New Convex Programs and Distributed Algorithms for Fisher Markets with Linear and Spending Constraint Utilities

Benjamin Birnbaum* University of Washington

Nikhil R. Devanur† Microsoft Research

Lin Xiao‡ Microsoft Research

Abstract

In this paper we shed new light on convex programs and distributed algorithms for Fisher markets with linear and spending constraint utilities.

- *We give a new convex program for the linear utilities case of Fisher markets. This program easily extends to the case of spending constraint utilities as well, thus resolving an open question raised by [Vaz10].*
- *We show that the gradient descent algorithm with respect to a Bregman divergence converges with rate $O(1/t)$ under a condition that is weaker than having Lipschitz continuous gradient (which is the usual assumption in the optimization literature for obtaining the same rate).*
- *We show that the* Proportional Response *dynamics recently introduced by Zhang [Zha09] is equivalent to a gradient descent algorithm for solving the new convex program. This insight also gives us better convergence rates, and helps us generalize it to spending constraint utilities.*

1 Introduction

Convex Programming based techniques have played an important role in the algorithmic study of market equilibria [DPSV08, Jai07, CPV05, Ye08] and (more recently) of Nash bargaining [Vaz09]. In this paper, we introduce a new convex program for Fisher markets with linear and spending constraint utilities. This is the first convex program known for spending constraint utilities, thus resolving an open problem of Vazirani [Vaz10]. We also show that this convex program explains the *Proportional Response* dynamics of Zhang [Zha09] as a generalized gradient descent algorithm with Bregman divergences (instead of Euclidean distance). Proportional Response is a simple distributed algorithm to compute market equilibrium. Designing such algorithms that converge quickly has been an important open problem, especially in the context where such algorithms are implemented by automated agents over a network such as the Internet. Our insight gives an intuitive understanding of why such dynamics work and also helps to improve previous results. Our work opens up many interesting directions for future research as well.

1.1 Fisher's Market Model with Linear and Spending Constraint Utilities

In the Fisher market model with linear utilities, there are n buyers and m perfectly divisible goods. There is a unit[1] supply of each good. Each buyer i has a budget B_i, and for each item j, a number u_{ij}, which represents the utility of that buyer for one unit of item j. Given an allocation vector $x \in \mathbb{R}_+^{n\times m}$, the utility of buyer i is $\sum_j u_{ij}x_{ij}$.

An allocation $x \in \mathbb{R}_+^{n\times m}$ and price vector $p \in \mathbb{R}_+^m$ are an *equilibrium* if two conditions are satisfied. The first, *buyer optimality*, is that given p, each player maximizes his utility subject to his budget constraint. In other words, for all i, the allocation x optimizes the following program (where p is a

*`birnbaum@cs.washington.edu`. Part of this work was performed while at Microsoft Research, Redmond.

†`nikdev@microsoft.com`.

‡`lin.xiao@microsoft.com`.

[1] This is without loss of generality, by the choice of unit, since the goods are divisible.

constant):

$$\begin{array}{ll} \text{maximize} & \sum_j u_{ij} x_{ij} \\ \text{subject to} & \sum_j p_j x_{ij} \leq B_i. \\ & x_{ij} \geq 0,\ \forall j. \end{array}$$

The second equilibrium condition, *market clearance*, is that $\sum_i x_{ij} = 1$ for all j.

In the *spending constraints* model, the utility of a buyer also depends on the prices, as follows. The utility is still additive across goods, that is, the total utility for a bundle of goods is the sum of utilities for each good separately. For a given good, the utility function is divided into segments; each segment l has a constant rate of utility u_{ij}^l, and a budget B_{ij}^l. The utility of the buyer for x_{ij}^l amount of good j under segment l is $u_{ij}^l x_{ij}^l$, but is subject to the constraint that $p_j x_{ij}^l \leq B_{ij}^l$. Hence the utility maximizing program for a given buyer is now as follows.

$$\begin{array}{ll} \text{maximize} & \sum_{j,l} u_{ij}^l x_{ij}^l \\ \text{subject to} & \sum_{j,l} p_j x_{ij}^l \leq B_i. \\ & 0 \leq x_{ij}^l \text{ and } p_j x_{ij}^l \leq B_{ij}^l\ \forall j, l. \end{array}$$

1.2 A New Convex Program

A well-known convex program called the Eisenberg-Gale [EG59] convex program captures equilibrium allocations of a linear Fisher market (see Appendix A for a formal definition). Here we introduce a new convex program that also does the same. As we shall see this convex program is closely related to the Proportional Response (PR) dynamics. The convex program is as follows.

$$\begin{array}{ll} \text{maximize} & \sum_{i,j} b_{ij} \log u_{ij} - \sum_j p_j \log p_j \\ \text{subject to} & \forall\ j, \sum_i b_{ij} = p_j, \\ & \forall\ i, \sum_j b_{ij} = B_i, \\ & \forall\ i,j,\ 0 \leq b_{ij}. \end{array} \tag{1}$$

The variable p_j corresponds to the price of good j. The variable b_{ij} represents the amount of money spent on good j by buyer i. Hence given a solution to the above program, the allocation x_{ij} is given by b_{ij}/p_j. It is easy to see that for any feasible solution to the program, the corresponding prices and allocations are such that the market always clears, and the buyers always exhaust their budgets. The only other condition required for equilibrium is that each buyer be allocated his optimal bundle of goods. We show in the following lemma that this indeed happens at equilibrium. The proof of this is in Appendix B.

Theorem 1. *Let $b^\star$ and $p^\star$ be an optimum solution to Convex Program (1). Then $p^\star$ and allocation $x_{ij}^\star = b_{ij}^\star / p_j^\star$ is an equilibrium for the Fisher market with linear utilities given by the u_{ij}'s and the B_i's.*

In fact, it is easy to generalize the above for markets with spending constraint utilities. This resolves an important open problem raised by Vazirani [Vaz10].

$$\begin{array}{ll} \text{maximize} & \sum_{i,j,l} b_{ij}^l \log u_{ij}^l - \sum_j p_j \log p_j \\ \text{subject to} & \forall\ j, \sum_{i,l} b_{ij}^l = p_j, \\ & \forall\ i, \sum_{j,l} b_{ij}^l = B_i, \\ & \forall\ i,j,l,\ 0 \leq b_{ij}^l \leq B_{ij}^l. \end{array} \tag{2}$$

As before, the allocation is given by $x_{ij}^l = b_{ij}^l / p_j$, and the only condition that needs verifying is buyer optimality.

Theorem 2. *An optimum solution to Convex Program (2) corresponds to an equilibrium for the Fisher market with spending constraint utilities given by the u_{ij}^l's, B_{ij}^l's and the B_i's.*

1.3 Generalized Gradient Descent with Bregman Divergence

Consider the constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in C \end{array} \tag{3}$$

where C is a compact convex set, and $f(x)$ is a differentiable convex function with $\operatorname{dom} f \supseteq C$ (i.e., f is finite on C). The conventional *projected gradient descent* method (e.g., [Ber99]) for solving (3) is

$$x(t+1) = \pi_C\big(x(t) - \mu_t \nabla f(x(t))\big) \tag{4}$$

where μ_t is an appropriate stepsize, and $\pi_C(\cdot)$ denotes Euclidean projection on the set C. Let

$$\ell_f(x; y) := f(y) + \langle \nabla f(y), x - y \rangle$$

be the linear approximation of f using the gradient at y. The method (4) is equivalent to

$$x(t+1) = \arg\min_{x \in C} \left\{ \ell_f(x; x(t)) + \frac{1}{2\mu_t} \|x - x(t)\|_2^2 \right\}, \tag{5}$$

where the *proximal* term $(1/2)\|x - x(t)\|_2^2$ is the squared Euclidean distance between two points. In the *generalized gradient descent* method the squared Euclidean distance in (5) is replaced by a generalized distance function $d(x, x(t))$, i.e.,

$$x(t+1) = \arg\min_{x \in C} \left\{ \ell_f(x; x(t)) + \gamma_t \, d(x, x(t)) \right\}, \tag{6}$$

where the parameter $\gamma_t > 0$, and its reciprocal $1/\gamma_t$ play the role of stepsize μ_t in the Euclidean case. In particular, we let $d(x, y)$ be the *Bregman divergence* [Bre67] of a differentiable convex function $h(x)$ (assuming $\operatorname{dom} h \supseteq C$), defined as

$$d_h(x, y) = h(x) - \ell_h(x; y), \quad \forall x \in C, \ y \in \operatorname{rint} C,$$

where $\operatorname{rint} C$ denotes the *relative interior* of C [Roc70, Section 6]. As an example, $d_h(x, y) = (1/2)\|x - y\|_2^2$ if $h(x) = (1/2)\|x\|_2^2$. We often drop the subscript h whenever there is no confusion from context. We say h is the *kernel* of d, and d is generated by h. We assume that both f and h are differentiable and they satisfy

$$f(x) \leq \ell_f(x; y) + \gamma d_h(x, y), \quad \forall x \in C, \ y \in \operatorname{rint} C, \tag{7}$$

for some constant $\gamma > 0$.

