair recommender systems or ad markets, interior-point methods encounter difficulties. This is because the linear system solved at each iteration becomes too slow since the solver does not take sufficient advantage of sparsity, or due to memory usage. To address this and keeping in mind that the purpose of the market formulations discussed throughout this chapter was being able to find equilibria, we now discuss methods suitable for large-scale problems. The approach we describe consists of two complementary elements. First, we will discuss *first-order methods* (FOMs), also known as *gradient-based methods*, for computing market equilibrium. The key selling point of such family of methods is that each iteration of the algorithm can be computed in roughly linear time, and storage costs are low. Second and as a complement to FOMs, we will discuss *abstraction methods*. The goal is to abstract a large Fisher market a smaller one that can be solved efficiently without a big loss in accuracy. Indeed, we propose a way to do this so the equilibria resulting from a reduced-size instance are at approximate equilibrium in the original market. This is crucial in cases where the original market instance is so large that we cannot efficiently store even the explicit iterates. Furthermore, abstraction methods can be used to deal with missing data by leveraging *low-rank models*.

5.1 Convex Programming Formulations for Fisher Markets

We will cover two different algorithms for computing market equilibria of a standard Fisher market model. We will then describe how these can be extended to handle quasilinear utilities. The methods will be based on the EG convex program, as well as its dual convex program:

$$
\begin{aligned}
\max_{x\ge 0,u} \quad & g_{EG}(u) := \sum_{i=1}^{n} B_i \log(u_i) \\
s.t. \quad & u_i \le \sum_{j=1}^{m} x_{ij} v_{ij}, \quad i \in N, \\
& \sum_{i=1}^{n} x_{ij} \le 1, \quad j \in M,
\end{aligned}
\qquad\qquad
\begin{aligned}
\min_{p\ge 0,\beta\ge 0} \quad & \sum_{j=1}^{m} p_j - \sum_{i=1}^{n} B_i \log(\beta_i) \\
s.t. \quad & p_j \ge v_{ij}\beta_i, \quad i \in N, j \in M .
\end{aligned}
$$

The convex program on the left is the linear-utility version of the EG program [Eisenberg and Gale, 1959]; see Section 2.1 for the general case and Section 3.1.2 for the quasi-linear utility case in the context of FPPE. The dual of EG is also of interest so we included it on the right. An interesting note on the dual of EG is that it optimizes over prices of goods p_j, and per-buyer *utility prices* β_i. At market equilibrium, β_i is exactly the rate at which buyer i derives utility, i.e., $\beta_i = \frac{B_i}{u_i}$.

The convex program below is called the *Shmyrev* formulation and it looks very different from EG. It optimizes over the *spends* b_{ij} (the amount of money that buyer i spends on good j) instead of over allocations directly.

$$
\begin{aligned}
\max_{b\geq 0, p} \quad & f_{sh}(b,p) := \sum_{i=1}^{n}\sum_{j=1}^{m} b_{ij}\log v_{ij} - \sum_{j=1}^{m} p_j \log p_j \\
s.t. \quad & \sum_{i=1}^{n} b_{ij} = p_j, \quad j \in M, \\
& \sum_{j=1}^{m} b_{ij} = B_i, \quad i \in N.
\end{aligned}
$$

The first constraint ensures that spends sum to the price of each good, while the second constraint ensures that each buyer spends their budget exactly. The objective is a value-weighted linear combination of spends plus an unscaled entropy regularizer on prices. While the Shmyrev convex program was introduced by Shmyrev [2009] as a new formulation for computing equilibria in Fisher markets, it turns out to be intimately related to EG. Cole et al. [2017] show that the Shmyrev program can be recovered from EG by first taking the dual of EG, applying a change of variables, and then taking the dual again. Despite this equivalence, we shall see that interesting and different algorithms result from solving each convex program.

5.2 First-Order Methods

We now describe two simple and scalable algorithms that arise from the convex programs mentioned above. The first algorithm we will describe is the *proportional response* (PR) algorithm. The PR algorithm is an iterative algorithm that can be viewed as a dynamic updating scheme between buyers and the seller. The buyers see current prices on goods and update their bids, while the seller sees these bids and in turn update the price. This can be summarized as follows:

Algorithm 2: Proportional response (PR)

Each buyer i submits a bid $b_{ij}^{t=1} \in \mathbb{R}_+^m$ for each good j.

for *every time step* $t = 1, \ldots$ **do**

- Given the bids, a price $p_j^t = \sum_i b_{ij}^t$ is computed for each good.
- Each buyer is assigned an allocation $x_{ij}^t = \frac{b_{ij}^t}{p_j^t}$ of each good.
- Each buyer submits a next bid on good j proportionally to the utility they received from good j in round t:

$$
b_{ij}^{t+1} = B_i \frac{x_{ij}^t v_{ij}}{\sum_{j'} x_{ij'}^t v_{ij'}} .
$$

As it is evident from the price and bid updating schemes, these updates are designed such that they alternatively correspond to each of the constraints in the Shmyrev program. The next theorem provides a convergence rate for this algorithm as a function of the size of the instance and the time period. It shows that PR has a reasonably attractive $1/T$ rate of convergence.

