arXiv:2411.03248v3 [cs.GT] 27 May 2026

Improved Hardness Results for Min-Max Optimization with Coupled Constraints

Martino Bernasconi[†] Matteo Castiglioni[‡] Andrea Celli[†] Gabriele Farina[*]

[†] Bocconi university
[‡] Politecnico di Milano
[*] Massachusetts Institute of Technology

{martino.bernasconi,andrea.celli2}@unibocconi.it, matteo.castiglioni@polimi.it, gfarina@mit.edu

Abstract

We investigate the computational complexity of min-max optimization under coupled constraints. The work of Daskalakis, Skoulakis, and Zampetakis [DSZ21] was the first to study min-max optimization through the lens of computational complexity, showing that min-max problems with nonconvex-nonconcave objectives are PPAD-hard under coupled constraints. By carefully exploiting the coupled constraints rather than the structure of the objective function, we are able to significantly simplify and strengthen the proof of the hardness result. More precisely, the first contribution of this paper is a fundamentally new proof of their main result, which improves it in multiple directions: it holds for degree-2 polynomials which are quadratic-linear, it improves the dependence on the parameters of the problem (also yielding constant inapproximability for gradient descent-ascent in ℓ_∞-norm), and it is much simpler than previous approaches. Second, we show that with general constraints (*i.e.*, the min player and max player have different constraints), even convex-concave (bilinear) min-max optimization becomes PPAD-hard. Along the way, we also provide PPAD-membership of a general problem related to quasi-variational inequalities, which has applications beyond our problem.

Contents

1 Introduction **2**

2 Preliminaries **8**

3 Hardness of Local Min-Max Optimization **14**

A Complexity Classes and Reductions **24**

B PPAD Membership **25**

C Reductions among LocalMinMax, GDAFixedPoint, QVI and VI **29**

D Missing Proofs for PPAD-membership **34**

E Omitted Proofs for Section 3 **44**

F Basic Mathematical Preliminaries **51**

1 Introduction

In this paper, we study the complexity of min-max optimization in the presence of constraints. This class of problems plays a key role in the development of game theory [Neu28], adversarial robustness in optimization, statistics, machine learning [BTN02; HR11; Mąd+17; SND18], and generative models such as Generative Adversarial Networks [Goo+14; ACB17]. In its simplest form, a constrained min-max optimization problem can be informally written as

$$\min_{x\in\mathbb{R}^d} \max_{y\in\mathbb{R}^d} f(x,y) \quad \text{s.t.} \quad g(x,y) \leq 0, \tag{1}$$

where $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, and $g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. Due to the strong connection with zero-sum games, x and y are often thought of as selected by *players* seeking to maximize and minimize the payoff function in a game. Although the benign setting in which f is convex-concave and g is convex is tractable and well understood, many fundamental questions remain open regarding more general settings. In this paper, we are specifically interested in these more general settings. Akin to regular (*i.e.*, non-saddle-point) optimization, in nonconvex domains it is too much to ask for a *global* solution. Instead, *local min-max solutions* (also known as *local equilibria*) serve as the target solution concept. Specifically, given $\epsilon, \delta > 0$, we will call (ϵ, δ)-*solution* of Problem (1) any pair $(x^\star, y^\star)$ such that $g(x^\star, y^\star) \leq 0$, and such that each player cannot improve their utility through any *feasible*, local, unilateral deviation, that is,

$$\begin{aligned} &f(x^\star, y^\star) \leq f(x, y^\star) + \epsilon \quad \text{for all } x \text{ s.t. } \|x - x^\star\| \leq \delta \text{ and } g(x, y^\star) \leq 0; \text{ and} \\ &f(x^\star, y^\star) \geq f(x^\star, y) - \epsilon \quad \text{for all } y \text{ s.t. } \|y - y^\star\| \leq \delta \text{ and } g(x^\star, y) \leq 0. \end{aligned} \tag{2}$$

Our work is not the first to study min-max optimization beyond the convex-concave setup (see, *e.g.*, [JNJ20; Jin+21; Das+23] and pointers in Section 1.3). For example, several papers have focused on *nonconvex-concave* objectives under constraints g that induce *product feasible sets* of the form $\mathcal{X} \times \mathcal{Y}$, where numerous algorithms find approximate solutions efficiently [Nou+19; LJJ20a; LJJ20b; KM21; Fie+21; OLR21]. On the other hand, in a recent landmark result, Daskalakis, Skoulakis, and Zampetakis [DSZ21] proved that finding solutions to (2), in the case of nonconvex-concave objective, is PPAD-complete.

In this paper, we consider a broader class of constraints g. In particular, we consider the case in which the min and max players may have different constraints. By focusing on the constraints, more than on the structure of the objective function, we show a much simpler and stronger hardness result with respect to Daskalakis et al. [DSZ21]. We obtain a significantly simpler reduction which holds for better approximation factors and for explicit degree-2 polynomial. In particular, our result holds when $f(x, y)$ is quadratic in x and linear in y. En route, we show that when the min and max players have different constraints, the problem is PPAD-hard even when $f(x, y)$ is bilinear. This result is not only a stepping stone for understanding the proof of our main theorem, but it has direct implications on the complexity of quasi-variational inequalities with *monotone* operators [FP03]. We discuss our contributions and their implications in more depth in the next subsection.

1.1 Contributions and Implications

A key factor in our taxonomy of optimization problems is the structure of the constraints. We consider the following classes, listed on the x-axis of Figure 1 in order of increasing generality.

- *Product constraints*: The feasible set $g(x, y) \leq 0$ is equal to $K_1 \times K_2$ for some appropriate $K_1, K_2 \subset \mathbb{R}^d$. In other words, there is *no coupling* between the domains of the players.
- *Jointly convex constraints*: Constraints for both the maximizing and minimizing player are encoded by a function g that is jointly convex, that is, convex as a function of (x, y). This implies that $g(x, y) \leq 0$ is a convex subset of $\mathbb{R}^d \times \mathbb{R}^d$. We remark that for our hardness result highlighted in Figure 1, it is enough to restrict to functions g such that $g(x, y) \leq 0$ defines a *convex polytope*. This is the class of constraints used in the results of Daskalakis et al. [DSZ21].
- *Independent constraints*: Finally, to be able to capture problems corresponding to *quasi*-variational inequalities, we also consider a more general setting in which the set of feasible local deviations of the

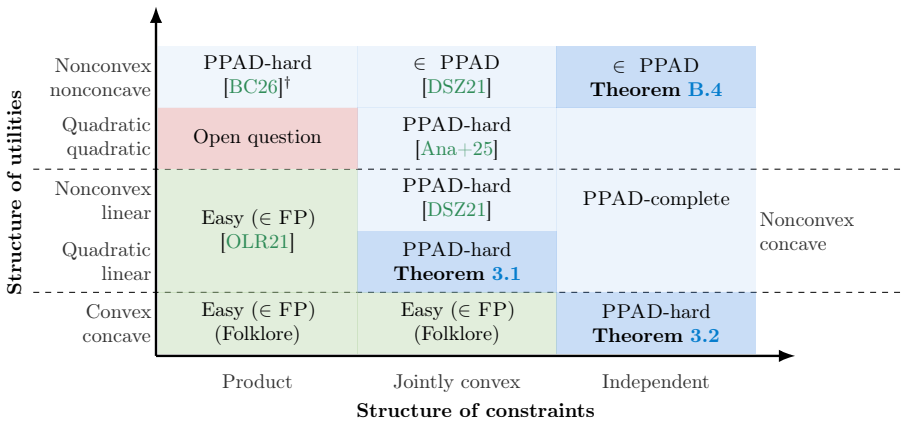


Figure 1: Summary of known results regarding the complexity of min-max optimization for a different combination of structures of utilities and constraints. The arrows point toward the direction of increased generality and thus hardness. †: [BC26] is subsequent and largely based on the insights of a preliminary version of this work.

players might differ and is controlled by two constraint functions $g_1, g_2 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, one for each player. This leads to the modified notion of (ϵ, δ)-solution (or local equilibrium) given as

$$\begin{aligned} &f(x^\star, y^\star) \leq f(x, y^\star) + \epsilon \quad \text{for all } x \text{ s.t. } \|x - x^\star\| \leq \delta \text{ and } g_1(x, y^\star) \leq 0; \text{ and} \\ &f(x^\star, y^\star) \geq f(x^\star, y) - \epsilon \quad \text{for all } y \text{ s.t. } \|y - y^\star\| \leq \delta \text{ and } g_2(x^\star, y) \leq 0. \end{aligned} \tag{3}$$

and where $g_i(x^\star, y^\star) \leq 0$ for $i \in \{1, 2\}$. In particular, we are interested in the case of *bilinear functions* g_1, g_2. Due to their bilinear nature, each of these functions defines a local convex polytope of feasible deviations for each player.

We will also consider different classes of objective functions: convex-concave, nonconvex-concave (further subdivided in nonconvex-linear and quadratic-linear), and nonconvex-nonconcave. We analyze how the problem's complexity evolves as the structure of the constraints changes. We categorize our results into hardness and membership results. Here, we focus especially on hardness results; more details about membership results are deferred to Section B.

Hardness results The main contribution of this paper is the following hardness result.

Theorem 1.1 (Informal version of Theorem 3.1)**.** *The problem of computing approximate local min-max equilibria with jointly convex constraints, and nonconvex-concave, G-Lipschitz, and L-smooth utilities is* PPAD*-hard for* $\epsilon, L = O(1)$*,* $\delta = \Omega(1)$*,* $G = \text{poly}(d)$ *and* $G_\infty = \sup_{x,y \in [0,1]^{2d}} \|\nabla f(x, y)\|_\infty = O(1)$*. Moreover the function* f *is a degree-2 polynomial quadratic in* x *and linear in* y*.*

This result has several implications with respect to the hardness result by Daskalakis et al. [DSZ21, Theorem 4.4]. Indeed, it is easy to prove that when $\delta \leq \frac{\epsilon}{G}$ then every point is a solution to the min-max problem and thus the problem is trivially in FP. On the other hand, Daskalakis et al. [DSZ21] proved that the problem is hard with $\delta = \Omega(d^{30} \frac{\epsilon}{G})$. Our results improve on this result by showing PPAD-hardness already for $\delta = \Omega(\sqrt{d} \frac{\epsilon}{G})$, significantly improving the gap between the polynomial regime and the hard one (we refer to Figure 2 for an illustration). This follows since we can set $\delta, \epsilon = \Theta(1)$ and $G = O(\sqrt{d})$.

While we leave the question of showing hardness for the regime $\delta = \Omega(\epsilon/G)$ as an interesting open question, we remark that our reduction (thanks to the fact that it holds for constant G_∞) shows tight (constant inapproximability) lower bounds for the related problem of computing approximate fixed points of gradient descent ascent with respect to the ℓ_∞ norm under joint constraints. We expand on this below and in Section 3.2. We conjecture that similar tight lower bounds can be derived for the problem of computing

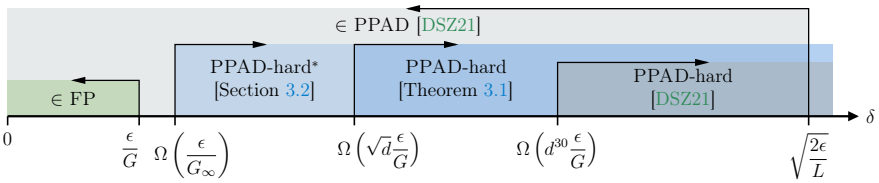


Figure 2: Our results nearly settle the complexity of finding local equilibria of nonconvex-nonconcave objective functions under jointly convex constraints as a function of the parameter δ. *The constant inapproximability result that holds for $\delta = \Omega(\epsilon/G_\infty)$ is for a natural variant of the problem in distances are measured in ℓ_∞ norm (which is polynomial for $\delta = O(\epsilon/G_\infty)$). See Section 3.2 for a detailed discussion.

approximate local min-max equilibria with jointly convex constraints, when the size of the local deviation is measured in norms other than the Euclidean one.

Moreover, the degree of the polynomial objective function f cannot be improved compared to what we obtain in our reduction. Indeed, it is well known that the case of bilinear—and thus convex-concave—objective admits polynomial time algorithms even with jointly convex constraints (Figure 1 second columns, last row). Thus, we need f to be a quadratic polynomial to hope to show hardness. We do this by constructing a function f that is quadratic only in one of the two players.

In Section 1.2, we give a first high-level overview of our new approach. In essence, we show that coupled constraints enable a simple embedding of a generic vector field into Equation (2), thereby solving a *variational inequality* (VI) associated with this vector field.[1]

We remark that in subsequent work, Bernasconi and Castiglioni [BC26] resolved the question about the complexity with product constraints, showing PPAD-hardness. A key idea was to use the specific instances in this paper (without enforcing coupled constraints) and to hide them within a Boolean PPAD-hard problem.

We complement the above result by proving PPAD-hardness for the setting where the objective function is restricted to only be convex-concave, but the constraints class is made more complex by allowing independent constraints. This marks a second "phase change" in the complexity of the problem due to constraints, since the convex-concave problem is clearly solvable with jointly convex constraints by using the monotone VI formulation of the problem (in Figure 1 this would correspond to the second and third column and last row).

Theorem 1.2 (Informal version of Theorem 3.2)**.** *The problem of computing approximate local min-max equilibria with independent constraints, and convex-concave, G-Lipschitz, and L-smooth utilities is* PPAD*-hard for $\epsilon, L = O(1)$, $\delta = \Omega(1)$, $G = \mathrm{poly}(d)$ and $G_\infty = \sup_{x,y\in[0,1]^{2d}} \|\nabla f(x,y)\|_\infty = O(1)$. Moreover, the function f is linear in x and y (separately).*

Furthermore, our reduction has an interesting consequence on the complexity of *quasi-variational inequalities* (QVIs). These are a generalization of VIs, whereby the VI domain Q is no longer a fixed set but rather a *correspondence*, which depends on the optimization variable. In this space, we show that even with linear constraints, monotone QVIs are unlikely to be polynomial-time solvable in general.

Corollary 1.3 (Informal version of Corollary 3.3)**.** *The problem of computing solutions to QVIs is* PPAD*-hard even when the set-valued fixed-point problem encoded by the QVI constraints is linear (and thus solvable in polynomial time) and the operator is monotone.*

Constant Inapproximability of Gradient Descent-Ascent in ℓ_∞-norm We are able to obtain a reduction that holds for $G_\infty = \sup_{x,y\in[0,1]^d} \|\nabla f(x,y)\|_\infty = O(1)$. This is significant, as it establishes tight (*i.e.* constant-factor inapproximability) lower bounds for the problem of computing fixed points of gradient

[1] A VI problem consists of finding a point $z \in Q$, for some convex set $Q \subset \mathbb{R}^d$, such that $F(z)^\top(z' - z) \geq 0$ for all $z' \in Q$, where $F : \mathbb{R}^d \to \mathbb{R}^d$ [Sta64]. VIs are a powerful framework for modeling optimization problems across diverse fields, including game theory [Roc70], economics [FP03], and machine learning [Goo+14], and they generalize classical optimization problems like complementarity problems [CD68].

descent-ascent under joint constraints, in the setting where the players feasibility sets (given the strategy of the other player) are products of intervals.

Local min-max points are tightly connected with fixed points of the gradient descent-ascent dynamics. In particular, Daskalakis et al. [DSZ21] proved a two-way reduction between these two problems, in the Euclidean norm. Our reduction proves an inapproximability result for the problem of computing a fixed point of gradient descent-ascent in the ℓ_∞-norm. In particular, consider the problem of computing (x, y) such that

$$\|x - \Pi_{\{x':g(x',y)\leq 0\}}(x - \eta\nabla_x f(x,y))\|_\infty + \|y - \Pi_{\{y':g(x,y')\leq 0\}}(y + \eta\nabla_y f(x,y))\|_\infty \leq \alpha.$$

Letting $G_\infty = \sup_{x,y\in[0,1]^d} \|\nabla f(x,y)\|_\infty$, the problem becomes trivially solvable in polynomial time for "small" η. In particular, in the special case in which the sets $\{x' : g(x',y) \leq 0\}, \{y' : g(x,y') \leq 0\}$ are product of intervals for each (x, y), and $\eta = O(\frac{\alpha}{G_\infty})$, then every (x, y) is a solution to the problem above. In particular this holds when the "unit-free" ratio $\eta/(\epsilon/G_\infty) = O(1)$. However, our reduction holds for $\eta, \alpha, G_\infty = O(1)$, and thus when $\eta/(\alpha/G_\infty) = \Omega(1)$, showing that there is no polynomial time algorithm for the problem in ℓ_∞. It is a natural question whether one could obtain the same inapproximability result in the ℓ_2 norm. We expand on this discussion in Section 3.2.

PPAD-membership results PPAD-membership for the case of jointly convex constraints is due to Daskalakis et al. [DSZ21, Theorem 5.2]. We look at the case of independent constraints. To do so, we prove PPAD-membership of a very general problem related to computing solutions to quasi-variational inequalities. Then, our PPAD-membership result for local min-max equilibria follows from the connection between QVIs, fixed points of gradient descent-ascent dynamics, and approximate local min-max equilibria.

Theorem 1.4 (Informal version of Theorem B.4)**.** *The problem of finding an approximate solution to quasi-variational inequalities is in* PPAD*.*

This result may be of independent interest, as the generality of the QVI problem could enable new PPAD-membership results. For example, we can easily give PPAD membership for generalized equilibrium [Ros65], which are more general than those recently considered in [FR+24] (see Section B).

1.2 Overview of Techniques

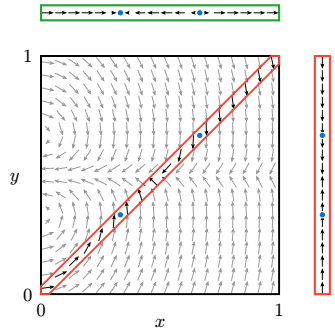
This section provides a high-level overview of the primary techniques employed throughout the paper by describing the main steps of the proof of our main result (Theorem 1.1). This is a concise summary of the discussion provided in Section 3.1.

Linear variational inequalities The first step in our reduction is to prove the hardness of an intermediate problem related to VI with linear operators defined on the hypercube, which we call LinVI. Specifically, given an affine operator $F(z) = Dz + c$ we consider the problem of finding a point z such that $F(z)^\top(z' - z) \geq 0$. We show that LinVI is PPAD-hard through a reduction from the problem of finding Nash equilibria in polymatrix games. This choice is instrumental in obtaining a hard problem on the hypercube and a stronger dependence on the parameters.

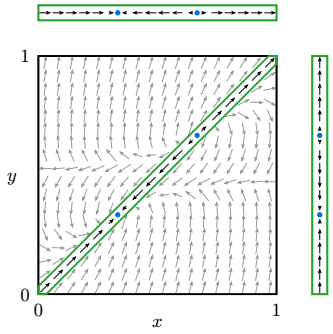
Then, we connect our problem to find fixed points of gradient descent-ascent dynamics, and we reformulate it as solving a variational inequality (VI) problem. Unfortunately, gradient descent-ascent dynamics are related to VI with a specific operator F, constructed by stacking the gradients $\nabla_x f$ and $-\nabla_y f$, which correspond to the utilities of the two players. In game theory this is sometimes called "pseudo-gradient" and we indicate it with the symbol

$$\widetilde{\nabla} f(x,y) \coloneqq (\nabla_x f(x,y), -\nabla_y f(x,y)).$$

It is not hard to see that a fixed point of the gradient descent-ascent map defined on f is a solution of the VI instance defined with the operator $\widetilde{\nabla} f$ (Theorem C.3).



(a) Embedding of $F(y) = (y-\frac{1}{3})\cdot(y-\frac{2}{3})$ with $f(x,y) = x^\top F(y)$.



(b) Embedding of $F(y) = (y-\frac{1}{3})\cdot(y-\frac{2}{3})$ with $f(x,y) = (x-y)^\top F(y)$.

Figure 3: The two embeddings of the field $F : \mathbb{R}^d \to \mathbb{R}^d$ into the pseudo-gradient of $f : \mathbb{R}^{2d} \to \mathbb{R}$. The box on the right is the field $x \mapsto \nabla_x f$, while the one on top represents the field $y \mapsto -\nabla_y f$. The embedded field is $F(y) = (y-\frac{1}{3})\cdot(y-\frac{2}{3})$; the blue dots are in correspondence with its zeros and are inserted only for reference. In both figures we have that the x-projection of the pseudo-gradient field on the $x = y$ subspace corresponds to the field F, while only in Figure 3b, both the x and y projections are aligned with the field F.

Pseudo-gradient fields It is clear that if we had to deal with standard gradients, then the problem would be much easier (and PPAD-hardness would be hopeless), as any local minima would satisfy the first-order optimality conditions given by the VI. Therefore, our approach hinges on the distinction between "pseudo-gradient" fields (vector fields generated by applying the pseudo-gradient operator to a differentiable function f) and gradient fields. Pseudo-gradient fields are more general than gradient fields, and allow for more complex behavior, *e.g.* gradient lines might form connected paths, while for conservative vector fields the gradient lines decrease the potential.

Perfect imitation and independent constraints Independent constraints provide a simple means of implementing the linear operator via imitation. Given the linear operator $z \mapsto Dz + c$ defining the LinVI instance, we embed it into the first half components $x \in [0,1]^d$ of the function f, while using the constraints to force the $y \in [0,1]^d$ components to lie on a specific d-subdomain of $[0,1]^{2d}$. Specifically, if we want to embed the operator $[0,1]^d \ni z \mapsto F(z) \in [0,1]^d$ into a pseudo-gradient $\widetilde{\nabla} f$, we can consider the following function $f(x,y) = x^\top F(y)$ and force y to imitate x (for a illustration with $d = 1$ see Figure 3a). Indeed, with independent constraints, we can build instances in which the y player must choose $y \in \{x\}$ while the x player is only constrained to play in the hypercube. Now, we have that $\nabla_x f(x,y) = F(y) = F(x)$ and thus, if x is a fixed point of gradient descent-ascent, we have that $F(x)^\top(x' - x) \geq 0$, implying that x is a solution to LinVI. This idea is enough to prove hardness of the convex-concave case with independent constraints (Theorem 1.2).