Theorem 3. *Suppose that the sequence $x(t)$ is defined by Equation (6), f and h are convex and differentiable functions and that they satisfy Condition (7). Then for all t,*

$$f(x(t)) - f(x^\star) \leq \frac{\gamma d(x^\star, x(0))}{t}.$$

In particular we show the following two lemmas from which the above theorem follows immediately.

Lemma 4. *The total deviation of f from the optimum throughout the run of the algorithm is bounded by a value that is independent of how long the algorithm is run.*

$$\sum_t \big(f(x(t)) - f(x^\star) \big) \leq \gamma d(x^\star, x(0)).$$

Lemma 5. *For all t, $f(x(t+1)) \leq f(x(t))$.*

1.4 Proportional Response Dynamics

The *Proportional Response (PR) Dynamics* is a distributed algorithm for computing market equilibrium and was introduced by Wu and Zhang [WZ07] in the context of a bandwidth trading market model for peer-to-peer networks and by Zhang [Zha09] in the current context of the Fisher markets. The algorithm proceeds in discrete time steps, and for each time $t = 0, 1, 2, \ldots$, buyer i submits a bid $b_{ij}(t)$ for each good j. Given the bids, a price vector $p(t)$ is computed by summing the bids for each item: $p_j(t) = \sum_i b_{ij}(t)$. Each buyer i is allocated an amount $x_{ij}(t) = b_{ij}(t)/p_j(t)$ of item j. The buyers then update their bids for the next time step so that their new bid for a good is proportional to the *utility obtained* from this good in the current round, that is $b_{ij}(t) \propto u_{ij} x_{ij}(t)$. In summary, the new bid vector is computed according to the following recursion:

$$b_{ij}(t) = B_i \frac{u_{ij} x_{ij}(t)}{\sum_{j'} u_{ij'} x_{ij'}(t)} \, . \tag{8}$$

One can immediately see the similarities between this and Convex Program (1) in the way the prices and allocations are defined. As before the market clears automatically and the only condition we need to worry about is buyer optimality. In fact as detailed below, we show that the PR dynamics is exactly the gradient descent algorithm given by (6) on Convex Program (1).

First of all, we can eliminate the p_j variables from the program; in fact, p_j can be thought of as a function of the b_{ij}'s, defined as $p_j(b) = \sum_i b_{ij}$. We stick to the notation p_j without the argument when there is no confusion[2] from the context. Let the objective function[3] of the convex program be denoted by $\varphi(b) = -\sum_{i,j} b_{ij} \log u_{ij} + \sum_j p_j(b) \log(p_j(b))$, and the feasible region by

$$S = \left\{ b \in \mathbb{R}^{n \times m} : \sum_j b_{ij} = B_i \ \forall \ i, b_{ij} \geq 0 \ \forall \ i, j \right\}.$$

[2]Similarly, we use $p_j(t)$ to denote $p_j(b(t))$ and $p_j^\star$ to denote $p_j(b^\star)$.

[3]To maintain consistency with our notation in the previous subsection, we rephrase the convex program as minimizing a convex function instead of maximizing a concave function.

Without loss of generality assume that $\sum_i B_i = 1$. Also define the Bregman divergence d with the kernel function being the unnormalized negative entropy, $h(b) = \sum_{i,j}(b_{ij} \log b_{ij} - b_{ij}) = \sum_{i,j} b_{ij} \log b_{ij} - \sum_i B_i$ for $b \in S$. A straightforward calculation shows that

$$d(a,b) = \sum_{i,j} a_{ij} \log \frac{a_{ij}}{b_{ij}} = D\left(a \,||\, b\right),$$

where $D\left(\cdot \,||\, \cdot\right)$ is the *KL-divergence* between two probability distributions.

Theorem 6. *The proportional response dynamics (8) is equivalent to the gradient descent algorithm (6) with $d(a,b) = D\left(a \,||\, b\right)$ as above, $f = \varphi$, $C = S$ and $\gamma = 1$.*

In order to use our results about convergence of the above algorithm (Theorem 3), we need to show that Condition (7) is satisfied, which we do in the following lemma.

Lemma 7. *For all $a, b \in S$,*

$$\varphi(b) \le \ell_\varphi(b; a) + D\left(b \,||\, a\right).$$

A more detailed analysis of the convergence is given in Section 4.

For markets with spending constraints, the algorithm (6) leads to an *iterative PR and capping* algorithm, which has the same convergence properties as the PR dynamics (see Section 3.1).

Moreover, Zhang [Zha09] also considers Constant Elasticity of Substitution (CES) utilities, $U_i = \sum_j (u_{ij} x_{ij})^{\rho_i}$ with $0 < \rho_i \le 1$. He gets a better bound on the convergence rate for these utilities: PR converges to an ϵ-approximate equilibrium in

$$O\left(\frac{L + \log(1/\epsilon)}{1 - (\max_i \rho_i)^2}\right)$$

rounds, where L is the bit-complexity of the input. Our techniques extend to CES utilities as well, and we strengthen Zhang's results as before, by obtaining the stronger form of convergence. The details of this will be presented in the full version.

1.5 Related Work and Motivation

The algorithmic study of equilibrium concepts in general, and of market equilibria in particular, are motivated by the question of whether markets can feasibly operate at equilibrium. After almost a decade since this line of research started [DPS02], the computational complexity of equilibrium concepts is fairly well understood. The hardness results [CT09, CDDT09, CD06, DGP06, CSVY06] point towards the unlikeliness that markets in general operate at equilibria, with recent results [CT09, CDDT09, VY09] showing that even for the simple class of separable piecewise linear functions, it is PPAD-hard to compute an equilibrium. The algorithmic results [DPSV08, Jai07, DV04, CPV05, CMV05, JV07, DK08, Ye08] provide hope for certain special classes of markets. A good survey of algorithms for market equilibria can be found in [Vaz07] and [CV07], with the summary being that we can find equilibria for markets with weak gross substitutes property, for certain resource allocation markets [JV07], and when there are a bounded number of goods [DK08].

An interesting special class of markets that is amenable to efficient algorithms is that of spending constraint utilities. The spending constraint utilities for the Fisher model was introduced by Vazirani [Vaz10] and extended to the more general Arrow-Debreu model by Devanur and Vazirani [DV04]. Vazirani [Vaz10] extended the primal-dual algorithm for linear utilities of Devanur et. al. [DPSV08] to the spending constraint utilities. The primal-dual interpretation of the algorithm of Devanur et. al. was based on the Eisenberg-Gale convex program, but no such program is known for spending constraints. Based on this, Vazirani [Vaz10] conjectured that there must exist a convex program that captures the spending constraint model as well. To quote,

> In our experience, non-trivial polynomial time algorithms for problems are rare and happen for a good reason - a deep mathematical structure intimately connected to the problem.

We finally resolve this conjecture positively in this paper.

The generalized gradient descent method, also known as *mirror-descent* has been well studied in the optimization community, see, e.g., [NY83, BT03]. In the optimization literature however, the kernel function h is usually assumed to be strongly convex. Then with a diminishing step-size $1/\gamma_t = O(1/\sqrt{t})$, the generalized gradient method (6) converges with rate $O(1/\sqrt{t})$, even if f is non-differentiable (e.g, [BT03, Nem05]). If f is further differentiable and ∇f is Lipschitz-continuous, then (6) converges with rate $O(1/t)$ using a constant γ that is no smaller than the Lipschitz constant. (e.g., [Ber99, Nes04]). In this case, more sophisticated variations of (6) can achieve a convergence rate $O(1/t^2)$; see, e.g., [Nes83, Nes04, Nes05, Tse08]. In this paper, we show that the method (6) can have convergence rate $O(1/t)$ with Condition (7). Condition (7) is implied by the assumptions that ∇f is Lipschitz continuous and h is strongly convex (e.g., [Nes04]) and is hence a weaker assumption. Without these stronger assumptions, Condition (7) often requires a close connection between the functions f and h, which is the case for our analysis of the PR dynamics in Section 3.

