
First-Order Methods for Large-Scale Market Equilibrium Computation

Yuan Gao

Department of IEOR, Columbia University
New York, NY, 10027
gao.yuan@columbia.edu

Christian Kroer

Department of IEOR, Columbia University
New York, NY, 10027
christian.kroer@columbia.edu

Abstract

Market equilibrium is a solution concept with many applications such as digital ad markets, fair division, and resource sharing. For many classes of utility functions, equilibria can be captured by convex programs. We develop simple first-order methods suitable for solving these convex programs for large-scale markets. We focus on three practically-relevant utility classes: linear, quasilinear, and Leontief utilities. Using structural properties of market equilibria under each utility class, we show that the corresponding convex programs can be reformulated as optimization of a structured smooth convex function over a polyhedral set, for which projected gradient achieves linear convergence. To do so, we utilize recent linear convergence results under weakened strong-convexity conditions, and further refine the relevant constants in existing convergence results. Then, we show that proximal gradient (a generalization of projected gradient) with a practical linesearch scheme achieves linear convergence under the *Proximal-PL* condition, a recently developed error bound condition for convex composite problems. For quasilinear utilities, we show that Mirror Descent applied to a new convex program achieves sublinear last-iterate convergence and yields a form of Proportional Response dynamics, an elegant, interpretable algorithm for computing market equilibria originally developed for linear utilities. Numerical experiments show that Proportional Response dynamics is highly efficient for computing approximate market equilibria, while projected gradient with linesearch can be much faster when higher-accuracy solutions are needed.

1 Introduction

Market equilibrium is a classical model from economics, where the decision of who gets what and why is based on finding a set of market-clearing prices such that every market participant gets allocated an optimal bundle given the prices and their budgets. In this paper, we study the *Fisher market* model, where a set of n buyers compete for m items. In the *competitive equilibrium from equal incomes* (CEEI) mechanism for fair division, a set of m divisible items are to be fairly divided among n agents. The mechanism works by giving every agent a budget of one unit of fake money. Then, a market equilibrium is computed, and the allocation from the equilibrium is used as the fair division of the items, while the prices are discarded since they are for fake money [62]. An allocation given by CEEI can be shown to satisfy various fairness desiderata, e.g., the allocation makes all agents envy-free, Pareto optimal and get at least their proportional share of utility [14, 52]. CEEI has recently been suggested for *fair recommender systems* [42, 43]. Divisible Fisher market equilibrium has also been shown equivalent to the *pacing equilibrium* solution concept for budget smoothing in Internet ad auctions [25, 24]. When items are indivisible, (approximate) CEEI has been used for allocating courses to students [14, 15], and the related max Nash welfare solution for fairly dividing estates and other goods (see, e.g., [17] and also the online fair division service platform

spliddit.org). The indivisible setting can be related to the divisible setting via lotteries, though the question of when lotteries can be resolved satisfactorily is intricate [16, 1].

Market equilibrium. In this paper we focus on computing Fisher market equilibrium in the divisible setting. Let the market consists of n buyers and m goods. Buyer i has budget $B_i > 0$ and utility function $u_i : \mathbb{R}_+^m \rightarrow \mathbb{R}_+$. As mentioned above, $B_i = 1$ for all i corresponds to the CEEI mechanism. Without loss of generality, assume each good j has *unit* supply. An (aggregate) *allocation* is a matrix $x = [x_1^\top; \dots; x_n^\top] \in \mathbb{R}_+^{n \times m}$, where x_{ij} is the amount of item j purchased by buyer i . We also refer to $x_i \in \mathbb{R}_+^m$ as the *bundle* for buyer i . Given prices $p \in \mathbb{R}_+^m$ of the goods, the *demand set* of buyer i is defined as the set of utility-maximizing allocations:

$$D_i(p) = \arg \max \{u_i(x_i) \mid x_i \in \mathbb{R}_+^m, \langle p, x_i \rangle \leq B_i\}.$$

A *competitive equilibrium* is a pair of prices and allocations (p, x) , $p \in \mathbb{R}_+^m$ that satisfy the following conditions [32, 39]:

- *Buyer optimality*: $x_i \in D_i(p)$ for all $i \in [n]$,
- *Market clearance*: $\sum_{i=1}^n x_{ij} \leq 1$ for all $j \in [m]$, and equality must hold if $p_j > 0$.

We say that u_i is *homogeneous* (with degree 1) if it satisfies $u_i(\alpha x_i) = \alpha u_i(x_i)$ for any $x_i \geq 0$ and $\alpha \geq 0$ [52, §6.2]. We assume that u_i are concave, continuous, nonnegative, and homogeneous utilities (CCNH). This captures many widely used utilities, such as linear, Leontief, Cobb-Douglas, and general *Constant Elasticity of Substitution* (CES) utilities (see, e.g., [52, §6.1.5] and [10]). For CCNH utilities, a market equilibrium can be computed using the following *Eisenberg-Gale convex program* (EG):

$$\max \sum_{i=1}^n B_i \log u_i(x_i) \quad \text{s.t.} \quad \sum_{i=1}^n x_{ij} \leq 1, \quad x \geq 0. \quad (1)$$

More precisely, we state the following theorem, which is well-known in various forms in the literature: see, e.g., [52, §6.2] for the case of differentiable u_i and Eisenberg’s original work [31, Theorem 4] for the convex program with a maximization objective of the Nash social welfare). For completeness, we present an elementary, self-contained proof (all proofs are in the Appendix).

Theorem 1 *For each i assume (i) u_i is finite, concave, continuous and homogeneous on \mathbb{R}_+^m ; (ii) there exists $x_i \in \mathbb{R}_+^m$ (component-wise) such that $u_i(x_i) > 0$. Then, (i) (1) has an optimal solution and (ii) any optimal solution x^* to (1) and the optimal Lagrangian multipliers $p^* \in \mathbb{R}_+^m$ associated with the constraints $\sum_i x_{ij} \leq 1$, $j \in [m]$ form a market equilibrium (x^*, p^*) . Furthermore, $\langle p^*, x_i^* \rangle = B_i$ for all i .*

By the homogeneity assumption, $u_i(\mathbf{0}) = 0$. The existence of $x_i \in \mathbb{R}_+^m$ such that $u_i(x_i) > 0$ is for Slater’s condition, a sufficient condition for strong duality. It is satisfied, for example, when $u_i(\mathbf{1}) > 0$, that is, buyer i derives strictly positive utility from all available items. Note that it is possible that $u_i(x_i) < 0$ at some $x_i \geq 0$.

We focus on linear, quasilinear (QL) and Leontief utilities. Under linear utilities, the utility of buyer i is $u_i(x_i) = \langle v_i, x_i \rangle$. As mentioned above, linear utilities are used in the CEEI mechanism in fair division and fair recommender systems. QL utilities have the form $u_i(x_i) = \sum_j (v_{ij} - p_j) x_{ij}$, which depends on the prices [19, 22]. It captures budget-smoothing problems in first- and second-price auction markets [25, 24]. Leontief utilities have the form $u_i(x_i) = \min_{j \in J_i} \frac{x_{ij}}{a_{ij}}$ (see §5). They model perfectly complementary goods and are suitable in resource sharing where an agent’s utility is capped by its *dominant resources*, such as allocating different types of compute resources to cloud computing tasks [35, 41].

Another notable convex program that captures market equilibrium under linear utilities is Shmyrev’s [59] (see Appendix B.1 and (19) there for more details). [9] shows that the well-known Proportional Response dynamics (PR) [68] for computing market equilibrium under linear utilities is in fact applying Mirror Descent to Shmyrev’s convex program (19). In §4, we will show that both Shmyrev’s convex program and PR generalize elegantly to QL utilities.

In principle, (1) can be solved using an interior-point method (IPM) that handles exponential cones [60, 26]. However, it does not scale to large markets since IPM requires solving a very large linear

system per iteration. This linear system can often be ill-conditioned and relatively dense, even when the problem itself is sparse. This makes IPM impractical for applications such as fair recommender systems and ad markets, where the numbers of buyers and items are typically extremely large. In this paper, we investigate iterative (gradient-based) first-order methods (FOMs) for computing equilibria of Fisher markets. Because each FOM iteration has low cost, which usually scales only with the number of nonzeros when the problem is sparse, these methods are suitable for scaling up to very large markets. Furthermore, there have been variants of FOMs that allow parallel and distributed updates, making them even more scalable [18, 58, 67, 46]. This is analogous to other equilibrium-computation settings such as zero-sum Nash equilibrium, where iterative first-order methods are also the state-of-the-art [11, 49, 44]. We will focus on the important classes of linear, quasi-linear and Leontief utilities, which are discussed in §3, 4 and 5, respectively. We also note that under CES utilities, PR yields linearly convergent prices and utilities [68, Theorem 4]. Meanwhile, market equilibrium under Cobb-Douglas utilities can be computed explicitly. See Appendix A.6 for more details. We note that there have been highly nontrivial algorithms that find market equilibria in time polynomial in $m, n, \log \frac{1}{\epsilon}$ in theory (where ϵ is a desired upper bound on error in prices) [27, 63, 7]. However, none of these are as easily implementable as the FOMs we consider here, which also achieves ϵ -equilibrium prices in $O(\log \frac{1}{\epsilon})$ time thanks to the linear convergence guarantees. Secondly, these methods all have prohibitively-expensive per-iteration costs from a large-scale perspective.

First-order methods. Consider optimization problems of the form

$$f^* = \min_{x \in \mathcal{X}} f(x) = h(Ax) + \langle q, x \rangle, \quad (2)$$

where \mathcal{X} is a bounded polyhedral set, $h : \mathbb{R}^r \rightarrow \mathbb{R}$ is μ -strongly convex with a L -Lipschitz continuous gradient on \mathcal{X} (or (μ, L) -s.c. for short), $A \in \mathbb{R}^{d \times r}$ and $q \in \mathbb{R}^d$. We say that an algorithm for (2) converges *linearly* (in objective value) with rate $\rho \in (0, 1)$ if its iterates x^t satisfy $f(x^t) - f^* \leq \rho^t(f(x^0) - f^*)$ for all t , where $x^0 \in \text{dom } \mathcal{X}$ is an initial iterate and f^* is the minimum objective value. Unless otherwise stated, \mathcal{X}^* denotes the set of optimal solutions to (2), which is always a bounded polyhedral set [65, Lemma 14].

Various FOMs, such as the following, are naturally suitable for (2).

- Projected gradient (PG): $x^{t+1} = \Pi_{\mathcal{X}}(x^t - \gamma_t \nabla f(x^t))$, where γ_t is the stepsize.
- Frank-Wolfe (FW): $x^{t+1} = x^t + \gamma_t(w^t - x^t)$, where $w^t \in \arg \min_{w \in \mathcal{X}} \langle \nabla f(x^t), w \rangle$ [34, 13, 4, 45]. Unless otherwise stated, w^t is chosen from the set of vertices of \mathcal{X} .
- Mirror Descent: $x^{t+1} = \arg \min_{x \in \mathcal{X}} \langle \nabla f(x^t), x - x^t \rangle + \gamma D(x \| x^t)$ [51, 5]. Here, D is the *Bregman divergence* of a differentiable convex function d (see, e.g., [8, §5.3]).

As is well-known, for a broad class of convex optimization problems, these FOMs and their variants achieve *sublinear* convergence, that is, $f(\tilde{x}^t) - f^* = O(t^{-k})$ for some $k > 0$, where \tilde{x}^t is either x^t or a weighted average of x^τ , $\tau \leq t$ [13, 8, 3]. When f is strongly convex, linear convergence can be derived. However, strong convexity is highly restrictive and relaxed sufficient conditions for linear convergence of FOMs have been considered. For example, the classical Polyak-Łojasiewicz (PL) condition ensures linear convergence of gradient descent [56, 47]. Also have been extensively studied are various *error bound* (EB) conditions, which, roughly speaking, say that the distance from a feasible solution x (possibly required to be close to \mathcal{X}^*) to \mathcal{X}^* is bounded by a constant multiple of a computable “residual” of x [48, 61, 65]. Another notable example is the *quadratic growth* (QG) condition, which essentially means that the objective grows at least quadratically in the distance to \mathcal{X}^* [57, 2, 29]. Furthermore, it can be seen to be equivalent to an EB under further assumptions on the problem [29]. Recently, [40] shows that, for convex composite “ $f + g$ ” problems, a so-called *Proximal-PL* condition, which generalizes the PL condition, is sufficient for linear convergence of proximal gradient. It is also shown to be equivalent to a few existing conditions under further assumptions on the problem.

Summary of contributions. We show that PG, FW and MD are suitable for computing various market equilibria via their convex optimization formulations. In terms of first-order methods, we prove that proximal gradient with a non-standard practical linesearch scheme (Algorithm 1 in the appendix) converges linearly, with a bounded number of backtracking steps per iteration, for the more general class of problems satisfying the Proximal-PL condition (Theorem 4). In terms of market equilibria under different utility classes, we establish simple bounds on equilibrium quantities by

exploiting properties of market equilibria (Lemmas 1, 2, 3). Through various problem reformulations exploiting convex optimization duality and using these bounds, we show that the convex programs for ME can be reformulated into ones with highly structured objectives over simple polyhedral sets (2). Then, we derive linear convergence of PG for these convex programs through establishing the Proximal-PL condition (Theorems 5,6,8). Specifically, for linear utilities, we show that PG for the EG convex program (1) converges linearly. For QL utilities, based on the relation between EG and Shmyrev’s convex programs, we derive a new “QL-Shmyrev” convex program (4) whose optimal solutions give equilibrium prices and bids. Similarly, PG for this convex program also achieves linear convergence. We also show that Mirror Descent for the same convex program (4) achieves sublinear last-iterate convergence with a small constant. MD for this convex program also leads to a form of Proportional Response dynamics (PR), a scalable and interpretable algorithm for computing market equilibrium, extending the a series of results for linear utilities [9, 68]. For Leontief utilities, we show that PG for a reformulated dual of (1), with variables being the prices, achieves linear convergence. For all utility classes, linear convergence of running iterates (e.g., prices p^t) to their corresponding equilibrium quantities (p^*) can be easily derived using linear convergence of the objective values. Extensive numerical experiments demonstrate that PR (and sometimes FW) can quickly compute approximate equilibrium allocations and prices, while PG with linesearch is more efficient for computing a higher-accuracy solution.

Notation. The j -th unit vector and vector of 1’s in \mathbb{R}^d are $\mathbf{e}^{j,(d)}$ and $\mathbf{1}^{(d)}$, respectively, where the superscript (d) is omitted when d is clear from the context. The d -dimensional simplex is $\Delta_d = \{x \in \mathbb{R}^d : \mathbf{1}^\top x = 1, x \geq 0\}$. For any vector x , x_+ denotes the vector with entries $\max\{x_i, 0\}$ and $\text{nnz}(x)$ denote its number of nonzeros. All unsubscripted vector norms and matrix norms are Euclidean 2-norms and matrix 2-norms (largest singular value), respectively. For a closed convex set $\mathcal{X} \subseteq \mathbb{R}^d$, $\text{int}(\mathcal{X})$ denotes the interior of \mathcal{X} , $\Pi_{\mathcal{X}}(x)$ denotes the Euclidean projection of $x \in \mathbb{R}^d$ onto \mathcal{X} and $\text{Diam}(\mathcal{X}) = \sup_{x,y \in \mathcal{X}} \|x - y\|$. For $p, q \in \mathbb{R}_+^d$, the (generalized) Kullback–Leibler (KL) divergence of p w.r.t. q is $D(p\|q) = \sum_i p_i \log \frac{p_i}{q_i} - \sum_i p_i + \sum_i q_i$. Denote the maximum and minimum *nonzero* singular values of A as $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$, respectively.

