To all Reviewers Thanks you for the insightful reviews. Reviewers 2 and 3 question the novelty and claim that we merely apply known results. Prior to our work, there were *no* first-order method (FOM) for computing market equilibria that converge linearly. Previous methods acheving (partial) linear convergence either (i) involve layers of oracle abstraction (with costly oracle calls) (Bei et al. - Ascending-Price Algorithms...) or (ii) only exhibits initial linear convergence under restrictive "large-market" assumptions (Cole & Tao - Balancing the Robustness...). Thus, the fact that we are able to achieve a linear rate with something as simple as proximal (projected) gradient (PG) is, in our view, a *strength* rather than weakness. Linear convergence of PG is not obvious either: applying the standard theory naively to Eisenberg-Gale only yields a $1/T$ rate, and even this already requires our Lemma 1 through 3 in order to get a Lipschitz constant. To get our linear rate (Thm 5), it is necessary to first realize that strong convexity and Lipschitz gradients can be proved in the EG utility space using properties of market equilibrium (nobody seems to have realized this crucial property before; a priori one only gets strict convexity), and utilizing our new bounds on equilibrium utilities (Lemma 1). This makes the reformulated problem satisfy the PL inequality (Eq. (13)). Finally, minor adjustments to existing theorems are needed to tackle our setting. Similar ideas are needed in the case of QL and Leontief utilities. For all cases, our new market-specific bounds on the equilibrium quantities (Lemma 1-3) are crucial as they ensure explicit Lipschitz constants on gradients. In our paper, we will also make sure to clarify that these steps are indeed necessary. Note also that for Thm 2 & 3, it was necessary to refine the known convergence guarantees, as explained in the texts (l. 590 & 631 in their proofs). Finally, we propose a practical linesearch scheme (Algorithm 1) which we implement in the experiments. We give convergence guarantees (Theorem 4 & 10) under the Proximal-PL condition beyond strong convexity. These are entirely new.

To Reviwer 1 Thanks for the suggestions. Our paper covers all well-known utility functions that can be handled within the Eisenberg-Gale framework (i.e. convex, continuous, nonnegative and homogeneous utility functions). See e.g. [9] which lists the class of standard utilities. The only standard ones missing from our paper are CES and Cobb-Douglas, which yield "easy" convex programs (see Appendix A.6). Other utilities do not always yield convex programs.

To Reviewer 2 [It seems straightforward to write down...] In fact, there is an entire literature on convex programming formulations of equilibrium problems. Many papers have been written on the celebrated Eisenberg-Gale convex program, which is a very non-obvious construction. See, e.g., the well-known book "Algorithmic Game Theory." Furthermore, our contributions are clearly *not* formulation of any convex program or *directly* calling any FOM. [The authors claim that this is one of the contributions...] As stated in the summary of contributions and reiterated above, when solving the convex programs using FOMs, in order to achieve linear convergence, we (need to and did) establish new bounds on various equilibrium quantities (Lemma 1-3). See "To all Reviewers" for other technical contributions along this line. Furthermore, utilizing the structure of problem (4) for QL utilities, Theorem 7 gives a $1/T$ last-iterate convergence rate for Mirror Descent (MD), a result much stronger than the general theory of MD. We also show that it yields highly efficient updates which are also interpretable dynamics (l. 284-287). With these clarifications, we hope that you can reevaluate our contributions.

To Reviewer 3 ["just" applying standard theory] Note that our new bounds on equilibrium quantities (Lemma 1-3) are crucial in establishing linear rates (Thm 5-7). Several other new techniques are needed as explained in "To all Reviewers." [choice of FOMs arbitrary] We choose them because (i) PG achieves a linear rate due to our theory, (ii) FW solution at iteration t has nt nonzeros (useful when computing a low accuracy solution with very large m), (iii) PR/MD exploits problem structure to get $1/T$ last-iterate convergence, gives desirable dynamics (284-287) and is fast when computing low-accuracy solutions (l. 326). To the best of our knowledge, dual averaging does not have particular theoretical advantages over standard settings. [MD rate on last iterate] Indeed, the convergence result for MD/PR (Thm 7) is nonstandard and specific to the Fisher market setting. It is because the entropy DGF on the prices aligns with the objective to give a form of relative Lipschitz continuity (Lemma 14 in Appendix C.3). [heavy-tailed distribution] Thanks for the suggestion! We ran experiments on Cauchy valuations: PGLS slows down significantly but PR is not affected (across many sizes and accuracy levels). We conjecture that the Hoffman constant, similar to the matrix condition number $\sigma_{\max}/\sigma_{\min}$, degrades significantly upon heavy-tail v (this seems extremely difficult to formalize). In contrast, the bound (Thm 7) for MD/PR is independent of v. Will include in the final version. [In practice ... $\tilde{h}$?] Quadratic extrapolations are used in the experiments. We hope that you can reevaluate our contributions given the clarifications.

To Reviewer 4 Many thanks for the detailed, helpful comments. We will take all into account but due to space constraints we respond only to a few here. The reason we do not compare to convex program solvers is that we are motivated by large-scale settings, which conic solvers do not scale to. For smaller problems, open-source solvers such as ECOS are much slower than our methods, while the industry-grade Mosek solver (high-performance C code; our code is in python) outperforms PGLS, but is slower than PR and FW when a low-accuracy solution is needed. We will point this out. In Theorem 1, existence of ME is guaranteed by existence of an EG optimal solution (which is then guaranteed by the Extreme Value Theorem; it is often simply assumed but we will clarify). We do implement Mirror Descent: the PR dynamics (6) are MD applied to the convex program (4) for QL markets.