Theorem 7 *The iterates of the PR algorithm converge at the rate of* $f_{sh}(b^*, p^*) - f_{sh}(b^t, p^t) \leq (\log nm)/t$, *where* b^* *and* p^* *denote any optimal solution to the Shmyrev convex program.*

From a practical perspective, the PR algorithm converges very rapidly to a medium-accuracy solution for most numerical examples. Thus, it is a very useful method in practice, since it has a very simple and lightweight implementation, requires no parameter tuning, and can be used for very large instances. This is especially the case if the valuations are *sparse*: we only need a variable b_{ij} corresponding to a buyer-good pair (i, j) if $v_{ij} > 0$.

While we have not described the mirror descent algorithm in this chapter, it is worth making a few comments on the equivalence between the PR and mirror descent algorithms. While this is not immediately apparent from the description of the algorithm, the PR algorithm is a first-order method because it is an application of mirror descent to a convex program. Even though there is no step size in PR (whereas FOMs, including mirror descent, typically have step sizes), the PR dynamics correspond to choosing a step size of one in mirror descent. This turns out to be a valid choice due to a strong connection between the unscaled entropy on prices in the Shmyrev objective, and the way distances are measured when using the negative entropy on the bids as a distance measure. A reader familiar with the typical convergence rates achieved by mirror descent would expect that one must average the iterates across all time steps in order to get convergence, and then this would typically converge at a rate of $1/\sqrt{T}$. Theorem 7 gives a guarantee on the *last iterates* b^t, p^t of PR without averaging, and the rate guarantee is of the order of $1/T$. From a theoretical perspective a rate improvement of $1/\sqrt{T}$ is very strong, and from a practical perspective the last iterate convergence is quite attractive. Both of these properties are again a consequence of the strong connection between the unscaled entropy on prices in the Shmyrev objective and the negative entropy distance measure.

As stated above, PR is a great algorithm for converging to medium-accuracy solutions. However, if one wants higher-accuracy solutions, then a method with a faster asymptotic rate of convergence is necessary. We now describe how this can be achieved via a *projected gradient descent* (PGD) algorithm. PGD operates on the original EG problem. It iterates the following update:

$$x^{t+1} = \Pi_{\mathcal{X}}(x^{t-1} - \gamma_t \nabla g_{EG}(x^t)),$$

where $\mathcal{X} = \{x \in \mathbb{R}_+^{n \times m} : \sum_i x_i = s\}$ is the set of feasible allocations that fully utilize every good[3], $\Pi_{\mathcal{X}}$ is the projection operator onto $\mathcal{X}$, and γ_t is the step size. Gao and Kroer [2020] show that PGD on EG converges at a linear rate $(1 - \delta)^t$ for a small constant δ. Thus, theoretically, the PGD algorithm should be preferred to PR when higher accuracy is needed. This result does come with some caveats however: first, the base of the exponent, $1 - \delta$, can be close to 1 as the term δ depends on some hard-to-compute constants, such as the *Hoffman* constant which makes δ very small. Second, the cost per iteration for PGD is slightly higher than for PR: it requires projecting onto $\mathcal{X}$, which can be done via sorting but leads to a per-iteration cost of $nm \log n$, as opposed to nm for PR. Here too the projection can take advantage of sparsity, thereby becoming much faster in the case where only a few buyers are interested in each good. Gao and Kroer [2020] also show numerically that PR is faster than PGD for medium-accuracy solutions, whereas PGD is faster when higher levels of accuracy are desired.

Both PR and PGD algorithms can also be extended to the case of quasilinear utilities, such as those used when modeling budget-smoothing in first price auctions as a market equilibrium problem. For PGD, this relies on extending

[3] Every optimal solution to EG must lie in this set, assuming that every good j has some i such that $v_{ij} > 0$; if this does not hold then that good can simply be preprocessed away.