Approximate imitation and jointly convex constraints Attempting to apply the same construction as in the case of jointly convex constraints quickly leads to seemingly insurmountable challenges. First, the function used in the previous construction was convex-concave, and equilibria for such functions are easy to compute on a jointly convex set K. Therefore, we must increase the complexity of f to be at least quadratic in x or y. Second, the most natural way to extend the imitation gadget from independent constraints to jointly convex ones is to consider the set $K := \{x, y : \|x - y\|_\infty \leq \Delta\}$, for "small" values of Δ. However, this comes at a high cost: we can no longer disregard player y and consider only the optimality condition of the x player. What if the solution x, y found by solving the min-max problem is on the boundary of the set K? Intuitively, we can see that the constraint $\{x, y : \|x - y\|_\infty \leq \Delta\}$ can only limit one player at a time, and either the x

player or the y player can move towards any deviation x'.[2] However, now we have to design f to align the gradient of the y player to the operator $F(z) = Dz + c$ of the LinVI instance we are reducing from. Otherwise, an optimality condition of the y player on f would be meaningless in terms of its implications for the original LinVI instance. A better embedding of an operator $F(z)$ is to consider the function $f(x, y) = (x - y)^\top F(x)$. We can see that this is a promising choice since $\nabla_x f(x, y) = F(x) + J_F(x)^\top (x - y)$, where $J_F(x)$ is the Jacobian of F at x, and $-\nabla_y f(x, y) = F(x)$. Given that the constraints enforce that $\|x - y\| \leq \Delta$, we have that approximately both gradients are aligned with $F(x)$, and thus a solution to min-max would be a solution to the original LinVI instance (Figure 3b). Incidentally, the set K is the same considered in the reduction by Daskalakis et al. [DSZ21]. We analyze the connections with their reduction in the following paragraph.

Relationship with the reduction of Daskalakis et al. [DSZ21] Our reduction follows a completely different approach compared to prior work. For example, we reduce from a continuous optimization problem, naturally defined on the hypercube, whereas Daskalakis et al. [DSZ21] reduces from a discrete optimization problem related to Sperner's lemma. This allows us to work with simple degree-two polynomials instead of building functions that are defined by circuits of some discrete problem. This simpler construction provides stronger results, such as simpler functions, constant smoothness, and hardness for larger approximations. More in detail, the reduction of Daskalakis et al. [DSZ21] starts from a problem (HighD-BiSperner) related to a generalization of Sperner's lemma [Spe28]. They partition the d-dimensional hypercube in cublets and color each vertex with d colors out of a total of $2d$ colors (instead of the $d+1$ usually used in Sperner's lemma). This problem is shown to be PPAD-hard by a reduction from Brouwer. Then they build a function f such that the fixed points of gradient descent-ascent dynamics are close to panchromatic cubelets of the simplicization, in which they interpret the d colors of each vertex as directions of the pseudo-gradient of f. To do so, they give a function value and direction based on the local coloring and then interpolate the function with a high-degree smooth step function. Their ingenious construction is extremely technical and requires careful techniques to ensure that no extra fixed points of gradient descent-ascent are created outside of those corresponding to panchromatic cubelets. It is precisely this last requirement that forces a coupling between the strategies of the minimizer and maximizer variables since the coefficients of the interpolation produce perturbations in the pseudo-gradient, which depend on the difference $x - y$. In contrast, our construction sidesteps any need for interpolating (or, in fact, the use of discrete Sperner-type problems altogether) by constructing a problem instance with an explicit quadratic-linear objective, and relying on the constraints to drive hardness from the start.

1.3 Further Related Work

Complexity of optimization problems We contribute to the line of work that shows lower bounds of optimization problems employing tools from computational complexity. This new line of research has achieved remarkable results with significant implications across complexity theory [Fea+22], game theory [BR21], and machine learning [DSZ21; HZ23]. Optimization problems belong naturally to the class of *total search problem* (TFNP) [JPY88; MP91; Pap94]. The bridge between optimization and search problems is not new and, for instance, it has been the main motivation for the definition of the PLS class [JPY88; Sch91; Kre89]. Fearnley et al. [Fea+22] showed that the problem of computing fixed points of gradient descent is complete for the class $\mathrm{PLS} \cap \mathrm{PPAD}$, and this result also proves that such class is equal to CLS. Therefore, it lies much lower in the hierarchy of TFNP problems then PPAD. Moreover, while we work inside the framework of "constrained" optimization (meaning that we work inside the hypercube rather than on the entire Euclidean space), there are some very recent works that consider the computational complexity of optimization in the unconstrained setting [HZ23; Kon+24].

Imitation The idea of imitation in recent years has been instrumental in various PPAD-hardness results [MT05; Rub15; Bab16; BR20; BR21]. These settings mainly focus on general-sum games, and imitation is

[2] Formally, this is not entirely accurate; we need to decompose a deviation direction into two distinct components, one corresponding to the x-player and the other to the y-player. See Section 3.1 for more details.

attained through carefully designed payoffs of one (or more) player. However, our setting is zero-sum, and adding an imitation component to the payoff would discourage one player from imitating just as much as it would encourage the other. Thus, our result can be seen as enforcing imitation through constraints rather than payoffs.

Most recently, Anagnostides et al. [Ana+25] proposed a reduction from symmetric equilibria in symmetric games, using ideas similar to our paper (*cf.* Section 1.2), to prove a weaker version of Theorem 3.1 via a simpler reduction. Several key differences set our result apart from theirs. First, our result shows hardness already for nonconvex-concave functions (more precisely, quadratic-linear), while their result only applies to nonconvex-nonconcave quadratic polynomials. Moreover their problem is for min-max optimization over the simplex (rather that on the hypercube), which admits a natural quasi-polynomial-time approximation scheme and thus can only lead to weaker hardness results.

Solving min-max problems A substantial body of recent research has focused on developing practical first-order and low-order methods for min-max optimization problems. A wide range of works illustrate divergent or cycling behavior when extending beyond minimization problems [MPP18; HMC21]. However, efficient algorithms are available for well-behaved objectives. In the convex-concave setup, solutions can be efficiently computed using convex programming techniques [Kor76; Azi+20; Gol+20; MRS20; HA21; DP19; MOP20; ALW21] and via algorithms for monotone VIs [Bru77; EB92; Tse95; Nem04; CLO17; GLG22].

Nonconvex-concave problems with "simple" constraints can also be solved efficiently [Das+18; Nou+19; LJJ20a; LJJ20b; KM21; OLR21]. In the nonconvex-nonconcave setting, motivated by the complexity results of [DSZ21], research has focused on identifying structural assumptions on the objective (*e.g.*, assuming the weak Minty variational inequality holds) that can help overcome the computational hardness barriers [DDJ21; Pet+22], and identifying different notions of local solutions and studying convergence to these points [JNJ20; MV21; Kes+22]. Finally, a more recent stream of works considers different notions of stationarity, such as Goldstein's stationarity to handle nonsmooth objectives [Zha+20; JLZ22; KPS24].

2 Preliminaries

We study local min-max solutions, or *equilibria*, of a function $f : [0,1]^d \times [0,1]^d \to [-B, B]$ with $B \in \mathbb{R}_+$ in which the two variables are subject to functional constraints $g_1, g_2 : [0,1]^d \times [0,1]^d \to [-1, 1]$. Formally, for any $\epsilon > 0$ and $\delta > 0$, an (ϵ, δ)-equilibrium is a tuple of points $(x^\star, y^\star) \in [0,1]^d \times [0,1]^d$ such that it holds $g_i(x^\star, y^\star) \leq 0$ for each $i \in \{1, 2\}$, and

$$f(x^\star, y^\star) \leq f(x, y^\star) + \epsilon \quad \forall x \in [0,1]^d \text{ such that } \|x - x^\star\| \leq \delta \text{ and } g_1(x, y^\star) \leq 0,$$
$$f(x^\star, y^\star) \geq f(x^\star, y) - \epsilon \quad \forall y \in [0,1]^d \text{ such that } \|y - y^\star\| \leq \delta \text{ and } g_2(x^\star, y) \leq 0.$$

In this paper, we use the symbol $\|\cdot\|$ to denote the Euclidean norm. We study the complexity of computing equilibria under a combination of assumptions on the structure of the payoff function and the constraint functions g_1, g_2. We detail the assumptions under consideration next. In Figure 1 we can see a table summarizing the payoffs and constraints considered in this paper.

Payoff function Throughout the paper, we assume that f is G-Lipschitz and L-smooth. Formally, a G-Lipschitz function satisfies

$$|f(x, y) - f(x', y')| \leq G \|(x, y) - (x', y')\| \quad \forall (x, y), (x', y') \in [0,1]^{2d},$$

while L-smoothness is defined as the L-Lipschitzness of the gradient, *i.e.*,

$$\|\nabla f(x, y) - \nabla f(x', y')\| \leq L \|(x, y) - (x', y')\| \quad \forall (x, y), (x', y') \in [0,1]^{2d}.$$

We study three different classes of payoff functions f. We present the classes from the most specific to the most general

- The simplest scenario we address is the *convex-concave* case, where the function $f(x, y)$ is convex in x over $[0, 1]^d$ for every fixed $y \in [0, 1]^d$ and concave in $y \in [0, 1]^d$ for every fixed $x \in [0, 1]^d$.
- Moving to a more complex scenario, we consider the *nonconvex-concave* case, where f remains concave in y over $[0, 1]$ for every fixed $x \in [0, 1]^d$, without imposing any specific properties on f for a fixed y.
- Finally, we consider the *nonconvex-nonconcave* case where no specific properties are imposed on f, other than being G-Lipschitz and L-smooth.

Constraints We consider the case in which g_1 and g_2 are *bilinear* functions. Let $[m] = \{1, \dots, m\}$. Formally, for each player i the function g_i is defined by a set of m bilinear functions $h_j^{(i)}$:

$$g_i(x, y) := \max_{j \in [m]} h_j^{(i)}(x, y). \tag{4}$$

More formally, for each $i \in \{1, 2\}$ and $j \in [m]$ there exists a matrix $B_j^{(i)} \in \mathbb{R}^{d \times d}$, vectors $b_{1,j}^{(i)}, b_{2,j}^{(i)} \in \mathbb{R}^d$, and constant $c_j^{(i)} \in \mathbb{R}$ such that $h_j^{(i)}(x, y) := x^\top B_j^{(i)} y + b_{1,j}^{(i),\top} x + b_{2,j}^{(i),\top} y + c_j^{(i)}$. The functions g_i induce a set of (approximately) feasible polytopes. In particular, for any $\nu \geq 0$ we can define the sets

$$K_1^\nu(y) := \{x' \in [0, 1]^d : g_1(x', y) \leq \nu\} \quad \text{and} \quad K_2^\nu(x) := \{y' \in [0, 1]^d : g_2(x, y') \leq \nu\}.$$

The bilinear structure of g_1 implies that for all $y \in [0, 1]^d$ the set $K_1^\nu(y)$ is a polytope defined by m inequalities. Similarly, for all $x \in [0, 1]^d$ the set $K_2^\nu(x)$ is a polytope. In the previous definition, we allow for a slight violation ν of the constraints. As will become evident in the following, this violation is needed for technical reasons and we will keep this violation negligible (arbitrarily small) or zero in all of our results. When the sets do not include the superscript ν, we imply $\nu = 0$.

We consider three different classes of constraints. We present them from the most specific to the most general:

- The simplest type of constraints are *product constraints*, where $g_1(x, y) = g_1(x, y')$ for all $x, y, y' \in [0, 1]^d$ (*i.e.*, g_1 is independent of y), and $g_2(x, y) = g_2(x', y)$ for all $x, x', y \in [0, 1]^d$ (*i.e.*, g_2 is independent of x). In this setting there exist two convex sets such that $x \in K_1 \subseteq [0, 1]^d$ (respectively, $y \in K_2 \subseteq [0, 1]^d$) if and only if $g_1(x, \cdot) \leq 0$ (respectively, $g_2(\cdot, y) \leq 0$).
- A more general class of constraints are *jointly convex constraints* where there exists a *convex* function $g : [0, 1]^{2d} \to [-1, 1]$ such that $g_1 = g_2 = g$. In particular, we assume g to be linear in (x, y), *i.e.*, without the bilinear component.[3] In this setting, we can define a polytope $K := \left\{(x, y) \in [0, 1]^{2d} : g(x, y) \leq 0\right\}$ and then for all $(x, y) \in K$ it holds:

  $$K_1(y) = \{x' : (x', y) \in K\}, \quad \text{and} \quad K_2(x) = \{y' : (x, y') \in K\}.$$

- Finally, we consider a broader class of constraints in which the feasible strategies of one player depend on the strategy chosen by the other player, with no additional conditions beyond that g_1 and g_2 be defined as in Equation (4). We refer to these as *independent constraints*.

  We remark that with independent constraints, the function g_i defining the coupling between the player i and player j strategies can be chosen independently of the function g_j defining the coupling between the player j and player i strategies.

All our hardness results apply with only linear constraints (*i.e.* g_1 and g_2 are defined without the bilinear components). This makes our results stronger since it is possible to find points that satisfy the constraints in polynomial time. On the other hand, we restrict ourselves to independent constraints (instead of considering

[3] This is the class of constraints studied in Daskalakis et al. [DSZ21].

more general ones or sets given implicitly by *e.g.* separation oracles as in Papadimitriou et al. [PVGZ23]) when proving our membership results. This is mainly done to guarantee some properties that simplify the exposition and analysis. For instance, it is possible to compute in polynomial time the projection on $K_1^\nu(y)$ and $K_2^\nu(x)$, and it is easier to "enlarge" the sets with a parameter ν. Our membership results, however, can be adapted for more general classes of constraints assuming the existence of an approximate projection oracle, as in Papadimitriou et al. [PVGZ23]. Indeed, the properties to guarantee such membership results are mainly related to convexity and the existence of strictly feasible points. Considering non-explicit definitions of the sets would add much complexity to the exposition without adding much to our contributions.

2.1 Gradient Descent-Ascent Dynamics and Min-Max Optimization

In this section, we introduce the main computational problems we deal with in this paper, while in Section A we formally introduce reductions between total problems. The formal problem of computing a local min-max equilibrium is defined as follows.

Problem 1 (LocalMinMax)**.** Given $\epsilon, \delta, L, G, \nu \in \mathbb{R}_+$, two circuits implementing a G-Lipschitz and L-smooth function f and its gradient ∇f, and an explicit representation of functions g_1 and g_2, find two points $(x^\star, y^\star) \in [0,1]^d$ such that $g_i(x^\star, y^\star) \leq \nu$ for all $i \in \{1,2\}$ and:

$$f(x^\star, y^\star) \leq f(x, y^\star) + \epsilon \quad \forall x \in [0,1]^d \text{ such that } \|x - x^\star\| \leq \delta \text{ and } g_1(x, y^\star) \leq \nu,$$
$$f(x^\star, y^\star) \geq f(x^\star, y) - \epsilon \quad \forall y \in [0,1]^d \text{ such that } \|y - y^\star\| \leq \delta \text{ and } g_2(x^\star, y) \leq \nu,$$

or output $\perp$ if no such point exists.

We call a solution $(x^\star, y^\star)$ of the LocalMinMax problem an (ϵ, δ, ν)-solution to LocalMinMax. If $\nu = 0$, we simply say that it is an (ϵ, δ)-solution to LocalMinMax.

The problem LocalMinMax defined above is not assured to have solutions. Following the approach of Daskalakis et al. [DSZ21], we specify LocalMinMax into *regimes*, and we focus on the *local regime*.

Problem 2 (LR-LocalMinMax)**.** We define LR-LocalMinMax (*local regime* min-max) as the class of LocalMinMax problems with $\delta < \sqrt{\frac{2\epsilon}{L}}$.

Relation to local min-max problems We observe that LocalMinMax generalizes the local min-max optimization problem studied by Daskalakis et al. [DSZ21] along two different directions. First, we consider a more general class of constraints, specifically bilinearly-coupled constraints that may be independent among players, whereas Daskalakis et al. [DSZ21] considers only jointly convex constraints. Moreover, we relax the problem allowing for slight violations of the constraints, meaning that we can violate the functions g_i by a margin $\nu \geq 0$. In most cases we will restrict to (ϵ, δ)-solutions and zero violation (*i.e.*, $\nu = 0$). However, when considering LocalMinMax in its full generality, an arbitrary (even exponentially) small relaxation $\nu > 0$ is needed to guarantee the existence of solutions (see Section B).

2.1.1 Fixed points of Gradient Descent-Ascent

Since the problem of verifying the local optimality of solutions is a hard problem, the local min-max problem is not a FNP problem. Therefore, we focus on a problem intimately related to approximate local min-max equilibria whose solutions are easy to verify, and hence the problem formally lies in TFNP. Specifically, we investigate the problem of finding a fixed point of gradient descent-ascent dynamics. This is essentially equivalent to the choice of Fearnley et al. [Fea+22] to focus on KKT solutions instead of local minima.[4]

Formally, given a smooth function f and any $\nu \geq 0$, we can define the gradient descent-ascent update as:

$$F_{\text{GDA}}^\nu(x, y) = \begin{pmatrix} \Pi_{K_1^\nu(y)}(x - \nabla_x f(x, y)) \\ \Pi_{K_2^\nu(x)}(y + \nabla_y f(x, y)) \end{pmatrix}.$$

[4]Daskalakis et al. [DSZ21] seems to overlook this technicality and thus treats LR-LocalMinMax as a FNP problem. However, this comes with little or no collateral damage as long as one is aware of this issue.

Then, the computational problem of finding a fixed point of the gradient descent-ascent dynamics is defined as follows.

Problem 3 (GDAFixedPoint)**.** Given α, L, G, $\nu \in \mathbb{R}_+$, two circuits implementing a G-Lipschitz and L-smooth function $f : [0,1]^d \times [0,1]^d \to [-B, B]$ and its gradient $\nabla f : [0,1]^d \times [0,1]^d \to \mathbb{R}^{2d}$, and an explicit representation of functions g_1 and g_2, find $(x^\star, y^\star) \in [0,1]^{2d}$ such that

$$g_i(x^\star, y^\star) \leq \nu \quad \forall i \in \{1,2\} \quad \text{and} \quad \|(x^\star, y^\star) - F^\nu_{\mathrm{GDA}}(x^\star, y^\star)\| \leq \alpha,$$

or output $\perp$ if no such point exists.

At the heart of Daskalakis et al. [DSZ21] lies the connection between GDAFixedPoint and LR-LocalMinMax. Indeed, Daskalakis et al. [DSZ21] showed that, in the local regime, any α-fixed point of GDAFixedPoint with $\alpha = \mathrm{poly}(\epsilon, \delta, G, L)$ is an (ϵ, δ)-solution to LR-LocalMinMax. The reduction between GDAFixedPoint and LR-LocalMinMax was proved by Daskalakis et al. [DSZ21] only for the jointly convex case, but the proof extends without changes to bilinear constraints since their proof simply exploits that $K^\nu_2(x)$ is convex for all x and $K^\nu_1(y)$ is convex for all y. This is verified when we deal with bilinear constraints.

Theorem 2.1 ([DSZ21, Theorem 5.1])**.** *Consider a* LR-LocalMinMax *problem, and assume that f is G-Lipschitz and L-smooth. Then:*

i) For any $\epsilon > 0$, $\delta > 0$ and any $\nu \geq 0$, define $\alpha := \frac{1}{2}\left(\sqrt{(G+\delta)^2 + 4\left(\epsilon - \frac{L}{2}\delta^2\right)} - (G+\delta)\right)$. Any α-approximate fixed point (x, y) of GDAFixedPoint *is an (ϵ, δ, ν) solution to* LocalMinMax*.*

ii) For any $\alpha > 0$, define $\epsilon := \frac{\alpha^2 L}{(5L+2)^2}$ and $\delta \leq \frac{\alpha}{5L+2}$. Any (ϵ, δ, ν) solution to LR-LocalMinMax *is an α-approximate fixed point of* GDAFixedPoint*.*

The previous theorem shows that there are solutions to LR-LocalMinMax that can be verified in polynomial time. However, this does not imply that *all* of them are checkable in polynomial time. Additionally, this also shows that there exists a reverse reduction that essentially shows that if finding an α-fixed point of F_{GDA} is PPAD-hard, then LocalMinMax is also PPAD-hard.

Promises Following the earlier discussion, LocalMinMax and GDAFixedPoint are formally defined as promise problems. All inputs satisfy the promise that the arithmetic circuits output function and gradient values that are consistent with the smoothness and Lipschitz constants specified in the input. Moreover, we make some minimal promises about the constraints that depend on whether the constraints are jointly convex constraints or independent constraints (see Section 2.1.2).

Function representation When defining a computational problem, we assume as input two circuits: one implementing a G-Lipschitz and L-smooth function f and the other implementing its gradient ∇f. Notably, this does not require the function f to be represented explicitly. Differently, we always assume that the functions g_i for each $i \in \{1,2\}$ are given in input by $\{B^{(i)}_j, b^{(i)}_{1,j}, b^{(i)}_{2,j}, c^{(i)}_j\}_{i \in \{1,2\}, j \in [m]}$ defining the constraints as per Equation (4).

2.1.2 Problems of interest

We further partition the instances of LocalMinMax depending on ν, ϵ, δ and the shape of the constraints g_1 and g_2. We focus on the local regime since, outside of this regime, solutions are not guaranteed to exist. Moreover, in the case of independent constraints, an arbitrarily small relaxation of the constraints $\nu > 0$ is required to guarantee the existence of solutions (details in Section B). Formally, we define the following problems.

Problem 4 (LR-LocalMinMax-JC)**.** We define as LR-LocalMinMax-JC (*local regime min-max with jointly convex constraints*) as the class of LocalMinMax problems with jointly convex constraints, $K \neq \emptyset$, $\nu = 0$, and $\delta < \sqrt{\frac{2\epsilon}{L}}$.