Algorithmic results in a centralized model of computation do not directly address the question of market *dynamics*: how might agents interacting in a market arrive at an equilibrium? Here, the quest is for simple and distributed algorithms that are guaranteed to converge fast. Such distributed algorithms are especially applicable when the agents involved are automated, and one has to prescribe a particular *protocol* for them to follow. This is true in many networking applications, and also in markets such as search advertising where often the bids are updated automatically. There is also a huge amount of work in the networking community on designing such distributed algorithms that achieve *proportional fairness* (e.g., [Kel97, KMT98]). It is known that proportional fairness is equivalent to market equilibrium in many settings (including linear utilities), via an Eisenberg-Gale type convex program [KV]. Our results could potentially be useful for these applications as well, and is a direction for future work.

Almost all dynamics considered in the literature are variants of the *tatonnement* process first defined by Leon Walras [Wal74]. The most general version of tatonnement just says that prices are increased when demand exceeds supply and are decreased when supply exceeds demand. Many versions of this [ABH59, Uza60, CMV05, FGK$^+$08, GK06] have been analyzed over the years, but none of these until recently gave simple, fast and distributed algorithms. Some of the early work in economics [ABH59, Uza60] ignored the convergence rate entirely. [CMV05] showed polynomial time convergence of a version of tatonnement for markets with weak gross substitutes, but had to transform the market *a priori* which then translates to passing around global information in every round. The auction algorithm of Garg and Kapoor [GK06] requires that it be started from a particular state. The algorithm of [FGK$^+$08] suffers from disadvantages such as having to take the average of all the prices and the average of all the allocations to get an equilibrium. The exception is the algorithm given by [CF08] which is distributed and asynchronous and converges fast for markets with weak gross substitutes property. The convergence time however depends on certain market parameters which tend to infinity for linear Fisher markets. One reason for the difficulty in designing tatonnement style processes for linear Fisher markets is that the demand of a buyer is not always uniquely determined. Thus at equilibrium one needs to specify the allocation in addition to the prices.

In contrast, the PR dynamics has many desirable properties that are also unique.

- The PR dynamics are simple, distributed and require no global communication.[4] In every step, a buyer only has to know the prices of the goods he is interested in, and a good only has to know the bids placed on it.

- PR dynamics is also *stateless*: the bids at any time step depend only on the bids of the previous step, and there is no special starting point.

- Unlike all other dynamics, which depend on choosing the right step size, PR dynamics does not need a step size to choose and adapt.

- Finally, PR dynamics elegantly takes care of the problem of non-uniqueness of demand, since the bids uniquely determine the prices and the allocations.

[4]Except for the fact that they need to be synchronized, in the sense that the dynamics happen in rounds and in every round, every bidder updates his bids.

PR was first proposed as a protocol for trading bandwidth on a peer-to-peer network by Wu and Zhang [WZ07] and has been shown to be an effective solution for BitTorrent [LLSB08]. PR seems to converge really fast to equilibrium in practice, and yet the reason for this remained a mystery. In a remarkable paper, Zhang [Zha09] shows convergence of PR dynamics in Fisher markets. The proof uses a potential function that gives little insight into why this happens, a trait that it shares with other distributed algorithms such as Cole and Fleischer's [CF08]. In this paper we *demystify* the PR dynamics by showing the equivalence to the generalized gradient descent algorithm for the new convex program. PR dynamics fit so well with this perspective that we can confidently say this is the "right" way of thinking about them.[5] Moreover, we get stronger convergence properties than obtained by Zhang [Zha09]. For instance, the convergence obtained by Zhang [Zha09] is of the form, "if τ is large enough, then there exists a $t < \tau$ such that $b(t)$ is close to optimal." We get the stronger form of convergence, of the form, "if t is large enough, then $b(t)$ is close to optimal." We give a detailed comparison to Zhang's results in Section 4.2. We are able to extend the results to spending constraint utilities as well. Clearly the extension was only possible because of our insights into the connection between PR dynamics and the convex program.

The design and analysis of distributed algorithms converging to equilibria in the context of games is also much studied, most commonly convergence of best response dynamics [AGM$^+$08, CS07, Ros73, AAE$^+$08]. Recently, similar questions have also been considered for network bargaining games [ABC$^+$09].

Organization: In Section 2 we prove Theorem 3. Theorem 6 is proved in Section 3. In Section 4, we define various measures of distance to equilibrium and show convergence of PR dynamics with respect to these measures. We also compare these in detail with Zhang's results in Section 4.2. We conclude with directions for future research in Section 5.

[5]It is natural to draw an analogy with the competitive analysis of online algorithms, many of which originally used a potential function. Later work [BN09] unified many of these as primal-dual algorithms, a framework which led to many new results.

2 Gradient Descent

In this section, we prove convergence properties of the generalized gradient method (6) based on the assumption (7) as stated in Theorem 3. We start by noting some key properties of Bregman divergences:

- $d(x, y) \geq 0$ for all $x, y \in \operatorname{dom} h$. If h is strictly convex, then $d(x, y) = 0$ if and only if $x = y$.
- In general $d(x, y) \neq d(y, x)$, and it does not satisfy the triangle inequality.
- The following *three-point identity* follows directly from definition:

$$d(x, z) = d(x, y) + d(y, z) + \langle \nabla h(y) - \nabla h(z),\, x - y \rangle. \tag{9}$$

If h is not strictly convex, then the solution to the minimization problem in (6) may not be unique. However, a solution always exists because we assume C is compact.

Recall that in order to prove Theorem 3, it is sufficient to prove Lemmas 5 and 4. We first prove Lemma 5.

Proof of Lemma 5. We have

$$\begin{aligned} f(x(t{+}1)) &\leq \ell_f(x(t{+}1); x(t)) + \gamma d(x(t{+}1), x(t)) \\ &\leq \ell_f(x(t); x(t)) + \gamma d(x(t), x(t)) \\ &= f(x(t)). \end{aligned}$$

The first inequality follows from (7). The second inequality is by definition of $x(t{+}1)$. The last equality is also by definition, since $d(x, x) = 0$. □

The key step in the proof of Lemma 4 is the following lemma. Let $x^\star := \arg\min_{x \in C}\{f(x)\}$ be an optimum solution to (3).

Lemma 8. *For all t,*

$$f(x(t{+}1)) - f(x^\star) \leq \gamma d(x^\star, x(t)) - \gamma d(x^\star, x(t{+}1)).$$

Before we prove Lemma 8, we show how Lemma 4 follows from it.

Proof of Lemma 4. The lemma follows by simply summing over all t the conclusion of Lemma 8 and using the fact that $d(x, y)$ is always greater than 0. □

In order to prove Lemma 8, we need a result on optimization with Bregman divergence. Consider the following optimization problem:

$$\begin{array}{ll} \text{minimize} & g(x) + d(x, y) \\ \text{subject to} & x \in C \end{array} \tag{10}$$

where $g(x)$ is a convex function and C is a compact convex set. The following lemma can be found in, e.g., [CT93]. We give the proof in the appendix for completeness.

Lemma 9. *If x^+ is the optimal solution to the optimization problem (10), then*

$$g(x) + d(x, y) \geq g(x^+) + d(x^+, y) + d(x, x^+). \tag{11}$$

We apply Lemma 9 with $g(x) = (1/\gamma)\ell_f(x; x(t))$ to obtain the following corollary.

Corollary 10. *For all $z \in C$,*

$$\begin{aligned} &\ell_f(x(t+1); x(t)) + \gamma d(x(t+1), x(t)) \\ &\quad \leq \ell_f(z; x(t)) + \gamma d(z, x(t)) - \gamma d(z, x(t+1)). \end{aligned}$$

Lemma 8 follows as a consequence of this corollary.

Proof of Lemma 8. We have

$$\begin{aligned} f(x(t+1)) &\leq \ell_f(x(t+1); x(t)) + \gamma d(x(t+1), x(t)) \\ &\leq \ell_f(x^\star; x(t)) + \gamma d(x^\star, x(t)) \\ &\quad - \gamma d(x^\star, x(t+1)) \\ &\leq f(x^\star) + \gamma d(x^\star, x(t)) - \gamma d(x^\star, x(t+1)). \end{aligned}$$

The first and third inequalities follow from (7), and the second inequality follows from Corollary 10. □

3 PR as Gradient Descent

In this section, we show the equivalence of the PR dynamics and the gradient descent algorithm on the new convex program, that is, we prove Theorem 6 and Lemma 7.

Recall the function $\varphi(b) = -\sum_{i,j} b_{ij} \log(u_{ij}/p_j)$ and feasible region S defined in the introduction. The components of the gradient of φ are given as

$$(\nabla\varphi(b))_{ij} = 1 - \log\left(\frac{u_{ij}}{p_j}\right).$$

It is clear that $\nabla\varphi$ is not Lipschitz continuous on S. Therefore we need the weaker assumption (7) to show $O(1/t)$ convergence.