2 Linear convergence of first-order methods

For PG, it has been shown that linear convergence can be achieved for non-strongly convex objectives under various relaxed conditions [48, 29, 40]. In particular, (2) can be shown to satisfy several such conditions and therefore guarantees linear convergence under PG. One notion of central significance in many such results is the *Hoffman constant*. For $A \in \mathbb{R}^{d \times r}$ and polyhedral set $\mathcal{X} \subseteq \mathbb{R}^d$, denote the (relative) Hoffman constant of A w.r.t. \mathcal{X} as $H_{\mathcal{X}}(A)$, which is the smallest $H > 0$ such that, for any $z \in \text{range}(A)$, $\mathcal{S} = \{x : Ax = z\}$, it holds that $\|x - \Pi_{\mathcal{X} \cap \mathcal{S}}(x)\| \leq H \|Ax - z\|$ for all $x \in \mathcal{X}$ (the Hoffman inequality). Note that $H_{\mathcal{X}}(A)$ is always well-defined, finite, and depends only on A and \mathcal{X} [37, 54]. In the optimization context, the intuition is as follows: the set of optimal solutions \mathcal{X}^* can often be expressed as $\mathcal{X}^* = \mathcal{X} \cap \mathcal{S}$, which means the distance to optimality can be bounded by a constant multiple of the residual $\|Ax - z\|$. Therefore, the Hoffman inequality can be viewed as an EB condition. Further details on this definition, as well as its connection to the classical definition and characterization, can be found in Appendix A.2. Using the Hoffman inequality, we have the following linear convergence guarantee. It is a special case of [40, Theorem 5]. A proof is in Appendix A.3, which is essentially the same as [40, Appendix F]. However, our proof is slightly refined to allow a weakened condition, provide more details on the relevant constants and conclude iterate convergence.

Theorem 2 For (2) with $q = 0$, PG with a constant stepsize $\gamma_t = \frac{1}{L\|A\|^2}$ generates x^t such that $\frac{\mu}{2H_{\mathcal{X}}(A)} \|x^t - \Pi_{\mathcal{X}^*}(x^t)\|^2 \leq f(x^t) - f^* \leq \left(1 - \frac{\mu}{\max\{\mu, LH_{\mathcal{X}}(A)^2\|A\|^2\}}\right)^t (f(x^0) - f^*)$ for all t .

For the case of $q \neq 0$, we have the following instead, which relies heavily on the boundedness of \mathcal{X} . The proof is in Appendix A.4. The key is to make use of the polyhedral structure of the set of optimal solutions and establish a QG condition similar to the one in [4, Lemma 2.3]. Here, we slightly refine the constants using the monotonicity of $f(x^t)$. We note that linear convergence with different rates can also be established through invoking different error bound conditions that also hold for (2) (see, e.g., [65, 29]).

Theorem 3 *There exists unique $z^* \in \mathbb{R}^r$ and $t^* \in \mathbb{R}$ such that $Ax^* = z^*$ and $\langle q, x^* \rangle = t^*$ for any $x^* \in \mathcal{X}^*$. Furthermore, PG for (2) with constant stepsize $\gamma_t = \frac{1}{L\|A\|^2}$, starting from $x^0 \in \mathcal{X}$, generates x^t such that $\frac{1}{\kappa}\|x^t - \Pi_{\mathcal{X}^*}(x^t)\| \leq f(x^t) - f^* \leq C\rho^t$, where $\rho = 1 - \frac{1}{\max\{1, 2\kappa L\|A\|^2\}}$, $\kappa = H_{\mathcal{X}}(A)^2 \left(C + 2GD_A + \frac{2(G^2+1)}{\mu} \right)$, $C = f(x^0) - f^*$, $G = \|\nabla h(z^*)\|$, $D_A = \sup_{x, x' \in \mathcal{X}} \|A(x - x')\|$.*

For FW, the classical sublinear convergence holds. Specifically, for (2), using FW with static stepsizes or exact linesearch, that is, $\gamma_t = \frac{2}{2+t}$ or $\gamma_t \in \arg \min_{\gamma \in [0,1]} f(x^t + \gamma(w^t - x^t))$, it holds that $f(x^t) - f^* \leq \frac{\text{Diam}(\mathcal{X})^2 L \|A\|^2}{k+2}$ for all t (see, e.g., [38, Theorem 1] and [21, Theorem 2.3]). The advantage of FW is that it can maintain highly sparse iterates when \mathcal{X} has sparse vertices. In fact, if we start with x^0 being a vertex, then x^t is a convex combination of $(t+1)$ vertices. For ME convex programs such as (1), the constraint set \mathcal{X} is a product of simplexes Δ_n . Therefore, any vertex of \mathcal{X} has the form $x \in \{0, 1\}^{n \times m}$, where each (x_{1j}, \dots, x_{nj}) , $i \in [m]$ is a unit vector. Hence, x^t is a linear combination of $(t+1)$ vertices and has at most $(t+1)m$ nonzeros. There have been recent theories on linear convergence of an away-step variant of FW for (2) [4, 45], in which a succinct “vertex representation” [3, §3.2] must be maintained through an additional, potentially costly reduction procedure. Initial trials suggest that it does not work well for our specific convex programs and are therefore not considered further.

PG with Linesearch. Linesearch is often essential for numerical efficiency of gradient-based methods, especially when the Lipschitz constant cannot be computed. Here, we consider a modified version of backtracking linesearch for PG (see, e.g., [3, §10.4.2] and [6, §1.4.3]). Specifically, given increment factor $\alpha \geq 1$, decrement factor $\beta \in (0, 1)$, and stepsize upper bound $\Gamma > 0$, at each iteration, set the (candidate) stepsize γ_t to either the previous one γ_{t-1} or $\min\{\Gamma, \alpha\gamma_{t-1}\}$, depending on whether backtracking occurs at all in the previous iteration. Then, perform the usual backtracking with factor β until a sufficient decrease condition is met. denote the linesearch subroutine as $(x^{t+1}, \gamma_t, k_t) \leftarrow \mathcal{LS}_{\alpha, \beta, \Gamma}(x^t, \gamma_{t-1}, k_{t-1})$, where k_{t-1} is the number of backtracking steps in the previous iteration, $\gamma_{-1} = \Gamma$, $k_{-1} = 0$. More details, as well as the proof of the following convergence result (in fact, a more general one that holds for the proximal gradient method), can be found in Appendix A.5. Let $L_f^{\mathcal{X}} = \sup_{x, y \in \mathcal{X}, x \neq y} \frac{\|\nabla f(x) - \nabla f(y)\|}{\|x - y\|} \leq L\|A\|^2$ be the (effective) Lipschitz constant of ∇f on \mathcal{X} . Note that, in addition to ensure linear convergence, linesearch can potentially improve the constant in the rate, since $L_f^{\mathcal{X}}$ can be much smaller than $L\|A\|^2$.

Theorem 4 *PG with linesearch $\mathcal{LS}_{\alpha, \beta, \Gamma}$ for (2) with $q = 0$ and $q \neq 0$ converges linearly with rate $1 - \frac{\mu}{\max\{\mu, H^2/\Gamma, H^2 L_f^{\mathcal{X}}/\beta\}}$ (where $H = H_{\mathcal{X}}(A)$) and $1 - \frac{1}{\max\{1, 2\kappa/\Gamma, 2\kappa L_f^{\mathcal{X}}/\beta\}}$, respectively. For both, the number of projections $\Pi_{\mathcal{X}}$ per iteration is at most $1 + \frac{\log \frac{\Gamma}{\tilde{\gamma}}}{\log \frac{1}{\beta}}$, where $\tilde{\gamma} := \min\{\Gamma, \beta/L_f^{\mathcal{X}}\}$.*

3 Linear utilities

Let $u_i(x_i) = \langle v_i, x_i \rangle$ and $v = [v_1, \dots, v_n] \in \mathbb{R}^{n \times m}$ be the matrix of all buyers’ valuations. Without loss of generality, from now on, we assume the following *nondegeneracy* condition, that is, v does not contain any zero row or column. Then, the inequality constraints in (1) can be replaced by equalities without affecting any optimal solution. Subsequently, EG with linear utilities can be written as (2) with $\mathcal{X} = (\Delta_n)^m$, $f(x) = h(Ax)$, where x can be viewed as a (nm) -dimensional vector, $A \in \mathbb{R}^{n \times nm}$ is a block-diagonal matrix with i -th block being v_i^\top , $h(u) = \sum_{i=1}^n h_i(u_i)$, $h_i(u_i) := -B_i \log u_i$. Clearly, $\|A\| = \max_i \|v_i\|$. However, in this way, h does not have a Lipschitz continuous gradient on the interior of $\mathcal{U} = \{Ax : x \in \mathcal{X}\}$: for $u \in \text{int}(\mathcal{U})$, each u_i can be arbitrarily close to 0. This can be circumvented by the following bounds on equilibrium utilities.

Lemma 1 *Let $\underline{u}_i = \frac{B_i \|v_i\|_1}{\|B\|_1}$ and $\bar{u}_i = \|v_i\|_1$, $i \in [n]$. Any feasible allocation $x \in \mathcal{X}$ satisfies $\langle v_i, x_i \rangle \leq \bar{u}_i$ for all i . For any equilibrium allocation x^* , we have $\langle v_i, x_i^* \rangle \geq \underline{u}_i$ for all i .*

Using Lemma 1, we can replace each h_i by its “quadratic extrapolation” when $\langle v_i, x_i \rangle \notin [u_i, \bar{u}_i]$. Specifically, let $\tilde{h}(z) := \sum_{i=1}^n \tilde{h}_i(u_i)$, where

$$\tilde{h}_i(u_i) := \begin{cases} \frac{1}{2} h_i''(u_i)(u_i - u_i)^2 + h_i'(u_i)(u_i - u_i) + h_i(u_i) & \text{if } u_i \leq \bar{u}_i, \\ h_i(u_i) = -B_i \log u_i & \text{otherwise.} \end{cases} \quad (3)$$

In the above, clearly, we have $h''(u_i) = \frac{B_i}{u_i^2} > 0$, $h'(u_i) = -\frac{B_i}{u_i}$ and $h_i(u_i) = -B_i \log u_i$. Therefore, \tilde{h} is μ -strongly convex and smooth with L -Lipschitz continuous gradient on \mathcal{U} , where $\mu = \min_i \frac{B_i}{\|v_i\|_1^2}$ and $L = \max_i \frac{B_i}{u_i^2} = \max_i \frac{\|B\|_1^2 \|v_i\|_1}{B_i}$. We can then use the minimization objective $f(x) = \tilde{h}(Ax)$ in (2) instead, without affecting any optimal solution. A direct application of Theorem 3 gives the following convergence guarantee regarding PG for the modified EG convex program. A similar conclusion can also be made using Theorem 4 on PG with linesearch for the same problem.

Theorem 5 *Let \mathcal{X} and A be as above. Then, PG for the problem $\min_{x \in \mathcal{X}} \tilde{h}(Ax)$ with constant stepsize $\gamma_t = \frac{1}{L\|A\|^2}$ converges linearly with rate $1 - \frac{\mu}{\max\{\mu, LH_{\mathcal{X}}(A)^2\|A\|^2\}}$, where $\|A\| = \max_i \|v_i\|$.*

Cost per iteration. In every iteration (or linesearch step), PG requires computing the projection $\Pi_{\mathcal{X}}$, which decomposes item-wise into Π_{Δ_n} . This can be computed via an efficient $O(n \log n)$ sorting-based algorithm [23, 20, 66] (in fact, only nonzero elements need to be sorted), while $O(n)$ algorithms are also available [30, 12]. Furthermore, when v is sparse, that is, only a small subset of buyers $I_j \subseteq [n]$ value each item j , the total cost for projection scales as $O(\sum_j |I_j| \log |I_j|) = O(\text{nnz}(v) \log \text{nnz}(v))$, without explicit dependence on n, m . Clearly, this is also true for computing the utilities $\langle v_i, x_i \rangle$, $i \in [n]$. Therefore, the cost per iteration scales with number of nonzeros $\text{nnz}(v)$ without explicit dependence on n, m . Furthermore, the gradient computation $\nabla f(x) = \sum_i \tilde{h}'_i(\langle v_i, x_i \rangle) v_i$ decomposes buyer-wise and $\Pi_{\mathcal{X}}(x) = (\Pi_{\Delta_n}(x_{:,j}))$ decomposes item-wise, allowing straightforward parallelization of PG. In fact, the same holds for subsequent QL and Lenontief utilities.

Convergent utilities and prices. The equilibrium utilities u^* are unique [31, Theorem 1]. So are the equilibrium prices p^* . Furthermore, $u_i^t = \langle v_i, x_i^t \rangle$ converges linearly to u^* by strong convexity of h . In addition, the simplex projection algorithm in each iteration yields a linearly convergent sequence of prices p^t in terms of relative price error $\eta^t := \max_j \frac{|p_j^t - p_j^*|}{p_j^*}$, a commonly used error measure in ME computation [68, 9]. See Appendix B.3 for details.

By Theorem 5 and the above discussion, we have the following, where u_i^* is the (unique) equilibrium utility of buyer i . Similar statements can also be made following the subsequent convergence results, i.e., Theorem 6 and 8.

Corollary 1 *PG for (1) computes an feasible allocation x^t such that $|\langle v_i, x_i^t \rangle - u_i^*| \leq \epsilon$ for all i in $t = O(\tilde{n} \log \tilde{n} \log(1/\epsilon))$ time, where $\tilde{n} = \text{nnz}(v)$ and the constant only depends on v_{ij} and B_i .*

Handling additional constraints. Although efficient computation of $\Pi_{\mathcal{X}}$ depends on the structure of \mathcal{X} , it is still arguably more flexible compared to combinatorial algorithms for computing market equilibria, which depend crucially on the specific market structures [64, 63]. Here, for example, the simplex projection algorithms can be modified easily to handle box constraints like $\bar{x}_{ij} \leq x_{ij} \leq \bar{x}_{ij}$ without affecting the time complexity [53, 12]. This allows additional at-most-one constraints $x_{ij} \leq 1$, useful in fair division applications [42]. Similar bounds on the individual bids b_{ij} (which denote how much each buyer spends), such as *spending constraints* (see, e.g., [9, Eq. (2)] and [64]) can also be incorporated without incurring additional cost (this requires solving the Shmyrev convex program (19) which we introduce in Appendix B.1).

4 Quasi-linear utilities

In quasi-linear (QL) utilities, the money has value outside the market, which means each buyer’s utility is the utility they derive from the items minus their payments, i.e., $u_i(x_i) = \sum_j (v_{ij} - p_j) x_{ij}$, where p_j is the price of item j . Note that QL utilities are not CCNH, as it depends on the prices. Nevertheless, based on EG and its dual for linear utilities, another pair of primal and dual convex programs can be derived to capture ME under QL utilities [22, Lemma 5]. In order to derive

Proportional Response dynamics for a market with QL utilities, we now introduce the following convex program, which we call QL-Shmyrev, for its similarity in structure to Shmyrev's convex program for linear utilities (19) [59]. It is the dual of a reformulation of a convex program in [22, Lemma 5]. Let the bids be $b = (b_1, \dots, b_n)$, where each $b_i = (b_{ij}) \in \mathbb{R}^m$. Denote $p_j(b) = \sum_i b_{ij}$ and $p(b) = (p_1(b), \dots, p_m(b))$. Introduce slack variables $\delta = (\delta_1, \dots, \delta_n)$ representing buyers' leftover budgets. Let $\mathcal{B} = \{(b, \delta) \in \mathbb{R}^{n \times m} \times \mathbb{R}^n : (b_i, \delta_i) \in \Delta_{n+1}, i \in [n]\}$. The convex program is

$$\varphi^* = \min_{\bar{b}=(b,\delta)} \varphi(b) = - \sum_{i,j} (1 + \log v_{ij}) b_{ij} + \sum_j p_j(b) \log p_j(b) \text{ s.t. } \bar{b} \in \mathcal{B}. \quad (4)$$

The derivation is in Appendix C.1. This convex program differs from Shmyrev's convex program (19) for linear utilities in the coefficients of b_{ij} and constraints on b_{ij} (allowing $\sum_j b_{ij} \leq B_i$ instead of $=$). For notational brevity, we assume that $v > 0$, although all results hold in the general case of nondegenerate v (no zero row or column) with summations like $\sum_{i,j} \log v_{ij}$ being over $(i, j) : v_{ij} > 0$ instead. Here, δ_i is interpreted as the "leftover" budget of buyer i : since u_i depends on the prices p_j , a buyer may not spend their entire budget B_i , if prices are too high relative to their valuations. In contrast, for CCNH u_i , budgets are always depleted at equilibrium. In order to apply PG to solve (4), we need a (μ, L) -s.c. objective, while the function $h(p) = \sum_j p_j \log p_j$ does not have Lipschitz continuous gradient on $\{p(b) : (b, \delta) \in \mathcal{B}\}$, since $p_j(b)$ can be arbitrarily small for $b \in \mathcal{B}$. To address this, we establish the following bounds on equilibrium prices. We also show that they are unique and equal to the sum of buyers' bids at equilibrium.