LR-LocalMinMax-JC corresponds to the *local-regime local min-max equilibrium computation problem* which is the main problem analyzed in [DSZ21]. Additionally, we define the corresponding problem with independent constraints.

Problem 5 (LR-LocalMinMax-Indep)**.** We define as LR-LocalMinMax-Indep (*local regime min-max with independent constraints*) as the class of LocalMinMax problems with $\nu > 0$ and $\delta < \sqrt{\frac{2\epsilon}{L}}$. Moreover, we require that $K_1(y) \neq \emptyset$ for all $y \in [0,1]^d$ and $K_2(x) \neq \emptyset$ for all $x \in [0,1]^d$.

Intuitively, the non-emptiness assumption is required to guarantee that each player has a feasible strategy for each opponent's strategy. This guarantees that the problem is well-posed and it always admits solutions. Unfortunately, as discussed above, even if there are certifiable solutions to these two problems, it is generally co-NP-hard to verify a generic candidate solution. Thus, we consider the following two problems, each respectively related to LR-LocalMinMax-JC and LR-LocalMinMax-Indep. These problems have the advantage of being formally in FNP, as the definition of GDAFixedPoint directly implies polynomial-time verification.

Problem 6 (GDAFixedPoint-JC)**.** We define as GDAFixedPoint-JC (fixed-point with *jointly convex constraints*) the class of GDAFixedPoint problems with jointly convex constraints, $K \neq \emptyset$, and $\nu = 0$.

Problem 7 (GDAFixedPoint-Indep)**.** We define as GDAFixedPoint-Indep (fixed point with *independent constraints*) the class of GDAFixedPoint problems with $\nu > 0$. Moreover, we require that $K_1(y) \neq \emptyset$ for all $y \in [0,1]^d$ and $K_2(x) \neq \emptyset$ for all $x \in [0,1]^d$.

While the PPAD-membership of GDAFixedPoint-JC was proven in [DSZ21, Theorem 4.3], the presence of independent constraints between the players introduces significant challenges to deriving existence results. We will prove the totality and PPAD-membership of GDAFixedPoint-Indep through the connection with (Quasi-)Variational inequalities (which might be of independent interest). We will exploit this connection also to show that GDAFixedPoint-Indep and LR-LocalMinMax-Indep might not admit solutions for $\nu = 0$.

2.2 Variational and Quasi-Variational Inequalities, and Kakutani Fixed Points

In this section, we introduce the required tools to understand our membership results. We will prove PPAD membership for a generalization of the GDAFixedPoint problem, which is related to quasi-variational inequalities (QVIs) [FP03]. This effort is not only directed toward our membership result as Variational Inequalities (VI) will also be useful to build our main hardness results. In Section B we will provide a clear picture of the relation among the computational versions of VIs, QVIs, and our problems of interests related to min-max optimization.

QVIs are a generalization of variational inequalities. A classical variational inequality problem is the problem of finding a vector $z \in Q$ such that:

$$F(z)^\top (z' - z) \geq 0 \quad \forall z' \in Q,$$

where $F : \mathbb{R}^d \to \mathbb{R}^d$ and $Q \subset \mathbb{R}^d$ is a convex set. QVIs generalize the problem by considering a non-fixed feasible set Q defined through a correspondence.

Definition 2.2 (Correspondence)**.** A correspondence between two topological spaces X and Y is a set-valued function that associates every point of X to an element of 2^Y (set of subsets of Y). We write $Q : X \rightrightarrows Y$ to denote a correspondence.

Given a correspondence $Q : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ and an operator $F : \mathbb{R}^d \to \mathbb{R}^d$, the QVI problem corresponds to finding a vector z such that

1. $z \in Q(z)$ and
2. $F(z)^\top (z' - z) \geq 0 \quad \forall z' \in Q(z)$.

The fact that feasible deviations $Q(z)$ change as z varies adds non-trivial complexity issues, and QVIs are, in general, much more challenging than VIs. Note that QVIs embed a set-valued fixed-point problem, *i.e.* find z such that $z \in Q(z)$. The existence of such fixed points is guaranteed, under suitable assumptions, by Kakutani's fixed-point theorem [Kak41].[5]

Now, we are ready to introduce some computationally meaningful problems related to QVIs.

Problem 8 (QuasiVI)**.** Given $G, L, \epsilon > 0, \nu > 0$, a circuit implementing a G-Lipschitz function $F : \mathbb{R}^d \to \mathbb{R}^d$, and two circuits implementing a L-Lipschitz continuous matrix valued function $A : [0,1]^d \to \mathbb{R}^{n \times d}$ and a L-Lipschitz continuous vector valued function $b : \mathbb{R}^n \to \mathbb{R}^d$ defining the correspondence $Q_\nu(\tilde{z}) := \{z \in [0,1]^d : A(\tilde{z})z \leq b(\tilde{z}) + \nu 1_d\}$, find a point $z \in Q_\nu(z)$ such that:

$$F(z)^\top (\tilde{z} - z) \geq -\epsilon,$$

for all $\tilde{z} \in Q_\nu(z)$. We assume that $Q(z) \neq \emptyset$ for all $z \in [0,1]^d$.[6]

Problem 9 (Kakutani)**.** Given $L, \nu \geq 0$, and two circuits implementing a L-Lipschitz continuous matrix valued function $A : [0,1]^d \to \mathbb{R}^{n \times d}$ and a L-Lipschitz continuous vector valued function $b : \mathbb{R}^n \to \mathbb{R}^d$ defining the correspondence $Q_\nu(\tilde{z}) := \{z \in [0,1]^d : A(\tilde{z})z \leq b(\tilde{z}) + \nu 1_d\}$, find a point z such that $z \in Q_\nu(z)$. We assume that $Q(z) \neq \emptyset$ for all $z \in [0,1]^d$.

Finally, we consider a problem that simplifies QuasiVI problem along two directions. First, we consider a fixed set, more specifically the hypercube, hence restricting to a VI problem. Second, we force the function F to be affine.[7]

Problem 10 (LinVI)**.** Given a matrix $D \in [-1,1]^{d \times d}$, a vector $c \in [-1,1]^d$, and an approximation parameter $\epsilon > 0$, find a point $z \in [0,1]^d$ such that:

$$(Dz + c)^\top (z' - z) \geq -\epsilon \quad \forall z' \in [0,1]^d.$$

Relation of QVIs to VIs and Kakutani Here we expand our discussion on how QuasiVI generalizes Kakutani and VIs. Clearly Kakutani is PPAD-hard, by a trivial reduction from Brouwer. This proves that QuasiVI is hard even when the operator F is the constant 0. Second, QVIs are a generalization of VIs, which are already hard by a trivial reduction from the problem of computing Nash equilibria in two-player games. Thus, it is straightforward to conclude that QVIs are hard even when the correspondence Q is fixed. An intriguing question arises here:

Is QuasiVI *hard even when both the corresponding VI and* Kakutani *problem are solvable in polynomial time?*

En route to proving our main result, we answer this affirmatively. In Theorem 3.2 and Corollary 3.3, we show that QuasiVI is PPAD-complete even restricting to instances in which: *i*) the set-valued fixed-point problem is easy (a vector such that $z \in Q(z)$ can be found in polynomial time) and *ii*) the operator F with the QuasiVI problem is monotone, *i.e.*

$$(F(z) - F(z'))^\top (z - z') \geq 0 \quad \forall z, z' \in \mathbb{R}^d.$$

We recall that a monotone operator implies that any VI over a convex set Q can be solved in polynomial time [Nem04].

[5] We defer all the details about continuity of correspondences and the statement of Kakutani theorem to Section D.1.

[6] We say that a matrix-valued function $A : [0,1]^d \to \mathbb{R}^{n \times d}$ is L-Lipschitz if its application to every $z \in [0,1]^d$ is L-Lipschitz, *i.e.*, if the function $\tilde{z} \mapsto A(\tilde{z})z$ is L-Lipschitz for every $z \in [0,1]^d$.

[7] In this work, we sometimes refer to functions with affine components as linear for simplicity.

Representation of sets Note that our correspondences are more restrictive than those employed by Papadimitriou et al. [PVGZ23], which, to our knowledge, is the only work studying a computational version of Kakutani's fixed-point theorem. Our primary goal is to demonstrate that these fixed-point problems are challenging even in basic settings, so using simpler correspondences aligns with this objective.

3 Hardness of Local Min-Max Optimization

In this section, we provide two fundamental hardness results on the problem of solving local min-max problems in the presence of constraints. For missing details about this section, we refer to Section E.

Nonconvex-concave function and jointly-convex constraints The first result shows that both LR-LocalMinMax-JC and GDAFixedPoint-JC are PPAD-hard, for much better dependency on the parameters of the problem, and even constant inapproximability for the problem of computing fixed points of gradient descent-ascent in ℓ_∞. Section 3.1 provides an overview of the proof, which is formally described in Section E.1.

Theorem 3.1. *Given a $\delta > 0$,* LR-LocalMinMax-JC *with nonconvex-concave objective is* PPAD*-hard for $G = O(\sqrt{d})$, $L, \epsilon = O(1)$,$\delta = \Omega(1)$, and $\sup_{x,y\in[0,1]^d} \|\nabla f(x,y)\|_\infty = G_\infty = O(1)$. Moreover,* GDAFixedPoint-JC *with nonconvex-concave objective is* PPAD*-hard for $\alpha_{\text{GDA}}, L = O(1)$, $G = O(\sqrt{d})$ and $\sup_{x,y\in[0,1]^d} \|\nabla f(x,y)\|_\infty = G_\infty = O(1)$. This holds even when f is quadratic in x and linear in y.*

Regime of the hardness result As pointed out by Daskalakis et al. [DSZ21], for $\delta \leq \epsilon/G$ the problem becomes easily solvable in polynomial time (indeed, when $\delta \leq \epsilon/G$ any pair (x, y) is a solution to LR-LocalMinMax). Our proof shows that the problem is PPAD-hard for any $\delta = \Omega\left(\sqrt{d}\frac{\epsilon}{G}\right)$. This significantly reduces the ratio between δ and ϵ/G for which the problem remains hard. Indeed, in previous result, the problem is hard for $\delta = \Omega(d^{30}\frac{\epsilon}{G})$. Moreover our reduction employs function with $G_\infty = O(1)$. The most important consequence of this result is that we are able to show constant inapproximability for the problem of computing fixed points of gradient descent-ascent when distances are measured in ℓ_∞. We point to Section 3.2 for more details about this point.

Convex-concave function and independent constraints The next result shows PPAD-hardness for the simplest case of payoff functions (*i.e.*, convex-concave, and more precisely linear-linear) and the most challenging case of constraints (*i.e.*, independent constraints). This highlights even more how constraints are capable of injecting hardness in otherwise computationally amenable problems since the convex-concave problem is solvable with jointly convex constraints. We remark that our result holds for linear constraints g_1 and g_2. This guarantees that the hardness result is not related only to solving the Kakutani-like problem. Indeed, a solution to the corresponding Kakutani problem can be found through linear programming. The proof of this result is provided in Section E.2.

Theorem 3.2. *Given a $\delta > 0$,* LR-LocalMinMax-Indep *with convex-concave (bilinear) objective and linear constraints is* PPAD*-hard for $G = O(\sqrt{d})$, $\sup_{x,y\in[0,1]^d} \|\nabla f(x,y)\|_\infty = O(1)$, $L, \epsilon = O(1)$, and $\delta = \Omega(1)$. Moreover,* GDAFixedPoint-Indep *with convex-concave objective and linear constraints is* PPAD*-hard for $\alpha_{\text{GDA}} = O(1)$, $G = O(\sqrt{d})$, $L = O(1)$ and $\sup_{x,y\in[0,1]^d} \|\nabla f(x,y)\|_\infty = O(1)$.*

Hardness of monotone QVIs The result holds when $f(x, y)$ is a bilinear function of x and y. Then, as a direct corollary of the previous theorem, we have that finding a solution to a monotone QVI is PPAD-hard. As we already remarked, the constraints in our reduction are linear. Hence, this result holds even when the set-valued fixed-point problem defined by the QuasiVI constraints is solvable in polynomial time, and the VI associated with the QVI operator is monotone.

Corollary 3.3. QuasiVI *is* PPAD*-hard even if the operator F is monotone, the constraints are linear.*

3.1 High-level Idea of the Proof: From Conservative Vector Fields to VIs

Given the well-known relationship between VIs and fixed points of gradient maps (we recall this relationship formally in Theorem C.3), we could potentially prove hardness for LocalMinMax by using its corresponding VI reformulation, *i.e.* find a point $z \in [0,1]^d$ such that

$$\widetilde{\nabla} f(z)^\top (z' - z) \geq 0 \quad \forall z' \in [0,1]^d,$$

where $\widetilde{\nabla} f = (\nabla_x f, -\nabla_y f)$ is the "pseudo-gradient" operator and $z = (x, y)$. Clearly, this idea has a chance of succeeding only because of the inverted sign in the $\widetilde{\nabla}$ operator. If instead of $\widetilde{\nabla}$ we had the classic ∇ operator, this would doom any approach since any (approximately) local minima of f would (approximately) satisfy the VI reformulation.

Moreover, observe that the reduction would be straightforward if we were free to design whatever VI operator we desire. Unfortunately, we will see in the following that such an approach is over-optimistic. However, we should not be entirely discouraged. Indeed, we know that we need to implement very simple operators F. A reduction from computing approximate Nash equilibria in polymatrix games shows that even linear vector fields $F(z) = Dz + c$ with $\|D\|_1, \|D\|_\infty \leq 1$ are PPAD-hard.[8,9]

Theorem 3.4. *There exists a constant ρ^* such that finding an ρ^*-approximation to* LinVI *is* PPAD*-complete. Moreover, the same result holds even when restricting to deviations along a single component and restricting to $\|D\|_1 \leq 1$ and $\|D\|_\infty \leq 1$, i.e., finding a z such that*

$$(Dz + c)_i \cdot (z'_i - z_i) \geq -\rho^* \quad \forall i \in [d],\ z'_i \in [0,1].$$

A first attempt The simplest construction would be to choose f such that $\widetilde{\nabla} f(z) = Dz + c$, where D and c define the linear vector field of a LinVI instance and $K = [0,1]^d$. However, this simple approach leads to a dead end. The linear form of $Dz + c$ suggests that a second-degree f should be sufficient. Indeed, the operator $\widetilde{\nabla}$ still maps n-degree polynomials into $n-1$ polynomials, exactly as the classic gradient ∇. However, the operator $\widetilde{\nabla}$ can only produce "pseudo-gradient fields", which are not conservative vector fields. Nevertheless, their structure remains too restrictive to implement any arbitrary linear field defined by D. Indeed, it is easy to check that there is no choice of second-degree polynomial f that makes $\widetilde{\nabla} f(z)$ equal to $(Dz + c)$ for a generic matrix D. There are settings in which this equivalence can be attained, for example if $D = -D^\top$, but with this condition LinVI is not hard as it becomes monotone.[10]

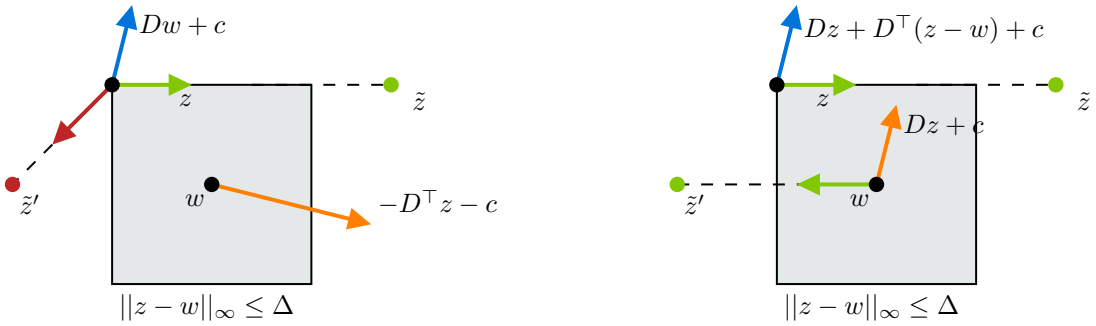
A path forward: Imitation gadgets In this paragraph, we temporarily relax the assumption of jointly convex constraints and instead consider independent. Our main idea to address the mentioned complications is to increase the dimensionality relative to the LinVI instance. In particular, we consider functions $f(z, w)$, where z matches the variables of the original LinVI instance, and w are additional auxiliary variables. The aim is to leverage the additional variables w to implement LinVI only in the equilibrium constraints of the minimizing player of f. The key ingredient here is the idea of *imitation*, according to which we force one player to imitate the other. However, to make this idea effective, we need to exploit the independence of the constraints between z and w. The construction works as follows.

Consider the function $f(z, w) = z^\top (Dw + c)$. We have that $\widetilde{\nabla} f(z, w) = (Dw + c, -D^\top z)$. Then, we can exploit the presence of constraints in our construction. Define the correspondence $\Omega(z, w) := K \times \{z\}$,

[8] For a matrix D, the $\|D\|_1$ norm represents the maximum of the ℓ_1 norms of its columns, while the $\|D\|_\infty$ norm corresponds to the maximum of the ℓ_1 norms of its rows. Additional details are provided in Section F.

[9] The reader can safely skip momentarily the second part of the statement, as it will only be used for obtaining better dependence on the parameters of the problem.

[10] Any second degree polynomial can be written as $f(z) = z^\top M z + h^\top z$. Applying to this form $\widetilde{\nabla}$ we obtain $\widetilde{\nabla} f(z) = \begin{bmatrix} M_{11} + M_{11}^\top & M_{12} + M_{21}^\top \\ -(M_{12}^\top + M_{21}) & -(M_{22} + M_{22}^\top) \end{bmatrix} z + \begin{bmatrix} h_1 \\ -h_2 \end{bmatrix}$, where we highlighted the block components of M and h. Solving the system $\widetilde{\nabla} f(z) = Dz + c$ it is easy to realize that there is no obvious class of D and c that can be implemented by $\widetilde{\nabla}$ and it is large enough to guarantee that LinVI is still hard.



(a) $f(z, w) = z^\top(Dw + c)$ (b) $f(z, w) = (z - w)^\top(Dz + c)$

Figure 4: (a) Two possible deviations of z to $\tilde{z}$ and $\tilde{z}'$. A deviation towards $\tilde{z}$ is possible without violating the constraint $\|z - w\|_\infty \leq \Delta$, while a deviation toward $\tilde{z}'$ is not feasible. (b) The deviation toward $\tilde{z}$ satisfying the constraint is achievable by either deviating from z or by deviating from w. The **blue** arrow represents the gradient $\nabla_z f(z, w)$ while the **orange** arrow represents the gradient $-\nabla_w f(z, w)$.

meaning that the constraints of the z player are independent of w, *i.e.*, the z player is free to deviate to any $z' \in K$, while the only available strategy for the w player is to imitate z, *i.e.*, the player w is forced to choose $w' = z$. Essentially, this means that the second player cannot deviate. Take any point that satisfies the VI, which operator is the pseudo-gradient of f, and consider only the component relative to z, *i.e.*,

$$\widetilde{\nabla} f(z, w)^\top (z' - z, w' - w) \geq 0 \quad \forall (z', w') \in \Omega(z, w) \implies (Dw + c)^\top (z' - z) \geq 0 \quad \forall z' \in K.$$

Finally, noticing that $w = z$ for each local min-max, *i.e.*, $(w, z) \in \Omega(z, w)$, we can conclude that z is a solution to LinVI. This result is interesting in itself, as it proves that monotone QVIs are hard even if the set inclusion problem $z \in Q(z)$ is easy and the operator is monotone. The formal statement of the theorem can be found in Theorem 3.2

Our result shows that the use of imitation gadgets is promising, but, to make it work in the setting of jointly convex constraints, we need to refine our construction. In particular, with independent constraints, we could completely ignore the player associated with the extra variables w. However, critically, this approach fails with joint convex constraints. Crucially, we will circumvent the related difficulties by exploiting deviations of both players to prove the soundness of our reduction from LinVI.

Overcoming pseudo-gradient fields using approximate imitation gadgets A natural imitation gadget for jointly convex constraints is obtained by defining the joint space $\Omega := \{(z, w) \in K \times K : \|z - w\|_\infty \leq \Delta\}$ for some small Δ. Differently from the case of independent constraints, also the deviations z' are forced to be close to w, *i.e.*, we can consider only deviations of the kind $(z', w) \in \Omega$. Therefore, there are some points of K toward which z cannot move. Indeed, if we try to use $f(z, w) = z^\top(Dw + c)$ as in the previous paragraph (which leads to $\widetilde{\nabla} f(z, w) = (Dw + c, -D^\top z)$), we inevitably fail to cover some directions. Figure 4a shows two examples of possible deviations. Deviation $\tilde{z}$ is covered by a deviation of the z player, but deviation $\tilde{z}'$ cannot be covered. This is because the z player cannot move even slightly in the direction of $\tilde{z}'$ due to the constraint that $\|z' - w\|_\infty \leq \Delta$ for any deviation z'. This means that a solution to the local min-max problem does *not* guarantee that $(Dz + c)^\top(\tilde{z}' - z) \geq 0$.

A closer look at Figure 4a suggests a new idea. It is inevitable that there are directions $\tilde{z} - z$, with $\tilde{z} \in K$, along which z cannot move. However, it is possible that even if z cannot, w might be able to move in that direction. We see that by choosing $f(z, w) = z^\top(Dw + c)$ the interests of w player (**orange** arrow in Figure 4a) are misaligned to the interests of the z player (**blue** arrow in Figure 4a). Indeed, $\widetilde{\nabla} f(z, w) = (Dw + c, -D^\top z)$. Even if $w \approx z$, clearly $Dw + c \not\approx -D^\top z$ and the optimality condition of player w along the direction $\tilde{z}'$ does *not* imply that optimality conditions of the LinVI instance are satisfied along that direction. This is unsurprising; indeed, this approach will fail unless the function f is modified. In this case, $\widetilde{\nabla} f$ yields a

monotone LinVI, which is solvable in polynomial time. Therefore, any function f we aim to construct must be at least quadratic (nonconvex) in either z or w. This is expected since the convex-concave problem can be solved in polynomial time over jointly convex domains [FP03].