The gradient descent algorithm (6) when applied to φ, S, with d being the KL-divergence and with $\gamma = 1$ is

$$b(t+1) := \arg\min_{a \in S} \left\{ \ell_\varphi(b(t); a) + D\left(a \,||\, b(t)\right) \right\}. \tag{12}$$

It is well known (and proved in Appendix B) that this update rule takes the following form:

$$\begin{aligned} b_{ij}(t) &= \frac{1}{Z_i'(t)} b_{ij}(t) \exp\left((\nabla\varphi(b(t)))_{ij} \right) \\ &= \frac{1}{Z_i(t)} b_{ij}(t) \left(\frac{u_{ij}}{p_j(t)} \right), \end{aligned}$$

where $Z_i(t)$ is chosen such that $\sum_{j=1}^n b_{ij}(t) = B_i$. This shows Theorem 6.

We now prove Lemma 7. The proof mainly relies on the following characterization of φ (whose proof is in Appendix B).

Lemma 11. *For all $a, b \in S$,*

$$\varphi(b) = \ell_\varphi(b; a) + D\left(p(b) \,||\, p(a)\right).$$

Proof of Lemma 7. The lemma follows from Lemma 11 and the fact that $D\left(p(b) \,||\, p(a)\right) \leq D\left(b \,||\, a\right)$, which we can prove using the convexity of the function $q(x, y) = x \log(x/y)$:

$$\begin{aligned} D(p(b) \,||\, p(a)) &= n \sum_j \left(\frac{1}{n} q(p_j(b), p_j(a)) \right) \\ &= n \sum_j q\left(\frac{1}{n} \sum_i b_{ij}, \frac{1}{n} \sum_i a_{ij} \right) \\ &\leq n \sum_j \frac{1}{n} \sum_i q(b_{ij}, a_{ij}) = D(b \,||\, a). \end{aligned}$$

This finishes the proof. □

3.1 Extension to Spending Constraint Utilities

For Fisher markets with spending constraint utilities, let φ be the negative objective function of the convex program (2):

$$\varphi(b) = -\sum_{i,j,l} b_{ij}^l \log\left(u_{ij}^l\right) + \sum_j p_j \log(p_j).$$

The components of its gradient is

$$\left(\nabla\varphi(b)\right)_{i,j,l} = 1 - \log\left(\frac{u_{ij}^l}{p_j}\right).$$

We use the same gradient descent method (12), but with the constraint set

$$S = \left\{a \,:\, \sum_{j,l} a_{ij}^l \le B_i,\ \forall\, i,\ \text{and } 0 \le a_{ij}^l \le B_{ij}^l,\ \forall\, i,j,l\right\}$$

Theorem 12. *Each bidding vector $b_i(t+1)$ can be computed separately by using the following iterative PR and capping algorithm.*

Algorithm: Iterative PR and Capping
input: B_i and B_{ij}^l, u_{ij}^l, and $b_{ij}^l(t)$ for all j and l.
initialize: let A_i be set of all pairs (j,l), and $C_i = \emptyset$.
repeat:

1. Let $A_i = A_i \setminus C_i$, $\bar{B}_i = \sum_{(j,l)\in A_i} B_{ij}^l$, and $Q_i = \sum_{(j,l)\in A_i}\left(u_{ij}^l/p_j(t)\right)b_{ij}^l(t)$.
2. Let
$$b_{ij}^l(t{+}1) = \begin{cases} \dfrac{\left(u_{ij}^l/p_j(t)\right)b_{ij}^l(t)}{Q_i}\bar{B}_i & \text{if } (j,l)\in A_i, \\ B_{ij}^l & \text{otherwise.}\end{cases}$$
3. Let $C_i = \left\{(j,l)\in A_i \,:\, b_{ij}^l(t+1) > B_{ij}^l\right\}$. If $C_i = \emptyset$, stop and return $b_i(t+1)$; otherwise continue.

This theorem is proved in the appendix.

4 Convergence Properties

The PR dynamics may never get to an exact equilibrium. Therefore we consider approximate equilibria, based on a measure of how close a given solution is to equilibrium. There are different ways to measure the "distance" from equilibria. We show that all of these can be related to $\varphi(b(t)) - \varphi(b^\star)$, the difference in the objective function between the current point and the optimum. (The additive difference is a reasonable measure since we normalized the sum of all budgets to 1.) Thus, any convergence bound we show with respect to the objective function can easily be translated to one for the particular measure of approximation.

The PR dynamics is defined for any initial bid vector b^0, and the convergence time is proportional to $D\left(b^\star \,||\, b(0)\right)$ (Theorem 3). For the sake of concreteness, we state our bounds assuming that $b_{ij}(0) = B^i/m$, that is each buyer initially divides his budget equally between all the items. With this, we have the following bound on $D\left(b^\star \,||\, b(0)\right)$, which is proved in Appendix B.

Lemma 13. *If $b_{ij}(0) = B_i/m$ for all i and j, then $D\left(b^\star \,||\, b(0)\right) \le \log(mn)$.*

We now have the following convergence results.

$$\varphi(b(t)) - \varphi(b^\star) \;\le\; \frac{\log(mn)}{t}. \qquad (13)$$

$$\sum_t \left(\varphi(b(t)) - \varphi(b^\star)\right) \;\le\; \log(mn). \qquad (14)$$

$$\varphi(b(t{+}1)) \;\le\; \varphi(b(t)). \qquad (15)$$

In other words, if $t \ge \frac{\log(mn)}{\epsilon}$ then $\varphi(b(t)) - \varphi(b^\star) \le \epsilon$.

4.1 Measures of Approximation

As mentioned before, we now show how to translate the convergence results mentioned above to other ways of measuring the distance from an equilibrium. Recall that we assumed without loss of generality, that the input has been scaled so that there is exactly one unit of each item and $\sum_i B_i = 1$. We may further assume that $\sum_j u_{ij} = 1$ for all i. Let $u_{\min} = \min_{i,j:u_{ij}>0} u_{ij}$ and $B_{\min} = \min_i B_i$.

- The KL-divergence is a well known (asymmetric) measure of dissimilarity between probability distributions, defined as $D\left(x \,||\, y\right) := \sum_i x_i \log(x_i/y_i)$. Since we normalized the total budget to be 1, the sum of the prices is also 1. In other words, one can think of the price vectors as probability distributions and use KL-divergence to measure the distance between them. Also it is well known that $\frac{1}{2}\left\|x-y\right\|_1^2 \le D\left(x \,||\, y\right)$. We show that
$$D\left(p(t) \,||\, p^\star\right) \le \varphi(b(t)) - \varphi(b^\star)\ . \qquad (16)$$
- Let ψ be the objective function of the Eisenberg-Gale convex program for x_{ij} =

b_{ij}/p_j. We can measure $\psi(b(t)) - \psi(b^\star) = \sum_i B_i \log(u_i^\star/u(t)_i)$, where $u_i^\star$ is the utility of buyer i at equilibrium. We show that

$$\psi(b(t)) - \psi(b^\star) \leq \varphi(b(t)) - \varphi(b^\star) \ . \quad (17)$$

- We can measure a relative notion of distance from the equilibrium price vector: $\eta := \max_j \left|\frac{p_j(t)-p_j^\star}{p_j^\star}\right|$. We show that

$$\eta^2 \leq O\left(\frac{n}{u_{\min}}\right)(\varphi(b(t)) - \varphi(b^\star)) \ . \quad (18)$$

- We can measure the maximum sub-optimality of the allocations. Let $u_i(t)$ be the utility of buyer i given his current allocation and $\tilde{u}_i^t$ be the maximum utility buyer i could have obtained if the prices were set to $p(t)$. We show that there exists a ζ such that $u_i(t) \geq (1-\zeta)\tilde{u}_i(t)$ for all i, and

$$\zeta^2 \leq O\left(\frac{n}{u_{\min}B_{\min}^2}\right)(\varphi(b(t)) - \varphi(b^\star)) \ . \quad (19)$$

- We can also measure an aggregate notion of suboptimality. To this end, let the vector δ be such that $u_i(t) = (1-\delta_i)\tilde{u}_i(t)$ for all i. Then we measure $\xi := \sum_i \delta_i B_i$. We show that

$$\xi^2 \leq O\left(\frac{n}{u_{\min}}\right)(\varphi(b(t)) - \varphi(b^\star)). \quad (20)$$

4.2 Comparison with Zhang's results

We now rephrase the main result in [Zha09] for ease of comparison. They follow the same framework as outlined earlier, first show convergence of a potential function and then relate that potential function to other measures of approximation. Whereas we show convergence of $\varphi(b(t)) - \varphi(b^\star)$, Zhang does the same for the quantity $\psi(b(t{+}1)) - \psi(b^\star) + D\left(p^\star \,||\, p(t)\right)$. His main result can be summarized by the following lemmas. (Recall that $\psi(b)$ is the value of the objective function of the Eisenberg-Gale convex program for $x_{ij} = b_{ij}/p_j$.) Let $p^\star_{\min} = \min_j p_j^\star$.