EG to the quasilinear case, which was shown by Chen et al. [2007], Cole et al. [2017]. Gao and Kroer [2020] show how to apply PGD to achieve a linear rate.

In this chapter, we focused on simple and scalable first-order methods for computing market equilibria. We believe these algorithms are very suitable for practical use, due to their easy implementation requirements and cheap per-iteration complexity. Beyond the algorithms covered here, there is a long history of theory-oriented polynomial-time algorithms for computing equilibria of Fisher markets. This line of work started with Devanur et al. [2008] who give a primal-dual algorithm. Later work offered simpler algorithms, e.g., Bei et al. [2019] put forward an algorithm with a linear rate that can be interpreted as a form of ascending-price auctions. However, these methods typically do not have cheap per-iteration complexity, and are thus less suitable for the market sizes considered in this chapter. Another interesting line of extensions studies algorithms for new utility classes such as, e.g., *spending constraint utilities*, which are additively separable utilities that have piecewise-linear utility per good [Vazirani, 2010, Birnbaum et al., 2011].

5.3 Market Abstractions

So far we have described scalable first-order methods for computing market equilibria. Still, these algorithms make a number of assumptions that may not hold in practice. To use the PR or PGD algorithms, one must be able to store the iterates which take nm space. If both the number of buyers and goods are of the order of 100,000, writing down an iterate using 64-bit floats requires about 80 GB of memory (assuming no sparsity). For applications in large internet companies such as ad markets, we might expect n, and especially m, to be even larger than that. Thus efficient computation is important but not enough. We may need to find a way to compress the instances to be solved down to some manageable size where we can at least hope to store iterates efficiently. Furthermore, we may not have access to all valuations v_{ij}. For instance, we may only have some samples and those values may be noisy. This means that we also need to somehow infer the remaining valuations. In a setting where we do not know all the true valuations, or we only have noisy estimates, it is important to understand how these misspecifications degrade the quality of computed equilibria.

The issues mentioned earlier motivated Kroer et al. [2019] to consider abstraction methods to solve those problems. For the purposes of abstraction, it will be useful to think of the set of valuations v_{ij} as a matrix V, where the i'th row corresponds to the valuation vector of buyer i. We will be interested in the outcome of computing a market equilibrium using a valuation matrix $\tilde{V} \approx V$, where $\tilde{V}$ would typically be obtained from an abstraction method. The goal is establishing that the market equilibria corresponding to V and $\tilde{V}$ are similar, which could be quantified by $\|\tilde{V} - V\|_F$. That would enable us to compute one equilibrium to approximate the other. Let us enumerate a couple of reasons why one might prefer to compute a market equilibrium for $\tilde{V}$ rather than V.

Low-rank markets: When there are missing valuations, we need to impute the missing values. Of course, if there is no relationship between the entries of V that are observed and those that are missing, then we have no hope of recovering V. However, in practice this is typically not the case. Valuations are often assumed to (approximately) belong to some low-dimensional space. A popular model is to assume that the valuations are *low rank*, meaning that every buyer i can be represented by some d-dimensional vector ϕ_i, every good j can also be represented by some d-dimensional vector ψ_j, and the valuation of buyer i for good j is $\tilde{v}_{ij} = \langle \phi_i, \psi_j \rangle$. One may interpret this model as every good having an associated set of *d features*, with ψ_j describing the value for each feature, and ϕ_i describing the value that i places on each feature. In a low-rank model, d is expected to be much smaller than $\min(n, m)$, meaning that V is far from full rank. If the real valuations are approximately of rank d (meaning that the remaining spectrum of V is very small), then $\tilde{V}$ will be close to V.

This model can also be motivated via the *singular-value decomposition* (SVD). Assume that we wish to find the matrix of rank d that is closest to V:

$$\begin{aligned} \min_{\tilde{V}} \quad & \|V - \tilde{V}\|_F^2 \quad := \sum_{ij} (v_{ij} - \tilde{v}_{ij})^2 \\ s.t. \quad & \text{rank}(\tilde{V}) \leq d\,. \end{aligned}$$

The optimal solution to this problem can be found easily via SVD. Letting $\sigma_1, \ldots, \sigma_d$ be the first d singular values of V, $\bar{u}_1, \ldots, \bar{u}_d$ the first left singular vectors, and $\bar{v}_1, \ldots, \bar{v}_d$ the first right singular vectors, the optimal solution is $\tilde{V} = \sum_{k=1}^d \sigma_k \bar{u}_k \bar{v}_k^T$. If the remaining singular values $\sigma_{k+1}, \ldots$ are small relative to the first k singular values, then this model captures most of the valuation structure.