A possible idea to realign the players' incentives is to design f so that $\widetilde{\nabla} f(z,w) \approx F_{\text{ideal}}(z,w) \coloneqq (Dz+c, Dw+c)$. While it is clear that inducing this pseudo-gradient exactly is not possible, we can attempt to obtain a good approximation by exploiting the imitation gadget implemented by the constraints. For example, the function $f(z,w) = (z-w)^\top (Dz+c)$ induces the operator $\widetilde{\nabla} f = (Dz + D^\top(z-w) + c, Dz + c)$ (see Figure 4b). This, together with the imitation gadget obtained through the constraints Ω that forces $w \approx z$, yields $\widetilde{\nabla} f(z,w) \approx F_{\text{ideal}}(z,w)$.

The importance of dealing with the hypercube In the proof of Theorem 3.1 we will rely heavily on the fact that LinVI remains PPAD-hard even when restricted to the hypercube. This has several important consequences for our reductions. The most important one is that we can focus on the individual components of a deviation $\tilde{z}$. Indeed, we only have to consider constraints that are independent among components, namely the hypercube constraint and the approximate imitation gadget constraint implemented by the infinity norm. In our reductions, for every deviation $\tilde{z}$, we analyze each of its i-th components independently. For all $i \in [d]$, we show that either z_i or w_i can be moved towards $\tilde{z}_i$ since at least one of the two deviations is feasible. Then, we identify two *orthogonal* deviations v_z and v_w, which are summed to define a vector along the direction $\tilde{z} - z$ (see Figure 5 for an intuitive picture). This approach would not have been feasible with constraints that are not independent among components. Indeed, this would impose joint constraints between variables and make it impossible to construct the deviations component-wise.

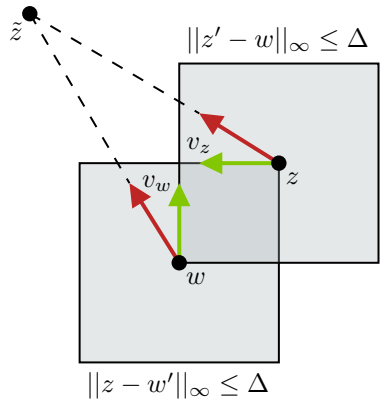


Figure 5: Neither w nor z can be moved towards $\tilde{z}$. However, we can decompose $\tilde{z} - z$ into v_w and v_z such that $v_w + v_z$ is parallel to $\tilde{z} - z$.

Single-components deviations and constant approximations for GDAFixedPoint Now we leverage another key feature of our hardness result for LinVI, specifically that the problem remains hard when considering single-component deviations (second part of Theorem 3.4). This will be key to obtain constant inapproximability for GDAFixedPoint in ℓ_∞-norm, as discussed in Section 3.2.

The main objective of this section is to show how we can obtain a reduction with $G_\infty = O(1)$ and a that works even for $\delta = \Omega(\sqrt{d}\frac{\epsilon}{G})$ instead of $\delta = \Omega(d\frac{\epsilon}{G})$.

Let's assume that for a deviation $\tilde{z}$ the direction $\tilde{z} - z$ is feasible for the z-player. The locality condition implies that the furthest we can move in this direction is $O(\delta/\|\tilde{z} - z\|)$ (which is $O(\delta/\sqrt{d})$ in the worst case). This is obtained by considering the local deviation $z' = z + O(\delta/\sqrt{d})(\tilde{z} - z)$. Hence, the locality condition would add a $\sqrt{d}$ factor in the approximation results:

$$f(z,w) \leq f(z',w) + \epsilon \implies (Dz+c)^\top (\tilde{z} - z) \geq -\frac{\epsilon\sqrt{d}}{\delta}.$$

Then, we would have to set $\delta = \Omega(\epsilon\sqrt{d})$ and thus $\delta = \Omega(d\frac{\epsilon}{G})$ to obtain a constant approximation for the LinVI instance. However, we can employ a simple trick that sidesteps this problem. If we consider only single-component deviations $\tilde{z} = (\tilde{z}_i, z_{-i})$, then $\|\tilde{z} - z\| \leq \delta$. Hence, the amount we can move z in the direction $\tilde{z} - z$ is exactly δ by taking $z' = z + \delta(\tilde{z}_i - z_i)$ (assuming as above that the direction $z' - z$ is feasible). This allows us to recover the instance-dependent factor, as simple calculations show that

$$f(z,w) \leq f(z',w) + \epsilon \implies (Dz+c)_i \cdot (\tilde{z}_i - z_i) \geq -\frac{\epsilon}{\delta},$$

even for $\epsilon, \delta = O(1)$ and thus $\delta = \Omega(\sqrt{d}\frac{\epsilon}{G})$.

Finally, we have to transfer this result to GDAFixedPoint. A direct application of the reduction from LR-LocalMinMax to GDAFixedPoint (introduced by [DSZ21] and rephrased in Theorem 2.1), would yield $\alpha = O(\frac{1}{G}(\epsilon - \frac{L}{2}\delta^2))$. Such α is not constant in our instance as $G = O(\sqrt{d})$. However, we can design a specialized reduction from LR-LocalMinMax to GDAFixedPoint that exploits that LR-LocalMinMax is PPAD-hard even for single-component deviations. Intuitively, the parameter α in our specialized reduction from LR-LocalMinMax to GDAFixedPoint (Lemma 3.5; see Section E for its proof) depends on $G_\infty = \sup_{x,y} \|\nabla f(x,y)\|_\infty = O(1)$ instead of $G = \sup_{x,y} \|\nabla f(x,y)\| = \Theta(\sqrt{d})$. This makes it possible to obtain hardness results for $\alpha = O(\frac{1}{G_\infty}(\epsilon - \frac{L}{2}\delta^2))$ which is constant since $L, G_\infty, \epsilon, \delta = O(1)$ and $\delta/(\epsilon/G) = \sqrt{n}$.

Lemma 3.5. *Assume that for all $y \in [0,1]^d$ there exist d intervals $\{K_{1,j}(y_j)\}_{j\in[d]}$ such that $K_1(y) = \bigtimes_{j\in[d]} K_{1,j}(y_j)$ and d intervals $\{K_{2,j}(x_j)\}_{j\in[d]}$ such that $K_2(x) = \bigtimes_{j\in[d]} K_{2,j}(x_j)$. Let $(x^\star, y^\star) \in K_1(y^\star) \times K_2(x^\star)$ satisfy*

$$\|(x^\star, y^\star) - F^\nu_{\mathrm{GDA}}(x^\star, y^\star)\| \le \alpha$$

Then, for all $i \in [d]$ and all single-component deviation $x' = (x'_i, x^\star_{-i}), y' = (y'_i, y^\star)$ it holds

$$f(x^\star, y^\star) \le f(x', y^\star) + \alpha(1 + G_\infty) + \frac{L}{2}\delta^2 \quad \forall x' \in K^\nu_1(y^\star) \quad \text{and} \quad \|x' - x^\star\| \le \delta,$$
$$f(x^\star, y^\star) \ge f(x^\star, y') - \alpha(1 + G_\infty) - \frac{L}{2}\delta^2 \quad \forall y' \in K^\nu_2(x^\star) \quad \text{and} \quad \|y' - y^\star\| \le \delta,$$

where $G_\infty = \sup_{x,y\in[0,1]^d} \|\nabla f(x,y)\|_\infty$.

The statement holds even if we only have $\|(x^\star, y^\star) - F^\nu_{\mathrm{GDA}}(x^\star, y^\star)\|_\infty \le \alpha$.

3.2 Constant Inapproximability of GDAFixedPoint-JC$_\infty$

We consider the following problem related to computing fixed points of F_{GDA} in ℓ_∞-norm.

Problem 11 (GDAFixedPoint-JC$_\infty$)**.** Given $\eta > 0$, α, L, G_∞, $\nu \in \mathbb{R}_+$, a circuit implementing a G-Lipschitz and L-smooth function $f : [0,1]^d \times [0,1]^d \to [-B, B]$ and its gradient $f : [0,1]^d \times [0,1]^d \to \mathbb{R}^{2d}$, where the ℓ_∞-norm of its gradients are upper bounded by G_∞. Moreover we are given a polytope $K \subset [0,1]^{2d}$ such that the slices $K_1(y)$ and $K_2(x)$ are product of intervals for every x and y. The problem is to find $(x^\star, y^\star) \in [0,1]^{2d}$ such that $(x^\star, y^\star) \in K$ and

$$\|x^\star - \Pi_{K_1(y^\star)}(x^\star - \eta\nabla_x f(x^\star, y^\star))\|_\infty \le \alpha \quad \text{and} \quad \|y^\star - \Pi_{K_2(x^\star)}(y^\star + \eta\nabla_y f(x^\star, y^\star))\|_\infty \le \alpha,$$

or output $\perp$ if no such point exists.

The problem becomes trivially solvable in polynomial time —since any (x, y) is a solution— if $\eta = 0$. More generally, even when $\eta = O(\alpha/G_\infty)$ any (x, y) is a solution to GDAFixedPoint-JC$_\infty$ provided that the slices $K_1(y)$ and $K_2(x)$ are product of intervals, *i.e.* $K_1(y) = \bigtimes_{j\in[d]} K_{1,j}(y_j)$ for some intervals $\{K_{1,j}(y_j)\}_{j\in[d]}$ and analougously for the y player. Indeed, by the non-expansiveness of the projection (which applies since the slices $K_1(y)$ are a product of intervals), we have:

$$\|x - \Pi_{K_1(y)}(x - \eta\nabla_x f(x,y))\|_\infty \le \eta\|\nabla_x f(x,y)\|_\infty \le \alpha,$$

and an analogous bound holds for the y player.

The instance used in the proof of Theorem 3.1 satisfies all the conditions of Problem 11. Moreover, it is clear from the proof that Lemma 3.5 still holds when the fixed points of F_{GDA} are measured in ℓ_∞-norm. Therefore, by slightly modifying the final part of the proof of Theorem 3.1, we can also prove that GDAFixedPoint-JC$_\infty$ is PPAD-hard when $\eta/(\epsilon/G_\infty) = \Omega(1)$. This result is tight since, as per the discussion above, the problem is in FP for $\eta/(\epsilon/G_\infty) = O(1)$.

Acknowledgments

This work is supported by an ERC grant (Project 101165466 — PLA-STEER). MB, MC are partially supported by the FAIR (Future Artificial Intelligence Research) project PE0000013, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, investment 1.3, line on Artificial Intelligence). MC is also partially supported by the EU Horizon project ELIAS (European Lighthouse of AI for Sustainability, No. 101120237). GF was supported in part by the National Science Foundation award CCF-2443068, the Office of Naval Research grant N000142512296, and an AI2050 Early Career Fellowship.

References

[AB09] Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[ACB17] Martin Arjovsky, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks". In: *International conference on machine learning*. PMLR. 2017, pp. 214–223.

[ALW21] Jacob Abernethy, Kevin A Lai, and Andre Wibisono. "Last-iterate convergence rates for min-max optimization: Convergence of hamiltonian gradient descent and consensus optimization". In: *Algorithmic Learning Theory*. PMLR. 2021, pp. 3–47.

[Ana+25] Ioannis Anagnostides, Ioannis Panageas, Tuomas Sandholm, and Jingming Yan. "The Complexity of Symmetric Equilibria in Min-Max Optimization and Team Zero-Sum Games". In: *arXiv preprint arXiv:2502.08519* (2025).

[Azi+20] Waiss Azizian, Ioannis Mitliagkas, Simon Lacoste-Julien, and Gauthier Gidel. "A tight and unified analysis of gradient-based methods for a whole spectrum of differentiable games". In: *International conference on artificial intelligence and statistics*. PMLR. 2020, pp. 2863–2873.

[Bab16] Yakov Babichenko. "Query complexity of approximate Nash equilibria". In: *Journal of the ACM (JACM)* 63.4 (2016), pp. 1–24.

[BC26] Martino Bernasconi and Matteo Castiglioni. "The Complexity of Min-Max Optimization with Product Constraints". In: *arXiv preprint arXiv:2602.04665* (2026).

[Ber77] Claude Berge. *Topological spaces: Including a treatment of multi-valued functions, vector spaces and convexity*. Oliver & Boyd, 1877.

[BR20] Yakov Babichenko and Aviad Rubinstein. "Communication complexity of Nash equilibrium in potential games". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2020, pp. 1439–1445.

[BR21] Yakov Babichenko and Aviad Rubinstein. "Settling the complexity of Nash equilibrium in congestion games". In: *Proceedings of the 53rd annual ACM SIGACT symposium on theory of computing*. 2021, pp. 1426–1437.

[Bro11] Luitzen Egbertus Jan Brouwer. "Über abbildung von mannigfaltigkeiten". In: *Mathematische annalen* 71.1 (1911), pp. 97–115.

[Bru77] Ronald E Bruck. "On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space". In: *J. Math. Anal. Appl* 61.1 (1977), pp. 159–164.

[BTN02] Aharon Ben-Tal and Arkadi Nemirovski. "Robust optimization–methodology and applications". In: *Mathematical programming* 92 (2002), pp. 453–480.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[CD68] Richard W Cottle and George B Dantzig. "Complementary pivot theory of mathematical programming". In: *Linear algebra and its applications* 1.1 (1968), pp. 103–125.

[CKK23] Xi Chen, Christian Kroer, and Rachitesh Kumar. "The complexity of pacing for second-price auctions". In: *Mathematics of Operations Research* (2023).

[CLO17] Yunmei Chen, Guanghui Lan, and Yuyuan Ouyang. "Accelerated schemes for a class of variational inequalities". In: *Mathematical Programming* 165 (2017), pp. 113–149.

[CP82] D Chan and Jong-Shi Pang. "The generalized quasi-variational inequality problem". In: *Mathematics of Operations Research* 7.2 (1982), pp. 211–222.

[Das+18] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. "Training GANs with Optimism". In: *International Conference on Learning Representations*. 2018.

[Das+23] Constantinos Daskalakis, Noah Golowich, Stratis Skoulakis, and Emmanouil Zampetakis. "Stay-on-the-ridge: Guaranteed convergence to local minimax equilibrium in nonconvex-nonconcave games". In: *The Thirty Sixth Annual Conference on Learning Theory*. PMLR. 2023, pp. 5146–5198.

[DDJ21] Jelena Diakonikolas, Constantinos Daskalakis, and Michael I Jordan. "Efficient methods for structured nonconvex-nonconcave min-max optimization". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2021, pp. 2746–2754.

[Deb52] Gerard Debreu. "A social equilibrium existence theorem". In: *Proceedings of the national academy of sciences* 38.10 (1952), pp. 886–893.

[Del+24] Argyrios Deligkas, John Fearnley, Alexandros Hollender, and Themistoklis Melissourgos. "Pure-Circuit: Tight Inapproximability for PPAD". In: *Journal of the ACM* 71.5 (2024), pp. 1–48.

[DP11] Constantinos Daskalakis and Christos Papadimitriou. "Continuous local search". In: *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*. SIAM. 2011, pp. 790–804.

[DP19] Constantinos Daskalakis and Ioannis Panageas. "Last-Iterate Convergence: Zero-Sum Games and Constrained Min-Max Optimization". In: *10th Innovations in Theoretical Computer Science* (2019).

[DSZ21] Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. "The complexity of constrained min-max optimization". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 1466–1478.

[EB92] Jonathan Eckstein and Dimitri P Bertsekas. "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators". In: *Mathematical programming* 55 (1992), pp. 293–318.

[EY10] Kousha Etessami and Mihalis Yannakakis. "On the complexity of Nash equilibria and other fixed points". In: *SIAM Journal on Computing* 39.6 (2010), pp. 2531–2597.

[Fea+22] John Fearnley, Paul Goldberg, Alexandros Hollender, and Rahul Savani. "The complexity of gradient descent: $\mathsf{CLS} = \mathsf{PPAD} \cap \mathsf{PLS}$". In: *Journal of the ACM* 70.1 (2022), pp. 1–74.

[Fie+21] Tanner Fiez, Lillian Ratliff, Eric Mazumdar, Evan Faulkner, and Adhyyan Narang. "Global convergence to local minmax equilibrium in classes of nonconvex zero-sum games". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 29049–29063.

[FK10] Francisco Facchinei and Christian Kanzow. "Generalized Nash equilibrium problems". In: *Annals of Operations Research* 175.1 (2010), pp. 177–211.

[FP03] Francisco Facchinei and Jong-Shi Pang. "Finite-dimensional variational inequalities and complementarity problems". In: *Springer Series in Operations Research and Financial Engineering* (2003).

[FR+23] Aris Filos-Ratsikas, Kristoffer A Hansen, Kasper Høgh, and Alexandros Hollender. "Fixp-membership via convex optimization: Games, cakes, and markets". In: *SIAM Journal on Computing* (2023), FOCS21–30.

[FR+24] Aris Filos-Ratsikas, Kristoffer Arnsfelt Hansen, Kasper Høgh, and Alexandros Hollender. "PPAD-membership for Problems with Exact Rational Solutions: A General Approach via Convex Optimization". In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. 2024, pp. 1204–1215.

[GLG22] Eduard Gorbunov, Nicolas Loizou, and Gauthier Gidel. "Extragradient method: O (1/k) last-iterate convergence for monotone variational inequalities and connections with cocoercivity". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2022, pp. 366–402.

[Gol+20] Noah Golowich, Sarath Pattathil, Constantinos Daskalakis, and Asuman Ozdaglar. "Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems". In: *Conference on Learning Theory*. PMLR. 2020, pp. 1758–1784.

[Goo+14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial nets". In: *Advances in neural information processing systems* 27 (2014).

[HA21] Erfan Yazdandoost Hamedani and Necdet Serhat Aybat. "A primal-dual algorithm with line search for general convex-concave saddle point problems". In: *SIAM Journal on Optimization* 31.2 (2021), pp. 1299–1329.

[HMC21] Ya-Ping Hsieh, Panayotis Mertikopoulos, and Volkan Cevher. "The limits of min-max optimization algorithms: Convergence to spurious non-critical sets". In: *International Conference on Machine Learning*. PMLR. 2021, pp. 4337–4348.

[HR11] Peter J Huber and Elvezio M Ronchetti. *Robust statistics*. John Wiley & Sons, 2011.

[HUL04] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of convex analysis*. Springer Science & Business Media, 2004.

[HZ23] Alexandros Hollender and Emmanouil Zampetakis. "The Computational Complexity of Finding Stationary Points in Non-Convex Optimization". In: *The Thirty Sixth Annual Conference on Learning Theory*. PMLR. 2023, pp. 5571–5572.

[Jin+21] Chi Jin, Praneeth Netrapalli, Rong Ge, Sham M Kakade, and Michael I Jordan. "On nonconvex optimization for machine learning: Gradients, stochasticity, and saddle points". In: *Journal of the ACM (JACM)* 68.2 (2021), pp. 1–29.

[JLZ22] Michael I Jordan, Tianyi Lin, and Manolis Zampetakis. "On the complexity of deterministic nonsmooth and nonconvex optimization". In: *arXiv preprint arXiv:2209.12463* (2022).

[JNJ20] Chi Jin, Praneeth Netrapalli, and Michael Jordan. "What is local optimality in nonconvex-nonconcave minimax optimization?" In: *International conference on machine learning*. PMLR. 2020, pp. 4880–4889.

[JPY88] David S Johnson, Christos H Papadimitriou, and Mihalis Yannakakis. "How easy is local search?" In: *Journal of computer and system sciences* 37.1 (1988), pp. 79–100.

[Kak41] Shizuo Kakutani. "A generalization of Brouwer's fixed point theorem". In: *Duke Mathematical Journal* 8.3 (1941), p. 457.

[Kes+22] Vijay Keswani, Oren Mangoubi, Sushant Sachdeva, and Nisheeth K Vishnoi. "A Convergent and Dimension-Independent Min-Max Optimization Algorithm". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 10939–10973.

[KM21] Weiwei Kong and Renato DC Monteiro. "An accelerated inexact proximal point method for solving nonconvex-concave min-max problems". In: *SIAM Journal on Optimization* 31.4 (2021), pp. 2558–2585.

[Kon+24] Andreas Kontogiannis, Vasilis Pollatos, Sotiris Kanellopoulos, Panayotis Mertikopoulos, Aris Pagourtzis, and Ioannis Panageas. "The computational complexity of finding second-order stationary points". In: *ICML 2024-41st International Conference on Machine Learning*. 2024, pp. 1–35.

[Kor76] Galina M Korpelevich. "The extragradient method for finding saddle points and other problems". In: *Matecon* 12 (1976), pp. 747–756.

[KPS24] Guy Kornowski, Swati Padmanabhan, and Ohad Shamir. "On the Hardness of Meaningful Local Guarantees in Nonsmooth Nonconvex Optimization". In: *arXiv preprint arXiv:2409.10323* (2024).

[Kre89] Mark W Krentel. "Structure in locally optimal solutions". In: *30th Annual Symposium on Foundations of Computer Science*. IEEE Computer Society. 1989, pp. 216–221.

[KS24] Bruce M Kapron and Koosha Samieefar. "The Computational Complexity of Variational Inequalities and Applications in Game Theory". In: *arXiv preprint arXiv:2411.04392* (2024).

[Kuh60] Harold W Kuhn. "Some combinatorial lemmas in topology". In: *IBM Journal of research and development* 4.5 (1960), pp. 518–524.

[LJJ20a] Tianyi Lin, Chi Jin, and Michael Jordan. "On gradient descent ascent for nonconvex-concave minimax problems". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 6083–6093.

[LJJ20b] Tianyi Lin, Chi Jin, and Michael I Jordan. "Near-optimal algorithms for minimax optimization". In: *Conference on Learning Theory*. PMLR. 2020, pp. 2738–2779.

[Mąd+17] Aleksander Mądry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. "Towards deep learning models resistant to adversarial attacks". In: *stat* 1050.9 (2017).

[MOP20] Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. "A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 1497–1507.

[MP91] Nimrod Megiddo and Christos H Papadimitriou. "On total functions, existence theorems and computational complexity". In: *Theoretical Computer Science* 81.2 (1991), pp. 317–324.