Lemma 14 ([Zha09])**.**

$$\sum_t \Big(\psi(b(t{+}1)) - \psi(b^\star) + D\left(p^\star \,||\, p(t)\right)\Big) \leq D\left(b^\star \,||\, b(0)\right).$$

Lemma 15 ([Zha09])**.** *Let* $\eta := \max_j \left|\frac{p_j(t)-p_j^\star}{p_j^\star}\right|$ *and assume* $\eta \leq 1$. *Then* $\eta^2 \leq 16\frac{D(p(t)\,||\,p^\star)}{p^\star_{\min}}$.

Lemma 16 ([Zha09])**.** *For all i, let $u_i(t)$ be buyer i's utility at time t, and let $\tilde{u}_i(t)$ be the maximum utility he can achieve given the prices $p(t)$. Then there exists a ζ such that $u_i(t) \geq (1-\zeta)\tilde{u}_i(t)$ for all i and*

$$\zeta^2 \leq \frac{256}{p^\star_{\min}B^2_{\min}}\Big(\psi(b(t)) - \psi(b^\star) + D\left(p(t) \,||\, p^\star\right)\Big) \ .$$

Lemma 17 ([Zha09])**.** $p^\star_{\min} \geq u_{\min}B_{\min}/m$.

However, it is not known if $\psi(b(t{+}1)) - \psi(b^\star) + D\left(p^\star \,||\, p(t)\right)$ is monotonically decreasing. So you simply obtain that for all τ, *there exists* a $t \leq \tau$ such that

$$\psi(b(t)) - \psi(b^\star) + D\left(p^\star \,||\, p(t)\right) \leq \frac{D\left(b^\star \,||\, b(0)\right)}{\tau}. \quad (21)$$

instead of obtaining that for all t

$$\psi(b(t)) - \psi(b^\star) + D\left(p^\star \,||\, p(t)\right) \leq \frac{D\left(b^\star \,||\, b(0)\right)}{t}.$$

In fact, we can derive all of Zhang's results in the stronger form by bounding each of $\psi(b(t)) - \psi(b^\star)$ and $D\left(p(t) \,||\, p^\star\right)$ separately by $\varphi(b(t)) - \varphi(b^\star)$ (and the fact that φ is monotonically decreasing).

Lemma 18. *For all t,* $D\left(p(t) \,||\, p^\star\right) \leq \varphi(b(t)) - \varphi(b^\star)$.

Lemma 19. *For all $b \in S$,* $\psi(b) - \psi(b^\star) \leq \varphi(b) - \varphi(b^\star)$.

We also have the following bound on the total deviation of the *symmetric* KL-divergence between $p^\star$ and $p(t)$.

Lemma 20.

$$\sum_t \Big(D\left(p^\star \,||\, p(t)\right) + D\left(p(t) \,||\, p^\star\right)\Big) \leq D\left(b^\star \,||\, b(0)\right) + \big(\varphi(b(0)) - \varphi(b^\star)\big).$$

Proofs are in Appendix B.

4.3 Improved bounds

The bounds mentioned in Section 4.1 (in particular, (18) and (19)) are stronger than the corresponding ones in [Zha09]. These are obtained by using Lemmas 15 and 16 with a better bound for $p^\star_{\min}$, which is as in the following lemma. (The proof is in Appendix B).

Lemma 21. $p^\star_{\min} \geq u_{\min}/n$.

Finally, we also give a bound on the aggregate notion of suboptimality of allocations (20) that is new.

Lemma 22. *Suppose* $\varphi(b(t)) - \varphi(b^\star) \leq u^2_{\min}/(8n^2)$. *Let the vector* δ *be such that*

$$u_i(t) = (1 - \delta_i)\tilde{u}_i(t),$$

where $u_i(t) = \sum_j u_{ij} b_{ij}(t)/p_j(t)$ *is the utility of buyer* i *at time* t *and* $\tilde{u}_i(t) = B_i \max_j(u_{ij}/p_j(t)) = B_i\beta_i(t)$ *is the maximum utility that buyer* i *could obtain given the prices* $p(t)$. *Then*

$$\left(\sum_i B_i\delta_i\right)^2 = O\left(\frac{n}{u_{\min}}\right)(\varphi(b(t)) - \varphi(b^\star)) \ .$$

Proof. To ease notation, we drop the argument t. For all i, j, let $\alpha_i = \log\beta_i = \max_j \log(u_{ij}/p_j)$, $\varepsilon_{ij} = \alpha_i - \log(u_{ij}/p_j)$, and $\varepsilon_i = \sum_j b_{ij}\varepsilon_{ij}/B_i$. Observe

$$\begin{aligned} u_i &= \sum_j b_{ij}\frac{u_{ij}}{p_j} = \sum_j b_{ij}\beta_i \exp(-\varepsilon_{ij}) \\ &\geq \sum_j b_{ij}\beta_i(1 - \varepsilon_{ij}) \\ &= B_i\beta_i - \beta_i\sum_j b_{ij}\varepsilon_{ij} \\ &= B_i\beta_i - B_i\beta_i\varepsilon_i = (1-\varepsilon_i)\tilde{u}'_i, \end{aligned}$$

and hence $\delta_i \leq \varepsilon_i$ for all i. Therefore,

$$\sum_i B_i\delta_i \leq \sum_i B_i\varepsilon_i = \sum_{i,j} b_{ij}\varepsilon_{ij} \ ,$$

and it suffices to bound $\sum_{i,j} b_{ij}\varepsilon_{ij}$. In particular, we show that $\sum_{i,j} b_{ij}\varepsilon_{ij} \leq O(\eta) + \varepsilon$. The proof then follows immediately from (18). To do this, we start by rewriting the objective function $\varphi(b)$ as

$$\begin{aligned} \varphi(b) &= -\sum_{i,j} b_{ij}\left(\log\frac{u_{ij}}{p_j} - \alpha_i + \alpha_i\right) \\ &= \sum_{i,j} b_{ij}\varepsilon_{ij} - \sum_{i,j} b_{ij}\alpha_i \\ &= \sum_{i,j} b_{ij}\varepsilon_{ij} - \sum_i B_i\alpha_i. \end{aligned}$$

Let $\alpha^\star_i = \max_j \log(u_{ij}/p^\star_j)$. The optimality condition states that $\varphi(b^\star) = -\sum_i B_i\alpha^\star_i$. Let $\varepsilon = \varphi(b) - \varphi(b^\star)$. Then

$$\sum_{i,j} b_{ij}\varepsilon_{ij} = \sum_i B_i(\alpha_i - \alpha^\star_i) + \varepsilon \ .$$

Since $\sum_i B_i = 1$, we have $\sum_i B_i(\alpha_i - \alpha^\star_i) \leq \max_i(\alpha_i - \alpha^\star_i)$. Furthermore, for all i,

$$\begin{aligned} \alpha_i - \alpha^\star_i &= \max_j \log\frac{u_{ij}}{p_j} - \max_j \log\frac{u_{ij}}{p^\star_j} \\ &= \max_j\left(\log\frac{u_{ij}}{p^\star_j} + \log\frac{p^\star_j}{p_j}\right) - \max_j \log\frac{u_{ij}}{p^\star_j} \\ &\leq \max_j \log\frac{u_{ij}}{p^\star_j} + \max_j \log\frac{p^\star_j}{p_j} - \max_j\log\frac{u_{ij}}{p^\star_j} \\ &= \max_j \log\frac{p^\star_j}{p_j} \leq O(\eta). \end{aligned}$$

□

5 Conclusion and Future Work

Our work opens up many interesting directions for future research.

- Eisenberg-Gale-type convex programs also capture equilibria for other classes of markets, especially for many resource allocation markets [JV07] and markets with Leontief utilities. It is not clear if there are analogs of our convex program for these markets. Such programs might lead to interesting algorithms and dynamics for these other markets as well. Also such programs have been used to design distributed algorithms to achieve proportional fairness. It is conceivable that our new convex program and/or the PR dynamics could get better results for the same.