Lemma 2 *For nondegenerate v , the equilibrium prices p^* under QL utilities are unique. Let $\underline{p}_j = \max_i \frac{v_{ij} B_i}{\|v_i\|_1 + B_i}$ and $\bar{p}_j = \max_i v_{ij}$ for all j . For any optimal solution (b^*, δ^*) to (4), we have $\underline{p}_j \leq p_j^* = p_j(b^*) \leq \bar{p}_j$.*

To cast (4) into the standard form (2), we take $\mathcal{X} = \mathcal{B}$, the linear map $A : (b, \delta) \mapsto (p_1(b), \dots, p_m(b))$, $h(p) = \sum_j p_j \log p_j$ and $q = (q_{ij})$, $q_{ij} = -1 - \log v_{ij}$. Viewing (b, δ) as a $n(m+1)$ -dimensional vector concatenating each b_i and δ , we have $A := [I, \dots, I, 0] \in \mathbb{R}^{n \times (n(m+1))}$ and $\|A\| = n$. Then, replace h with a (μ, L) -s.c. function \tilde{h} via a smooth extrapolation similar to that in §3, where $\mu = \frac{1}{\max_j \bar{p}_j}$ and $L = \frac{1}{\min_j \underline{p}_j}$. Let $\tilde{\varphi}(b) = -\sum_{i,j} (1 + \log v_{ij}) b_{ij} + \sum_j \tilde{h}(p_j(b))$. Clearly, $\varphi(b) = \tilde{\varphi}(b)$ as long as $p(b) \in [\underline{p}, \bar{p}]$, since $h = \tilde{h}$ on $[\underline{p}, \bar{p}]$. Combining Theorem 3 and Lemma 2, we have the following.

Theorem 6 *Let (b^0, δ^0) satisfies $p(b^0) \in [\underline{p}, \bar{p}]$. Then, PG with stepsize $\gamma_t = \frac{1}{Ln^2}$ for the problem $\min_{(b,\delta) \in \mathcal{B}} \tilde{\varphi}(b)$ converges linearly with rate $1 - \frac{1}{\max\{1, 2\kappa Ln^2\}}$, where $\kappa = H_{\mathcal{X}}(A)^2 \left(C + 2GD + \frac{2(G^2+1)}{\mu} \right)$, $C = \varphi(b^0) - \varphi^*$, $G = \|\nabla \tilde{h}(p^*)\|$, $D = \sup_{(b,\delta), (b',\delta') \in \mathcal{B}} \|p(b) - p(b')\|$.*

Remark. Some constants can be bounded explicitly. Clearly, $D \leq \sqrt{\sum_j (\max_i v_{ij})^2}$. By Lemma 2, $\tilde{h}(p^*) = h(p^*)$, since $p^* \in [\underline{p}, \bar{p}]$. Therefore, $G = \sqrt{\sum_j (1 + \log p_j^*)^2} \leq \sqrt{\sum_j (1 + \log \max_i v_{ij})^2}$.

MD for (4) as Proportional Response dynamics. Similar to [9], we can also apply MD to (4) and obtain a PR dynamics under QL utilities with $O(1/T)$ convergence in both objective value $\varphi(b^t) - \varphi^*$ and price error $D(p^t \| p^*)$. Recall that MD for (22) with unit stepsize $\gamma_t = 1$ performs the following update:

$$(b^{t+1}, \delta^{t+1}) = \arg \min_{(b,\delta) \in \mathcal{B}} \langle \nabla \varphi(b^t), b - b^t \rangle + D(b, \delta \| b^t, \delta^t), \quad (5)$$

where $D(b', \delta' \| b, \delta) = \sum_{i,j} b'_{ij} \log \frac{b'_{ij}}{b_{ij}} + \sum_i \delta'_i \log \frac{\delta'_i}{\delta_i}$ is the usual KL divergence. Extending (and slightly strengthening) [9, Eq. (13) and (16)], we first establish the following last-iterate convergence in objective value and price error. Its proof is in Appendix C.3.

Theorem 7 *Let $b_{ij}^0 = \frac{B_i}{m+1}$, $\delta_i^0 = \frac{B_i}{m+1}$ for all i, j . Then, MD applied to (4) generates iterates (b^t, δ^t) , $t = 1, 2, \dots$ such that $D(p(b^t) \| p^*) \leq \varphi(b^t) - \varphi^* \leq \frac{\|B\|_1 \log(m+1)}{t}$ for all t .*

The MD update (5) leads to a form of PR dynamics [68, 9] as follows (see Appendix C.4 for the derivation). At time t , buyers first submit their bids $b^t = (b_{ij}^t)$. Then, item prices are computed via $p_j^t = \sum_j b_{ij}^t$. Next, each buyer is allocated $x_{ij}^t = b_{ij}^t / p_j^t$ amount of item j . Finally, the bids and leftover budgets are updated via

$$b_{ij}^{t+1} = B_i \cdot \frac{v_{ij} x_{ij}^t}{\sum_{\ell} v_{i\ell} x_{i\ell}^t + \delta_i^t}, \quad \delta_i^{t+1} = B_i \cdot \frac{\delta_i^t}{\sum_{\ell} v_{i\ell} x_{i\ell}^t + \delta_i^t}. \quad (6)$$

Convergence of prices and a computable bound. Similar to the linear case, PG for QL utilities also yields prices p^t with relative error η^t converging linearly to 0. Furthermore, it can also be bounded explicitly by computable quantities. See Appendix C.5 for details.

5 Leontief utilities

Leontief utilities model perfectly complementary items and are suitable for resource sharing scenarios where utilities are capped by dominant resources. The utility function is $u_i(x_i) = \min_{j \in J_i} \frac{x_{ij}}{a_{ij}}$, where $a_{ij} > 0$ for all $j \in J_i \neq \emptyset$. Denote $I_j = \{i \in [n] : j \in J_i\}$ and assume that $I_j \neq \emptyset$ for all j , without loss of generality. Denote $a_i = (a_{i1}, \dots, a_{im}) \in \mathbb{R}_+^m$, where $a_{ij} := 0$ if $j \notin J_i$ and $a = [a_1^\top; \dots, a_n^\top] \in \mathbb{R}^{n \times m}$. Under Leontief utilities, EG (1) can be written in terms of the utilities $u \in \mathbb{R}_+^n$ (see (33) in Appendix D.1), whose dual, after reformulation, is

$$\min f(p) = - \sum_i B_i \log \langle a_i, p \rangle \quad \text{s.t. } p \in \mathcal{P}, \quad (7)$$

where $\mathcal{P} = \{p \in \mathbb{R}_+^m : \sum_j p_j = \|B\|_1\}$. The derivation is in Appendix D.1. In particular, we use the fact that $\|p^*\|_1 = \|B\|_1$ at equilibrium. Similar to the case of linear and QL utilities, we have the following bounds on $\langle a_i, p \rangle$ for Leontief utilities.

Lemma 3 *For any $p \in \mathcal{P}$, it holds that $\langle a_i, p \rangle \leq \bar{r}_i := \|B\|_1 \|a_i\|_\infty$ for all i . Furthermore, for any equilibrium prices p^* , it holds that $\langle a_i, p^* \rangle \geq r_i := B_i \|a_i\|_\infty$ for all i .*

Let $h(u) = - \sum_i B_i \log u_i$. Again, using Lemma 3, we can use a simple quadratic extrapolation to construct a (μ, L) -s.c. function \tilde{h} with $\mu = \min_i \frac{B_i}{\bar{r}_i^2}$, $L = \max_i \frac{B_i}{r_i^2}$ without affecting any optimal solution. Clearly, a Lipschitz constant of the gradient of $f(p) = \tilde{h}(ap)$ is $L\|a\|^2$. Analogous to Theorem 5, we have the following convergence guarantee for Leontief utilities. Here, equilibrium prices may not be unique, but the equilibrium utilities u^* are (e.g., by (33) in the appendix). We can also easily construct u^t that converges linearly to u^* . See Appendix D.3 for details.

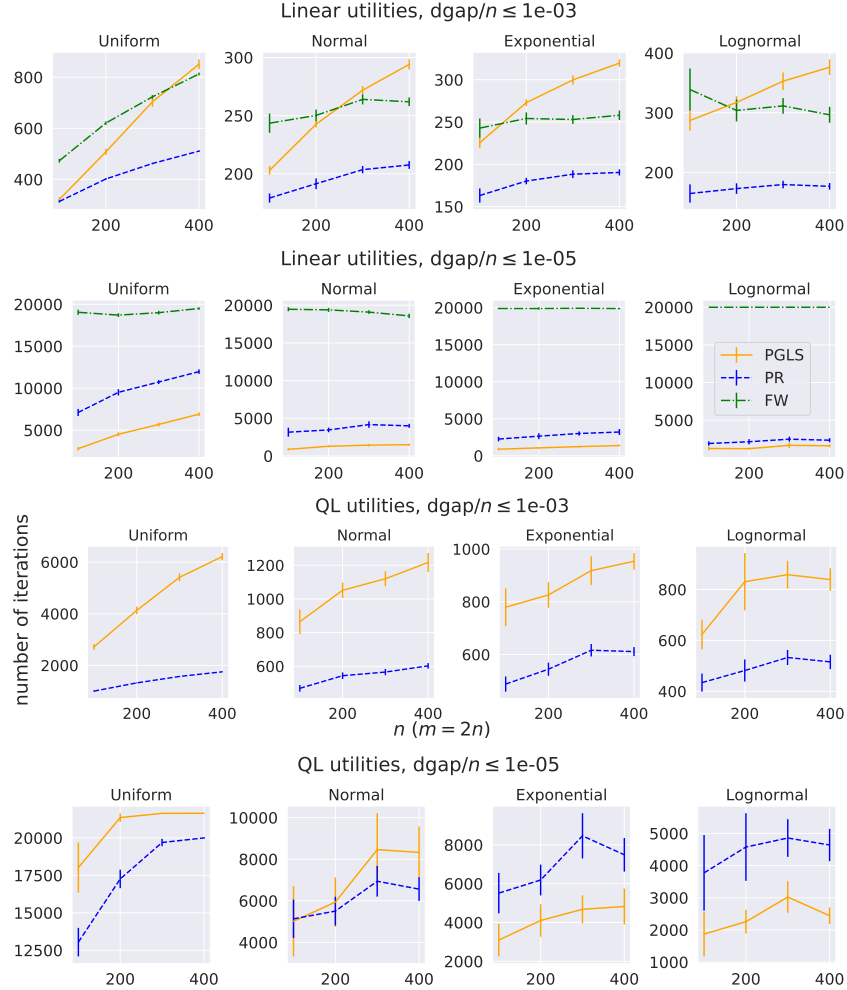
Theorem 8 *PG with fixed stepsize $\gamma_t = \frac{1}{L\|a\|^2}$ for the problem $\min_{p \in \mathcal{P}} \tilde{h}(ap)$ converges linearly at a rate $1 - \frac{\mu}{\max\{\mu, L H_{\mathcal{P}}(a)^2 \|a\|^2\}}$.*

6 Experiments

We perform numerical experiments on market instances under all three utilities with various generated parameters. The algorithms are PGLS (PG with linesearch $\mathcal{LS}_{\alpha, \beta, \Gamma}$, see Appendix A.5), PR and FW (with exact linesearch). For linear utilities, we generate market data $v = (v_{ij})$ where v_{ij} are i.i.d. from standard Gaussian, uniform, exponential, or lognormal distribution. For each of the sizes $n = 50, 100, 150, 200$ (on the horizontal axis) and $m = 2n$, we generate 30 instances with unit budgets $B_i = 1$ and random budgets $B_i = 0.5 + \tilde{B}_i$ (where \tilde{B}_i follows the same distribution as v_{ij}). For QL utilities, we repeat the above (same random v , same sizes and termination conditions) using budgets $B_i = 5(1 + \tilde{B}_i)$. The termination criterion is either (i) $\epsilon(p^t, p^*) = \max_i \frac{|p_i^t - p_i^*|}{p_i^*} \leq \eta$, where p^* are the prices computed by CVXPY+Mosek [28, 50, 26], or (ii) average duality gap $\text{dgap}_t / n \leq \eta$, for various thresholds values η . For PGLS, we report the number of linesearch iterations (that is, the total number of projection computations). For other algorithms, we report the number of iterations. As a fair comparison, we use the same parameters α, β, Γ for PGLS throughout without handpicking. The plots report average numbers of iterations to reach the termination condition and their standard

errors across $k = 30$ repeats. In Appendix E we show: more details on the setup, additional plots with different termination criteria and for Leontief utilities. Codes for the numerical experiments are available at <https://github.com/CoffeeAndConvexity/fom-for-me-codes>.

As can be seen, for linear utilities, PR is more efficient for obtaining an approximate solution (i.e., termination at $\epsilon(p^t, p^*) \leq 10^{-2}$ or $\text{dgap}_t/n \leq 10^{-3}$). When higher accuracy is required, PGLS takes far fewer iterations. For QL utilities, PR is more efficient in most cases, except when very high accuracy is required ($\text{dgap}_t/n \leq 5 \times 10^{-6}$). We do not show FW for the QL case, as it performed extremely badly. In the appendix, we also see that for Leontief utilities, PGLS with linesearch terminates within tens of iterations in all cases.



7 Conclusions

We investigate the computation of market equilibria under different buyer utility assumptions using simple first-order methods. Through convex optimization duality and properties of market equilibria, we show that the associated convex programs can be reformulated into structured forms suitable for FOMs. We show that projected gradient achieves linear convergence for these reformulations, through weakened strong-convexity conditions. As a technical contribution on FOMs, we also prove a more general linear convergence result of proximal gradient with linesearch under the Proximal-PL condition. For QL utilities, we derive a form of Proportional Response through Mirror Descent on a convex optimization formulation, extending the classical convergence results for linear utilities. Finally, extensive numerical experiments compare the efficiency of the FOMs for various low- and high-accuracy termination criteria.

8 Broader Impact

As mentioned in the introduction, large-scale market equilibrium computation problems arise in important applications such as Internet advertising markets, course assignment at universities, fair recommender systems and compute resource allocation. As such, this work has the following potential positive societal impact: Based on this work, resource allocation schemes, especially those with desirable fairness properties previously deemed hard to solve at scale (and are thus simplified or disregarded), can be implemented in reasonable time through computing a market equilibrium by a first-order method. Progress of a FOM can be easily monitored via the duality gap while feasibility is guaranteed. Our work also enables greater scalability of certain Internet advertising market equilibrium models. This could be used for greater market efficiency or better monetization of such markets. Whether this is viewed as a positive or negative thing is beyond the scope of our paper.