[MPP18] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. "Cycles in adversarial regularized learning". In: *Proceedings of the twenty-ninth annual ACM-SIAM symposium on discrete algorithms*. SIAM. 2018, pp. 2703–2717.

[MRS20] Eric Mazumdar, Lillian J Ratliff, and S Shankar Sastry. "On gradient-based learning in continuous games". In: *SIAM Journal on Mathematics of Data Science* 2.1 (2020), pp. 103–131.

[MT05] Andrew McLennan and Rabee Tourky. "From imitation games to Kakutani". In: *Manuscript, available at http://www. econ. umn. edu/mclennan/Papers/papers. html* 37 (2005), p. 41.

[MV21] Oren Mangoubi and Nisheeth K Vishnoi. "Greedy adversarial equilibrium: an efficient alternative to nonconvex-nonconcave min-max optimization". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 896–909.

[Nem04] Arkadi Nemirovski. "Prox-method with rate of convergence O (1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems". In: *SIAM Journal on Optimization* 15.1 (2004), pp. 229–251.

[Neu28] J v. Neumann. "Zur theorie der gesellschaftsspiele". In: *Mathematische annalen* 100.1 (1928), pp. 295–320.

[Nou+19] Maher Nouiehed, Maziar Sanjabi, Tianjian Huang, Jason D Lee, and Meisam Razaviyayn. "Solving a class of nonconvex min-max games using iterative first order methods". In: *Advances in Neural Information Processing Systems* 32 (2019).

[NS06] Yurii Nesterov and Laura Scrimali. "Solving strongly monotone variational and quasi-variational inequalities". In: (2006).

[OLR21] Dmitrii M Ostrovskii, Andrew Lowy, and Meisam Razaviyayn. "Efficient search of first-order nash equilibria in nonconvex-concave smooth min-max problems". In: *SIAM Journal on Optimization* 31.4 (2021), pp. 2508–2538.

[Pap94] Christos H Papadimitriou. "On the complexity of the parity argument and other inefficient proofs of existence". In: *Journal of Computer and system Sciences* 48.3 (1994), pp. 498–532.

[PB14] Neal Parikh and Stephen Boyd. "Proximal Algorithms". In: *Foundations and Trends in Optimization* 1.3 (2014), pp. 127–239.

[Pet+22] Thomas Pethick, Puya Latafat, Panagiotis Patrinos, Olivier Fercoq, and Volkan Cevher. "Escaping limit cycles: Global convergence for constrained nonconvex-nonconcave minimax problems". In: *International Conference on Learning Representations*. 2022.

[PVGZ23] Christos Papadimitriou, Emmanouil-Vasileios Vlatakis-Gkaragkounis, and Manolis Zampetakis. "The Computational Complexity of Multi-player Concave Games and Kakutani Fixed Points". In: *Proceedings of the 24th ACM Conference on Economics and Computation*. 2023, pp. 1045–1045.

[Roc70] R Tyrrell Rockafellar. "Monotone operators associated with saddle-functions and minimax problems". In: *Proceedings of Symposia in Pure Mathematics*. Vol. 18. American Mathematical Society. 1970, pp. 241–250.

[Ros65] J Ben Rosen. "Existence and uniqueness of equilibrium points for concave n-person games". In: *Econometrica: Journal of the Econometric Society* (1965), pp. 520–534.

[Rub15] Aviad Rubinstein. "Inapproximability of Nash equilibrium". In: *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*. 2015, pp. 409–418.

[Sch91] Alejandro A Schäffer. "Simple local search problems that are hard to solve". In: *SIAM journal on Computing* 20.1 (1991), pp. 56–87.

[SND18] Aman Sinha, Hongseok Namkoong, and John Duchi. "Certifying Some Distributional Robustness with Principled Adversarial Training". In: *International Conference on Learning Representations*. 2018.

[Spe28] E Sperner. "Neuer beweis für die invarianz der dimensionszahl und des gebietes". In: *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*. Vol. 6. Springer Science and Business Media LLC. 1928, pp. 265–272.

[Sta64] Guido Stampacchia. "Formes bilineaires coercitives sur les ensembles convexes". In: *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* 258.18 (1964), p. 4413.

[Tse95] Paul Tseng. "On linear convergence of iterative methods for the variational inequality problem". In: *Journal of Computational and Applied Mathematics* 60.1-2 (1995), pp. 237–252.

[Zha+20] Jingzhao Zhang, Hongzhou Lin, Stefanie Jegelka, Suvrit Sra, and Ali Jadbabaie. "Complexity of finding stationary points of nonconvex nonsmooth functions". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 11173–11182.

Appendix

A Complexity Classes and Reductions

Model of computation We work in the standard Turing model of computation in which inputs are represented by irreducible fractions and natural numbers are represented in binary [AB09]. Functions are represented by algebraic (also known as arithmetic) circuits. An algebraic circuit is a directed acyclic graph in which each node represents one of the operations $\{+, *, \max, \times c\}$ where for $c \in \mathbb{Q}$ the symbol $\times c$ denotes the multiplication by some constant c. All nodes have in-degree 2 and out-degree 1 except for the $*c$ nodes with in- and out-degree 1. The tree's leaves are the inputs, and nodes without children are the outputs. For rational numbers x, we denote with $\text{size}(x)$ the number of bits used to write down the numerator and denominator, while for a circuit $\mathcal{C}$, we use $\text{size}(\mathcal{C})$ to denote the number of bits used to write down its gates and constants.[11]

Total search problems A search problem is defined by a binary relation P, where for each input $x \in \{0,1\}^*$ the objective is to find a $y \in \{0,1\}^*$ such that $P(x,y)$ evaluates to `True`, and returns "$\perp$" if $P(x,y)$ evaluates to `False` for all $y \in \{0,1\}^*$. FNP is the class of search problems in which $P(x,y)$ can be evaluated in polynomial time in $\text{size}(x) + \text{size}(y)$, and there exists some polynomial p such that if there exists y such that $P(x,y) = \texttt{True}$ then there exists z such that $P(x,z) = \texttt{True}$ and $\text{size}(z) \leq p(\text{size}(x))$.

We will work with a subset of problems in FNP for which a solution is guaranteed to exist, specifically within the class TFNP (*total search problems in NP*). This class contains the search problems in FNP for which, for every input x, there exists a y such that $P(x,y)$ evaluates to `True`.

Let Π_1 and Π_2 be two problems in TFNP. We say that Π_1 *reduces to* Π_2 if there exist two polynomial-time functions a and b such that for any instance x of Π_1, if we can find a solution y to instance $a(x)$ of Π_2, then $b(y)$ yields a solution to instance x for Π_1. In symbols, Π_1 reduces to Π_2 if, for all x, y,

$$\Pi_2(a(x), y) = \texttt{True} \quad \Longrightarrow \quad \Pi_1(x, b(y)) = \texttt{True}.$$

This indicates that Π_2 is "more difficult" than Π_1. Moreover, if we could efficiently find a solution y for $a(x)$, we would also be able to efficiently find a solution $b(y)$ for Π_1.

The class PPAD It is widely believed that TFNP has no complete problems [MP91]. Hence, its analysis has been advanced by identifying and studying interesting subclasses. These subclasses are characterized by the type of existence argument used to prove their membership in TFNP. One of the most extensively studied subclasses is PPAD [Pap94]. The complexity class PPAD $\subseteq$ TFNP consists of all problems which can be reduced to the End-of-the-Line problem, which we recall next.

Problem 12 (End-of-the-Line)**.** Given in input two boolean circuits $S : \{0,1\}^n \to \{0,1\}^n$ and $P : \{0,1\}^n \to \{0,1\}^n$, such that $S(0^n) \neq 0^n$ and $P(0^n) = 0^n$, find $\omega \in \{0,1\}^n$ such that $P(S(\omega)) \neq \omega$ or $S(P(\omega)) \neq \omega \neq 0^n$.

The two circuits P and S implicitly define a directed graph with an exponential number of vertices, each corresponding to an element of $\{0,1\}^n$, where each vertex has in- and out-degree of at most 1. The goal of End-of-the-Line is to find a non-trivial source or a sink.

An important PPAD-complete problem is finding fixed points of continuous functions, whose existence is guaranteed by Brouwer's fixed point theorem [Bro11].

Problem 13 (Brouwer)**.** Given a precision parameter $\epsilon \geq 0$ and a circuit implementing an L-Lipschitz function $F : [0,1]^d \to [0,1]^d$, find a point $z \in [0,1]^d$ such that $\|F(z) - z\| \leq \epsilon$.

[11] We adopt the assumption from Fearnley et al. [Fea+22] that algebraic circuits are *well-behaved*, meaning that any directed path leading to an output contains at most $\log(\text{size}(\mathcal{C}))$ multiplication gates. This condition ensures efficient evaluation of the circuit by preventing iterated squaring, which could otherwise produce doubly exponential numbers with an exponential representation size relative to the input and circuit size.

Another PPAD-complete problem that we will use extensively in this paper is related to computing equilibria in polymatrix games. In these games, agents are represented as nodes of an undirected graph, where each agent plays a bi-matrix game with every other node to whom they are connected.

Problem 14 (PolyMatrix)**.** Given a graph (V, E), a matrix $A_{i,j} \in [0,1]^{n\times n}$ for each $(i,j) \in E$, and an approximation $\epsilon > 0$, find vectors $x_i \in \Delta_n$, $i \in V$ such that:

$$\sum_{j\in V:(i,j)\in E} x_i^\top A_{i,j} x_j \geq \sum_{j\in V:(i,j)\in E} (A_{i,j}x_j)_k - \epsilon \quad \forall i \in V, k \in [n].$$

PolyMatrix was at the center of a landmark result of Rubinstein [Rub15], which first proved constant inapproximability of equilibria.

Proposition A.1 ([Rub15, Theorem 1])**.** *There exists a constant ϵ^* such that* PolyMatrix *is* PPAD*-complete for ϵ^*, even when the graph has degree 3, and each player has $n = 2$ actions.*

Recently, Deligkas et al. [Del+24] improved the constant to $\epsilon^* \approx 0.088$.[12]

Promise problems Formally, problems in TFNP are expected to guarantee the existence of a solution for any input. However, many of our problems admit solutions only when certain conditions are met, placing them in TFNP only for a restricted subset of inputs. The community refers to these conditions as *promises*. For instance, we may promise that a circuit indeed defines an L-Lipschitz function. One recent approach to circumvent this issue and obtain problems that are formally in TFNP is to add a certificate of promise violations to the solutions of the problems [DP11]. However, verifying violations of the promises may introduce additional complexities in solving the problem. For instance, verifying whether a circuit computes an L-Lipschitz function could be significantly more challenging than finding a solution to the original problem. A principled approach proposed by Fearnley et al. [Fea+22] to address this issue is to utilize *promise-preserving reductions*, where any violation of certain promises corresponds to a violation of the promise of the original problem from which we are reducing. In our specific setting, while this may be more formally accurate, it complicates the statements and proofs without significantly adding merits to our results. Therefore, we concentrate on the promise version of our problems. Moreover, when proving hardness results, the instances generated by our reductions will enforce the promises syntactically, and thus our hardness results will not rely on the hardness of checking the satisfaction of the promises.

B PPAD Membership

There are a few recent works that give general techniques to prove PPAD-membership of TFNP problems. Papadimitriou et al. [PVGZ23] is the first to define a meaningful computational problem related to Kakutani's theorem. Their main contribution is to enable a highly general representation of sets (*e.g.*, via separation oracles). Their main interest is Kakutani's problem in the context of concave games, whereas our focus is on QVIs, which are more general. More details about this relationship can be found in Section D.3. More recently, Filos-Ratsikas et al. [FR+24] extended the "pseudogate" technique of Filos-Ratsikas et al. [FR+23] (originally used for FIXP-membership results) to the PPAD setting. They designed a general technique, based on pseudogates, that can be used to easily prove membership of an impressive number of problems with exact rational solutions. Independently and concurrently, Kapron and Samieefar [KS24] extended the result by Papadimitriou et al. [PVGZ23], and showed PPAD-membership of QVIs when the correspondences are described via separation oracles (and thus capture more general settings than ours). One difference is that Kapron and Samieefar [KS24] assumes that the set-valued function of the QVI is Lipschitz with respect to the Hausdorff distance, whereas we do not. However, as part of our proof, we show that a particular relaxation of our QVI instance is Lipschitz with respect to the Hausdorff distance. See Section D for more details.

[12] The result by Deligkas et al. [Del+24] applies even in the case of normalized payoffs. Since our payoffs are not normalized, we could, in principle, use a larger constant. We choose not to, as we do not focus on optimizing constants.

In this section, we describe the PPAD-membership results for the case of independent constraints. The PPAD membership of GDAFixedPoint-Indep is proved by showing the PPAD membership of QuasiVI, which is very general and can be of independent interest.[13]

B.1 Existence of Solutions With Jointly Convex Constraints

We begin by briefly reviewing how Daskalakis et al. [DSZ21] established PPAD membership for LocalMinMax under jointly convex constraints and explaining why this approach does not readily extend to independent constraints. Daskalakis et al. [DSZ21] introduced the *safe gradient descent-ascent* map F_{sGDA} in which the projection step is joint and thus "safe", *i.e.*, the resulting point is guaranteed to lie in K. Formally, F_{sGDA} is defined as follows:

$$F_{\text{sGDA}}(x, y) = \Pi_K \begin{pmatrix} x - \nabla_x f(x, y) \\ y + \nabla_y f(x, y) \end{pmatrix}.$$

Note that this joint projection step can only be defined for jointly convex constraints. In Section C we give a thorough overview of the relationship of QVIs, VIs, and fixed points of various gradient descent-ascent maps.

The membership proof by Daskalakis et al. [DSZ21] proceeds roughly as follows: fixed points of F_{sGDA} are fixed points of F_{GDA}, which are, by Theorem 2.1, solutions to LR-LocalMinMax-JC. Additionally, the problem of computing fixed points of F_{sGDA} is in PPAD through a standard reduction to Brouwer.

Theorem B.1 ([DSZ21, Theorem 5.2])**.** GDAFixedPoint-JC *is in* PPAD.

Extending this result to the case of independent constraints introduces fundamental complications, which can be understood by looking at the well-known equivalence between VIs and approximate fixed points of the safe gradient steps (see Facchinei and Pang [FP03] and Theorem C.5 for details). In particular, when the constraints are jointly convex, the problem reduces to a single player optimizing a "complicated" function on the joint set K. However, with independent constraints, there is no straightforward single-player formulation or VI formulation of the problem. As we already pointed out, F_{GDA} does not guarantee the resulting points to be feasible. Moreover, even with jointly convex constraints, solutions to F_{GDA} and F_{sGDA} are not equivalent as fixed points of F_{GDA} are not necessarily fixed points of F_{sGDA} (see Theorem C.4).

This suggests that we should look to a larger class of problems to provide existence results for independent constraints. As we demonstrate in the following section, min-max problems with independent constraints are equivalent to QVIs rather than VIs, making the original approach inapplicable.

B.2 Existence of Solutions With Independent Constraints

In this section, we discuss the existence of solutions and PPAD-membership of GDAFixedPoint-Indep. Moreover, we show why it is necessary to relax the constraint satisfaction. In particular, we can easily show that, without relaxing the constraints, a solution to QuasiVI and LocalMinMax might not exist even for independent constraints (the proof of this proposition and the proof of the other results on the PPAD-membership with independent constraints is provided in Section D).

Proposition B.2. QuasiVI, LR-LocalMinMax, *and* GDAFixedPoint *with* $\nu = 0$ *do not always admit a solution. On the other hand,* Kakutani *always admits a (potentially irrational) solution even when* $\nu = 0$.[14]

This should not come as a surprise. Indeed, the standard result regarding the existence of solutions to QVI [CP82] requires the correspondence Q to be continuous.[15] However, our correspondences are of the form $Q(z) = \{z' \in [0,1]^d : A(z)z' - b(z) \leq 0\}$ for an L-Lipschitz continuous matrix $A : [0,1]^d \to [0,1]^{m \times d}$ and an L-Lipschitz continuous vector $b : [0,1]^d \to [0,1]^m$. This generates upper semicontinuous correspondences but not continuous ones. In contrast, Kakutani theorem [Kak41] requires only upper semicontinuous correspondences

[13] For the reasons discussed in Section 2.1.1, we cannot formally provide membership results for LR-LocalMinMax since such problems are not in TFNP.

[14] We have slightly abused notation, as QuasiVI is formally defined only for $\nu > 0$.

[15] Different definitions of continuity for correspondences are recalled in Section F.

to guarantee the existence of solutions. This is precisely the reason why, in Proposition B.2, we could prove that KAKUTANI always admits solutions. The dichotomy between QUASIVI and KAKUTANI highlighted in Proposition B.2 shows that the class of correspondences chosen in our problems is not trivial, and has interesting structural properties.

The following theorem relates specific QUASIVI problems to LR-LOCALMINMAX problems and, by Theorem 2.1, to GDAFIXEDPOINT (the proof is provided in Section C.2). We observe that even an arbitrarily small relaxation of the constraints is enough to establish the result.

Theorem B.3 (Two-way reduction between QUASIVI and LR-LOCALMINMAX)**.** *Consider a* LR-LOCALMINMAX *problem where f is G-Lipschitz and L-smooth. Then:*

i) For any $\epsilon, \delta > 0$, and $\nu \geq 0$, define $\alpha = \frac{1}{8}\left(\sqrt{(G+\delta)^2 + 4(\epsilon - \frac{L}{2}\delta^2)} - (G+\delta)\right)^2$. Any (x, y) such that $g_i(x, y) \leq \nu$ for $i \in \{1, 2\}$ and

$$(\nabla_x f(x, y), -\nabla_y f(x, y))^\top \ (x' - x, y' - y) \geq -\alpha \ \text{ for all } (x', y') \in K_1^\nu(y) \times K_2^\nu(x),$$

is an (ϵ, δ, ν)-solution to LR-LOCALMINMAX*.*

ii) Any (ϵ, δ, ν)-solution (x, y) to LR-LOCALMINMAX *satisfies $g_i(x, y) \leq \nu$ for $i \in \{1, 2\}$ and*

$$(\nabla_x f(x, y), -\nabla_y f(x, y))^\top (x' - x, y' - y) \geq -\sqrt{d}\left(3L\delta + 2\frac{\epsilon}{\delta}\right) \ \text{ for all } (x', y') \in K_1^\nu(y) \times K_2^\nu(x).$$

Our main membership result shows that QUASIVI is in PPAD, and hence the same result holds for GDAFIXEDPOINT-INDEP by the previous discussion (the proof is provided in Section D.3).

Theorem B.4. QUASIVI *and* GDAFIXEDPOINT-INDEP *are in* PPAD*.*

The above theorem is proven by showing a reduction from QUASIVI to a high-dimensional version of Sperner's Lemma. In Section D.3, we provide all the details of this result. In particular, we discuss in detail why we need to reduce to Sperner rather than to BROUWER. Intuitively, we cannot find a continuous selector function for the correspondence Q that is Lipschitz and for which fixed points are approximate solutions of QUASIVI.

Theorem B.3, together with Theorem B.4, also shows that there exists a reduction from LR-LOCALMINMAX-INDEP to END-OF-THE-LINE (or any other PPAD-complete problem). However, as already discussed in Section 2.1.1, this alone is not enough to conclude that the problem is in PPAD, as we lack FNP membership, which we ruled out. Notably, the existence of this reduction highlights that certain solutions to LR-LOCALMINMAX-INDEP have polynomial witnesses.

B.3 Globalization

To be concise, we only gave existence results for nonconvex-nonconcave problems. Clearly, these results can hold for the local regime (*i.e.*, $\delta \leq \sqrt{2\epsilon/L}$). However, as expected, when the payoff function of the minimizing player is convex (or, equivalently, when it's concave for the maximizing player), we can "globalize" the solution and extend the existence result to $\delta \geq \sqrt{d}$, which is the diameter of the hypercube and thus encodes global deviations. In Section D.3 we provide the proof for the case in which the minimizer (*i.e.*, the x player) has a convex function (a similar statement would hold for y if $f(x, \cdot)$ was concave).

Proposition B.5. *For any $(x^\star, y^\star)$ which is an (ϵ, δ, ν)-solution to* LOCALMINMAX*, if the function $f(\cdot, y)$ is convex for any $y \in [0, 1]^d$, then $f(x^\star, y^\star) \leq f(x', y^\star) + \frac{\epsilon\sqrt{d}}{\delta} \quad \forall x' \in K_1^\nu(y^\star)$.*

In particular, in the case of convex-concave objectives, given any $\tilde{\epsilon} > 0$, we can take $\epsilon = \frac{\tilde{\epsilon}^2}{dL}$ and $\delta = \frac{\tilde{\epsilon}}{L\sqrt{d}}$. Then, since there exists an (ϵ, δ, ν)-solution for any $\delta < \sqrt{\frac{2\epsilon}{L}}$, we have that there exists an $(\tilde{\epsilon}, \tilde{\delta}, \nu)$-solution where $\tilde{\delta} = \sqrt{d}$. The same reasoning can be easily extended to cases where only one between the minimizer or maximizer is convex/concave.

B.4 Application to Generalized Equilibria

Our PPAD membership result has implications beyond local min-max equilibria. It shows that we can work with highly general correspondences while still ensuring the existence of *rational* (and verifiable) solutions. At a high level, we show that a slight relaxation of the correspondence is sufficient to guarantee the existence of solutions when no solution exists in the non-relaxed problem. Moreover, this ensures the existence of rational solutions even when all solutions of the original correspondence are irrational. Alternatively, by looking at our proof, we observe that this relaxation is not strictly required and can be replaced by the assumption that the correspondence $Q(z)$ has always non-zero volume. To illustrate the generality of our approach, we include an application to generalized equilibria.

In a recent work about a general convex optimization approach to PPAD membership results [FR+24], the authors considered a natural computational version of Generalized Equilibrium Problems (GNEPs) [Deb52; FK10]. Using a reduction to QuasiVI, we can extend the membership result of Filos-Ratsikas et al. [FR+24, Theorem 4.4] to a broader class of constraints by relaxing the correspondences.