- Can we get a better convergence bound for PR dynamics? While it seems difficult to get a convergence rate better than $O(1/\epsilon)$ w.r.t φ, it is possible one could get faster rates w.r.t other notions of approximate equilibria that we discuss.

- Can we design other similar dynamics that converge faster? We note that under the stronger condition of Lipschitz continuity of gradients, there are gradient algorithms [Nes83, Nes05, Tse08] that converge in time $O(1/\sqrt{\epsilon})$ (as opposed to the one we use which converges in time $O(1/\epsilon)$). However, these algorithms do not seem to work with the weaker assumption that we have. It would be very interesting (from a more general convex programming point of view) to get an $O(1/\sqrt{\epsilon})$ time algorithm with our weaker assumption.

- Is there any relation to the primal-dual algorithms of [DPSV08] and [Vaz10] with our convex program?

- There is a natural asynchronous version of the PR dynamics: in each iteration, the bids of a single buyer and the corresponding prices are updated, with some non-saturation condition saying that every buyer updates his bids frequently enough. There are also randomized versions where a buyer wakes up with a Poisson clock and updates his bids and the corresponding prices. We believe our techniques (perhaps combined with some of the techniques from the book by Bertsekas and Tsitsiklis [BT97]) could be extended to handle these cases as well.

References

[AAE+08] Baruch Awerbuch, Yossi Azar, Amir Epstein, Vahab S. Mirrokni, and Alexander Skopalik. Fast convergence to nearly optimal solutions in potential games. In *ACM Conference on Electronic Commerce*, pages 264–273, 2008.

[ABC+09] Yossi Azar, Benjamin E. Birnbaum, L. Elisa Celis, Nikhil R. Devanur, and Yuval Peres. Convergence of local dynamics to balanced outcomes in exchange networks. In *FOCS*, 2009.

[ABH59] K. Arrow, H. Block, and L. Hurwicz. On the stability of the competitive equilibrium: II. *Econometrica*, 27(1):82–109, 1959.

[AGM+08] Heiner Ackermann, Paul W. Goldberg, Vahab S. Mirrokni, Heiko Röglin, and Berthold Vöcking. Uncoordinated two-sided matching markets. In *ACM Conference on Electronic Commerce*, pages 256–263, 2008.

[Ber99] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 2nd edition, 1999.

[BN09] Niv Buchbinder and Joseph Naor. The design of competitive online algorithms via a primal-dual approach. *Foundations and Trends in Theoretical Computer Science*, 3(2-3):93–263, 2009.

[Bre67] L. Bregman. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *U.S.S.R. Comput. Math. and Math. Phys.*, 7:200–217, 1967.

[BT97] Dimitri P. Bertsekas and John N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods.* Athena Scientific, 1997.

[BT03] A. Beck and M. Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31:167–175, 2003.

[CD06] Xi Chen and Xiaotie Deng. Settling the complexity of two-player Nash equilibrium. In *FOCS*, pages 261–272, 2006.

[CDDT09] Xi Chen, Decheng Dai, Ye Du, and Shang-Hua Teng. Settling the complexity of Arrow-Debreu equilibria in markets with additively separable utilities. In *FOCS*, 2009.

[CF08] Richard Cole and Lisa Fleischer. Fast-converging tatonnement algorithms for one-time and ongoing market problems. In *STOC*, pages 315–324, 2008.

[CMV05] Bruno Codenotti, Benton McCune, and Kasturi R. Varadarajan. Market equilibrium via the excess demand function. In *STOC*, pages 74–83, 2005.

[CPV05] Bruno Codenotti, Sriram V. Pemmaraju, and Kasturi R. Varadarajan. On the polynomial time computation of equilibria for certain exchange economies. In *SODA*, pages 72–81, 2005.

[CS07] Steve Chien and Alistair Sinclair. Convergence to approximate Nash equilibria in congestion games. In *SODA*, pages 169–178, 2007.

[CSVY06] Bruno Codenotti, Amin Saberi, Kasturi R. Varadarajan, and Yinyu Ye. Leontief economies encode nonzero sum two-player games. In *SODA*, pages 659–667, 2006.

[CT93] G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, August 1993.

[CT09] Xi Chen and Shang-Hua Teng. Spending is not easier than trading: On the computational equivalence of Fisher and Arrow-Debreu equilibria. *CoRR*, abs/0907.4130, 2009.

[CV07] Bruno Codenotti and Kasturi Varadarajan. *Algorithmic Game Theory*, chapter Computation of Market Equilibria by Convex Programming. Cambridge University Press, 2007.

[DGP06] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. In *STOC*, pages 71–78, 2006.

[DK08] Nikhil R. Devanur and Ravi Kannan. Market equilibria in polynomial time for fixed number of goods or agents. In *FOCS*, pages 45–53, 2008.

[DPS02] Xiaotie Deng, Christos H. Papadimitriou, and Shmuel Safra. On the complexity of equilibria. In *STOC*, pages 67–71, 2002.

[DPSV08] Nikhil R. Devanur, Christos H. Papadimitriou, Amin Saberi, and Vijay V. Vazirani. Market equilibrium via a primal–dual algorithm for a convex program. *J. ACM*, 55(5), 2008.

[DV04] Nikhil R. Devanur and Vijay V. Vazirani. The spending constraint model for market equilibrium: algorithmic, existence and uniqueness results. In *STOC*, pages 519–528, 2004.

[EG59] Edmund Eisenberg and David Gale. Consensus of subjective probabilities: The pari-mutuel method. *Ann. Math. Statist.*, 30(1):165–168, 1959.

[FGK+08] Lisa Fleischer, Rahul Garg, Sanjiv Kapoor, Rohit Khandekar, and Amin Saberi. A fast and simple algorithm for computing market equilibria. In *WINE*, pages 19–30, 2008.

[GK06] Rahul Garg and Sanjiv Kapoor. Auction algorithms for market equilibrium. *Math. Oper. Res.*, 31(4):714–729, 2006.

[Jai07] Kamal Jain. A polynomial time algorithm for computing an Arrow-Debreu market equilibrium for linear utilities. *SIAM J. Comput.*, 37(1):303–318, 2007.

[JV07] Kamal Jain and Vijay V. Vazirani. Eisenberg-Gale markets: algorithms and structural properties. In *STOC*, pages 364–373, 2007.

[Kel97] F. P. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8:33–37, 1997.

[KMT98] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[KV] F. P. Kelly and V. V. Vazirani. Rate control as a market equilibrium. Manuscript, 2002.

[LLSB08] Dave Levin, Katrina LaCurts, Neil Spring, and Bobby Bhattacharjee. Bittorrent is an auction: analyzing and improving bittorrent's incentives. In *SIGCOMM '08: Proceedings of the ACM SIGCOMM 2008 conference on Data communication*, pages 243–254, New York, NY, USA, 2008. ACM.

[Nem05] A. Nemirovski. Lectures on modern convex optimization. Lecture notes, Georgia Institute of Technology, 2005.

[Nes83] Yu. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$. *Soviet Math. Docl. (Doklady AN SSSR)*, 269:543–547, 1983.

[Nes04] Yu. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course.* Kluwer, Boston, 2004.

[Nes05] Yu. Nesterov. Smooth minimization of nonsmooth functions. *Mathematical Programming*, 103(1):127–152, 2005.

[NY83] A. Nemirovsky and D. Yudin. *Problem Complexity and Method Efficiency in Optimization.* J. Wiley & Sons, New York, 1983.

[Roc70] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[Ros73] R.W. Rosenthal. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.

[Tse08] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. Submitted to *SIAM Journal on Optimization*, 2008.

[Uza60] H. Uzawa. Walras' tatonnement in the theory of exchange. *Review of Economic Studies*, 27:182–94, 1960.

[Vaz07] Vijay V. Vazirani. *Algorithmic Game Theory*, chapter Combinatorial Algorithms for Market Equilibria. Cambridge University Press, 2007.

[Vaz09] Vijay V. Vazirani. Nash bargaining via flexible budget markets. Manuscript available from the author's homepage, 2009.

[Vaz10] Vijay V. Vazirani. Spending constriant utilities, with applications to the adwords market. *To appear in Math of OR*, 2010. A manuscript has been publicly available since 2003.

[VY09] Vijay V. Vazirani and Mihalis Yannakakis. Market equilibrium under separable, piecewise-linear, concave utilities. In *ICS*, 2009.

[Wal74] L. Walras. *Éléments d'économie politique pure ou théorie de la richesse sociale (Elements of Pure Economics, or the theory of social wealth)*. Lausanne, Paris, 1874. (1899, 4th ed.; 1926, rev ed., 1954, Engl. transl.).