9 Funding Transparency Statement

Over the past 36 months, Yuan Gao was a PhD student at Columbia University and was fully funded by Fu Foundation School of Engineering and Applied Sciences (through Department of Industrial Engineering and Operations Research) throughout the development of the present work.

Over the past 36 months, Christian Kroer was employed full time for a year at Facebook Inc., and was at times employed part-time with Facebook Inc. That employment did not fund the present work.

Part of the numerical experiments were run on the computing server of Columbia University Data Science Institute

<https://datascience.columbia.edu/about-us/work-with-us/computing-resources/>.

References

- [1] M. AKBARPOUR AND A. NIKZAD, *Approximate random allocation mechanisms*, Available at SSRN 2422777, (2019).
- [2] M. ANITESCU, *Degenerate nonlinear programming with a quadratic growth condition*, SIAM Journal on Optimization, 10 (2000), pp. 1116–1135.
- [3] A. BECK, *First-order methods in optimization*, vol. 25, SIAM, 2017.
- [4] A. BECK AND S. SHTERN, *Linearly convergent away-step conditional gradient for non-strongly convex functions*, Mathematical Programming, 164 (2017), pp. 1–27.
- [5] A. BECK AND M. TEBoulLE, *Mirror descent and nonlinear projected subgradient methods for convex optimization*, Operations Research Letters, 31 (2003), pp. 167–175.
- [6] ———, *Gradient-based algorithms with applications to signal recovery*, Convex optimization in signal processing and communications, (2009), pp. 42–88.
- [7] X. BEI, J. GARG, AND M. HOEFER, *Ascending-price algorithms for unknown markets*, ACM Transactions on Algorithms (TALG), 15 (2019), pp. 1–33.
- [8] A. BEN-TAL AND A. NEMIROVSKI, *Lectures on modern convex optimization*, Online version: http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt, (2019).
- [9] B. BIRNBAUM, N. R. DEVANUR, AND L. XIAO, *Distributed algorithms via gradient descent for Fisher markets*, in Proceedings of the 12th ACM conference on Electronic commerce, 2011, pp. 127–136.
- [10] S. BRÂNZEI, Y. CHEN, X. DENG, A. FILOS-RATSIKAS, S. K. S. FREDERIKSEN, AND J. ZHANG, *The Fisher market game: equilibrium and welfare*, in Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014.
- [11] N. BROWN AND T. SANDHOLM, *Superhuman ai for heads-up no-limit poker: Libratus beats top professionals*, Science, 359 (2018), pp. 418–424.
- [12] P. BRUCKER, *An $o(n)$ algorithm for quadratic knapsack problems*, Operations Research Letters, 3 (1984), pp. 163–166.
- [13] S. BUBECK ET AL., *Convex optimization: Algorithms and complexity*, Foundations and Trends® in Machine Learning, 8 (2015), pp. 231–357.
- [14] E. BUDISH, *The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes*, Journal of Political Economy, 119 (2011), pp. 1061–1103.
- [15] E. BUDISH, G. P. CACHON, J. B. KESSLER, AND A. OTHMAN, *Course match: A large-scale implementation of approximate competitive equilibrium from equal incomes for combinatorial allocation*, Operations Research, 65 (2016), pp. 314–336.
- [16] E. BUDISH, Y.-K. CHE, F. KOJIMA, AND P. MILGROM, *Designing random allocation mechanisms: Theory and applications*, American Economic Review, 103 (2013), pp. 585–623.
- [17] I. CARAGIANNIS, D. KUROKAWA, H. MOULIN, A. D. PROCACCIA, N. SHAH, AND J. WANG, *The unreasonable fairness of maximum nash welfare*, ACM Transactions on Economics and Computation (TEAC), 7 (2019), pp. 1–32.
- [18] A. I. CHEN AND A. OZDAGLAR, *A fast distributed proximal-gradient method*, in 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), IEEE, 2012, pp. 601–608.
- [19] L. CHEN, Y. YE, AND J. ZHANG, *A note on equilibrium pricing as convex optimization*, in International Workshop on Web and Internet Economics, Springer, 2007, pp. 7–16.
- [20] Y. CHEN AND X. YE, *Projection onto a simplex*, arXiv preprint arXiv:1101.6081, (2011).
- [21] K. L. CLARKSON, *Coresets, sparse greedy approximation, and the frank-wolfe algorithm*, ACM Transactions on Algorithms (TALG), 6 (2010), pp. 1–30.
- [22] R. COLE, N. R. DEVANUR, V. GKATZELIS, K. JAIN, T. MAI, V. V. VAZIRANI, AND S. YAZDANBOD, *Convex program duality, fisher markets, and Nash social welfare*, in 18th ACM Conference on Economics and Computation, EC 2017, Association for Computing Machinery, Inc, 2017.

- [23] L. CONDAT, *Fast projection onto the simplex and the ℓ_1 ball*, Mathematical Programming, 158 (2016), pp. 575–585.
- [24] V. CONITZER, C. KROER, D. PANIGRAHI, O. SCHRIJVERS, E. SODOMKA, N. E. STIER-MOSES, AND C. WILKENS, *Pacing equilibrium in first-price auction markets*, in Proceedings of the 2019 ACM Conference on Economics and Computation, ACM, 2019.
- [25] V. CONITZER, C. KROER, E. SODOMKA, AND N. E. STIER-MOSES, *Multiplicative pacing equilibria in auction markets*, in International Conference on Web and Internet Economics, 2018.
- [26] J. DAHL AND E. D. ANDERSEN, *A primal-dual interior-point algorithm for nonsymmetric exponential-cone optimization*, Optimization Online, (2019).
- [27] N. R. DEVANUR, C. H. PAPADIMITRIOU, A. SABERI, AND V. V. VAZIRANI, *Market equilibrium via a primal-dual-type algorithm*, in Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on, IEEE, 2002, pp. 389–395.
- [28] S. DIAMOND AND S. BOYD, *Cvxpy: A python-embedded modeling language for convex optimization*, The Journal of Machine Learning Research, 17 (2016), pp. 2909–2913.
- [29] D. DRUSVYATSKIY AND A. S. LEWIS, *Error bounds, quadratic growth, and linear convergence of proximal methods*, Mathematics of Operations Research, 43 (2018), pp. 919–948.
- [30] J. DUCHI, S. SHALEV-SHWARTZ, Y. SINGER, AND T. CHANDRA, *Efficient projections onto the l_1 -ball for learning in high dimensions*, in Proceedings of the 25th international conference on Machine learning, 2008, pp. 272–279.
- [31] E. EISENBERG, *Aggregation of utility functions*, Management Science, 7 (1961), pp. 337–350.
- [32] E. EISENBERG AND D. GALE, *Consensus of subjective probabilities: The pari-mutuel method*, The Annals of Mathematical Statistics, 30 (1959), pp. 165–168.
- [33] K. FAN AND G. PALL, *Imbedding conditions for hermitian and normal matrices*, Canadian Journal of Mathematics, 9 (1957), pp. 298–304.
- [34] M. FRANK AND P. WOLFE, *An algorithm for quadratic programming*, Naval research logistics quarterly, 3 (1956), pp. 95–110.
- [35] A. GHODSI, M. ZAHARIA, B. HINDMAN, A. KONWINSKI, S. SHENKER, AND I. STOICA, *Dominant resource fairness: Fair allocation of multiple resource types.*, in Nsdi, vol. 11, 2011, pp. 24–24.
- [36] O. GÜLER, *Foundations of optimization*, vol. 258, Springer Science & Business Media, 2010.
- [37] A. J. HOFFMAN, *On approximate solutions of systems of linear inequalities*, Journal of Research of the National Bureau of Standards, 49 (1952).
- [38] M. JAGGI, *Revisiting frank-wolfe: Projection-free sparse convex optimization.*, in Proceedings of the 30th international conference on machine learning, 2013, pp. 427–435.
- [39] K. JAIN AND V. V. VAZIRANI, *Eisenberg-Gale markets: Algorithms and game-theoretic properties*, Games and Economic Behavior, 70 (2010), pp. 84–106.
- [40] H. KARIMI, J. NUTINI, AND M. SCHMIDT, *Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition*, arXiv preprint arXiv:1608.04636, (2016).
- [41] I. KASH, A. D. PROCACCIA, AND N. SHAH, *No agent left behind: Dynamic fair division of multiple resources*, Journal of Artificial Intelligence Research, 51 (2014), pp. 579–603.
- [42] C. KROER AND A. PEYSAKHOVICH, *Scalable fair division for 'at most one' preferences*, arXiv preprint arXiv:1909.10925, (2019).
- [43] C. KROER, A. PEYSAKHOVICH, E. SODOMKA, AND N. E. STIER-MOSES, *Computing large market equilibria using abstractions*, in Proceedings of the 2019 ACM Conference on Economics and Computation, 2019, pp. 745–746.
- [44] C. KROER, K. WAUGH, F. KILINÇ-KARZAN, AND T. SANDHOLM, *Faster algorithms for extensive-form game solving via improved smoothing functions*, Mathematical Programming, (2020), pp. 1–33.
- [45] S. LACOSTE-JULIEN AND M. JAGGI, *On the global linear convergence of frank-wolfe optimization variants*, in Advances in Neural Information Processing Systems, 2015, pp. 496–504.

- [46] Q. LIN, Z. LU, AND L. XIAO, *An accelerated proximal coordinate gradient method*, in Advances in Neural Information Processing Systems, 2014, pp. 3059–3067.
- [47] S. LOJASIEWICZ, *A topological property of real analytic subsets*, Coll. du CNRS, Les équations aux dérivées partielles, 117 (1963), pp. 87–89.
- [48] Z.-Q. LUO AND P. TSENG, *Error bounds and convergence analysis of feasible descent methods: a general approach*, Annals of Operations Research, 46 (1993), pp. 157–178.
- [49] M. MORAVČÍK, M. SCHMID, N. BURCH, V. LISÝ, D. MORRILL, N. BARD, T. DAVIS, K. WAUGH, M. JOHANSON, AND M. BOWLING, *Deepstack: Expert-level artificial intelligence in heads-up no-limit poker*, Science, 356 (2017), pp. 508–513.
- [50] A. MOSEK, *The mosek optimization software*, Online at <http://www.mosek.com>, 54 (2010), p. 5.
- [51] A. NEMIROVSKI AND D. YUDIN, *Information-based complexity of mathematical programming*, Izvestia AN SSSR, Ser. Tekhnicheskaya Kibernetika, 1 (1983).
- [52] N. NISAN, T. ROUGHGARDEN, E. TARDOS, AND V. V. VAZIRANI, *Algorithmic game theory*, Cambridge University Press, 2007.
- [53] P. M. PARDALOS AND N. KOVOOR, *An algorithm for a singly constrained class of quadratic programs subject to upper and lower bounds*, Mathematical Programming, 46 (1990), pp. 321–328.
- [54] J. PEÑA AND J. C. VERA, *New characterizations of hoffman constants for systems of linear constraints*, Mathematical Programming (Published Online), (2020).
- [55] J. PLESNÍK, *Finding the orthogonal projection of a point onto an affine subspace*, Linear algebra and its applications, 422 (2007), pp. 455–470.
- [56] B. T. POLYAK, *Gradient methods for minimizing functionals*, Zhurnal Vychislitel’noi Matematiki i Matematicheskoi Fiziki, 3 (1963), pp. 643–653.
- [57] R. T. ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, SIAM journal on control and optimization, 14 (1976), pp. 877–898.
- [58] W. SHI, Q. LING, G. WU, AND W. YIN, *A proximal gradient algorithm for decentralized composite optimization*, IEEE Transactions on Signal Processing, 63 (2015), pp. 6013–6023.
- [59] V. I. SHMYREV, *An algorithm for finding equilibrium in the linear exchange model with fixed budgets*, Journal of Applied and Industrial Mathematics, 3 (2009), p. 505.
- [60] A. SKAJAA AND Y. YE, *A homogeneous interior-point algorithm for nonsymmetric convex conic optimization*, Mathematical Programming, 150 (2015), pp. 391–422.
- [61] P. TSENG, *Approximation accuracy, gradient methods, and error bound for structured convex optimization*, Mathematical Programming, 125 (2010), pp. 263–295.
- [62] H. R. VARIAN ET AL., *Equity, envy, and efficiency*, Journal of Economic Theory, 9 (1974), pp. 63–91.
- [63] V. V. VAZIRANI, *Combinatorial algorithms for market equilibria*, Algorithmic Game Theory, (2007), pp. 103–134.
- [64] —, *Spending constraint utilities with applications to the adwords market*, Mathematics of Operations Research, 35 (2010), pp. 458–478.
- [65] P.-W. WANG AND C.-J. LIN, *Iteration complexity of feasible descent methods for convex optimization*, The Journal of Machine Learning Research, 15 (2014), pp. 1523–1548.
- [66] W. WANG AND M. A. CARREIRA-PERPINÁN, *Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application*, arXiv preprint arXiv:1309.1541, (2013).
- [67] Y.-X. WANG, V. SADHANALA, W. DAI, W. NEISWANGER, S. SRA, AND E. XING, *Parallel and distributed block-coordinate frank-wolfe algorithms*, in International Conference on Machine Learning, 2016, pp. 1548–1557.
- [68] L. ZHANG, *Proportional response dynamics in the fisher market*, Theoretical Computer Science, 412 (2011), pp. 2691–2698.

A Preliminaries

We list and prove a few elementary lemmas used in subsequent proofs and discussions.