The problem is defined by a set of n players, and each player i has a set of actions $X_i \subseteq [0,1]^l$, and its utility is represented through a circuit computing a concave L_u-Lipschitz and G-smooth function $u_i : \bigtimes_{j\in[n]} X_j \to \mathbb{R}$, its gradient $\nabla_{x_i} u_i : \bigtimes_{j\in[n]} X_j \to \mathbb{R}^l$, and a correspondence $\Gamma_i : X_{-i} \rightrightarrows X_i$.[16] We assume that the correspondences have the form $\Gamma_i^\nu(x_{-i}) = \left\{x_i \in X_i : A_i(x_{-i})x_i \le b(x_{-i}) + \nu \cdot 1_{m_j}\right\}$ for each $i \in [n]$. These are polytopes with m_i inequalities for each x_{-i}. We assume that the matrix $A_i : X_{-i} \to \mathbb{R}^{m_i \times l}$ and vector $b_i : X_{-i} \to \mathbb{R}^{m_i}$ are generic circuits implementing L-Lipschitz functions and $\nu > 0$ is the relaxation of the correspondences. As standard in GNEPs we assume that, for all $i \in [n]$ and $x_{-i} \in [0,1]^{l(n-1)}$, the set of available strategies is not empty, *i.e.*, $\Gamma_i^\nu(x_{-i}) \neq \emptyset$.

A generalized equilibrium is a tuple $x = (x_i)_{i\in[n]}$ that is feasible for each player $i \in [n]$, and such that each player is best-responding, *i.e.*

$$x_i \in \bigtimes_{j\in[n]} \Gamma_j^\nu(x_{-j}) \quad \forall i \in [n] \quad \text{and} \quad u_i(x) \ge u_i(x_i', x_{-i}) - \epsilon \quad \forall x_i' \in \Gamma_i^\nu(x_{-i}),$$

for a given approximation $\epsilon > 0$.

We can reduce this problem to QuasiVI by mimicking the existence proof of GNEP in [CP82]. Define the search space as $[0,1]^d$, where $d = nl$, and the correspondence as $Q_\nu(z) = \bigtimes_{j\in[n]} \Gamma_j^\nu(z_{-j})$. Moreover, define the operator as $F(z) = (-\nabla_1 u_1(z), \dots, -\nabla_n u_n(z))$.

We now prove that the constructed problem is indeed a QuasiVI problem. First, since each u_i is smooth, the operator F is nG-Lipschitz. Moreover, $Q(z)$ is nonempty for all z since it is the Cartesian product of non-empty sets. Finally, each set $Q(z)$ is the Cartesian product of polytopes and thus is itself a polytope and can be written in the form $Q(z) = \{z' \in [0,1]^d : \tilde{A}(z)z' \le \tilde{b}(z)\}$.[17] This proves the PPAD-membership of GNEP.

The ingenious framework developed by Filos-Ratsikas et al. [FR+24] only considers exact problems with exact rational solutions. This allows them to create a "pseudogate" that, when inserted into arithmetic circuits, automatically computes fixed points and provides PPAD membership results by reduction to Brouwer.[18] This approach is extremely powerful, as it proves the PPAD-membership of a large class of problems with relative ease. However, it can only work with problems for which exact rational solutions exist. For example, in the particular case of generalized equilibria, the authors considered the case in which $\Gamma_i = \{x_i \in X_i : A_i x_i \le b(x_{-i})\}$, *i.e.*, the case in which the dependency on the other players in the correspondence only appears in the right-hand side vector of the polytope, which guarantees the existence of exact rational solutions [FR+24, Remark 3.4]. We remark that, in our definition, the matrix A_i is not fixed, and we cannot guarantee the existence of rational solutions (also see Proposition B.2). We overcome this

[16] We denote with x_{-i} the tuple $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ and, similarly, we denote with X_{-i} the set $\bigtimes_{j\in[n],j\neq i} X_j$.

[17] Formally, $\tilde{A}(z)$ is the block diagonal matrix with $\{A_i(z_{-i})\}_{i\in[n]}$ on the diagonal and $\tilde{b}(z)$ is the concatenation of the vectors $\{b_i(z_{-i})\}_{i\in[n]}$.

[18] Formally, their main result is that any circuit with pseudogates can be mimicked with normal gates, by adding extra variables, in the sense that from a fixed point of the normal circuit one can build a fixed point of the one with pseudogates.

difficulty at the expenses (which is necessary, also by Proposition B.2) of relaxing the correspondence (or equivalently assuming positive volume). It seems that our QuasiVI problem manages to handle approximate fixed points naturally. Even if our main interest in this paper was about hardness results, it would be interesting to explore whether our approach could be adapted to show membership for problems without exact rational solutions. For example, this argument seems to be extendable to the class of Piecewise-Linear best response oracles (PLBRO) games defined in Filos-Ratsikas et al. [FR+24].

C Reductions among LocalMinMax, GDAFixedPoint, QVI and VI

In this section, we give a clear overview of the relationships between QVIs, VIs, LocalMinMax, and fixed points of gradient descent-ascent dynamics. For the sake of completeness, we include some folklore results. Here we consider the computational versions of the problems, and, therefore, such results can be considered as polynomial-time reductions between problems. Figure 6 gives a summary of these results.

Finally, Section D.1 provides a proof for Proposition B.2, showing why constraints have to be relaxed to guarantee the existence of a solution.

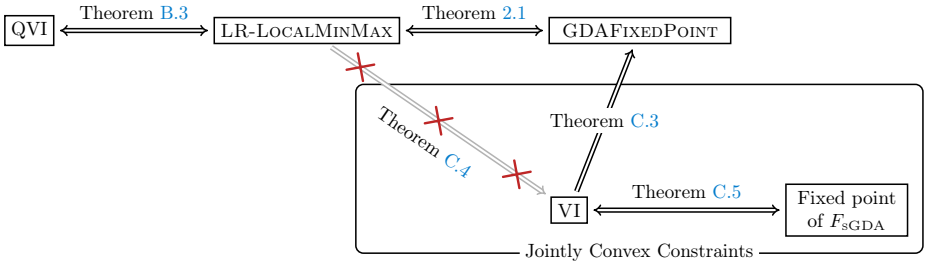


Figure 6: A connection between two problems means that they can be reduced to one another in polynomial time. The crossed connection between LR-LocalMinMax and VI means that there are solutions of LR-LocalMinMax that are not solutions to VI. We remark that in the picture, VI and QVI do not refer to the computational problems but refer to the specific instances associated with the local min-max problem. Theorem C.3, Theorem C.4 and Theorem C.5 holds only for jointly convex constraints, while Theorem 2.1 and Theorem B.3 holds even for independent constraints.

C.1 Separation Between Fixed Points of F_{GDA} and Fixed Points of F_{sGDA}

It is interesting to analyze the relation between the gradient updates of F_{GDA} and F_{sGDA} in the case of jointly convex constraints (*i.e.*, when both are well-defined). One might be induced to believe that fixed points of F_{sGDA} and F_{GDA} are equivalent. We show that this is not the case, and only one direction holds. In particular, all fixed points of F_{sGDA} are fixed points of F_{GDA} (Theorem C.3), but the opposite does not hold (Theorem C.4). This observation highlights the fundamental difference between VIs and QVIs. Indeed, while we can associate both a VI and a QVI to a LR-LocalMinMax-JC problem, the QVI has a strictly larger set of solutions.

Theorem C.3, together with Theorem 2.1, states that we can look for solutions of VI in order to find a solutions to LR-LocalMinMax-JC. For instance, if the payoff function is convex-concave, then there are polynomial-time algorithms that find solutions to the VI since it becomes monotone [NS06].

Before proving the theorem, we provide two technical lemmas.

Lemma C.1 ([HUL04, Theorem 3.1.1]). *For any $x \in \mathbb{R}^d$, and convex set K in $\mathbb{R}^d$ it holds:*

$$(y - \Pi_K(x))^\top (\Pi_K(x) - x) \geq 0 \quad \forall y \in K.$$

Lemma C.2. *Consider a pair $(x, y) \in K_1^\nu(y) \times K_2^\nu(x)$ and a $\beta \geq 0$. If $\nabla_x f(x, y)^\top (x' - x) \geq -\beta$ for all $x' \in K_1^\nu(y)$ then:*

$$\|x - \Pi_{K_1^\nu(y)}(x - \nabla_x f(x, y))\|^2 \leq \beta.$$

Similarly, if $-\nabla_y f(x, y)^\top (y' - y) \geq -\beta$ for all $y' \in K_2^\nu(x)$ then:

$$\|y - \Pi_{K_2^\nu(x)}(y + \nabla_y f(x, y))\|^2 \leq \beta.$$

Proof. We are going to prove only the first part of the statement, as the second can be proven with a symmetric argument.

Take $x' = \Pi_{K_1^\nu(y)}(x - \nabla_x f(x, y))$ which clearly lies in $K_1^\nu(y)$. By Lemma C.1 we know that for all $\tilde{x} \in K_1(y)$ and $\hat{x} \in \mathbb{R}^d$

$$(\tilde{x} - \Pi_{K_1^\nu(y)}(\hat{x}))^\top (\Pi_{K_1^\nu(y)}(\hat{x}) - \hat{x}) \geq 0.$$

By taking $\tilde{x} = x$, $\hat{x} = x - \nabla_x f(x, y)$, it holds $x' = \Pi_{K_1^\nu(y)}(\hat{x})$ and

$$(x - x')^\top (x' - (x - \nabla_x f(x, y))) \geq 0. \tag{5}$$

Thus it follows that:

$$\begin{aligned}
\beta &\geq \nabla_x f(x, y)^\top (x - x') \\
&= (x - (x - \nabla_x f(x, y))^\top (x - x') \\
&= (x - x')^\top (x - x') + (x' - (x - \nabla_x f(x, y)))^\top (x - x') \\
&\geq \|x - x'\|^2,
\end{aligned}$$

where the first inequality holds by assumption as $x' \in K_1^\nu(y)$ and in the last inequality we used Equation (5). □

Theorem C.3 (From VIs to GDAFixedPoint-JC)**.** *Consider an instance of* GDAFixedPoint-JC *with jointly convex constraints $K := \{x, y \in [0, 1]^d : g(x, y) \leq 0\}$. Then, any $(x, y) \in K$ such that:*

$$\begin{pmatrix} \nabla_x f(x, y) \\ -\nabla_y f(x, y) \end{pmatrix}^\top \begin{pmatrix} x' - x \\ y' - y \end{pmatrix} \geq -\frac{\beta^2}{2}, \text{ for all } (x', y') \in K,$$

is a β-fixed point of GDAFixedPoint.

Proof. The statement follows easily from Lemma C.2. Indeed by assumption, we have that for $y' = y$ it holds that:

$$\nabla_x f(x, y)^\top (x' - x) \geq \frac{-\beta^2}{2}$$

for all $(x', y) \in K$. The inequality holds for all (x', y) such that $g(x', y) \leq 0$. Thus we can apply Lemma C.2 by suitably redefining β and find that:

$$\frac{\beta^2}{2} \geq \|x - \Pi_{K_1^\nu(y)}(x - \nabla_x f(x, y))\|^2.$$

We can repeat the same argument for the second player and obtain:

$$\|(x, y) - F_{\text{GDA}}(x, y)\|^2 = \|y - \Pi_{K_2(x)}(y + \nabla_y f(x, y))\|^2 + \|x - \Pi_{K_1(y)}(x - \nabla_x f(x, y))\|^2 \leq \beta^2,$$

which concludes the proof. □

The following example demonstrates that the converse of Theorem C.3 does not hold. Indeed, the set of solutions of VI is strictly smaller than the set of solutions to LR-LocalMinMax-JC.

Theorem C.4. *There exists a payoff function f and a jointly convex constraint function g such that:*

i) there exists a point $(x^\star, y^\star)$ *which is a solution to* LocalMinMax *for any* $\epsilon > 0$ *and* $\delta > 0$,

ii) $(x^\star, y^\star)$ *is not a* $\bar{\beta}$*-approximate solution of the associated VI even for a constant* $\bar{\beta}$*, i.e., there exists a pair* $(x', y') \in K$ *such that*

$$(\nabla_x f(x^\star, y^\star), -\nabla_y f(x^\star, y^\star))^\top (x' - x^\star, y' - y^\star) < -\bar{\beta}. \tag{6}$$

Proof. Consider the payoff function $f : [0,1]^2 \to [0,1]$ given by:

$$f(x,y) = \frac{4}{5}\left((x-1)^2 - \left(y - \frac{1}{2}\right)^2 + \frac{1}{4}\right)$$

and the constraint function g given by:

$$g(x,y) = x + y - 1.$$

The function f is $8/5$-smooth and $4\sqrt{5}/5$-Lipschitz. Let $(x^\star, y^\star) = (1, 0)$. This is a (ϵ, δ)-solution to LocalMinMax for any ϵ and δ. Indeed,

$$f(x^\star, y^\star) = 0$$

and for all $x \in [0,1]$ we have that:

$$f(x, y^\star) \geq 0.$$

On the other hand, independently from $\delta > 0$, the second player cannot deviate as $g(x^\star, y) \leq 0$ if and only if $y = y^\star = 0$. This proves that, for any $\epsilon \geq 0$ and $\delta \geq 0$, the point $(x^\star, y^\star)$ is a solution to LocalMinMax.

However, there exists a constant $\bar{\beta} = \frac{4}{5}$ such that the point $(x^\star, y^\star)$ does not solve the following VI for any $\beta < \bar{\beta}$:

$$\begin{pmatrix} \nabla_x f(x^\star, y^\star) \\ -\nabla_y f(x^\star, y^\star) \end{pmatrix}^\top \begin{pmatrix} x' - x^\star \\ y' - y^\star \end{pmatrix} \geq -\beta, \text{ for all } (x', y') \in K.$$

Indeed, we have $\nabla_x f(x,y) = \frac{4}{5}(2x - 2)$ and $\nabla_x f(x^\star, y^\star) = 0$. On the other hand, $\nabla_y f(x,y) = -\frac{4}{5}(2y - 1)$ and thus: $\nabla_y f(x^\star, y^\star) = \frac{4}{5}$. This yields

$$\begin{pmatrix} \nabla_x f(x^\star, y^\star) \\ -\nabla_y f(x^\star, y^\star) \end{pmatrix}^\top \begin{pmatrix} x' - x^\star \\ y' - y^\star \end{pmatrix} = -\frac{4}{5} y'. \tag{7}$$

Hence, for $x' = 0$ and $y' = 1$ we have that $(x', y') \in K$ and

$$\begin{pmatrix} \nabla_x f(x^\star, y^\star) \\ -\nabla_y f(x^\star, y^\star) \end{pmatrix}^\top \begin{pmatrix} x' - x^\star \\ y' - y^\star \end{pmatrix} = -\frac{4}{5}.$$

This concludes the proof. □

C.2 Equivalence Between QVIs and LocalMinMax With Independent Constraints

We now shift our focus to proving the main result of this section: the equivalence between QuasiVI and LocalMinMax.

Theorem B.3 (Two-way reduction between QuasiVI and LR-LocalMinMax)**.** *Consider a* LR-LocalMinMax *problem where* f *is* G*-Lipschitz and* L*-smooth. Then:*

i) For any $\epsilon, \delta > 0$*, and* $\nu \geq 0$*, define* $\alpha = \frac{1}{8}\left(\sqrt{(G+\delta)^2 + 4(\epsilon - \frac{L}{2}\delta^2)} - (G+\delta)\right)^2$*. Any* (x, y) *such that* $g_i(x, y) \leq \nu$ *for* $i \in \{1, 2\}$ *and*

$$(\nabla_x f(x,y), -\nabla_y f(x,y))^\top \ (x' - x, y' - y) \geq -\alpha \text{ for all } (x', y') \in K_1^\nu(y) \times K_2^\nu(x),$$

is an (ϵ, δ, ν)*-solution to* LR-LocalMinMax*.*

ii) Any (ϵ, δ, ν)*-solution* (x, y) *to* LR-LocalMinMax *satisfies* $g_i(x, y) \leq \nu$ *for* $i \in \{1, 2\}$ *and*

$$(\nabla_x f(x, y), -\nabla_y f(x, y))^\top (x' - x, y' - y) \geq -\sqrt{d}\left(3L\delta + 2\frac{\epsilon}{\delta}\right) \text{ for all } (x', y') \in K_1^\nu(y) \times K_2^\nu(x).$$

Proof. Part *i*): Let $\alpha := \left(\frac{1}{2\sqrt{2}}\left(\sqrt{(G+\delta)^2 + 4(\epsilon - \frac{L}{2}\delta^2)} - (G+\delta)\right)\right)^2$. By applying Lemma C.2 we can show that

$$\|(x, y) - F_{\mathrm{GDA}}^\nu(x, y)\| \leq \sqrt{2\alpha}.$$

Indeed, it holds $\|\Pi_{K_1^\nu(y)}(x - \nabla_x f(x, y))\|^2 \leq \alpha$ and $\|\Pi_{K_2^\nu(x)}(y - \nabla_y f(x, y))\|^2 \leq \alpha$. The proof of *i*) is concluded by using the first point of Theorem 2.1.

Part *ii*): Let (x, y) be a (ϵ, δ, ν) solution to LR-LocalMinMax. Define $\hat{x}$ as any solution in

$$\arg \min_{\tilde{x} \in K_1^\nu(y) \cap \{\bar{x} : \|\bar{x} - x\| \leq \delta\}} f(\tilde{x}, y).$$

Then, the following VI holds for all $x' \in K_1^\nu(y)$ such that $\|x' - x\| \leq \delta$

$$\nabla_x f(\hat{x}, y)^\top (x' - \hat{x}) \geq 0. \tag{8}$$

Moreover, the L-smoothness of f implies that

$$\begin{aligned}
\nabla_x f(\hat{x}, y)^\top (x - \hat{x}) &\leq \frac{L}{2}\|x - \hat{x}\|^2 + f(x, y) - f(\hat{x}, y) \\
&\leq \frac{L}{2}\|x - \hat{x}\|^2 + \epsilon \\
&\leq \frac{L}{2}\delta^2 + \epsilon.
\end{aligned} \tag{9}$$

Then, combining Equation (8) and Equation (9) we have the for all $x' \in K_1^\nu(y)$ such that $\|x' - x\| \leq \delta$ it holds:

$$\begin{aligned}
0 &\leq \nabla_x f(\hat{x}, y)^\top (x' - \hat{x}) && \text{(Equation (8))} \\
&= \nabla_x f(\hat{x}, y)^\top (x' - x) + \nabla_x f(\hat{x}, y)^\top (x - \hat{x}) \\
&\leq \nabla_x f(\hat{x}, y)^\top (x' - x) + \frac{L}{2}\delta^2 + \epsilon && \text{(Equation (9))} \\
&= \nabla_x f(x, y)^\top (x' - x) + (\nabla_x f(\hat{x}, y) - \nabla_x f(x, y))^\top (x' - x) + \frac{L}{2}\delta^2 + \epsilon \\
&\leq \nabla_x f(x, y)^\top (x' - x) + \|\nabla_x f(\hat{x}, y) - \nabla_x f(x, y)\| \cdot \|x' - x\| + \frac{L}{2}\delta^2 + \epsilon && \text{(Cauchy–Schwarz)} \\
&\leq \nabla_x f(x, y)^\top (x' - x) + L\|\hat{x} - x\| \cdot \|x' - x\| + \frac{L}{2}\delta^2 + \epsilon && (L\text{-smoothness}) \\
&\leq \nabla_x f(x, y)^\top (x' - x) + L\delta^2 + \frac{L}{2}\delta^2 + \epsilon && (\|\hat{x} - x\|, \|x' - x\| \leq \delta) \\
&= \nabla_x f(x, y)^\top (x' - x) + \frac{3}{2}L\delta^2 + \epsilon.
\end{aligned}$$

Take any $x'' \in K_1^\nu(y)$ and consider $x' := x + \delta\frac{x'' - x}{\|x'' - x\|}$. Clearly, since $K_1^\nu(y)$ is convex, we have that $x' \in K_1^\nu(y)$ and $\|x' - x\| \leq \delta$. Therefore,

$$\begin{aligned}
-\frac{3}{2}L\delta^2 - \epsilon &\leq \nabla_x f(x, y)^\top (x' - x) \\
&= \frac{\delta}{\|x'' - x\|}\nabla_x f(x, y)^\top (x'' - x).
\end{aligned}$$

Thus, for all $x'' \in K_1^\nu(y)$:

$$\nabla_x f(x,y)^\top (x'' - x) \geq -\frac{\|x'' - x\|}{\delta}\left(\frac{3}{2}L\delta^2 + \epsilon\right) \geq -\frac{\sqrt{d}}{\delta}\left(\frac{3}{2}L\delta^2 + \epsilon\right).$$

Similarly we can prove that for all $y'' \in K_2^\nu(x)$:

$$\nabla_y f(x,y)^\top (y'' - y) \leq \frac{\sqrt{d}}{\delta}\left(\frac{3}{2}L\delta^2 + \epsilon\right).$$

Combining the two inequalities, we complete the proof. □

C.3 Equivalence Between VIs and Fixed Points Under Jointly Convex Constraints

In order to complete the picture for jointly convex constraints, we provide the following folklore theorem (adapted from Facchinei and Pang [FP03] in order to account for approximations) that states the well-known equivalence between fixed points and solutions to VI.

Theorem C.5 (Equivalence between VIs and Fixed Points)**.** *Given a function $F : [0,1]^d \to [0,1]^d$ and a convex set $K \subseteq [0,1]^d$ the following holds*

i) Any point $x \in K$ which is α-approximate fixed point of $x \mapsto \Pi_K(x - F(x))$ satisfy the following VI:

$$F(x)^\top (x' - x) \geq -2\alpha\sqrt{d} \text{ for all } x' \in K.$$

ii) Any point $x \in K$ which satisfies the following VI:

$$F(x)^\top (x' - x) \geq -\alpha \text{ for all } x' \in K,$$

is a $\sqrt{\alpha}$-approximate fixed point of $x \mapsto \Pi_K(x - F(x))$.