In practice, since the matrix V might not be known exactly, we cannot solve this problem to get $\tilde{V}$. Instead, we search for a low-rank model that minimizes some loss on the observed entries OBS, e.g., $\sum_{i,j \in \text{OBS}} (v_{ij} - \langle \phi_i, \psi_j \rangle)^2$ (this objective is typically also regularized by the Frobenius norm of the low-rank matrices). Under the assumption that V is generated from a true low-rank model via some simple distribution, it is possible to recover the original matrix with only samples of entries by minimizing the loss on observed entries. In practice this approach is also known to perform extremely well, and it is used extensively at tech companies. The hypothesis is that in practice data is approximately low rank, so one does not lose much accuracy from a rank-d model.

Representative Markets: It is also convenient to generate a smaller set of representative buyers, where each original buyer i maps to some particular representative buyer $r(i)$. Similarly, we may generate representative goods that correspond to many non-identical but similar goods from the original market. These representative buyers and goods may be generated using clustering techniques. In this case, our approximate valuation matrix $\tilde{V}$ has as row i the valuation vector of the representative buyer $r(i)$. This means that all i, i' such that $r(i) = r(i')$ have the same valuation vector in $\tilde{V}$, and thus they can be treated as a single buyer for equilibrium-computation purposes. The same grouping can also be applied to the goods. If the number of buyers and goods is reduced by a factor of q, then the resulting problem size is reduced by a factor of q^2, since we have $n \times m$ variables.

We now analyze what happens when we compute a market equilibrium under $\tilde{V}$ rather than V. Throughout this subsection we will let $(\tilde{x}, \tilde{p})$ be a market equilibrium for $\tilde{V}$. We use the error matrix $\Delta V = V - \tilde{V}$ to quantify the solution quality, and we measure the size of ΔV using the ℓ_1-ℓ_∞ matrix norm $\|\Delta V\|_{1,\infty} = \max_i \|\Delta v_i\|_1$. We will also use the norm of the error vector for an individual buyer $\|\Delta v_i\|_1 = \|v_i - \tilde{v}_i\|_1$.

The next proposition turns out to be useful in proving guarantees on the approximate equilibrium. It establishes that under linear utility functions the change in utility when going from v_i to $\tilde{v}_i$ is linear in Δv_i.

Proposition 1 *If* $\langle \tilde{v}_i, x_i \rangle + \epsilon \geq \langle \tilde{v}_i, x_i' \rangle$ *then* $\langle v_i, x_i \rangle + \epsilon + \|\Delta v_i\|_1 \geq \langle v_i, x_i' \rangle$

This proposition can be used to immediately derive bounds on envy, proportionality, and regret (how far each buyer is from achieving the utility of their demand bundle). For example, we know that under $\tilde{V}$, each buyer i has no envy towards any other buyer k: $\langle \tilde{v}_i, \tilde{x}_i \rangle \geq \langle \tilde{v}_i, \tilde{x}_k \rangle$. By Proposition 1 each buyer i has envy at most $\|\Delta v_i\|_1$ under V when using $(\tilde{x}, \tilde{p})$. All envies are thus bounded by $\|\Delta V\|_{1,\infty}$. Regret and proportionality can be bounded similarly which implies guarantees under $\tilde{V}$.

Market equilibria also guarantee Pareto optimality. Unfortunately, we cannot give any meaningful guarantee on how much social welfare improves under Pareto-improving allocations for $\tilde{V}$. The following real and abstracted matrices provide an example of it:

$$V = \begin{bmatrix} 1 & \epsilon & \epsilon \\ 0 & 1 & \epsilon \end{bmatrix}, \quad \tilde{V} = \begin{bmatrix} 1 & \epsilon & 0 \\ 0 & 1 & \epsilon \end{bmatrix}.$$

If we set $B_1 = B_2 = 1$, then for supply-aware market equilibrium, we end up with competition only on good 2, and we get prices $\tilde{p} = (0, 2, 0)$ and allocation $\tilde{x}_1 = (1, 0.5, 0), \tilde{x}_2 = (0, 0.5, 1)$. Under V this is a terrible allocation, and we can Pareto improve by using $x_1 = (1, 0, 0.5), x_2 = (0, 1, 0.5)$, which increases overall social welfare by $\frac{1}{2} - \epsilon$, in spite of $\|\Delta V\|_1 = \epsilon$.