Lemma 4 *Let $n \geq m$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ be such that $\mathcal{S} = \{x \in \mathbb{R}^n : Ax = b\} \neq \emptyset$. Let the singular value decomposition (SVD) of A^\top be $A^\top = U\Sigma V^\top$, where $U \in \mathbb{R}^{n \times r}$ and $V \in \mathbb{R}^{m \times r}$ have orthonormal columns, $\Sigma = \text{Diag}(\sigma_1, \dots, \sigma_r)$, $\sigma_1 \geq \dots \sigma_r > 0$, $r = \text{rank}(A)$. For any $x \in \mathbb{R}^n$, the projection of x onto \mathcal{S} can be expressed as*

$$\Pi_{\mathcal{S}}(x) = (I - UU^\top)x + U\Sigma^{-1}V^\top b.$$

In particular, when A has full rank,

$$\Pi_{\mathcal{S}}(x) = (I - A^\top(AA^\top)^{-1}A)x + A^\top(AA^\top)^{-1}b.$$

Proof. The case of full rank A is well-known, see, e.g., [55, Eq. (1)]. For general A , $Ax = b \Leftrightarrow V\Sigma U^\top x = b$. Since $V\Sigma$ has full column rank, there exists a unique $w \in \mathbb{R}^r$ such that $V\Sigma w = b$. In fact, $w = \Sigma^{-1}V^\top b$. Therefore,

$$Ax = b \Leftrightarrow U^\top x = \Sigma^{-1}V^\top b.$$

Since U^\top has full row rank, the formula follows directly from the full rank case. \square

Lemma 5 *Under the same assumptions as Lemma 4, for any $x \in \mathbb{R}^n$, it holds that*

$$\sigma_r \cdot \|x - \Pi_{\mathcal{S}}(x)\| \leq \|Ax - b\|.$$

Proof. By Lemma (4) and the fact that U and V have orthonormal columns,

$$\begin{aligned} \|x - \Pi_{\mathcal{S}}(x)\| &= \|U(U^\top x - \Sigma^{-1}V^\top b)\| = \|U^\top x - \Sigma^{-1}V^\top b\| \\ &= \|\Sigma^{-1}V^\top(Ax - b)\| \leq \|\Sigma^{-1}\| \|Ax - b\|, \end{aligned}$$

where $\|\Sigma^{-1}\| = \frac{1}{\sigma_r}$. \square

Lemma 6 *For $B > 0$ and $g \in \mathbb{R}^d$, let*

$$x^* = \arg \min \left\{ \langle g, x \rangle + \sum_{i=1}^d x_i \log x_i : x \geq 0, \mathbf{1}^\top x = B \right\}.$$

Then, $x_i^ = B \cdot \frac{e^{-g_i}}{\sum_{\ell} e^{-g_\ell}}$, $i \in [d]$.*

Proof. It can be easily verified via KKT optimality conditions. The Lagrangian is

$$L(x, \lambda) = \langle g, x \rangle + \sum_i x_i \log x_i - \lambda(\mathbf{1}^\top x - B).$$

By the first-order condition, for any λ , the minimizer $x(\lambda)$ of $L(x, \lambda)$ has $x_i(\lambda) = e^{\lambda-1-g_i}$. Primal feasibility implies $\sum_i x_i(\lambda) = B \Rightarrow e^\lambda = \frac{B}{\sum_i e^{-(1+g_i)}}$. Therefore,

$$x_i^* = B \cdot \frac{e^{-g_i}}{\sum_{\ell} e^{-g_\ell}}.$$

\square

A.1 Proof of Theorem 1

Existence of an optimal solution. Since u_i is concave and continuous on $[0, 1]^m$, we have

$$\sup_{x_i \in [0, 1]^m} u_i(x_i) = M_i < \infty.$$

For any $x \in \mathcal{X} = \{x \in \mathbb{R}_+^{n \times m} : \sum_i x_{ij} \leq 1, \forall j\}$,

$$F(x) := \sum_i B_i \log u_i(x_i) \leq \sum_i B_i M_i < \infty.$$

For each i , since there exists $x'_i > 0$ s.t. $u_i(x'_i) > 0$, by homogeneity, $u_i(\alpha x'_i) > 0$ for any $\alpha > 0$. Hence, there exists $x \in \text{int}(\mathcal{X})$, i.e., $x > 0$ and $\sum_i x_{ij} < 1$ for all j , such that $u_i(x_i) > 0$ for all i . Therefore,

$$-\infty < F^* = \sup_{x \in \mathcal{X}} \sum_i B_i \log u_i(x_i) < \infty.$$

Note that the sup is not affected if we change it to $\sup_{x \in \mathcal{X}'}$, where

$$\mathcal{X}' = \{x \in \mathcal{X} : u_i(x_i) \geq \epsilon\}$$

for some sufficiently small $\epsilon > 0$. Here, \mathcal{X}' is a compact set, on which F is continuous. Therefore, $F^* = F(x^*)$ for some $x^* \in \mathcal{X}$.

Next, we show that an optimal solution of (1) gives a ME. Consider the minimization form of (1). Let $p_j \geq 0$ be the dual variable associated with constraint $\sum_i x_{ij} \leq 1$. The Lagrangian is

$$\mathcal{L}(x, p) = \left[-\sum_i B_i \log u_i(x) + \left\langle p, \sum_i x_i \right\rangle - \sum_j p_j \right].$$

The Lagrangian dual is

$$\max_{p \geq 0} g(p) := \min_{x \geq 0} \mathcal{L}(x, p). \quad (8)$$

As shown above, Since u_i are homogeneous, (1) has a strictly feasible solution with finite objective value. (The dual (8) clearly has a strictly feasible solution $p > 0$ with finite objective value $g(p) = \min_{x \geq 0} \mathcal{L}(x, p)$.) Therefore, strong duality holds by Slater's condition and the KKT conditions are necessary and sufficient for (primal and dual) optimality of a solution pair (x, p) (see, e.g., [8, Appendix D]). Let x^* and p^* be optimal solutions to the primal (1) and dual (8), respectively. Clearly, $u_i(x_i^*) > 0$ for all i (since F^* is finite). By Lagrange duality, we have

$$x^* \in \arg \min_{x \geq 0} \mathcal{L}(x, p^*).$$

In other words, each x_i^* maximizes

$$r_i(x_i, p^*) := B_i \log u_i(x_i) - \langle p^*, x_i \rangle$$

on $x_i \geq 0$. We show that (x^*, p^*) is a market equilibrium.

Buyer optimality First, we verify that $\langle p^*, x_i^* \rangle = B_i$ for all i . Consider the smooth function

$$\phi(\epsilon) = B_i \log u_i((1 + \epsilon)x_i^*) - \langle p^*, (1 + \epsilon)x_i^* \rangle = r_i(x_i^*, p^*) + B_i \log(1 + \epsilon) - \epsilon \langle p^*, x_i^* \rangle.$$

Note that ϕ is differentiable on $(0, 1)$ and

$$\phi'(\epsilon) = \frac{B_i}{1 + \epsilon} - \langle p^*, x_i^* \rangle.$$

Assume that $\langle p^*, x_i^* \rangle < B_i$ for some i . Then, $\phi'(\epsilon) > 0$ for $\epsilon \in \left[0, \frac{B_i}{\langle p^*, x_i^* \rangle} - 1\right)$. Hence, replacing x_i^* with $(1 + \epsilon)x_i^*$ for a sufficiently small $\epsilon > 0$ strictly increases $r_i(x_i, p^*)$, contradicting to the choice of x_i^* . Hence, $\langle p^*, x_i^* \rangle \leq B_i$. Completely analogously, $\langle p^*, x_i^* \rangle \geq B_i$. Therefore, for each buyer i , x_i^* is feasible and depletes its budget B_i under prices p^* . Hence, for any $x_i \in \mathbb{R}_+^m$ such that $\langle p^*, x_i \rangle \leq B_i$ (a budget-feasible bundle), since x_i^* maximizes $r_i(x_i, p^*)$, we have

$$B_i \log u_i(x_i^*) - \langle p^*, x_i^* \rangle \geq B_i \log u_i(x_i) - \langle p^*, x_i \rangle.$$

Since $\langle p^*, x_i \rangle \leq B_i = \langle p^*, x_i^* \rangle$, the above implies

$$B_i \log u_i(x_i^*) \geq B_i \log u_i(x_i).$$

Therefore, $u_i(x_i^*) \geq u_i(x_i)$. In other words, $x_i^* \in D_i(p^*)$ (buyer i is optimal) for all i .

Market clearance By the complementary slackness condition regarding the optimal (x^*, p^*) , for item j such that $\sum_i x_{ij}^* < 1$, it must hold that $p_j^* = 0$.

A.2 Characterizations of Hoffman constant

We compare our definition of Hoffman constant and another common, explicit characterization. Recall that $H_{\mathcal{X}}(A)$ is the smallest H such that, for any b , $\mathcal{S} = \{x : Ax = b\}$,

$$\|x - \Pi_{\mathcal{X} \cap \mathcal{S}}(x)\| \leq H \|Ax - b\|, \forall x \in \mathcal{X}.$$

For any matrix M , let $\mathcal{B}(M)$ be the set of *nonsingular* submatrices consisting of rows of M . Define

$$H(M) = \max_{B \in \mathcal{B}(M)} \frac{1}{\sigma_{\min}(B)} < \infty. \quad (9)$$

The following fact is known (see, e.g., [36, §11.8] and [4, §2.1]).

Lemma 7 *Suppose the reference polyhedral set can be represented by inequality constraints $\mathcal{X} = \{x : Cx \leq d\}$. Then,*

$$H_{\mathcal{X}}(A) \leq H \left(\begin{bmatrix} A \\ C \end{bmatrix} \right).$$

Clearly, $H(M)$ is finite for any M . In fact, this is the most well-known characterization of Hoffman constant, and is tight in the following sense: let $\mathcal{S} = \{x : Ax = b\}$ for some arbitrary right hand side b , then it is the smallest constant H such that

$$\|x - \Pi_{\mathcal{X} \cap \mathcal{S}}(x)\| \leq H \left\| \begin{bmatrix} Ax - b \\ (Cx - d)_+ \end{bmatrix} \right\|$$

for all x (not necessarily $\in \mathcal{X}$). However, for all of our purposes, that is, analysis of PG, x is always restricted to be $\in \mathcal{X}$. Therefore, we choose to define $H_{\mathcal{X}}(A)$ as such, consistent with [5] and [54]. Meanwhile, the following is clear.

Lemma 8 *For any matrices $A \in \mathbb{R}^{m \times n}$, $m \leq n$ and $C \in \mathbb{R}^{\ell \times n}$, it holds that*

$$H \left(\begin{bmatrix} A \\ C \end{bmatrix} \right) \geq \max \left\{ \frac{1}{\sigma_{\min}(A)}, H(A) \right\}.$$

Proof. By definition (9), $H' := H \left(\begin{bmatrix} A \\ C^\top \end{bmatrix} \right) \geq H(A)$. If $\text{rank}(A) = m$, then $H' \geq \frac{1}{\sigma_{\min}(A)}$

because $A \in \mathcal{B} \left(\begin{bmatrix} A \\ C \end{bmatrix} \right)$. If $r = \text{rank}(A) < m$, let the (nonzero) singular values of A be $\sigma_1 \geq \dots \geq$

$\sigma_r = \sigma_{\min}(A) > 0$. Consider any $B \in \mathcal{B}(A) \subseteq \mathcal{B} \left(\begin{bmatrix} A \\ C \end{bmatrix} \right)$ with rank r (having exactly r rows), let its nonzero singular values be $\sigma'_1 \geq \dots \geq \sigma'_r = \sigma_{\min}(B) > 0$. Applying Cauchy's Interlacing Theorem (see, e.g., [33, Theorem 1]) on AA^\top and its principal submatrix BB^\top , we have

$$\sigma_1 \geq \sigma'_1 \geq \dots \geq \sigma_r \geq \sigma'_r.$$

Therefore, $H' \geq \frac{1}{\sigma_{\min}(B)} \geq \frac{1}{\sigma_{\min}(A)}$. □

A.3 Proof of Theorem 2

We follow the development in [40, §4 & Appendix F] and further articulate the constants. There, the authors show that proximal gradient achieves linear convergence under the so-called *Proximal-PL* inequality. Consider the following general nonsmooth problem

$$F^* = \min_x F(x) = f(x) + g(x) \quad (10)$$

where f is smooth convex with L_f -Lipschitz continuous gradient, g is simple closed proper convex and $\text{dom } g \subseteq \text{dom } f$. One iteration of the proximal gradient method with stepsize $\gamma > 0$ is as follows:

$$x^{t+1} = \text{Prox}_g(x^t - \gamma \nabla f(x^t)) = \arg \min_x \left[\langle \gamma \nabla f(x), x - x^t \rangle + \frac{1}{2} \|x - x^t\|^2 + g(x) \right]. \quad (11)$$

For any $\alpha > 0$ and any $x \in \text{dom } g$, define

$$\mathcal{D}(x, \alpha) = -2\alpha \min_{x'} \left[\langle \nabla f(x), x' - x \rangle + \frac{\alpha}{2} \|x' - x\|^2 + g(x') - g(x) \right]. \quad (12)$$

Say that $F = f + g$ satisfies the proximal-PŁ inequality at x w.r.t. $\Lambda \geq \lambda > 0$ if

$$\frac{1}{2} \mathcal{D}(x, \Lambda) \geq \lambda(F(x) - F^*), \quad (13)$$

Below is essentially [40, Theorem 5], which shows that the so-called Proximal-PŁ condition is sufficient for linear convergence. Note that, different from [40, Theorem 5], we only require (13) to hold for $x \in \mathcal{X}$ such that $F(x) \leq F(x^0)$ instead of all $x \in \mathcal{X}$. In addition, we note that in some cases (13) may hold with $\Lambda > L_f$, in which case the rate needs to be slightly adjusted. Since $\mathcal{D}(x, \cdot)$ is monotone [40, Lemma 1], (13) holds when Γ is replaced by $\Gamma' \geq \Gamma$. The statement and proof are the same as [40, pp. 9] otherwise.

Theorem 9 *Let $x^0 \in \text{dom } g$. If f and g satisfies (13) for all $x \in \text{dom } g$ such that $F(x) \leq F(x^0)$, then x^t defined by (11) starting from x^0 with constant stepsize $\gamma = 1/L_f$ converges linearly with rate $1 - \frac{\lambda}{\bar{L}}$, where $\bar{L} = \max\{\Lambda, L_f\}$. In other words,*

$$F(x^t) - F^* \leq \left(1 - \frac{\lambda}{\bar{L}}\right)^t (F(x^0) - F^*), \quad t = 1, 2, \dots$$

Proof. By assumption, (13) holds for all $x \in \text{dom } g$, $x \leq F(x^0)$. In particular, it holds for x^t , $t = 1, 2, \dots$, since proximal gradient is a *descent* method, i.e., $F(x^0) \geq F(x^1) \geq \dots$ (see, e.g., [3, Corollary 10.18]). Therefore, by L_f -Lipschitz continuity of ∇f , proximal gradient update (11), definition of $\mathcal{D}(x, \cdot)$, its monotonicity, and (13) for all x^t ,

$$\begin{aligned} F(x^{t+1}) &\leq F(x^t) + \langle \nabla f(x^t), x^{t+1} - x^t \rangle + \frac{L_f}{2} \|x^{t+1} - x^t\|^2 + g(x^{t+1}) - g(x^t) \\ &\leq F(x^t) + \left[\langle \nabla f(x^t), x^{t+1} - x^t \rangle + \frac{\bar{L}}{2} \|x^{t+1} - x^t\|^2 + g(x^{t+1}) - g(x^t) \right] \\ &\leq F(x^t) - \frac{1}{2\bar{L}} \mathcal{D}(x^t, \bar{L}) \\ &\leq F(x^t) - \frac{\lambda}{\bar{L}} (F(x^t) - F^*) \\ \Rightarrow F(x^{t+1}) - F^* &\leq \left(1 - \frac{\lambda}{\bar{L}}\right) (F(x^t) - F^*). \end{aligned}$$

Repeatedly applying the above inequality completes the proof. \square

Then, we prove Theorem 2. Clearly, problem (2) is (10) with $g(x) = \delta_{\mathcal{X}}(x)$ and PG is a special case of proximal gradient. By Theorem 9, in order to prove Theorem 2, it suffices to establish the Proximal-PŁ condition (13) (for all $x \in \mathcal{X}$, $f(x) \leq f(x^0)$ for some initial iterate x^0). Let \mathcal{X}^* be the set of optimal solutions to (2) and f^* be the optimal objective value. Since h is μ -strongly convex and $f(x) = h(Ax)$, there exists $z^* \in \text{dom } f$ such that $\mathcal{S} = \{x : Ax = z^*\}$ and $\mathcal{X}^* = \mathcal{X} \cap \mathcal{S}$. Therefore, for any $x \in \mathcal{X}$, $x_p := \Pi_{\mathcal{X}^*}(x)$, we have

$$f(x_p) = h(Ax_p) \geq h(Ax) + \langle \nabla h(Ax), Ax_p - x \rangle + \frac{\mu}{2} \|Ax_p - x\|^2.$$

Note that

$$\langle \nabla h(Ax), Ax_p - x \rangle = \langle A^\top \nabla h(Ax), x_p - x \rangle = \langle \nabla f(x), x_p - x \rangle.$$

Hence, for any $x \in \mathcal{X}$, by strong convexity of h and definition of $H = H_{\mathcal{X}^*}(A)$, we have

$$\begin{aligned} f(x_p) &\geq f(x) + \langle \nabla f(x), x_p - x \rangle + \frac{\mu}{2} \|Ax - x_p\|^2 \\ &= f(x) + \langle \nabla f(x), x_p - x \rangle + \frac{\mu}{2} \|Ax - z^*\|^2 \\ &\geq f(x) + \langle \nabla f(x), x_p - x \rangle + \frac{\mu}{2H^2} \|x - x_p\|^2, \end{aligned}$$