[WZ07] Fang Wu and Li Zhang. Proportional response dynamics leads to market equilibrium. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 354–363, New York, NY, USA, 2007. ACM.

[Ye08] Y. Ye. A path to the Arrow-Debreu competitive market equilibrium. *Mathematical Programming*, 111:315–348, 2008.

[Zha09] Li Zhang. Proportional response dynamics in the Fisher market. In *ICALP '09: Proceedings of the 36th Internatilonal Colloqquium on Automata, Languages and Programming*, pages 583–594, Berlin, Heidelberg, 2009. Springer-Verlag.

A The Eisenberg-Gale Convex Program

The Eisenberg-Gale Convex Program is as follows

$$\begin{array}{ll} \text{minimize} & -\sum_i B_i \log(u_i) \\ \text{subject to} & \forall\, i, u_i = \sum_j u_{ij} x_{ij}. \\ & \forall\, j, \sum_i x_{ij} \le 1. \\ & \forall\, i, j, x_{ij} \ge 0. \end{array}$$

We denote by $\psi(b)$ the objective function of the Eisenberg-Gale Convex Program as a function of the bids, by setting $x_{ij} = b_{ij}/p_j$. Note that this satisfieses $\sum_i x_{ij} \le 1$ for all j. Also u_i is then $\sum_j \frac{u_{ij} b_{ij}}{p_j}$. Hence

$$\psi(b) = -\sum_i B_i \log\left(\sum_j \frac{u_{ij} b_{ij}}{p_j}\right).$$

B Missing Proofs

Proof of Theorem 1. By definition $\sum_i x_{ij} = \sum_i b^\star_{ij}/p_j = 1$, so the market clears. A straightforward calculation shows that

$$\frac{\partial \varphi(b)}{\partial b_{ij}} = 1 - \log \frac{u_{ij}}{p_j}.$$

From this, we can derive using the KKT conditions for $b^\star$ that for each buyer i, there is a number λ_i such that $\log(u_{ij}/p_j) - 1 \le \lambda_i$ for all j, and if $b^\star_{ij} > 0$, then $\log(u_{ij}/p_j) - 1 = \lambda_i$. Since there must be at least one good j such that $b^\star_{ij} > 0$, this implies that $\lambda_i = \max_j(\log(u_{ij}/p_j) - 1)$ and $\exp(\lambda_i + 1) = \beta_i$, the maximum bang-for-buck that buyer i can get. The KKT conditions can now be reinterpreted as saying that each buyer only bids on goods that maximize his bang for buck, which implies the buyer optimality condition for market equilibria. Hence, both conditions are satisfied, and (x, p) is a market equilibrium. □

Proof of Theorem 2. This is almost identical to the Proof of Theorem 1. □

Proof of Lemma 9. The optimality condition for x^+ states that there is a subgradient $\nu \in \partial\psi(x^+)$ such that

$$\langle \nu + \nabla d(x^+, y),\, x - x^+ \rangle \ge 0, \qquad \forall x \in C,$$

where $\nabla d(x^+, y) = \nabla h(x^+) - \nabla h(y)$. Now using the triangle identity, we have for all $x \in C$,

$$\begin{aligned} g(x) + d(x, y) &= g(x) + \langle \nabla h(x^+) - \nabla h(y),\, x - x^+\rangle \\ &\quad + d(x, x^+) + d(x^+, y) \\ &\geq g(x^+) + \langle \nu + \nabla h(x^+) - \nabla h(y),\, x - x^+\rangle \\ &\quad + d(x, x^+) + d(x^+, y) \\ &\geq g(x^+) + d(x^+, y) + d(x, x^+), \end{aligned}$$

where in the first inequality above we used convexity of g, and in the second inequality we used the optimality condition of x^+. □

Proof of Lemma 11. We have

$$\begin{aligned} \varphi(b) - \ell_\varphi(b; a) &= \varphi(b) - \varphi(a) - \langle \nabla\varphi(a),\, b - a\rangle \\ &= -\sum_{i,j} b_{ij} \log \frac{u_{ij}}{p_j(b)} + \sum_{i,j} a_{ij} \log \frac{u_{ij}}{p_j(a)} \\ &\quad - \sum_{i,j} \left(1 - \log \frac{u_{ij}}{p_j(a)}\right)(b_{ij} - a_{ij}) \\ &= \sum_{i,j} b_{ij} \log \frac{p_j(b)}{p_j(a)} = D\left(p(b)\,||\,p(a)\right), \end{aligned}$$

where in the third equality we used $\sum_{i,j} b_{ij} = \sum_{i,j} a_{ij} = 1$. □

Proof of Theorem 6. The following standard proof can be found in, for example [Nes05]. For each $i = 1, \ldots, n$,

$$\begin{aligned} b_i(t+1) = \operatorname*{arg\,min}_{a_i : \sum_j a_{ij} = B_i} \Bigg\{ &\sum_j a_{ij} \big(\nabla\varphi(b(t))\big)_{ij} \\ &+ \sum_j a_{ij} \log \frac{a_{ij}}{b_{ij}(t)} \Bigg\}. \end{aligned}$$

Note that we do not need the nonnegative constraints because log can serve as the natural barrier for nonnegativity. Introducing λ_i as the Lagrange multiplier for the constraint $\sum_j a_{ij} = B_i$, we have the Lagrangian

$$\begin{aligned} L(a_i, \lambda) = &\sum_j a_{ij} \big(\nabla\varphi(b(t))\big)_{ij} \sum_j a_{ij} \log \frac{a_{ij}}{b_{ij}(t)} \\ &+ \lambda_i \left(\sum_j a_{ij} - 1\right). \end{aligned}$$

Taking the derivative with respect to a_{ij} and setting it to zero yields

$$\big(\nabla\varphi(b(t))\big)_{ij} + \log \frac{a_{ij}}{b_{ij}(t)} + 1 + \lambda_i = 0,$$

then solving for a_{ij} gives

$$\begin{aligned} a_{ij} &= b_{ij}(t) \exp\left(-\big(\nabla\varphi(b(t))\big)_{ij} - 1 - \lambda_i\right) \\ &= b_{ij}(t) \exp\left(-\left(1 - \log \frac{u_{ij}}{p_j}\right) - 1 - \lambda_i\right) \\ &= \frac{1}{\exp(2 + \lambda_i)} b_{ij}(t) \frac{u_{ij}}{p_j} \end{aligned}$$

Then $b_{ij}(t)$ is the a_{ij} when λ_i is adjusted such that the sum over j equals B_i. □

Proof of Theorem 12. The minimization problem (12) can be solved separately for each agent i,

$$\begin{aligned} b_i(t+1) = \operatorname*{arg\,min}_{\substack{\sum_{j,l} a_{ij}^l \leq B_i, \\ 0 \leq a_{ij}^l \leq B_{ij}^l,\ \forall j,l}} \Bigg\{ &\sum_{j,l} a_{ij}^l \big(\nabla\varphi(b(t))\big)_{i,j,l} \\ &+ \sum_{j,l} a_{ij}^l \log\left(\frac{a_{ij}^l}{b_{ij}^l(t)}\right) \Bigg\}. \end{aligned}$$

To solve this problem, we introduce Lagrange multipliers λ_i and μ_{ij}^l. The Lagrangian is

$$\begin{aligned} &L_i(a_i, \lambda_i, \mu_i) \\ &= \sum_{j,l} a_{ij}^l \big(\nabla\varphi(b(t))\big)_{i,j,l} + \sum_{j,l} a_{ij}^l \log\left(\frac{a_{ij}^l}{b_{ij}^l(t)}\right) \\ &\quad + \lambda_i \left(\sum_{j,l} a_{ij}^l - B_i\right) + \sum_{j,l} \mu_{ij}^l \left(a_{ij}^l - B_{ij}^l\right) \\ &= \sum_{j,l} a_{ij}^l \left(\big(\nabla\varphi(b(t))\big)_{i,j,l} + \lambda_i + \mu_{ij}^l\right) \\ &\quad + \sum_{j,l} a_{ij}^l \log\left(\frac{a_{ij}^l}{b_{ij}^l(t)}\right) - \lambda_i B_i - \sum_{j,l} \mu_{ij}^l B_{ij}^l. \end{aligned}$$

Taking derivatives with respect to a_{ij}^l and setting them to be zero, i.e., let $\big(\nabla L_i(a_i, \lambda_i, \mu_i)\big)_{jl} = 0$, we have

$$\big(\nabla\varphi(b(t))\big)_{i,j,l} + \lambda_i + \mu_{ij}^l + \log\left(\frac{a_{ij}^l}{b_{ij}^l(t)}\right) + 1 = 0.$$