On the other hand, we can show that under any Pareto-improving allocation, some buyer i improves by at most $\|\Delta V\|_{1,\infty}$. To see this, note that for any Pareto improving allocation x, under $\tilde{V}$ there existed at least one buyer i such that $\langle \tilde{v}_i, \tilde{x}_i - x_i \rangle \geq 0$, and so this buyer must improve by at most $\|\Delta v_i\|_1$ under V.

5.4 Historical Notes

The proportional response (PR) algorithm was introduced by Wu and Zhang [2007]. It was later shown to be effective for BitTorrent sharing dynamics [Levin et al., 2008], and it was eventually shown to be an instantiation of the *mirror descent* algorithm by Birnbaum et al. [2011], who also show the last-iterate $\frac{1}{T}$ rate. Birnbaum et al. [2011] also show how mirror descent can be applied to the Shmyrev formulation. For a general introduction to mirror descent, see e.g., Bubeck [2015] or the lecture notes of Ben-Tal and Nemirovski [2001]. Alternatively, the lectures notes from Kroer [2020] also cover this derivation, as well as the general duality derivations required to obtain Shmyrev from EG.

Abstraction in the context of Fisher market equilibrium was introduced by Kroer et al. [2019]. However, abstraction has been applied in other related problems. Candogan et al. [2016] consider replacing sets of agents in trading networks with representative buyers, in order to do comparative statics. There has also been work on abstraction for non-market-based allocation problems [Walsh et al., 2010, Lu and Boutilier, 2015, Peng and Sandholm, 2016], where the results largely center around good abstraction of LP or MIP-based allocation problems. Abstraction has also been studied in the computation of game-theoretic equilibria, where it has been used extensively in practice [Gilpin et al., 2007, Brown et al., 2015, Brown and Sandholm, 2018], and studied from a theoretical perspective [Lanctot et al., 2012, Kroer and Sandholm, 2014, 2016, 2018].

A brief introduction to low-rank models can be found in Udell [2019]. Udell et al. [2016] gives a more thorough exposition and describes more general model types. Beyond these papers, there is a large theory of low-rank models that show a number of interesting results. There is a class of nuclear-norm-regularized convex optimization problems that can recover the original matrix with only a small number of entry samples [Candès and Recht, 2009, Recht, 2011]. One might think that this would then be the preferred method in practice, but surprisingly non-convex models are often preferred instead. These non-convex methods also have interesting guarantees on statistical recovery under certain assumptions. An overview of non-convex methods is given in Chi et al. [2019]. Low-rank market equilibrium models were also studied in Kroer and Peysakhovich [2019], where it is shown that large low-rank markets enjoy a number of properties not satisfied by small-scale markets.

6 Conclusion

We have described how market models can be used to generate insights and compute solutions that are relevant to internet platforms. We have given examples in the context of ad auctions, recommendation systems and fair division problems. The general area is very rich and there is ample research published about these applications. For additional context on online advertising, particularly as it relates to display ads, we direct the reader to a survey by Choi et al. [2020]. For some context on market equilibrium in electricity markets, see Azizan et al. [2020] and references therein. A somewhat dated overview of work in the theoretical computer science community on computing market equilibrium can be found in Nisan et al. [2007]. The references in Section 5.4 provide some examples of more recent work. For more context on fair division, we recommend several surveys covering various aspects of this problem. Aleksandrov and Walsh [2020] gives a very recent overview of *online* fair division, a problem highly related to the topics covered here. For more general fair division coverage, see, e.g., Procaccia [2013], Brandt et al. [2016], Walsh [2020].

To use the ideas discussed in this chapter in practice, it is important to feed models with the right input. Some data may be available from historical information, logs, and other measurements, while other data may need to be estimated. To do the latter, one can rely on several statistical and machine learning techniques. In relation to advertising, clustering, recommender systems, and particularly to calibrate values of goods to buyers, we refer the reader to the chapter by Bastani et al. [2021] in this volume about the interplay between machine learning and operations.

To conclude we mention some open problems and directions, as of this article. For the general problem of computing market equilibrium, there are several interesting open questions. While the general Fisher market equilibrium can be computed efficiently, as discussed in earlier chapters, the refinement of market equilibria that is needed for SPPE is harder to handle. It was shown by Conitzer et al. [2018] that maximizing various objectives over the set of SPPE equilibria is NP-complete, but the complexity of finding any SPPE is currently unknown. In a similar vein, one might investigate the existence of approximate methods for finding SPPE, even if exact SPPE are hard to find. A related question is how to reconcile these potential hardness results with the fact that under independence assumptions, Balseiro and Gur [2019] show that it is possible to have buyers converge to an SPPE-like equilibrium in an online setting. Understanding the exact boundaries of what can be learned online is an important practical question.