Proof. First, we prove *i)*. Define $\hat{x} := \Pi_K(x - F(x))$. We know that, by Lemma C.1, the following holds

$$(x' - \hat{x})^\top (\hat{x} - x + F(x)) \geq 0.$$

Simple algebraic manipulations lead to:

$$F(x)^\top (x' - x) + F(x)^\top (x - \hat{x}) + (x' - x)^\top (\hat{x} - x) - \|\hat{x} - x\|^2 \geq 0. \tag{10}$$

Since x is an α-approximate fixed point of $x \mapsto \Pi_K(x - F(x))$ it means that $\|\hat{x} - x\| \leq \alpha$. Thus:

$$\begin{aligned}
F(x)^\top (x' - x) &\geq \|\hat{x} - x\|^2 + F(x)^\top (\hat{x} - x) + (x - x')^\top (\hat{x} - x) && \text{(Equation (10))} \\
&\geq \|\hat{x} - x\|^2 - \|F(x)\| \cdot \|\hat{x} - x\| - \|x - x'\| \cdot \|\hat{x} - x\| && \text{(Cauchy-Schwartz)} \\
&\geq \|\hat{x} - x\|^2 - \sqrt{d} \cdot \|\hat{x} - x\| - \sqrt{d} \cdot \|\hat{x} - x\| && (F(x)_j, x_j, x'_j \in [0,1]) \\
&\geq -\sqrt{d} \cdot \|\hat{x} - x\| - \sqrt{d} \cdot \|\hat{x} - x\| && (F(x)_j, x_j, x'_j \in [0,1]) \\
&\geq -2\alpha\sqrt{d}.
\end{aligned}$$

Now we prove *ii)*. Let $\hat{x} := \Pi_K(x - F(x))$, again by Lemma C.1

$$-\|\hat{x} - x\| + F(x)^\top (x - \hat{x}) = (x - \hat{x})^\top (\hat{x} - x + F(x)) \geq 0,$$

since $x \in K$. Moreover, by assumption we have that $F(x)^\top (x' - x) \geq -\alpha$ for all $x' \in K$ and thus also for $x' = \hat{x}$. Thus, combining the two, we have:

$$-\|x - \hat{x}\|^2 + \alpha \geq -\|x - \hat{x}\|^2 + F(x)^\top (x - \hat{x}) \geq 0, \tag{11}$$

which proves that $\|x - \hat{x}\| \leq \sqrt{\alpha}$. □

Daskalakis et al. [DSZ21] proved the PPAD membership of GDAFixedPoint-JC, *i.e.*, with *jointly convex* constraints (and therefore also for the case of *box constraints*), for $\nu = 0$ and $\delta < \sqrt{2\epsilon/L}$.

Equivalently, we can observe that this result follows since Theorem C.3, along with the known equivalence between VIs and fixed points (formally restated in Theorem C.5), implies that every fixed point of F_{sGDA} is a solution to GDAFixedPoint-JC. Moreover, a fixed point of F_{sGDA} is guaranteed to exist due to a direct reduction from the problem of finding a fixed point of F_{sGDA} to Brouwer (which also proves that the problem of finding fixed points of F_{sGDA} is in PPAD).[19]

D Missing Proofs for PPAD-membership

We present the proofs for our membership results. First, in Section D.1, we show that without relaxing the constraints, solutions might not exist. Then, in Section D.2, we introduce preliminary results on the continuity of the correspondence Q, which are then used in Section D.3 to establish our main membership result. Finally, in Section D.4, we show how our result can be globalized for convex or concave functions.

D.1 Proof of Proposition B.2

The following statement proves that, without relaxing the constraints, a solution to QuasiVI and LR-LocalMinMax might not exist even for independent constraints.

Proposition B.2. QuasiVI, LR-LocalMinMax, *and* GDAFixedPoint *with* $\nu = 0$ *do not always admit a solution. On the other hand,* Kakutani *always admits a (potentially irrational) solution even when* $\nu = 0$.[20]

Proof. ▷ **Kakutani *admits a solution.*** To apply Kakutani's fixed-point theorem (see Theorem F.4) we need to show that $Q(z) = \{z' \in [0,1]^d : (A(z)z' - b(z))_i \leq 0 \quad \forall i \in [n]\}$, given by two circuits $A : [0,1]^d \to [0,1]^{d\times n}$ and $b : [0,1]^d \to [0,1]^n$ which are guaranteed to be L-Lipschitz and such that $Q(z) \neq \emptyset$ for all $z \in [0,1]^d$, satisfies that:

- $Q(z)$ is convex for all $z \in [0,1]^d$;
- $Q(z)$ is closed for all $z \in [0,1]^d$;
- $Q(z)$ is non-empty for all $z \in [0,1]^d$;
- Q is upper-semicontinuous in $[0,1]^d$.

The only non-trivial condition is that Q is upper-semicontinuous (see Definition F.3). It is known that if the graph of Q (defined as $\{(z,y) \in [0,1]^{2d} : y \in Q(z)\}$) is closed, then the correspondence Q is upper-semicontinuous [Ber77, Corollary of Theorem 7, Chapter VI]. In our case, we can define the function:

$$g(z,z') \coloneqq (A(z)z' - b(z))_i \quad \forall i \in [n],$$

which is continuous in $[0,1]^{2d}$ and thus $\{z' \in [0,1]^d : g(z,z') \leq 0\}$ is closed for each z.

▷ ***All solutions might be irrational.*** We will show an instance of the problem of finding a point $z^\star \in [0,1]^2$ such that $z^\star \in Q(z^\star)$ has only irrational solutions, where $Q(z)$ is given by the points contained in a polytope $\{z' : A(z)z' \leq b(z)\}$.

We write x for the first component and y for the second component of $z \in \mathbb{R}^2$. Let $A : [0,1]^2 \to \mathbb{R}^{4\times 2}$ and $b : [0,1]^2 \to \mathbb{R}^4$ defined by

$$A(x,y) \coloneqq \begin{bmatrix} +1 & 0 \\ -1 & 0 \\ 0 & +x \\ 0 & -x \end{bmatrix}, \quad b(x,y) \coloneqq \begin{bmatrix} +y \\ -y \\ +\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}.$$

[19] Actually, Daskalakis et al. [DSZ21] directly prove that a fixed point of F_{sGDA} is a fixed point of F_{GDA} without going trough the VI formulation of the problem.

[20] We have slightly abused notation, as QuasiVI is formally defined only for $\nu > 0$.

Clearly a point (x', y') is in $Q(x, y)$ if $x' = y$ and $y' = \frac{1}{2x}$. Thus we are looking for solutions in $[0, 1]^2$ to the system

$$\begin{cases} x = y \\ y = \frac{1}{2x}, \end{cases}$$

which has as the only solution $x = y = \frac{1}{\sqrt{2}}$, which is irrational.

▷ ***Non-existence of solutions.*** We will prove the result by providing an instance for LR-LocalMinMax for which there is no solution. The non-existence of solutions for QuasiVI with $\nu = 0$ follows by writing the QVI associated with the LR-LocalMinMax instance. In particular, Theorem B.3 states that from a solution to such a QVI we would been able to find a solution to LR-LocalMinMax, and thus there is no such solution of the QVI. Similarly, thanks to Theorem 2.1, we can show that GDAFixedPoint does not admit solutions.

Let $n = d = 1$ and let the constraints be defined by $g_1(x, y) = \{x - y \leq 0 \wedge y - x \leq 0\}$, $g_2(x, y) = \{yx \leq 0\}$, while the payoff function be given by $f(x, y) = \frac{1}{2}y^2$. The approximations ϵ, δ are such that $\delta = \frac{5}{4}\epsilon$. It can be easily checked that this is a valid LR-LocalMinMax instance as the function f is L-smooth for all L and thus $\delta \leq \sqrt{2\epsilon/L}$, is always verified by our choice of δ and $\epsilon \leq 1$. However, the only point (x, y) for which $g_1(x, y) \leq 0$ and $g_2(x, y) \leq 0$ is given by $(0, 0)$, but this is not an equilibrium, as the second player can deviate to $y' = \delta$ and gain an extra utility δ, and thus, it is not an equilibrium as $\delta > \epsilon$. □

D.2 Hausdorff Continuity of the Correspondence Q

In this section, we provide some results on the Hausdorff continuity of correspondence $Q_\beta(z)$ (see Problem 9) both with respect to z and β.

Definition D.1. A correspondence $Q : X \rightrightarrows Y$ between two topological spaces is K-Lipschitz with respect to the Hausdorff metric if

$$d_H(Q(z), Q(z')) \leq K \cdot d(z, z'),$$

where, for two set $A, B \subseteq Y$, the Hausdorff distance $d_H(A, B)$ is defined as

$$d_H(A, B) := \max\left\{\sup_{a\in A}\inf_{b\in B} d(a, b), \sup_{b\in B}\inf_{a\in A} d(b, a)\right\},$$

and $d(\cdot, \cdot)$ is the distance of the topological space Y. Equivalently, $d_H(A, B) = \inf\{\epsilon > 0 : A \subseteq B_\epsilon \quad \text{and} \quad B \subseteq A_\epsilon\}$, where $A_\epsilon := \cap_{a\in A}\{x : d(x, a) \leq \epsilon\}$.

In Euclidean geometry, for a K-Lipschitz correspondence Q we have that each point x in the set $Q(z)$ can be projected onto the set $Q(z')$ and the distance between x and $\Pi_{Q(z')}(x)$ is at most $K\|z - z'\|$.

We start by showing the Hausdorff continuity of $Q_\beta(z)$ with respect to β. Notice that our result holds only for (even arbitrary small) relaxations, *i.e.*, $\beta > 0$.

Lemma D.2. *For any $z \in [0, 1]^d$ and for any $\bar{\beta} > 0$, the correspondence $\gamma \to Q_\gamma(z)$ is $(\sqrt{d}/\bar{\beta})$-Lipschitz in the Hausdorff metric on $\gamma \in [\bar{\beta}, \infty)$, i.e.,*

$$d_H(Q_\gamma(z), Q_{\gamma'}(z)) \leq \frac{\sqrt{d}}{\bar{\beta}}|\gamma - \gamma'| \quad \forall \gamma, \gamma' \geq \bar{\beta}.$$

Proof. Consider two $\gamma \geq \gamma' \geq \bar{\beta}$. Clearly, $Q_{\gamma'}(z) \subseteq Q_\gamma(z)$. Hence, to prove the statement is sufficient to consider any $x \in Q_\gamma(z)$ and show that $\|\Pi_{Q_{\gamma'}(z)}(x) - x\| \leq \sqrt{d}\frac{\gamma - \gamma'}{\bar{\beta}}$.

Consider an $x^0 \in Q(z)$, which exists by our assumption on the non-emptiness of $Q(z)$. Let $x^* = (1 - \alpha)x + \alpha x^0$ where α is set in the following. Then, we have that

$$\begin{aligned} A(z)x^* - b(z) &= A(z)\left((1 - \alpha)x + \alpha x^0\right) - b(z) \\ &= (1 - \alpha)\left(A(z)x - b(z)\right) + \alpha x^0\left(A(z)x - b(z)\right) \\ &\leq (1 - \alpha)\gamma. \end{aligned}$$

Hence, for $\alpha = \frac{\gamma - \gamma'}{\gamma}$ we have that

$$A(z)x^* - b(z) \leq (1-\alpha)\gamma \leq \gamma'.$$

This implies that

$$||\Pi_{Q_\beta(z)}(x) - x|| \leq ||x^* - x|| \leq \alpha ||x^0 - x|| \leq \sqrt{d}\frac{\gamma - \gamma'}{\gamma} \leq \sqrt{d}\frac{\gamma - \gamma'}{\bar{\beta}}.$$

Hence, the function is $\frac{\sqrt{d}}{\bar{\beta}}$-Lipschitz. □

Now, we show the Hausdorff continuity of $Q_\beta(z)$ with respect to z. Notice that also in this case our result holds only for (even arbitrarily small) relaxations, *i.e.*, $\beta > 0$ We start by recalling the following result about the Lipschitz continuity of the maximum of Lipschitz functions. We provide a proof for completeness.

Lemma D.3. *Let $h_i : \mathbb{R}^d \to \mathbb{R}$ be L_i-Lipschitz functions for $i \in [n]$. Then $g(\cdot) = \max_{i\in[n]} h_i(\cdot)$ is L-Lipschitz, where $L = \max_{i\in[n]} L_i$.*

Proof. For each $i \in [n]$ we have that $h_i(x) \leq h_i(y) + L_i\|x - y\|$ by L_i-Lipschitzness of h_i. Consider now x and $i(x)$ such that $g(x) = h_{i(x)}(x)$. Then

$$\begin{aligned} g(x) &= h_{i(x)}(x) \\ &\leq h_{i(x)}(y) + L_{i(x)}\|x - y\| \\ &\leq g(y) + L_{i(x)}\|x - y\| \\ &\leq g(y) + L\|x - y\|. \end{aligned}$$

Similarly, consider y and $i(y)$ such that $g(y) = h_{i(y)}(y)$. Then

$$\begin{aligned} g(y) &= h_{i(y)}(y) \\ &\leq h_{i(y)}(x) + L_{i(y)}\|x - y\| \\ &\leq g(x) + L_{i(y)}\|x - y\| \\ &\leq g(x) + L\|x - y\|. \end{aligned}$$

This proves that g is L-Lipschitz. □

We continue by proving some properties related to the constraint function.

Lemma D.4. *Given an L-Lipschitz matrix valued function $A : [0,1]^d \to [0,1]^{n\times d}$ and a vector valued L-Lipschitz function $b : [0,1]^d \to [0,1]^n$ let $G_i(z|z')$ be defined as:*

$$G_i(z|z') := [A(z')^\top z - b(z')]_i,$$

and let $G(z|z') := \max_{i\in[n]} G_i(z|z')$. Then the following holds:

i) G is $2L$-Lipschitz in the second argument;

ii) G is convex in the first argument.

Proof. First, we prove the first point. We prove that for all $i \in [n]$ the function $G_i(z|\cdot) : [0,1]^d \to \mathbb{R}$ is $2L$-Lipschitz for each z. Formally:

$$\begin{aligned} |G_i(z|\tilde{z}_1) - G_i(z|\tilde{z}_2)| &= |(A(\tilde{z}_1)z)_i - b_i(\tilde{z}_1) - (A(\tilde{z}_2)z)_i + b_i(\tilde{z}_2)| \\ &\leq |(A(\tilde{z}_1)z)_i - (A(\tilde{z}_1)z)_i| + |b_i(\tilde{z}_1) - b_i(\tilde{z}_2)| \\ &\leq 2L\|\tilde{z}_1 - \tilde{z}_2\|. \end{aligned}$$

Applying Lemma D.3 we conclude that G is $2L$-Lipschitz in the second argument.

The second point is proven by noting that $G_i(\cdot|z)$ is a linear function for each fixed z, and thus $G(\cdot|z)$ is the maximum over finitely many linear functions, which thus is convex. □

Lemma D.5. *For any $\beta > 0$, the correspondence $Q_\beta : [0,1]^d \to 2^{[0,1]^d}$ is $\frac{2L\sqrt{d}}{\beta}$-Lipschitz in the Hausdorff metric.*

Proof. Consider any $z, z' \in [0,1]^d$ and any $\tilde{z} \in Q_\beta(z)$. We have to prove that

$$d(\tilde{z}, Q_\beta(z')) := \|\tilde{z} - \Pi_{Q_\beta(z')}(\tilde{z})\| \leq K\|z - z'\|$$

for some $K > 0$. This would show that $d(\tilde{z}, Q_\beta(z')) \leq K\|z - z'\|$ for all $\tilde{z} \in Q_\beta(z)$. By this result, exchanging the roles of z and z' we prove that also $d(\tilde{z}, Q_\beta(z)) \leq K\|z - z'\|$ for all $\tilde{z} \in Q_\beta(z')$ and thus that

$$d_H(Q_\beta(z), Q_\beta(z')) \leq K\|z' - z\|.$$

Consider $\bar{z} := \alpha\tilde{z} + (1-\alpha)z_0$ where z_0 is any point in $Q(z')$ (and thus $G(z_0|z') \leq 0$) and $\alpha \in [0,1]$ will be specified in the following. Now consider the following inequalities:

$$\begin{aligned} G(\bar{z}|z') &\leq \alpha G(\tilde{z}|z') + (1-\alpha)G(z_0|z') \\ &\leq \alpha\left(2L\|z' - z\| + G(\tilde{z}|z)\right) + (1-\alpha)G(z_0|z') \\ &\leq 2\alpha L\|z' - z\| + \alpha\beta, \end{aligned}$$

where in the first inequality we used ii) of Lemma D.4, in the second inequality we used i) of Lemma D.4, and in the last one $G(\tilde{z}|z) \leq 0$ and $G(z_0|z') \leq 0$. By taking $\alpha = \beta/(2L\|z' - z\| + \beta)$ we get that $G(\bar{z}|z') \leq \beta$ and thus $\bar{z} \in Q_\beta(z')$. This implies that:

$$\begin{aligned} \|\tilde{z} - \Pi_{Q_\beta(z')}(\tilde{z})\| &\leq \|\tilde{z} - \bar{z}\| \\ &= (1-\alpha)\|z_0 - \tilde{z}\| \\ &\leq (1-\alpha)\sqrt{d} \\ &= \frac{2L\|z' - z\|}{2L\|z' - z\| + \beta}\sqrt{d} \\ &\leq \frac{2L\sqrt{d}}{\beta}\|z' - z\|. \end{aligned}$$

By the discussion above, this proves that the correspondence Q_β is $\left(\frac{2L\sqrt{d}}{\beta}\right)$-Lipschitz in the Hausdorff metric. □

By the previous analysis, we can prove the Lipschitzness-like condition on $\min_{z' \in Q_\nu(z)} F(z)^T z'$ for a fixed z and varying ν. Formally,

Lemma D.6. *Given z, $\nu > 0$, and $\beta > 0$, it holds*

$$\min_{z' \in Q_\nu(z)} F(z)^T z' \leq \min_{z' \in Q_{\nu+\beta}(z)} F(z)^T z' + \frac{d\beta}{\nu}.$$

Proof. Let $\bar{z} \in \arg\min_{z' \in Q_{\nu+\beta}(z)} F(z)^T z'$, let $\tilde{z} = \Pi_{Q_\nu}(\bar{z})$, and let $z^\star \in \arg\min_{z' \in Q_\nu(z)} F(z)^T z'$. Then:

$$\begin{aligned} F(z)^\top \bar{z} &= F(z)^\top(\bar{z} - \tilde{z}) + F(z)^\top \tilde{z} \\ &\geq F(z)^\top(\bar{z} - \tilde{z}) + F(z)^\top z^\star \\ &\geq -||F(z)|| \cdot ||\bar{z} - \tilde{z}|| + F(z)^\top z^\star \\ &\geq -\sqrt{d}\frac{\sqrt{d}}{\nu}\beta + F(z)^\top z^\star \\ &= -\frac{d}{\nu}\beta + F(z)^\top z^\star \end{aligned}$$

where the last follow by Lemma D.2. □

D.3 Proof of Theorem B.4

In this section, we provide a proof of the following theorem.

Theorem B.4. QuasiVI *and* GDAFixedPoint-Indep *are in* PPAD.

Our main technical contribution is to show that QuasiVI is in PPAD. The membership in PPAD of GDAFixedPoint-Indep is then given by Theorem B.3 and Theorem 2.1. To see that, let us consider the problem of finding an α_{GDA}-fixed point of GDAFixedPoint-Indep with a given ν. Then, by Theorem 2.1 any (ϵ, δ, ν)-solution to LR-LocalMinMax is an α_{GDA}-fixed point of GDAFixedPoint-Indep, where $\epsilon = \frac{\alpha_{\text{GDA}}^2 L}{(5L+2)^2}$ and $\delta = \frac{\alpha_{\text{GDA}}}{5L+2}$. Finally, let size$(I)$ be the size of the instance. Then, thanks to Theorem B.3 any solution to QuasiVI with

$$\alpha = \left(\frac{1}{2\sqrt{2}}\left(\sqrt{(G+\delta)^2 + 4(\epsilon - \frac{L}{2}\delta^2)} - (G+\delta)\right)\right)^2 \geq \text{poly}\left(\frac{1}{\text{size}(I)}\right)$$

is a (ϵ, δ, ν)-solution to LR-LocalMinMax and thus an α_{GDA}-fixed point of GDAFixedPoint-Indep.

We proceede proving the PPAD-membership of QuasiVI. We reduce to a problem related to Sperner's Lemma called HighD-Sperner.

HighD-Sperner Consider the d-dimensional hypercube $[0,1]^d$. A canonical way to partition the hypercube in simplexes is to divide the hypercube into cubelets of side-length and then divide each cublet into simplices with the Coxeter-Freudenthal-Kuhn triangulation [Kuh60]. We call V the set of vertices. A Sperner coloring T of a simplicization is the assignment of one of $d+1$ colors $[d] \cup \{0\}$ to each vertex of the simplicization. The assignment must satisfy the following boundary conditions:

i) for each $i \in [d]$, none of the vertices $v \in V$ on the face with $v_i = 0$ has color i;

ii) color 0 is not assigned to a vertex $v \in V$ on the face with $v_i = 1$ for some $i \in [d]$.

Figure 7 shows a valid application of these rules in the 2-dimensional hypercube. A panchromatic simplex induced by T is a simplex whose vertices have all the $d+1$ colors. Such simplex is guaranteed to exist by a parity argument [Spe28; Kuh60]. It is well known that not only every Sperner coloring of the hypercube has a panchromatic simplex but that the search problem of finding one can be reduced to End-of-the-Line [PVGZ23; CKK23; EY10].