Therefore,

$$\begin{aligned}
f^* &\geq f(x) + \langle \nabla f(x), x_p - x \rangle + \frac{\mu}{2H^2} \|x - x_p\|^2 \\
&\geq f(x) + \min_{y \in \mathcal{X}} \left\{ \langle \nabla f(x), y - x \rangle + \frac{\mu}{2H^2} \|y - x\|^2 \right\} \\
&\geq f(x) - \frac{H^2}{2\mu} \mathcal{D}\left(x, \frac{\mu}{H^2}\right) \\
&\Rightarrow \frac{1}{2} \mathcal{D}\left(x, \frac{\mu}{H^2}\right) \geq \frac{\mu}{H^2} (f(x) - f^*).
\end{aligned}$$

Thus, (13) holds for all $x \in \mathcal{X}$, $f(x) \leq f(x^0)$ with

$$\Lambda = \lambda = \frac{\mu}{H^2}.$$

Since $\nabla f(x) = A^\top \nabla h(Ax)$ and h is (μ, L) -s.c., its Lipschitz constant can be chosen as

$$L_f = L\|A\|^2.$$

By Theorem 9, PG with stepsize $\gamma = \frac{1}{L_f}$ converges linearly with rate

$$1 - \frac{\frac{\mu}{H^2}}{\max\left\{\frac{\mu}{H^2}, L\|A\|^2\right\}} = 1 - \frac{\mu}{\max\{\mu, LH^2\|A\|^2\}}.$$

Finally, convergence of the distance to optimality $\|x^t - \Pi_{\mathcal{X}^*}(x^t)\|$ is straightforward: for any $x \in \mathcal{X}$, by the strong convexity of h and definition of H ,

$$f(x) - f^* = h(Ax) - h(Ax_p) \geq \frac{\mu}{2} \|Ax - Ax_p\|^2 = \frac{\mu}{2} \|Ax - z^*\|^2 \geq \frac{\mu}{2H} \|x - x_p\|^2.$$

□

Remark A special case is when $d \geq r$ (recall that $A \in \mathbb{R}^{d \times r}$) and $\text{rank}(A) = r$. In this case, $f(x) = h(Ax)$ itself is strongly convex with modulus $\mu\sigma_{\min}(A)^2$. In this case, classical analysis (e.g., [3, §10.6]) implies linear convergence with rate $1 - \frac{\mu\sigma_{\min}(A)^2}{L\|A\|^2}$. Meanwhile, in the above analysis, we have $\mathcal{X}^* = \{x^*\} = \mathcal{S} = \{x : Ax = z^*\} = \mathcal{X} \cap \mathcal{S}$ (since x^*, z^* are unique and $\text{rank}(A) = r$). By Lemma 5, for any x , it holds that

$$\|x - \Pi_{\mathcal{X}^*}(x)\| \leq \frac{1}{\sigma_{\min}(A)^2} \|Ax - z^*\|.$$

Therefore, by the definition of Hoffman constant, $H_{\mathcal{X}^*}(A) \leq \frac{1}{\sigma_{\min}(A)^2}$ and the classical rate under strong convexity is recovered.

A.4 Proof of Theorem 3

Let \mathcal{X}^* be the set of optimal solutions to (2). First, recall the following lemma [65, Lemma 14], which ensures the first part of the theorem, that is, uniqueness of Ax^* and $q^\top x^*$ for all $x^* \in \mathcal{X}^*$.

Lemma 9 *There exist unique $z^* \in \mathbb{R}^r$ and $w^* \in \mathbb{R}$ such that for any $x^* \in \mathcal{X}^*$,*

$$Ax^* = z^*, \quad \langle q^*, x^* \rangle = w^*.$$

The next lemma is essentially [4, Lemma 2.5]. Different from the statement of [4, Lemma 2.5], we keep $\|\nabla h(z^*)\|$ instead of bounding it by $\sup_{x \in \mathcal{X}} \|\nabla h(Ax)\|$. We also define $C = f(x^0) - f^*$ instead of $C = \sup_{x \in \mathcal{X}} f(x) - f^*$, since subsequent application of the lemma only involves PG iterates x^t , which have monotone decreasing objective values $f(x^0) \geq f(x^1) \geq \dots$. The proof remains unchanged otherwise.

Lemma 10 Let z^* be as in Lemma 9 and $x^0 \in \mathcal{X}$. For any $x \in \mathcal{X}$ such that $f(x) \leq f(x^0)$, it holds that

$$\|x - \Pi_{\mathcal{X}^*}(x)\|^2 \leq \kappa(f(x) - f^*),$$

where, same as in Theorem 3, $\kappa = H_{\mathcal{X}}(A)^2 \left(C + 2GD_A + \frac{2(G^2+1)}{\mu} \right)$, $C = f(x^0) - f^*$, $G = \|\nabla h(z^*)\|$, $D_A = \sup_{x,y \in \mathcal{X}} \|A(x-y)\|$.

Finally, take $L_f = L\|A\|^2$ as a Lipschitz constant of ∇f . By Lemma 10 and [40, §4.1], it holds that (2) satisfies the proximal-PŁ inequality (13) with

$$\Lambda = \lambda = \frac{1}{2\kappa}$$

for all $x \in \mathcal{X}$ such that $f(x) \leq f(x^0)$ (in particular, for all x^t , $t = 1, 2, \dots$). By Theorem 9, PG converges linearly with rate $1 - \frac{\lambda}{\max\{\Lambda, L_f\}} = 1 - \frac{1}{\max\{1, 2\kappa L\|A\|^2\}}$.

Remark Lemma 10 shows that QG holds. Similar convergence guarantees can also be derived from other QG-based analysis, e.g., [29, Corollary 3.7].

A.5 Linear convergence of PG with linesearch

First, we consider the more general proximal gradient setup (10). Let L_f be a Lipschitz constant of ∇f and the Proximal-PŁ inequality 13 holds with $\Lambda \geq \lambda \geq 0$ for all $x \in \text{dom } g$ such that $F(x) \leq F(x^0)$. Let $\alpha \geq 1$, $\beta \in (0, 1)$, $\Gamma > 0$ (increment factor, decrement factor, upper bound on stepsize, respectively). The linesearch subroutine $\mathcal{LS}_{\alpha, \beta, \Gamma}$ is defined in Algorithm 1.

Algorithm 1 $x_{t+1}, \gamma_t, k_t \leftarrow \mathcal{LS}_{\alpha, \beta, \Gamma}(x, \gamma, k_{\text{prev}})$ with parameters $\alpha \geq 1$, $\beta \in (0, 1)$, $\Gamma > 0$.

If $k_{\text{prev}} = 0$, set $\gamma^{(0)} = \min\{\alpha\gamma, \Gamma\}$. Otherwise, set $\gamma^{(0)} = \gamma$.

For $k = 0, 1, 2, \dots$

1. Compute $x^{(k)} = \text{Prox}_{\lambda^{(k)}g}(x - \gamma^{(k)}\nabla f(x))$.
2. Break if

$$f(x^{(k)}) \leq f(x) + \langle \nabla f(x), x^{(k)} - x \rangle + \frac{1}{2\gamma^{(k)}} \|x^{(k)} - x\|^2. \quad (14)$$

3. Set $\gamma^{(k+1)} = \beta\gamma^{(k)}$ and continue to $k+1$.

Return $x_{t+1} = x^{(k)}$, $\gamma_t = \gamma^{(k)}$, $k_t = k$.

In this way, proximal gradient with linesearch can be described formally as follows: starting from $x^0 \in \text{dom } f$, $\gamma_{-1} = \Gamma$, $k_{-1} = 0$, perform the following iterations

$$(x^{t+1}, \gamma_t, k_t) \leftarrow \mathcal{LS}_{\alpha, \beta, \Gamma}(x^t, \gamma_{t-1}, k_{t-1}), \quad t = 1, 2, \dots$$

Note that (14) holds for any $\gamma^{(k)} \leq \frac{1}{L_f}$ (see, e.g., [3, Theorem 10.16]). Therefore, Algorithm 1 terminates when $\gamma^{(0)}\beta^k \leq \frac{1}{L_f}$. This means

$$\gamma_t \geq \tilde{\gamma} := \min \left\{ \Gamma, \frac{\beta}{L_f} \right\}. \quad (15)$$

for all t . Note that we explicitly include the case of $\Gamma \leq \frac{1}{L_f}$, although in practice Γ is often set very large. Clearly,

$$\Gamma\beta^k \leq \tilde{\gamma} \Leftrightarrow k \geq \frac{\log \frac{\Gamma}{\tilde{\gamma}}}{\log \frac{1}{\beta}}.$$

Therefore, in Algorithm 1, the backtracking iteration index satisfies $k_t \leq \frac{\log \frac{\Gamma}{\tilde{\gamma}}}{\log \frac{1}{\beta}}$ for all t . Note that if the loop breaks at k_t , the number of Prox evaluations is exactly $k_t + 1$.

Let

$$\bar{L} = \max \left\{ \frac{1}{\bar{\gamma}}, \Lambda \right\} = \max \left\{ \frac{1}{\Gamma}, \frac{L_f}{\beta}, \Lambda \right\}. \quad (16)$$

Then, monotonicity of $D(x, \cdot)$ implies, for all $x \in \text{dom } g$ such that $F(x) \leq F(x^0)$,

$$\frac{1}{2} \mathcal{D}(x, \bar{L}) \geq \frac{1}{2} \mathcal{D}(x, \Lambda) \geq \lambda (F(x) - F^*).$$

Following the proof of Theorem 9 (or that of [40, Theorem 5]), we have

$$\begin{aligned} F(x^{t+1}) &\leq F(x^t) + \langle \nabla f(x^t), x^{t+1} - x^t \rangle + \frac{L_f}{2} \|x^{t+1} - x^t\|^2 + g(x^{t+1}) - g(x^t) \\ &\leq F(x^t) + \langle \nabla f(x^t), x^{t+1} - x^t \rangle + \frac{\bar{L}}{2} \|x^{t+1} - x^t\|^2 + g(x^{t+1}) - g(x^t) \\ &\leq F(x^t) - \frac{1}{2\bar{L}} \mathcal{D}(x^t, \bar{L}) \\ &\leq F(x^t) - \frac{\lambda}{\bar{L}} (F(x^t) - F^*) \\ \Rightarrow F(x^{t+1}) - F^* &\leq \left(1 - \frac{\lambda}{\bar{L}}\right) (F(x^t) - F^*). \end{aligned}$$

Summarizing the above discussion, we have the following convergence guarantee for PG with linesearch.

Theorem 10 *Let $\alpha \geq 1$, $\beta \in (0, 1)$ and $\Gamma > 0$. For problem (10) satisfying the Proximal-PL inequality with $\Lambda \geq \lambda > 0$ for all $x \in \text{dom } g$ such that $F(x) \leq F(x^0)$, proximal gradient (11) with linesearch subroutine $\mathcal{LS}_{\alpha, \beta, \Gamma}$ described in Algorithm 1 generates iterates x^t such that*

$$F(x^{t+1}) - F^* \leq \left(1 - \frac{\lambda}{\bar{L}}\right)^t (F(x^0) - F^*), \quad t = 1, 2, \dots, \quad (17)$$

where \bar{L} is defined in (16). Furthermore, each iteration requires at most $1 + \frac{\log \frac{\Gamma}{\bar{\gamma}}}{\log \frac{1}{\beta}}$ number of Prox evaluations.

Proof of Theorem 4. In the above discussion, when $g(x) = \delta_{\mathcal{X}}(x)$, we can replace the Lipschitz constant L_f by the restricted one $L_f^{\mathcal{X}}$ throughout, since Algorithm 1 ensures $x^t \in \mathcal{X}$ for all t . It remains to apply Theorem 10. For $q = 0$, $\Lambda = \lambda = \frac{\mu}{H^2}$ and $\bar{L} = \max \left\{ \frac{1}{\Gamma}, \frac{L_f^{\mathcal{X}}}{\beta}, \frac{\mu}{H^2} \right\}$. Therefore, the rate is

$$1 - \frac{\lambda}{\bar{L}} = 1 - \frac{\mu}{\max\{\mu, H^2/\Gamma, H^2 L_f^{\mathcal{X}}/\beta\}}.$$

For $q \neq 0$, $\Lambda = \lambda = \frac{1}{2\kappa}$ and $\bar{L} = \max \left\{ \frac{1}{\Gamma}, \frac{L_f^{\mathcal{X}}}{\beta}, \frac{1}{2\kappa} \right\}$. Therefore, the rate is

$$1 - \frac{1}{\max\{1, 2\kappa L_f^{\mathcal{X}}/\beta, 2\kappa/\Gamma\}}.$$

□

A.6 Other utility functions

Recall that, by Theorem (1), for any CCNH utilities u_i , optimal solutions to the EG convex program (1) correspond to equilibrium allocation and prices.

CES utilities are parametrized by a nondegenerate v and exponent $\rho \in (-\infty, 1] \setminus \{0\}$:

$$u_i(x_i) = \left(\sum_{j=1}^m v_{ij} x_{ij}^{\rho} \right)^{1/\rho}.$$

Clearly, $\rho = 1$ gives linear utilities. For $\rho < 1$, it has been shown that Proportional Response dynamics achieves linear convergence in prices and utilities [68, Theorem 4] under their notion of ϵ -approximate market equilibrium [68, pp. 2693].

Cobb-Douglas utilities represent substitutive items and take the following form, for parameters $\lambda = (\lambda_i), \lambda_i \in \Delta_m$:

$$u_i(x_i) = \prod_j x_{ij}^{\lambda_{ij}}.$$

In this case, EG (1) decomposes item-wise into simple problems with explicit solutions. Specifically, for each item j , the minimization problem is

$$\min_{x_{:,j} \in \Delta_n} - \sum_i B_i \lambda_{ij} \log x_{ij}.$$

Let p_j be the Lagrangian multiplier associated with constraint $\sum_i x_{ij} = 1$. The Lagrangian is

$$\mathcal{L}(x_{:,j}, p_j) = - \sum_i B_i \lambda_{ij} \log x_{ij} + p_j \left(\sum_i x_{ij} - 1 \right).$$

By first-order stationarity condition, for any $p_j \in \mathbb{R}$, $\mathcal{L}(x_{:,j}, p_j)$ is minimized when

$$x_{ij} = \frac{B_i \lambda_{ij}}{p_j}. \quad (18)$$

Substituting it into \mathcal{L} and discarding the constants w.r.t. p_j , we have

$$g(p_j) = \left(\sum_i B_i \lambda_{ij} \right) \log p_j - p_j,$$

which is maximized at equilibrium prices

$$p_j^* = \sum_i B_i \lambda_{ij}.$$

Therefore, by 18, the equilibrium x^* under Cobb-Douglas utilities is given by

$$x_{ij}^* = \frac{B_i \lambda_{ij}}{\sum_i B_i \lambda_{ij}}, \forall i, j.$$

B Linear utilities

B.1 Shmyrev's convex program

Under linear utilities, it turns out that we can also compute market equilibrium via the following convex program due to Shmyrev [59, 9]. In this convex program, the variables are the *bids* $b_{ij}, i \in [n], j \in [m]$ and prices $p_j, j \in [m]$.

$$\max \sum_{i,j} b_{ij} \log v_{ij} - \sum_j p_j \log p_j \quad \text{s.t.} \quad \sum_i b_{ij} = p_j, j \in [m], \sum_j b_{ij} = B_i, i \in [n], b \geq 0. \quad (19)$$

Given an optimal solution b^* , equilibrium prices and allocations are then given by $p_j^* = \sum_i b_{ij}^*$ and $x_{ij}^* = \frac{b_{ij}^*}{p_j^*}$, respectively.