The solution is

$$a_{ij}^l = \frac{1}{\exp\bigl(2 + \lambda_i + \mu_{ij}^l\bigr)} \frac{u_{ij}^l b_{ij}^l(t)}{p_j(t)}.$$

Using the optimality conditions (KKT conditions),

$$\mu_{ij}^l \left(a_{ij}^l - B_{ij}^l \right) = 0, \quad \forall\, i, j, l,$$

we can find the optimal solution $b_i(t+1)$ using an iterative PR and capping algorithm. □

Proof of Lemma 14 in our framework. First by using Corollary 10, we have

$$\begin{aligned}\varphi(b(t{+}1)) &\le \ell_\varphi(b(t{+}1); b(t)) + D\left(b(t{+}1) \,||\, b(t)\right)\\ &\le \ell_\varphi(b^\star; b(t)) + D\left(b^\star \,||\, b(t)\right) - D\left(b^\star \,||\, b(t{+}1)\right).\end{aligned}$$

Next using Lemma 11, i.e., $\varphi(b^\star) = \ell_\varphi(b^\star; b(t)) + D\left(p^\star \,||\, p(t)\right)$, we get

$$\begin{aligned}&\varphi(b(t{+}1)) - \varphi(b^\star) + D\left(p^\star \,||\, p(t)\right)\\ &\qquad \le D\left(b^\star \,||\, b(t)\right) - D\left(b^\star \,||\, b(t{+}1)\right).\end{aligned}$$

Summing the above inequality over all t and noticing $D(b^\star||b(t+1)) \ge 0$ lead to

$$\sum_t \Big(\varphi(b(t{+}1)) - \varphi(b^\star) + D\left(p^\star \,||\, p(t)\right) \Big) \le D\left(b^\star \,||\, b(0)\right).$$

Further using Lemma 19 finishes the proof. □

Proof of Lemma 13. We have

$$\begin{aligned}D\left(b^\star \,||\, b(0)\right) &= \sum_{i,j} b_{ij}^\star \log \frac{b_{ij}^\star}{b(0)_{ij}} = \sum_{i,j} b_{ij}^\star \log \frac{m b_{ij}^\star}{B_i}\\ &= \log m + \sum_{i,j} b_{ij}^\star \log b_{ij}^\star - \sum_{i,j} b_{ij}^\star \log B_i\\ &= \log m - \sum_i B_i \log B_i + \sum_{i,j} b_{ij}^\star \log b_{ij}^\star\\ &\le \log m - \sum_i B_i \log B_i\\ &\le \log m + \log n.\end{aligned}$$

□

Proof of Lemma 18. By Lemma 11,

$$\begin{aligned}\varphi(b(t)) &= \ell_\varphi(b(t); b^\star) + D\left(p(t) \,||\, p^\star\right)\\ &= \varphi(b^\star) + \langle \nabla\varphi(b^\star),\, b(t) - b^\star \rangle + D\left(p(t) \,||\, p^\star\right)\end{aligned}$$

By optimality, $\langle \nabla\varphi(b^\star),\, b(t) - b^\star \rangle \ge 0$, so the lemma follows. □

Proof of Lemma 19. It suffices to show

$$\psi(b) \le \varphi(b) - \sum_i B_i \log B_i$$

and

$$\psi(b^\star) = \varphi(b^\star) - \sum_i B_i \log B_i.$$

For the inequality above, we use convexity of $-\log$:

$$\begin{aligned}\psi(b) &= -\sum_i B_i \log \left(\sum_j u_{ij} \frac{b_{ij}}{p_j} \right)\\ &= -\sum_i B_i \log \left(\sum_j \frac{u_{ij}}{p_j} \frac{b_{ij}}{B_i} \right) - \sum_i B_i \log B_i\\ &\le -\sum_i B_i \sum_j \frac{b_{ij}}{B_i} \log \frac{u_{ij}}{p_j} - \sum_i B_i \log B_i\\ &= -\sum_{i,j} b_{ij} \log \frac{u_{ij}}{p_j} - \sum_i B_i \log B_i.\end{aligned}$$

For the equality at $b^\star$, we use the optimality conditions:

$$\begin{aligned}\psi(b^\star) &= -\sum_i B_i \log \left(\sum_j \frac{u_{ij}}{p_j^\star} \frac{b_{ij}^\star}{B_i} \right) - \sum_i B_i \log B_i\\ &= -\sum_i B_i \log \left(\max_j \frac{u_{ij}}{p_j^\star} \right) - \sum_i B_i \log B_i\\ &= -\sum_{i,j} b_{ij}^\star \log \frac{u_{ij}}{p_{ij}^\star} - \sum_i B_i \log B_i.\end{aligned}$$

□

Proof of Lemma 20. First using Corollary 10, we have

$$\begin{aligned}\varphi(b(t{+}1)) &\le \ell_\varphi(b(t{+}1); b(t)) + D\left(b(t{+}1) \,||\, b(t)\right)\\ &\le \ell_\varphi(b^\star; b(t)) + D\left(b^\star \,||\, b(t)\right) - D\left(b^\star \,||\, b(t{+}1)\right).\end{aligned}$$

Plugging in $\ell_\varphi(b^\star; b(t)) = \varphi(b(t)) + \langle \nabla\varphi(b(t)), b^\star - b(t) \rangle$ gives

$$\begin{aligned}\langle \nabla\varphi(b(t)), b(t) - b^\star \rangle &\le D\left(b^\star \,||\, b(t)\right) - D\left(b^\star \,||\, b(t{+}1)\right)\\ &\quad + \varphi(b(t)) - \varphi(b(t{+}1)).\end{aligned}$$

The optimality of $b^\star$ implies $\langle \nabla\varphi(b^\star), b(t) - b^\star \rangle \ge 0$. Subtracting $\langle \nabla\varphi(b^\star), b(t) - b^\star \rangle$ from the left hand gives

$$\begin{aligned}&\langle \nabla\varphi(b(t)) - \nabla\varphi(b^\star), b(t) - b^\star \rangle\\ &\le D\left(b^\star \,||\, b(t)\right) - D\left(b^\star \,||\, b(t{+}1)\right) + \varphi(b(t)) - \varphi(b(t{+}1)).\end{aligned}$$

Using $\nabla\varphi(b) = 1 - \log(u_{ij}/p_j)$, we have

$$\begin{aligned}
&\langle \nabla\varphi(b(t)) - \nabla\varphi(b^\star), b(t) - b^\star \rangle \\
=&\sum_{i,j} (b_{ij}(t) - b^\star_{ij}) \log \frac{p_j(t)}{p^\star_j} \\
=&\sum_{j} (p_j(t) - p^\star_j) \log \frac{p_j(t)}{p^\star_j} \\
=&D\left(p(t) \,||\, p^\star\right) + D\left(p^\star \,||\, p(t)\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&D\left(p(t) \,||\, p^\star\right) + D\left(p^\star \,||\, p(t)\right) \\
\le&D\left(b^\star \,||\, b(t)\right) - D\left(b^\star \,||\, b(t{+}1)\right) + \varphi(b(t)) - \varphi(b(t{+}1)).
\end{aligned}$$

Summing over all t gives the desired result. $\square$

Proof of Lemma 21. Let $b^\star$ be an equilibrium bid vector, and let $\beta^\star_i = \max_j u_{ij}/p^\star_j$. Suppose for contradiction that there exists an item $\tilde{j}$ such that $p^\star_{\tilde{j}} < u_{\min}/n$. Then for all bidders i,

$$\beta^\star_i \ge \frac{u_{i\tilde{j}}}{p^\star_{\tilde{j}}} \ge \frac{u_{\min}}{p^\star_{\tilde{j}}} > n \ .$$

By the buyer optimality condition, $b^\star_{ij} > 0$ implies $u_{ij}/p^\star_j = \beta^\star_i$. Combining this with the above yields the condition

$$b^\star_{ij} > 0 \implies p^\star_j < \frac{u_{ij}}{n} \ .$$

We now obtain the contradiction

$$\begin{aligned}
1 = &\sum_{i,j \,:\, b^\star_{ij}>0} b^\star_{ij} \le \sum_{i,j \,:\, b^\star_{ij}>0} p^\star_j \\
< &\sum_{i,j \,:\, b^\star_{ij}>0} \frac{u_{ij}}{n} \le \frac{1}{n} \sum_{i,j} u_{i,j} = 1.
\end{aligned}$$

Hence, we conclude that there can be no such item $\tilde{j}$, and therefore $p^\star_{\min} \ge u_{\min}/n$. $\square$