Another interesting line of work would be to generalize pacing equilibria to more realistic allocation mechanisms. In practice, each item is not sold via independent first or second price auctions. Instead, the number of items sold per auction is the number of slots available in a given impression, and a single ad is typically only allowed to win one of those slots. This breaks the correspondence with market equilibrium, but an appropriate notion of pacing still exists. It would be interesting to understand which, if any, results from market equilibrium carry over to this setting. There is a rich space of possible questions to ask for this problem, based on the type of multi-item allocation mechanism being run, the presence of reserve prices, and so on.

Finally, we mentioned in the introduction that various mechanisms have been used in practice to allocate goods to buyers. It would be interesting to further explore the relations between the models presented in this chapter on market mechanisms and their underlying situations in the technology industry to bond auctions, supply function equilibria, and other market approaches.

References

Martin Aleksandrov and Toby Walsh. Online fair division: A survey. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(09):13557–13562, 2020.

Amine Allouah, Vashish Avadhanula, Christian Kroer, Sergey Pupyrev, and Parikshit Shah. The robust content review problem and market equilibrium. *In preparation*, 2021.

Navid Azizan, Yu Su, Krishnamurthy Dvijotham, and Adam Wierman. Optimal pricing in markets with nonconvex costs. *Operations Research*, 68(2):480–496, 2020.

Santiago R Balseiro and Yonatan Gur. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968, 2019.

Santiago R Balseiro, Omar Besbes, and Gabriel Y Weintraub. Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Science*, 61(4):864–884, 2015.

Santiago R Balseiro, Anthony Kim, Mohammad Mahdian, and Vahab Mirrokni. Budget management strategies in repeated auctions. In *Proceedings of the 26th International Conference on World Wide Web (WWW'17)*, pages 15–23, 2017.

Hamsa Bastani, Dennis J. Zhang, and Heng Zhang. Applied machine learning in operations management. In Birge Babich V. and Hilary J. G., editors, *Innovative Technology at the interface of Finance and Operations*, Series in Supply Chain Management. Springer, 2021. Springer Nature.

Xiaohui Bei, Jugal Garg, and Martin Hoefer. Ascending-price algorithms for unknown markets. *ACM Transactions on Algorithms*, 15(3):1–33, 2019.

Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001. Lastest version can be found at `https://www2.isye.gatech.edu/~nemirovs/LMCOLN2020WithSol.pdf`. Last accessed: November 2020.

Dimitri P Bertsekas, Angelia Nedic, and Asuman Ozdaglar. Convex analysis and optimization. 2003. *Athena Scientific*, 2003.

Benjamin Birnbaum, Nikhil R Devanur, and Lin Xiao. Distributed algorithms via gradient descent for Fisher markets. In *Proceedings of the ACM conference on Electronic Commerce (EC'11)*, pages 127–136. ACM, 2011.

Christian Borgs, Jennifer Chayes, Nicole Immorlica, Kamal Jain, Omid Etesami, and Mohammad Mahdian. Dynamics of bid optimization in online advertisement auctions. In *Proceedings of the 16th international conference on World Wide Web (WWW'07)*, pages 531–540, 2007.

Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of computational social choice*. Cambridge University Press, 2016.

Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018.

Noam Brown, Sam Ganzfried, and Tuomas Sandholm. Hierarchical abstraction, distributed equilibrium computation, and post-processing, with application to a champion no-limit Texas hold'em agent. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems (AAMAS'15)*, pages 7–15. International Foundation for Autonomous Agents and Multiagent Systems, 2015.

Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

Eric Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103, 2011.

Eric Budish, Yeon-Koo Che, Fuhito Kojima, and Paul Milgrom. Designing random allocation mechanisms: Theory and applications. *American Economic Review*, 103(2):585–623, 2013.

Eric Budish, Gérard P Cachon, Judd B Kessler, and Abraham Othman. Course match: A large-scale implementation of approximate competitive equilibrium from equal incomes for combinatorial allocation. *Operations Research*, 65 (2):314–336, 2016.

Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717, 2009.

Ozan Candogan, Markos Epitropou, and Rakesh V Vohra. Competitive equilibrium and trading networks: A network flow approach. In *Proceedings of the ACM Conference on Economics and Computation (EC'16)*, pages 701–702, 2016.