B.2 Proof of Lemma 1

Any $x \in \mathcal{X}$ satisfies $x \leq 1$. Therefore, $\langle v_i, x_i \rangle \leq \|v_i\|_1 \|x_i^*\|_\infty \leq \|v_i\|_1 = \bar{u}_i$. For the lower bound, recall that at an equilibrium allocation x^* ensures that every buyer gets at least the utility of the proportional share, that is,

$$\langle v_i, x_i^* \rangle \geq \left\langle v_i, \frac{B_i}{\|B\|_1} \mathbf{1} \right\rangle = \frac{B_i \|v_i\|_1}{\|B\|_1} = \bar{u}_i.$$

B.3 Uniqueness of equilibrium quantities and convergence of u^t, p^t

Convergence of u^t to u^* can be easily seen as follows. Let x^t be the PG iterates and $\tilde{h}, A, f = \tilde{h}(Ax)$, μ be defined as in §3 and $f^* = \min_{x \in \mathcal{X}} f(x)$. Since \tilde{h} is μ -strongly convex, we have

$$\frac{\mu}{2} \|u^t - u^*\|^2 \leq \tilde{h}(u^t) - \tilde{h}(u^*) \leq \tilde{h}(Ax^t) - f^*,$$

which converges linearly. Next, we show uniqueness of p^* via simple arguments and construct a sequence of linearly convergent prices p^t .

Lemma 11 *Assume that v is nondegenerate. Then, the equilibrium prices p^* under linear utilities are unique.*

Proof. By Theorem 1 and [22, Lemma 3], p^* is an optimal solution (together with some β^*) to the following problem (dual of (1) with linear utilities): \square

$$\min_{p, \beta} \sum_j p_j - \sum_i B_i \log \beta_i \quad \text{s.t. } p \geq 0, \beta \geq 0, p_j \geq v_{ij} \beta_i, \forall i, j. \quad (20)$$

Here, strong duality holds since there clearly exist primal and dual strictly feasible solutions with finite objective values given nondegenerate v (c.f. Theorem 1 and Appendix A.1). We can eliminate p by letting $p_j = \max_i v_{ij} \beta_i$ for all j and rewrite (20) as

$$\min_{\beta} \sum_j \max_i v_{ij} \beta_j - \sum_i B_i \log \beta_i \quad \text{s.t. } \beta \geq 0.$$

In the above, since the objective is strongly convex and the feasible region is $\beta \geq 0$, the optimal solution β^* is clearly unique. Furthermore, it must hold that $\beta^* > 0$ (since the optimal objective value is finite and strong duality holds). For p^* optimal to (20), it must hold that $p_j^* = \max_i v_{ij} \beta_i^*$. In fact, $p_j^* \geq \max_i v_{ij} \beta_i^*$ by feasibility and, for any strict inequality, decreasing the corresponding p_j^* strictly decreases the objective. \square

The following lemma provides simple upper and lower bounds on feasible and equilibrium prices, respectively. The lower bounds are slightly strengthened over the existing one [9, Lemma 17].

Lemma 12 *Let p^* be equilibrium prices under linear utilities with nondegenerate valuations v . Then, $\underline{p}_j \leq p_j^* \leq \bar{p}_j$ for all j , where $\underline{p}_j = \max_i \frac{v_{ij} B_i}{\|v_i\|_1}$ and $\bar{p}_j = \|B\|_1$.*

Proof. It is essentially the same as the proof of Lemma 2, except that, at optimality, $u_i \leq \|v_i\|_1 + B_i$ can be strengthened to $u_i \leq \|v_i\|_1$ (utility of each buyer is at most that of having a unit of every item). \square

A linearly convergent sequence of p^t . Here, all norms are vector norms. Note that each step of PG is of the form $x^{t+1} = \Pi_{\mathcal{X}}(\tilde{x}^t)$, where $\tilde{x}^t = x^t - \gamma \nabla f(x^t)$. Since ∇f is L_f -Lipschitz, the mapping

$$\phi_1 : x \mapsto x - \gamma \nabla f(x^t)$$

is Lipschitz continuous (w.r.t. $\|\cdot\|_2$) with constant $1 + \gamma L_f = 2$ (where $\gamma = \frac{1}{L\|A\|^2}$ is the fixed stepsize). Meanwhile, we have the following.

Lemma 13 *Let $y \in \mathbb{R}^n$ and $y^* = \Pi_{\Delta^n}(y)$. There exists a unique multiplier $\lambda \in \mathbb{R}$, which can be computed in $O(n \log n)$ time, such that*

$$\sum_{i=1}^n (y_i - \lambda)_+ = 1. \quad (21)$$

Moreover, the mapping $\phi_2 : y \mapsto \lambda$ is piecewise linear and 1-Lipschitz continuous w.r.t. $\|\cdot\|_1$.

Proof. By the KKT conditions for simplex projection (see, e.g., [66, §3]), it holds that there exists unique λ such that

$$y^* = (y - \lambda \mathbf{1})_+.$$

Suppose there exists $\lambda_1 < \lambda_2$ that satisfy (21). Then, since the left-hand side of (21), denoted as $w(\lambda)$, is monotone decreasing in λ , it must hold that $w(\lambda) = 1$ for all $\lambda \in [\lambda_1, \lambda_2]$. In other words, $w(\cdot)$ is *constant* on $[w_1, w_2]$. This further implies $w(\lambda) = 0$ for all $w \in [w_1, w_2]$, a contradiction. Therefore, $\lambda = \phi_2(y)$ is uniquely defined. Let $I^+(y)$, $I^0(y)$, $I^-(y)$ denote the set of indices $i \in [n]$ such that $y_i > \lambda$, $y_i = \lambda$, $y_i < \lambda$, respectively (where $\lambda = \phi_2(y)$). We have

$$\lambda = \frac{\sum_{i \in I^+(y)} y_i - 1}{|I^+(y)|} = \frac{\sum_{i \in I^+(y) \cup I^0(y)} y_i - 1}{|I^+(y)| + |I^0(y)|},$$

which is piecewise linear in y since there are only finitely many index possible sets of indices and $I^+(y)$ is always nonempty (otherwise $\sum_i (y_i - \lambda)_+ = 0$). To see Lipschitz continuity, let y' be such that $\|y' - y\|_1 \leq \epsilon$, where $0 < \epsilon < \min\{|y_i - y_j| : i, j \in [n], y_i \neq y_j\}$. It must hold that $I^+(y) \subseteq I^+(y')$. In other words, $\lambda' = \phi_2(y')$ does not deactivate any $i \in I^+(y)$, only bringing new $i \in I^0(y)$. Hence, it holds that $|\lambda' - \lambda| \leq \frac{\|y - y'\|_1}{|I^+(y)|} \leq \|y - y'\|_1$. In other words, ϕ_2 is 1-Lipschitz continuous w.r.t. $\|\cdot\|_1$.

Finally, [66, Algorithm 1]) computes λ and y^* in $O(n \log n)$ time. \square

Abusing the notation, let ϕ_2 also denote the mapping from $x \in \mathbb{R}^{n \times m}$ to $\lambda \in \mathbb{R}^m$, that is, $\lambda_j = \phi_2(x_{1j}, \dots, x_{nj})$. Let

$$\phi(x) = \phi_2(\phi_1(x)) / \gamma$$

and $p^t = \phi(x^t)$. Here, ϕ_1 is 2-Lipschitz continuous and ϕ_2 is 1-Lipschitz continuous w.r.t. $\|\cdot\|_1$. For any optimal solution $x^* \in \mathcal{X}^*$, by $x^* = \Pi_{\mathcal{X}}(x^*)$ and KKT conditions for (1) and (20), it can be seen that

$$p^* = \phi(x^*).$$

Using the Lipschitz continuity properties of ϕ_1 , ϕ_2 and Theorem 2, we have

$$\begin{aligned} \|p^t - p^*\|_1 &= \|\phi(x^t) - \phi(\Pi_{\mathcal{X}^*}(x^t))\|_1 \leq \frac{1}{\gamma} \|\phi_1(x^t) - \phi_1(\Pi_{\mathcal{X}^*}(x^t))\|_1 \\ &\leq \frac{n}{\gamma} \|\phi_1(x^t) - \phi_1(\Pi_{\mathcal{X}^*}(x^t))\| \leq \frac{2n}{\gamma} \cdot \|x^t - \Pi_{\mathcal{X}^*}(x^t)\| \\ &\leq \frac{2n}{\gamma} \cdot \sqrt{\frac{2H_{\mathcal{X}}(A)}{\mu}} (f(x^t) - f^*) \\ &\leq \frac{2n}{\gamma} \sqrt{\frac{2H_{\mathcal{X}}(A)}{\mu}} \cdot \left(1 - \frac{\mu}{2HL\|A\|^2}\right)^{t/2} \cdot \sqrt{f(x^0) - f^*}. \end{aligned}$$

Therefore, we can take $C = \frac{2n}{\gamma} \sqrt{\frac{2H_{\mathcal{X}}(A)}{\mu}} \cdot \sqrt{f(x^0) - f^*}$ and $\rho = \sqrt{1 - \frac{\mu}{2HL\|A\|^2}} \in (0, 1)$.

Since $p^* \geq \underline{p} > 0$, we can bound the maximum relative price error $\eta^t = \max_j \frac{|p_j^t - p_j^*|}{p_j^*}$ as follows, where $\underline{p}_{\min} = \min_j \underline{p}_j$.

$$\eta^t \leq \frac{\|p^t - p^*\|_1}{\underline{p}_{\min}} \leq \frac{C}{\underline{p}_{\min}} \cdot \rho^t.$$

In other words, η^t converges (R-)linearly to zero.

C QL utilities

C.1 Derivation of the QL-Shmyrev convex program (4)

In [22, Lemma 5], the convex program for the equilibrium prices is as follows:

$$\min \sum_j p_j - \sum_i B_i \log \beta_i \quad \text{s.t. } v_{ij} \beta_i \leq p_j, \forall i, j, \quad 0 \leq \beta \leq 1. \quad (22)$$

Note that it is simply the dual of EG under linear utilities (20) with additional constraints $\beta \leq 1$. Assuming v is nondegenerate, by a change of variable and Lagrange duality, we can derive the dual

of (22). First, at optimality, it must hold that $\beta_i > 0$ for all i . Therefore, by nondegeneracy of v , $p_j > 0$ for all j at optimality. Let $p_j = e^{q_j}$ and $\beta_i = e^{-\gamma_i}$. The above problem is equivalent to

$$\begin{aligned} \min \quad & \sum_j e^{q_j} + \sum_i B_i \gamma_i \\ \text{s.t.} \quad & q_j + \gamma_i \geq \log v_{ij}, \forall i, j, \\ & \gamma_i \geq 0. \end{aligned} \tag{23}$$

Let $b_{ij} \geq 0$ be the dual variable associated with constraint $q_j + \gamma_i \geq \log v_{ij}$. The Lagrangian is

$$\begin{aligned} L(q, \gamma, b) &:= \sum_j e^{q_j} + \sum_i B_i \gamma_i - \sum_{i,j} b_{ij} (q_j + \gamma_i - \log v_{ij}) \\ &= \sum_j \left(e^{q_j} - \left(\sum_i b_{ij} \right) q_j \right) + \sum_i (B_i - \sum_j b_{ij}) \gamma_i + \sum_{i,j} (\log v_{ij}) b_{ij}. \end{aligned}$$

Clearly, when $\sum_j b_{ij} \leq B_i$ for all i , $\gamma_i \geq 0$, $L(q, \gamma, b)$ is minimized at $q_j = \log \sum_i b_{ij}$ and $\gamma_i = 0$. When $\sum_j b_{ij} > B_i$ for some i , $L \rightarrow -\infty$ as $\gamma_i \rightarrow \infty$. Therefore, when $\sum_j b_{ij} \leq B_i$ for all i , we have

$$g(b) = \sum_j \left[\sum_{i,j} b_{ij} - \left(\sum_i b_{ij} \right) \log \sum_i b_{ij} \right] + \sum_{i,j} (\log v_{ij}) b_{ij}.$$

Therefore, the dual is

$$\max g(b) \quad \text{s.t. } b \geq 0, \sum_j b_{ij} \leq B_i, \forall i.$$

Adding slack variables $\delta = (\delta_1, \dots, \delta_n)$ and writing it in minimization form yield (4).

Remark. As mentioned in §4, when some $v_{ij} = 0$ (but v is still nondegenerate), by the above derivation, the first summation in (4) should be replaced by $\sum_{(i,j) \in \mathcal{E}}$, where $\mathcal{E} = \{(i, j) : v_{ij} > 0\}$. The dual remains the same otherwise.

C.2 Proof of Lemma 2

Similar to the proof of Lemma 11, this can be seen via the uniqueness of the optimal solution (p^*, β^*) of (22), that is, from uniqueness of β^* to that of $p_j^* = \max_i v_{ij} \beta_i^*$.

Let (b^*, δ^*) be an optimal solution to (4). Note that strong duality holds for (23) and (4), since there exist simple strictly feasible solutions. By the derivation in Appendix C.1, it holds that $q_j^* = \log \sum_i b_{ij}^*$ gives an optimal solution to (23) (the first-order optimality condition). Therefore,

$$p_j^* = e^{q_j^*} = \sum_i b_{ij}^*.$$

Next we establish the upper and lower bounds on p^* . By the derivation in Appendix C.1 and Lagrange duality, for any optimal solution b^* to (4), it holds that $p_j^* := \sum_i b_{ij}^*$ and $\beta_j^* = \min_{j \in J_i} \frac{p_j^*}{v_{ij}}$ give the (unique) optimal solution to (22). Clearly, $\beta^* \leq 1$ and therefore

$$p_j^* = \max_i v_{ij} \beta_i^* \leq \max_i v_{ij} = \bar{p}_j.$$

By [22, Lemma 5], the dual of (22) is (c.f. the original EG primal 1)

$$\begin{aligned} \max_{u, x, s} \quad & \sum_i B_i \log u_i - s_i \\ \text{s.t.} \quad & u_i \leq v_i^\top x_i + s_i, \forall i, \\ & \sum_i x_{ij} \leq 1, \forall j, \\ & x, s \geq 0. \end{aligned} \tag{24}$$

Clearly, strong duality holds for (22) and (24). Furthermore, notice the following.

- $\beta_i^* = \frac{B_i}{u_i^*}$ at optimality, where u_i^* is the amount of utility of buyer i . This is by the stationarity condition in the KKT optimality conditions.
- $u_i^* \leq \|v_i\|_1 + B_i$, where the right hand side is the amount of utility of all items and the entire budget. This can also be seen as follows. When $s_i > B_i$, decreasing s_i strictly increases the objective of (24). Therefore, the optimal s^* must satisfy $s_i^* \leq B_i$. It then follows from the constraint $u_i \leq v_i^\top x_i + s_i$.

Therefore,

$$p_j^* \geq \max_i v_{ij} \beta_i^* \geq \max_i \frac{v_{ij} B_i}{\|v_i\|_1 + B_i} = \underline{p}_j.$$

C.3 Proof of Theorem 7

Similar to [9, Lemma 7], we first establish the following “generalized Lipschitz condition” for φ , which is key to the claimed last-iterate convergence.

Lemma 14 *For all $(b, \delta), (b', \delta') \in \mathcal{B}$, it holds that*

$$\varphi(b') \leq \varphi(b) + \langle \nabla \varphi(b), b' - b \rangle + D(b', \delta' \| b, \delta). \quad (25)$$

Proof. Recall that $p_j(b) = \sum_i b_{ij}$, $\frac{\partial}{\partial b_{ij}} \varphi(b) = \log \frac{p_j(b)}{v_{ij}}$. For $(a, \delta^a), (b, \delta^b) \in \mathcal{B}$, we have

$$\begin{aligned} & \varphi(b) - \varphi(a) - \langle \nabla \varphi(a), b - a \rangle \\ &= - \sum_{i,j} (1 + \log v_{ij})(b_{ij} - a_{ij}) + \sum_j p_j(b) \log p_j(b) - \sum_j p_j(a) \log p_j(a) \\ & \quad - \sum_{i,j} (b_{ij} - a_{ij}) \log \frac{p_j(a)}{v_{ij}} \\ &= - \sum_{i,j} (b_{ij} - a_{ij}) + \sum_j p_j(b) \log \frac{p_j(b)}{p_j(a)} \\ &= \sum_i (\delta_i^b - \delta_i^a) + \sum_j p_j(b) \log \frac{p_j(b)}{p_j(a)}. \end{aligned} \quad (26)$$

Note that convexity and smoothness of $x \mapsto x \log \frac{x}{y}$ ($y > 0$) implies

$$\delta_i^b - \delta_i^a \leq \delta_i^b \log \frac{\delta_i^b}{\delta_i^a}. \quad (27)$$

As in the proof of [9, Lemma 7], by convexity of $q(x, y) = x \log \frac{x}{y}$, it holds that

$$\sum_j p_j(b) \log \frac{p_j(b)}{p_j(a)} \leq \sum_{i,j} b_{ij} \log \frac{b_{ij}}{a_{ij}}. \quad (28)$$

By (27) and (28), the right hand side of (26) can be bounded by $D(b, \delta^b \| a, \delta^a)$. Therefore, (25) holds. \square

Next, we prove the inequality on the right. Clearly, $(b^0, \delta^0) \in \mathcal{B}$. By [9, Theorem 3] (with objective $f = \varphi$, constraint set $C = \mathcal{B}$ and stepsize γ), we have

$$\varphi(b^t) - \varphi(b^*) \leq \frac{D(b^*, \delta^* \| b^0, \delta^0)}{t}.$$

Similar to the proof of [9, Lemma 13], we can bound the Bregman divergence on the right hand side as follows, where $b_{ij} = \delta_i = \frac{B_i}{m+1}$.

$$\begin{aligned} D(b^*, \delta^* \| b^0, \delta^0) &= \sum_{i,j} b_{ij}^* \log \frac{b_{ij}^*}{B_i} + \sum_i \delta_i^* \log \frac{\delta_i^*}{B_i} + \sum_{i,j} b_{ij}^* \log(m+1) + \sum_i \delta_i^* \log(m+1) \\ &\leq \sum_{i,j} b_{ij}^* \log(m+1) + \sum_i \delta_i^* \log(m+1) \\ &\leq \|B\|_1 \log(m+1), \end{aligned}$$

where the first inequality is because $\frac{b_{ij}^*}{B_i} \leq 1$. Combining the above yields the desired inequality.¹

Finally, we show the inequality on the left. By optimality of (b^*, δ^*) , we have

$$\langle \nabla \varphi(b^*), b - b^* \rangle \geq 0, \quad \forall (b, \delta) \in \mathcal{B}.$$

Recall that $p_j(b) = \sum_i b_{ij}$. By (26), we have

$$D(p^t \| p^*) = - \sum_{i,j} (b_{ij}^t - b_{ij}^*) + \sum_j p_j(b^t) \log \frac{p_j(b^t)}{p_j(b^*)} \leq \varphi(b^t) - \varphi^*.$$

□

C.4 Details from MD (5) to PR (6)

Note that (5) is buyer-wise separable: for each i , we have (where $\frac{\partial}{\partial b_{ij}} \varphi_b(b) = \log \frac{p_j(b)}{v_{ij}}$ and $B_i = B_i \cdot \Delta_{m+1}$)

$$\begin{aligned} (b_i^{t+1}, \delta_i^{t+1}) &= \arg \min_{(b_i, \delta_i) \in \mathcal{B}_i} \sum_j \left(\log \frac{p_j(b^t)}{v_{ij}} - \log b_{ij}^t \right) b_{ij} - (\log \delta_i^t) \delta_i + \sum_j b_{ij} \log b_{ij} + \delta_i \log \delta_i \\ &= \arg \min_{(b_i, \delta_i) \in \mathcal{B}_i} - \sum_j (\log b_{ij}^t) b_{ij} - (\log \delta_i^t) \delta_i + \sum_j b_{ij} \log b_{ij} + \delta_i \log \delta_i. \end{aligned} \quad (29)$$

By Lemma 6, for all i, j ,

$$b_{ij}^{t+1} = B_i \cdot \frac{\frac{v_{ij} b_{ij}^t}{p_j(b^t)}}{\sum_\ell \frac{v_{i\ell} b_{i\ell}^t}{p_\ell(b^t)} + \delta_i^t}, \quad \delta_i^{t+1} = B_i \cdot \frac{\delta_i^t}{\sum_\ell \frac{v_{i\ell} b_{i\ell}^t}{p_\ell(b^t)} + \delta_i^t}. \quad (30)$$

Let $p_j^t = p_j(b^t)$. Then, (30) can be written in terms of the allocations $x_{ij}^t = b_{ij}^t / p_j^t$ (which sum up to 1 over buyers i for any item j) and leftover δ_i^t , thus giving (6).

C.5 Convergence of prices

Let $\eta^t = \max_j \frac{|p_j^t - p_j^*|}{p_j^*}$ be the relative price error, which can clearly be bounded by $\frac{\|p^t - p^*\|_1}{p_{\min}}$, where $p_{\min} = \min_j p_j > 0$ is given in Lemma 2. By Theorem 7 and strong convexity of KL divergence (w.r.t. $\|\cdot\|_1$), for b^t and $p^t = p(b^t)$ generated by either PG or PR,

$$\frac{1}{2} \|p^t - p^*\|_1^2 \leq D(p^t \| p^*) \leq \varphi(b^t) - \varphi^*. \quad (31)$$

Therefore, for PG, the quantities η^t , $\|p^t - p^*\|$ and $D(p^t \| p^*)$ all converge linearly to 0. For PR, they converge at $O(1/T)$.

We can further bound $\varphi(b^t) - \varphi^*$ by the duality gap. Specifically, given b^t , $p^t = p(b^t)$, let

$$b_i^t = \min \left\{ \min_j \frac{p_j^t}{v_{ij}}, 1 \right\}.$$

Then, (p^t, β^t) is feasible to (22). By weak duality,

$$\varphi(b^t) - \varphi^* \leq \varphi(b^t) + g(p^t, \beta^t), \quad (32)$$

where $g(p, \beta)$ is the (minimization) objective of (22). Combining the above, we have

$$\eta^t \leq \frac{\sqrt{2(\varphi(b^t) + g(p^t, \beta^t))}}{p_{\min}}.$$

Note that the above holds for b^t from either PG or PR. Although neat in theory, numerical experiments suggest that the above bound can be loose and is not suitable as a termination criteria.

¹In fact, the bound $\log(mn)$ in [9, Lemma 13] (which assumes $\|B\|_1 = 1$) can be easily strengthened to $\log m$ via the above derivation. In other words, it does not depend explicitly on the number of buyers (but implicitly through $\|B\|_1$ in general).

D Leontief utilities

D.1 Derivation of (7)

The primal EG (1) under Leontief utilities $u_i(x_i) = \min_{j \in J_i} \frac{x_{ij}}{a_{ij}}$ can be written in both x and u :

$$\begin{aligned} \min_{u, x} \quad & - \sum_i B_i \log u_i \\ \text{s.t.} \quad & u_i \leq \frac{x_{ij}}{a_{ij}}, \forall j \in J_i, \forall i \in [n], \\ & \sum_i x_{ij} \leq 1, \forall j \in [m], \\ & x \geq 0, u \geq 0. \end{aligned}$$

Clearly, it can also be written in terms of u_i only as follows:

$$\min - \sum_i B_i \log u_i \quad \text{s.t.} \quad \sum_{i \in I_j} a_{ij} u_i \leq 1, \forall j, u \geq 0. \quad (33)$$

Let $p_j \geq 0$ be the dual variable associated with constraint $\sum_{i \in I_j} a_{ij} u_i \leq 1$. The Lagrangian is

$$\begin{aligned} \mathcal{L}(u, p) &= - \sum_i B_i \log u_i + \sum_j p_j \left(\sum_{i \in I_j} a_{ij} u_i - 1 \right) \\ &= - \sum_j p_j + \sum_i [-B_i \log u_i + \langle a_i, p \rangle u_i]. \end{aligned}$$

Note that minimizing \mathcal{L} w.r.t. u can be performed separably for each u_i . For any i such that $\sum_{j \in J_i} p_j > 0$, by first-order stationarity condition, the term $-B_i \log u_i + \langle a_i, p \rangle u_i$ is minimized at $u_i^*(p) = \frac{B_i}{\langle a_i, p \rangle}$ with minimum value $B_i(1 - \log B_i) + B_i \log \langle a_i, p \rangle$. If $\sum_{j \in J_i} p_j = 0$, the term approaches $-\infty$ as $u_i \rightarrow \infty$. Therefore, the dual objective is

$$g(p) = \begin{cases} - \sum_j p_j + \sum_i B_i \log \langle a_i, p \rangle + \sum_i B_i(1 - \log B_i) & \text{if } p \geq 0 \text{ and } \sum_{j \in J_i} a_{ij} p_j > 0 \\ -\infty & \text{o.w.} \end{cases}$$

Hence the (Lagrangian) dual problem is $\max_p g(p)$. Its minimization form, up to the constant $-\sum_i B_i(1 - \log B_i)$, is

$$\min \left[\sum_j p_j - \sum_i B_i \log \langle a_i, p \rangle \right] \quad \text{s.t.} \quad p \geq 0. \quad (34)$$

By Theorem 1, we have the following.

- An optimal solution to (34) gives equilibrium prices.
- A market equilibrium (x^*, p^*) satisfies $\langle p^*, x_i \rangle = B_i$ for all i and $\sum_i x_{ij}^* = 1$ for all j . Therefore, we have $\sum_j p_j^* = \|B\|_1$.

Therefore, we can add the constraint $\sum_j p_j = \|B\|_1$ to (34) without affecting any optimal (equilibrium) solution. This leads to (7).

D.2 Proof of Lemma 3

let p be any feasible solution to (7). Since $\sum_j p_j = \|B\|_1$, we have $\langle a_i, p \rangle \leq \|a_i\|_\infty \|p\|_1 = \|a_i\|_\infty \|B\|_1$ for all i . By Appendix D.1, at equilibrium, p^* and primal variables u_i^* satisfy $u_i^* = \frac{B_i}{\langle a_i, p^* \rangle}$ (the stationarity condition) and

$$u_i^* \leq \text{utility of getting one unit of every item} = \min_{j \in J_i} \frac{1}{a_{ij}} = \frac{1}{\|a_i\|_\infty}$$

for all i . Therefore $\langle a_i, p^* \rangle = \frac{B_i}{u_i^*} \leq \|a_i\|_\infty \|B\|_1$.

D.3 Linear convergence of utilities

Note that the equilibrium utilities u^* are clearly unique by (33). By the KKT stationary condition,

$$u_i^* = \frac{B_i}{\langle a_i, p^* \rangle}, \quad \forall i$$

for equilibrium prices p^* . Therefore, an intuitive construction of u^t is as follows. Let p^t be the current iterate, $r_i^t = \langle a_i, p^t \rangle$. First compute $\tilde{u}_i^t = \frac{B_i}{r_i^t}$. Then, to satisfy the primal constraints $\sum_i u_i a_{ij} \leq 1$, take

$$u^t = \frac{\tilde{u}^t}{\max_j \sum_i u_i a_{ij}} = \frac{\tilde{u}^t}{\|a^\top \tilde{u}\|_\infty}.$$

Let $r^* = \langle a_i, p^* \rangle = \frac{B_i}{u_i^*}$ and $f^* = \arg \min_{p \in \mathcal{P}} \tilde{h}(ap) = \tilde{h}(r^*) = h(r^*)$. Strong convexity of \tilde{h} implies $\frac{\mu}{2} \|r^t - r^*\|^2 \leq \tilde{h}(r^t) - f^*$. Furthermore, the mapping $r^t \mapsto \tilde{u}^t \mapsto u^t$ is Lipschitz continuous on $r^t \in [\underline{r}, \bar{r}]$. Therefore, $\|u^t - u^*\|$ converges to 0 linearly as well.

E Additional details on numerical experiments

For linear utilities, we generate market data $v = (v_{ij})$ where v_{ij} are i.i.d. from standard Gaussian, uniform, exponential, or lognormal distribution. For each of the sizes $n = 50, 100, 150, 200$ (on the horizontal axis) and $m = 2n$, we generate 30 instances with unit budgets $B_i = 1$ and random budgets $B_i = 0.5 + \tilde{B}_i$ (where \tilde{B}_i follows the same distribution as v_{ij}). See §6 for plots under random budgets and below for those under uniform budgets.

The termination conditions (on the vertical axis) are

$$\epsilon(p^t, p^*) \leq \eta, \quad \eta = 10^{-2}, 10^{-3},$$

where p^* is the optimal Lagrange multipliers of (1) computed by CVXPY+Mosek. Then, for $n = 100, 200, 300, 400$ and $n = 2m$, we repeat the above with termination conditions

$$\text{dgap}_t/n \leq \eta, \quad \eta = 10^{-3}, 10^{-4}, 10^{-5}, 5 \times 10^{-6}.$$

For QL utilities, we repeat the above (same random v , same sizes and termination conditions) using budgets $B_i = 5(1 + \tilde{B}_i)$. This is to make buyers have nonzero bids and leftovers (i.e., $0 < \delta_i^* < B_i$) at equilibrium in most scenarios. In this case, $p^* = p(b^*)$, where b^* is the optimal solution to (4) computed by CVXPY+Mosek. For QL, FW does not perform well in initial trials and is excluded in subsequent experiments.

For the linesearch subroutine $\mathcal{LS}_{\alpha, \beta, \Gamma}$ in PG (see Appendix A.5), we use parameters $\alpha = 1.02$, $\beta = 0.8$ and $\Gamma = 100L\|A\|^2$ throughout.

For Leontief utilities, in addition to $\text{dgap}_t/n \leq \eta$, we also use the termination condition $\epsilon(u^t, u^*) = \max_j \frac{|u_j^t - u_j^*|}{u_j^*} \leq \eta$, where u^* is the optimal solution to EG under Leontief utilities (33) computed by CVXPY+Mosek.

Computing the duality gap. For linear utilities, the objective of the original Shmyrev's convex program (19) is

$$\varphi(b) = - \sum_{i,j} (\log v_{ij}) b_{ij} + \sum_j p_j(b) \log p_j(b)$$

where $p_j(b) = \sum_i b_{ij}$. Recall the objective of the (EG) dual (20), equivalent to the dual of Shmyrev's (19),

$$g(p, \beta) = \sum_j p_j - \sum_i B_i \log \beta_i.$$

Given iterate b^t , let $p_j^t = p_j(b^t)$ and $\beta_i^t = \min_j \frac{p_j^t}{v_{ij}}$, which is finite since v is nondegenerate and $p^t > 0$. The duality gap is computed via

$$\text{dgap}_t = \varphi(b^t) + g(p^t, \beta^t).$$

For QL utilities, it is computed similarly, that is, through (32). For Leontief utilities, it is computed using the construction in Appendix D.3.

Additional plots. In §6, the plots for linear utilities are generated under random B_i . Here we present an augmented set of plots under different random market data, different utilities, unit v.s. varying, random budgets B_i and different termination conditions ($\text{dgap}_t/n \leq \eta$ or $\epsilon(p^t, p^*) \leq \eta$). All horizontal axes are market sizes n (with $m = 2n$) and all vertical axes are number of iterations. The error vertical bars are the standard errors of the number of iterations across 30 repeats. The legends are in the subplot “Linear utilities, $\text{dgap}/n \leq 1e-5$ ”.