Ioannis Caragiannis, David Kurokawa, Hervé Moulin, Ariel D Procaccia, Nisarg Shah, and Junxing Wang. The unreasonable fairness of maximum Nash welfare. *ACM Transactions on Economics and Computation*, 7(3):1–32, 2019.

Lihua Chen, Yinyu Ye, and Jiawei Zhang. A note on equilibrium pricing as convex optimization. In *International Workshop on Web and Internet Economics (WINE'07)*, pages 7–16. Springer, 2007.

Yuejie Chi, Yue M Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269, 2019.

Hana Choi, Carl F. Mela, Santiago R. Balseiro, and Adam Leary. Online display advertising markets: A literature review and future directions. *Information Systems Research*, 31(2):556–575, 2020.

Richard Cole, Nikhil R Devanur, Vasilis Gkatzelis, Kamal Jain, Tung Mai, Vijay V Vazirani, and Sadra Yazdanbod. Convex program duality, Fisher markets, and Nash social welfare. In *ACM Conference on Economics and Computation (EC'17)*. Association for Computing Machinery, Inc, 2017.

Vincent Conitzer, Christian Kroer, Eric Sodomka, and Nicolas E Stier-Moses. Multiplicative pacing equilibria in auction markets. In *International Conference on Web and Internet Economics (WINE'18)*, 2018.

Vincent Conitzer, Christian Kroer, Debmalya Panigrahi, Okke Schrijvers, Eric Sodomka, Nicolas E Stier-Moses, and Chris Wilkens. Pacing equilibrium in first-price auction markets. In *Proceedings of the ACM Conference on Economics and Computation (EC'19)*. ACM, 2019.

Gerard Debreu. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences*, 38(10): 886–893, 1952.

Nikhil R Devanur, Christos H Papadimitriou, Amin Saberi, and Vijay V Vazirani. Market equilibrium via a primal–dual algorithm for a convex program. *Journal of the ACM*, 55(5):1–18, 2008.

Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

Edmund Eisenberg. Aggregation of utility functions. *Management Science*, 7(4):337–350, 1961.

Edmund Eisenberg and David Gale. Consensus of subjective probabilities: The pari-mutuel method. *The Annals of Mathematical Statistics*, 30(1):165–168, 1959.

Ky Fan. Fixed-point and minimax theorems in locally convex topological linear spaces. *Proceedings of the National Academy of Sciences of the United States of America*, 38(2):121, 1952.

Arthur Flajolet and Patrick Jaillet. Real-time bidding with side information. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS'17)*, pages 5168–5178. Curran Associates Inc., 2017.

Rupert Freeman and Nisarg Shah. AAAI 2020 Tutorial on Recent Advances in Fair Resource Allocation, 2020. https://users.cs.duke.edu/~rupert/fair-division-aaai20/Tutorial-Slides.pdf. Last accessed: November 2020.

Yuan Gao and Christian Kroer. First-order methods for large-scale market equilibrium computation. In *Advances in Neural Information Processing Systems (NeurIPS'20)*, 2020.

Mohammad Ghodsi, MohammadTaghi HajiAghayi, Masoud Seddighin, Saeed Seddighin, and Hadi Yami. Fair allocation of indivisible goods: Improvements and generalizations. In *Proceedings of the ACM Conference on Economics and Computation (EC'18)*, pages 539–556, 2018.

Andrew Gilpin, Tuomas Sandholm, and Troels Bjerre Sørensen. Potential-aware automated abstraction of sequential games, and holistic equilibrium analysis of Texas hold'em poker. In *Proceedings of the 22nd national conference on Artificial Intelligence*, volume 1, pages 50–57. AAAI Press, 2007.

Irving L Glicksberg. A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium points. *Proceedings of the American Mathematical Society*, 3(1):170–174, 1952.

Jonathan Goldman and Ariel D Procaccia. Spliddit: Unleashing fair division algorithms. *ACM SIGecom Exchanges*, 13(2):41–46, 2015.

Christian Kroer. IEOR8100: Economics, AI, and Optimization Lecture Note 14: Large-Scale Fisher Market Equilibrium, 2020.

Christian Kroer and Alexander Peysakhovich. Scalable fair division for 'at most one' preferences, 2019. arXiv:1909.10925.

Christian Kroer and Tuomas Sandholm. Extensive-form game abstraction with bounds. In *Proceedings of the ACM conference on Economics and computation (EC'14)*, pages 621–638. ACM, 2014.

Christian Kroer and Tuomas Sandholm. Imperfect-recall abstractions with bounds in games. In *Proceedings of the ACM Conference on Economics and Computation (EC'16)*, pages 459–476. ACM, 2016.

Christian Kroer and Tuomas Sandholm. A unified framework for extensive-form game abstraction with bounds. In *Advances in Neural Information Processing Systems (NeurIPS'18)*, pages 612–623, 2018.

Christian Kroer, Alexander Peysakhovich, Eric Sodomka, and Nicolas E Stier-Moses. Computing large market equilibria using abstractions. In *Proceedings of the ACM Conference on Economics and Computation (EC'19)*, pages 745–746, 2019.

Marc Lanctot, Richard Gibson, Neil Burch, Martin Zinkevich, and Michael Bowling. No-regret learning in extensive-form games with imperfect recall. In *Proceedings of the 29th International Coference on Machine Learning (ICML'12)*, pages 1035–1042. Omnipress, 2012.

Dave Levin, Katrina LaCurts, Neil Spring, and Bobby Bhattacharjee. Bittorrent is an auction: analyzing and improving bittorrent's incentives. In *Proceedings of the ACM SIGCOMM 2008 conference on Data communication*, pages 243–254, 2008.

Tyler Lu and Craig Boutilier. Value-directed compression of large-scale assignment problems. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*. AAAI Press, 2015.

Duncan C McElfresh, Christian Kroer, Sergey Pupyrev, Eric Sodomka, Karthik Abinav Sankararaman, Zack Chauvin, Neil Dexter, and John P Dickerson. Matching algorithms for blood donation. In *Proceedings of the ACM Conference on Economics and Computation (EC'20)*, pages 463–464, 2020.

Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *Proceedings of the 27th acm international conference on information and knowledge management*, pages 2243–2251, 2018.

APS Mosek. The MOSEK optimization software, 2010. `http://www.mosek.com`.

Riley Murray, Christian Kroer, Alex Peysakhovich, and Parikshit Shah. Robust market equilibria with uncertain preferences. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020a.

Riley Murray, Christian Kroer, Alex Peysakhovich, and Parikshit Shah. Robust market equilibria: How to model uncertain buyer preferences, 2020b. `https://research.fb.com/blog/2020/09/robust-market-equilibria-how-to-model-uncertain-buyer-preferences`. Last accessed Dec, 2020.

Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic game theory*. Cambridge University Press, 2007.

Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068, June 2016.

Renato Paes Leme, Balasubramanian Sivan, and Yifeng Teng. Why do competitive markets converge to first-price auctions? In *Proceedings of The Web Conference 2020*, WWW '20, page 596–605, New York, NY, USA, 2020. Association for Computing Machinery.

David C Parkes, Ariel D Procaccia, and Nisarg Shah. Beyond dominant resource fairness: Extensions, limitations, and indivisibilities. *ACM Transactions on Economics and Computation*, 3(1):1–22, 2015.

Fei Peng and Tuomas Sandholm. Scalable segment abstraction method for advertising campaign admission and inventory allocation optimization. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)*, pages 655–661. AAAI Press, 2016.

Ariel D Procaccia. Cake cutting: not just child's play. *Communications of the ACM*, 56(7):78–87, 2013.

Benjamin Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12:3413–3430, 2011.

Vadim I Shmyrev. An algorithm for finding equilibrium in the linear exchange model with fixed budgets. *Journal of Applied and Industrial Mathematics*, 3(4):505, 2009.

Kalyan Talluri and Garrett Van Ryzin. An analysis of bid-price controls for network revenue management. *Management Science*, 44:1577–1593, 1998.

Madeleine Udell. Big data is low rank. *SIAG/OPT Views and News*, 2019.

Madeleine Udell, Corinne Horn, Reza Zadeh, and Stephen Boyd. Generalized low rank models. *Foundations and Trends® in Machine Learning*, 9(1):1–118, 2016.

Hal R Varian et al. Equity, envy, and efficiency. *Journal of Economic Theory*, 9(1):63–91, 1974.

Vijay V Vazirani. Spending constraint utilities with applications to the adwords market. *Mathematics of Operations Research*, 35(2):458–478, 2010.

Toby Walsh. Fair division: The computer scientist's perspective, 2020. arXiv:2005.04855.

William E Walsh, Craig Boutilier, Tuomas Sandholm, Rob Shields, George Nemhauser, and David C Parkes. Automated channel abstraction for advertising auctions. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, pages 887–894. AAAI Press, 2010.

Fang Wu and Li Zhang. Proportional response dynamics leads to market equilibrium. In *Proceedings of the thirty-ninth annual ACM Symposium on Theory of Computing (STOC'07)*, pages 354–363, 2007.