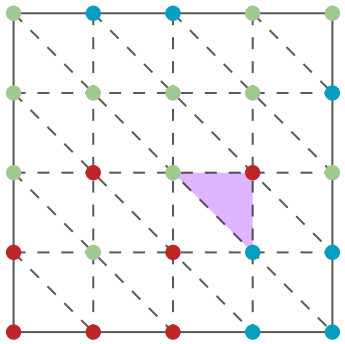


Figure 7: Example of a valid coloring of the 2-dimensional cube. The color **red** corresponds to the color 0 while the color **green** and **blue** corresponds to colors 1 and 2, respectively. The box on the boundaries are colored according to the allowed coloring on those edges. We highlighted in **purple** a panchromatic simplex.

Theorem D.7. *Given a boolean circuit $\mathcal{C} : V \to [d] \cup \{0\}$ that encodes the Sperner's coloring of a Kuhn's triangulation, the problem of finding a panchromatic simplex is in* PPAD.

Note that the coloring is given implicitly by the circuit $\mathcal{C}$ and that the domain, *i.e.*, set of vertices, is exponential in the input size. Hence, the problem is not trivially easy to solve, as one cannot simply check all the simplexes in the coloring.

Overview The high-level idea of the proof is simple: a fixed point of the correspondence that maps z to the set $\tilde{Q}(z) := \arg\min_{\tilde{z}\in Q_\nu(z)} F(z)^\top(\tilde{z}-z)$ would be a solution to QuasiVI. To show that this lies in PPAD we would need to find a Lipschitz continuous selector function of the correspondence $\tilde{Q}$ (a continuous function $f : [0,1]^d \to [0,1]^d$ such that $f(z) \in \tilde{Q}(z)$) and show a way of constructing efficiently this function as a circuit to feed to Brouwer. A simple way to define a selector function is to regularize slightly the objective $\tilde{z} \mapsto F(z)^\top(\tilde{z}-z) + \frac{\eta}{2}\|\tilde{z}-z\|^2$ to make it strongly convex. This makes the correspondence a singleton (convex optimization problems with strongly convex objective functions have a unique solution) without modifying too much the set of solutions (if η is small enough). Moreover, by Berge's maximum theorem, we know that this function is continuous [Ber77]. We can thus define the function

$$p_{\gamma,\eta}(z) := \arg\min_{\tilde{z}\in Q_\gamma(z)} \left\{ F(z)^\top \tilde{z} + \frac{\eta}{2}\|\tilde{z}-z\|^2 \right\}.$$

We need to make three important observations about the function $p_{\gamma,\eta}$. First we know that $p_{\gamma,\eta}(z)$ is simply the projection onto the set $Q_\gamma(z)$ of the point $z - \frac{1}{\eta}F(z)$ (this is a standard result about proximal mapping [PB14]) and thus we can write

$$p_{\gamma,\eta}(z) = \Pi_{Q_\gamma(z)}\left(z - \frac{1}{\eta}F(z)\right).$$

Second, we know that we can compute exactly and efficiently this projection by means of the ellipsoid method, as the set $Q_\gamma(z)$ is a polytope. The third observation is that, due to the fact that the set onto we project depends on z, even though the function $p_{\gamma,\eta}$ is continuous, it is easy to see that it is not Lipschitz continuous and thus we cannot reduce directly to Brouwer (differently, the function $p_{\gamma,\eta}$ would be trivially Lipschitz continuous if the set Q_γ would not depend on z). For this reason, we have to reduce to HighD-Sperner instead. This approach is not too dissimilar to the one used by [PVGZ23] as they also reduce to HighD-Sperner rather than to Brouwer for similar reasons.

However, the main challenges of our proof are essentially disjoint from Papadimitriou et al. [PVGZ23]. The first challenge is to prove that the correspondence Q is continuous (see Section D.2), while continuous correspondences are assumed in Papadimitriou et al. [PVGZ23]. Moreover, it is even more challenging to show that approximate solutions of the Brouwer-like problem, *i.e.*, (z, z') such that $z' = f(z)$ and $\|z - z'\|$ is small, can be used to build solutions of QuasiVI in polynomial time, *i.e.*, an exact fixed point of the correspondence. A similar result is not presented in [PVGZ23].

Construction of the Sperner's instance Consider an instance of QuasiVI in d dimensions, with approximation parameters $\nu, \epsilon > 0$ and Lipschitzness constant $L > 0$. We then build an instance of HighD-Sperner in which each cublet has side-length μ. To simplify computations, we define the following parameters that depend on η, μ, γ:

$$\ell := 2d\mu\left(\frac{4L\sqrt{d}}{\gamma} + L + \eta\right), \quad \kappa := \frac{2Ld\mu}{\gamma}, \quad \omega := \sqrt{d}\left(\mu + \kappa + \sqrt{\frac{2\ell}{\eta}}\right).$$

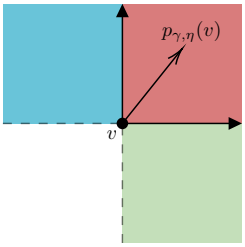


Figure 8: Example of Sperner's coloring induced by $p_{\gamma,\eta}(\cdot)$. The color **red** corresponds to color 0 while the colors **green** and **blue** correspond to colors 1 and 2, respectively.

Our construction works for any η, μ, γ that satisfy the following inequalities:

$$\gamma + \sqrt{d}\omega \le \nu, \quad \sqrt{d}(1+\eta)\omega + \frac{d\sqrt{d}\omega}{\gamma} \le \epsilon.$$

In Lemma D.12, we prove that there exists a choice for η, μ, γ in which all have a value bounded polynomially in ϵ, ν, L, d such that the above inequalities hold. Let $V \subset [0,1]^d$ be the set of vertices of the simplicial decomposition of the hypercube $[0,1]^d$. As is customary in reductions related to Sperner's lemma and fixed-point theorems, we construct the coloring circuit $\mathcal{C}$ by leveraging the fact that the function $p_{\gamma,\eta}$ maps the hypercube to itself. Specifically, the color $i = 0$ is assigned to a vertex $v \in V$ only if $(p_{\gamma,\eta}(v) - v)_i \ge 0$ for all

$i \in [d]$. Conversely, color $i \in [d]$ is assigned to v only if $(p_{\gamma,\eta}(v) - v)_i \leq 0$. Ties are broken arbitrarily as long as they do not violate Sperner's boundary conditions. Note that this circuit can be constructed in polynomial time using the polynomial-time algorithm that computes $p_{\gamma,\eta}(v)$. Figure 8 illustrates a valid Sperner coloring from the function $p_{\gamma,\eta}(\cdot)$ when $d = 2$. Moreover, since $p_{\gamma,\eta}$ maps the hypercube to itself, the boundary conditions for Sperner's coloring are automatically satisfied. Note that the instance of HIGHD-SPERNER is polynomial in the size of the QUASIVI instance. Indeed, $p_{\gamma,\eta}$ can be computed in polynomial time (by means of the ellipsoid algorithm), and the size of the input of the circuit C is $O(d \cdot \log(1/\mu))$.

From a solution to Sperner to a solution of QUASIVI Consider a solution $\{v^i\}_{i\in[d]\cup\{0\}}$ of HIGHD-SPERNER (without loss of generality we can assume that the vertices of the panchromatic simplex are numbered by their color, *i.e.* $\mathcal{C}(v^i) = i$ for all $i \in [d] \cup \{0\}$). The difficult part of the proof is to show that a solution to HIGHD-SPERNER can be used to build solutions to QUASIVI efficiently.

Let $z^i = p_{\gamma,\eta}(v^i)$ for each i. Then, by the definition of a solution to HIGHD-SPERNER it holds $(z^0 - v^0)_i \geq 0$ for all $i \in [d]$, and $(z^i - v^i)_i \leq 0$ for each $i \in [d]$. In the remaining of the proof, we show that v^0 is a solution to QUASIVI. We start by showing that all the z^i are close to z^0. This is not obvious due to the dependence of the set $Q_\gamma(v^i)$ on v^i. Our proof employs that the correspondence $Q_\gamma(\cdot)$ is Lipschitz with respect to the Hausdorff metric (Lemma D.5).

Lemma D.8. *For all $i \in [d]$ we have that $\|z^0 - z^i\| \leq \kappa + \sqrt{\frac{2\ell}{\eta}}$.*

Proof. For each $i \in [d]$, let $\bar{z}^i = \Pi_{Q_\gamma(v^0)}(z^i)$ be the projection on $Q_\gamma(v^0)$ of z^i. We bound first $\|z^i - \bar{z}^i\|$, then we bound $\|z^0 - \bar{z}^i\|$. Finally, we conclude the proof by a simple triangular inequality.

Before proceeding with the proof, we notice that for any two vertexes $v, v' \in \{v^i\}_{i\in[d]\cup\{0\}}$ it holds

$$\|v - v'\| \leq \sqrt{d}\mu \tag{12}$$

since both vertexes belong to the same cubelet.

▷ ***Bounding*** $\|z^i - \bar{z}^i\|$**.** First let us bound the difference between z^i and $\bar{z}^i := \Pi_{Q_\gamma(v^0)}(z^i)$. By Lemma D.5 and Definition D.1, we have that

$$\begin{aligned}
\|z^i - \bar{z}^i\| &= \|\Pi_{Q_\gamma(v^0)}(z^i) - z^i\| \\
&\leq \frac{2L\sqrt{d}}{\gamma}\|v^0 - v^i\| && (z^i \in Q_\gamma(v^i)) \\
&\leq \frac{2L\sqrt{d}}{\gamma}\mu\sqrt{d} \\
&= \frac{2Ld\mu}{\gamma} \\
&:= \kappa.
\end{aligned} \tag{13}$$

▷ ***Bounding*** $\|z^0 - \bar{z}^i\|$**.** We define the function $\Phi_\eta(\tilde{z}, z) := F(z)^\top \tilde{z} + \frac{\eta}{2}\|\tilde{z} - z\|^2$ and thus we can write

$$p_{\gamma,\eta}(z) = \Pi_{Q_\gamma(z)}(z) = \arg\min_{\tilde{z}\in Q_\gamma(z)} \Phi_\eta(\tilde{z}, z). \tag{14}$$

Now we can use the properties of strongly convex functions to bound the distance between $p_{\gamma,\eta}(z)$ and $\tilde{z}$ in terms of the residual error of $\Phi_\eta(\tilde{z}, z)$. Indeed, since the function $\tilde{z} \mapsto \Phi(\tilde{z}, z)$ is η-strongly convex we know that:

$$\|p_{\gamma,\eta}(z) - \tilde{z}\| \leq \sqrt{\frac{2}{\eta}(\Phi_\eta(\tilde{z}, z) - \Phi_\eta(p_{\gamma,\eta}(z), z))}. \tag{15}$$

Moreover, we observe that for all $z \in [0, 1]^d$ the function $\tilde{z} \mapsto \Phi_\eta(\tilde{z}, z)$ is $\sqrt{d}(1+\eta)$-Lipschitz as $\tilde{z} \mapsto F(z)^\top \tilde{z}$ is ($\|F(z)\| \leq \sqrt{d}$)-Lipschitz and $\tilde{z} \mapsto \|\tilde{z} - z\|^2$ is $2\sqrt{d}$-Lipschitz. Similarly the function $z \mapsto \Phi_\eta(\tilde{z}, z)$ is $\sqrt{d}(L + \eta)$-Lipschitz.

Consider now the following inequalities:

$$\begin{aligned}\Phi_\eta(\bar{z}^i, v^0) &\leq \Phi_\eta(z^i, v^0) + \sqrt{d}(1+\eta)\|\bar{z}^i - z^i\| \\ &\leq \Phi_\eta(z^i, v^i) + \sqrt{d}(1+\eta)\|\bar{z}^i - z^i\| + \sqrt{d}(L+\eta)\|v^i - v^0\| \\ &\leq \Phi_\eta(z^i, v^i) + \frac{2L\mu d^{3/2}}{\gamma}(1+\eta) + d(L+\eta)\mu, \end{aligned} \tag{16}$$

where in the first inequality we used the Lipschitzness of $\tilde{z} \mapsto \Phi_\eta(\tilde{z}, v^0)$, in the second inequality we used the Lipschitzness of $v \mapsto \Phi_\eta(z^i, v)$, and in the last inequality we used Equation (13) and Equation (12).

Define $\tilde{z}^i = \Pi_{Q_\gamma(v^i)}(z^0)$. Similarly to the first part of the proof, by Lemma D.5 and Definition D.1 we have

$$\|\tilde{z}^i - z^0\| \leq \frac{2L\sqrt{d}}{\gamma}\|v^i - v^0\| \leq \frac{2L\sqrt{d}}{\gamma}\mu\sqrt{d} = \frac{2Ld\mu}{\gamma}. \tag{17}$$

Now we combine Equation (16) and the observation that $\Phi_\eta(z^i, v^i) \leq \Phi_\eta(\tilde{z}^i, v^i)$ (since $\tilde{z}^i \in Q_\gamma(v^i)$ by construction) and we use again the Lipschitzness of Φ_η in both its argument to show that

$$\begin{aligned}\Phi_\eta(\bar{z}^i, v^0) &\leq \Phi_\eta(\tilde{z}^i, v^i) + \frac{2L\mu d^{3/2}}{\gamma}(1+\eta) + d(L+\eta)\mu \\ &\leq \Phi_\eta(\tilde{z}^i, v^0) + \frac{2L\mu d^{3/2}}{\gamma}(1+\eta) + 2d(L+\eta)\mu \\ &\leq \Phi_\eta(z^0, v^0) + \frac{4L\mu d^{3/2}}{\gamma}(1+\eta) + 2d(L+\eta)\mu \\ &\leq \Phi_\eta(z^0, v^0) + \ell.\end{aligned}$$

Now, in light of Equation (14), we can bound the distance between z^0 and $\bar{z}^i$ as follows:

$$\begin{aligned}\|z^0 - \bar{z}^i\| &= \|p_{\gamma,\eta}(v^0) - \bar{z}^i\| \\ &\leq \sqrt{\frac{2}{\eta}(\Phi_\eta(\bar{z}^i, v^0) - \Phi_\eta(\bar{z}^0, v^0))} \leq \sqrt{\frac{2\ell}{\eta}}\end{aligned}$$

This concludes the proof of the lemma.

□

We can exploit the previous lemma to show that v^0 is an approximate fixed-point of $p_{\gamma,\eta}(\cdot)$.

Lemma D.9. *It holds that* $\|z^0 - v^0\| \leq \omega$.

Proof. First, we recall that by the definition of a solution to HighD-Sperner, it holds $(z^0 - v^0)_i \geq 0$ for all $i \in [d]$, and $(z^i - v^i)_i \leq 0$ for each $i \in [d]$.

Then, take an $i \in [d]$ and consider the following chain of inequalities:

$$\begin{aligned}(z^0 - v^0)_i &\leq (z^i - v^0)_i + \kappa + \sqrt{\frac{2\ell}{\eta}} \\ &\leq (z^i - v^i)_i + \mu + \kappa + \sqrt{\frac{2\ell}{\eta}} \\ &\leq \mu + \kappa + \sqrt{\frac{2\ell}{\eta}},\end{aligned}$$

where in the first inequality we used again Lemma D.8, in the third one $|(v^i - v^0)_i| \leq \mu$, and the last one from $(z^i - v^i)_i \leq 0$ by the properties of the Sperner's coloring. Hence, $\|z^0 - x^0\| \leq \sqrt{d}\left(\mu + \kappa + \sqrt{\frac{2\ell}{\eta}}\right) := \omega$. □

Then, we show that v^0 belong to the set $Q_\nu(v^0)$.

Lemma D.10. *It holds $v_0 \in Q_\nu(v^0)$.*

Proof. We have that $z^0 = p_{\gamma,\eta}(v^0) \in Q_\gamma(v^0)$. Hence, by definition of the set $Q_\gamma(v^0)$, we have that $A(v^0)z^0 \le b(v^0) + \gamma 1_d$. This implies that

$$A(v^0)v^0 \le b(v^0) + \gamma 1_d + \sqrt{d}||z^0 - v^0||1_d \le b(v^0) + (\gamma + \sqrt{d}\omega)1_d.$$

Hence, $v^0 \in Q_{\gamma+\sqrt{d}\omega}(v^0)$. Since $\nu \ge \gamma + \sqrt{d}\omega$, this concludes the proof. □

Now, we can show that v_0 is also a approximate solution of $\arg\min_{z\in Q_\nu(v^0)} F(v^0)^\top z$.

Lemma D.11. *Let v^0 be the vertex of a panchromatic simplex of the* HighD-Sperner *instance colored with the color 0. Then,*

$$F(v^0)^\top v^0 \le \min_{z\in Q_\nu(v^0)} F(v^0)^\top z + \epsilon.$$

Proof. It holds

$$\begin{aligned}
F(v^0)^\top v^0 &= F(v^0)^\top v^0 + \frac{\eta}{2}\|v^0 - v^0\|^2 \\
&\le \Phi_\eta(z^0, v^0) + \sqrt{d}(1+\eta)\|v^0 - z^0\| && (\sqrt{d}(1+\eta)\text{-Lip. of } \Phi_\eta(\cdot, v^0)) \\
&= F(v^0)^\top z^0 + \frac{\eta}{2}\|z^0 - v^0\|^2 + \sqrt{d}(1+\eta)\|z^0 - v^0\| \\
&= \min_{z\in Q_\gamma(v^0)} \left\{F(v^0)^\top z + \frac{\eta}{2}\|z - v^0\|^2\right\} + \sqrt{d}(1+\eta)\|z^0 - v^0\| \\
&\le \min_{z\in Q_\gamma(v^0)} F(v^0)^\top z + \frac{\eta}{2}\|z^0 - v^0\|^2 + \sqrt{d}(1+\eta)\|z^0 - v^0\| \\
&\le \min_{z\in Q_\gamma(v^0)} F(v^0)^\top z + \frac{\eta}{2}\omega^2 + \sqrt{d}(1+\eta)\omega && (\text{Lemma D.9}) \\
&\le \min_{z\in Q_{\gamma+\sqrt{d}\omega}(v^0)} F(v^0)^\top z + \frac{\eta}{2}\omega^2 + \sqrt{d}(1+\eta)\omega + \frac{d\sqrt{d}\omega}{\gamma} && (\text{Lemma D.6}) \\
&\le \min_{z\in Q_\nu(v^0)} F(v^0)^\top z + \epsilon.
\end{aligned}$$

where in the last inequality we use $\gamma + \sqrt{d}\omega \le \nu$, $\frac{\eta}{2}\omega^2 + \sqrt{d}(1+\eta)\omega + \frac{d\nu}{\gamma} \le \epsilon$, and that $Q_{\gamma+\sqrt{d}\omega}(v^0) \subseteq Q_\nu(v^0)$. □

To conclude the proof and show that v^0 is a solution to QuasiVI, we only need to observe that:

$$F(v^0)^\top(z - v^0) \ge F(v^0)^\top(v^0 - v^0) - \epsilon = -\epsilon \quad \forall z \in Q_\nu(v^0),$$

where the first inequality follows from Lemma D.11.

We are only left to show that there exists a set of sufficiently large parameters η, μ and γ that satisfy the required inequalities.

Lemma D.12. *For any given $d \in \mathbb{N}$, any $\epsilon, \nu, L > 0$ and any*

$$\eta \le \min\left(1, L, d^{1/4}, \sqrt{\frac{\nu}{2}}, \sqrt{2dL}, \frac{1}{L^{1/8}}, \frac{1}{8d^2\sqrt{2L}}, \frac{1}{2(d\sqrt{2dL})^{1/3}}, \frac{\epsilon}{24d^3\sqrt{2dL}}\right),$$

let $\mu = \eta^9, \gamma = \eta^2$, and let ℓ, κ, and ω be defined as above. Then:

$$\gamma + \sqrt{d}\omega \le \nu, \quad \frac{\eta}{2}\omega^2 + \sqrt{d}(1+\eta)\omega + \frac{d\sqrt{d}\omega}{\gamma} \le \epsilon$$

Proof. It is easy to see that $\kappa = 2Ld\eta^7$ and that

$$\begin{aligned}\ell &= 2d\eta^9\left(\frac{4L\sqrt{d}}{\eta^2} + L + \eta\right)\\ &\leq 4d\eta^9 L\left(\frac{2\sqrt{d}}{\eta^2} + 1\right)\\ &\leq 16\eta^7 d^2 L,\end{aligned}$$

where in the first inequality we used that $\eta \leq L$ and in the second that $\eta \leq d^{1/4}$. Plugging this upper bounds into ω we obtain:

$$\begin{aligned}\omega =& \sqrt{d}\left(\mu + \kappa + \sqrt{\frac{2\ell}{\eta}}\right)\\ &= \sqrt{d}\left(\eta^9 + 2\eta^7 dL + 4\eta^3 d\sqrt{2L}\right)\\ &\leq \sqrt{d}\left(4\eta^7 dL + 4\eta^3 d\sqrt{2L}\right)\\ &\leq 8\eta^3 d\sqrt{2dL},\end{aligned}$$

where in the first inequality we used the upper bound on ℓ computed above, in the second inequality we used that $\eta \leq \sqrt{2dL}$, the third that $\eta \leq \frac{1}{L^{1/8}}$. Now the first inequality is readily satisfied:

$$\begin{aligned}\gamma + \sqrt{d}\omega &\leq \eta^2 + 8\eta^3 d^2\sqrt{2L}\\ &\leq 2\eta^2\\ &\leq \nu,\end{aligned}$$

where the first inequality holds by the definition of γ and the upper bound on ω computed above, the second follows from the fact that $\eta \leq \frac{1}{8d^2\sqrt{2L}}$ and the last from $\eta \leq \sqrt{\nu/2}$. Moreover, the second inequality is also satisfied. First observe that for $\eta \leq \frac{1}{2(d\sqrt{2dL})^{1/3}}$ then $\omega \leq 1$ and

$$\begin{aligned}\frac{\eta}{2}\omega^2 + \sqrt{d}(1+\eta)\omega + \frac{d\sqrt{d}\omega}{\gamma} &\leq \omega\left(\frac{\eta}{2} + 2\sqrt{d} + \frac{d\sqrt{d}}{\gamma}\right)\\ &= \omega\left(\frac{\eta}{2} + 2\sqrt{d} + \frac{d\sqrt{d}}{\eta^2}\right)\\ &\leq 3\omega\frac{d^2}{\eta^2}\\ &\leq 24\eta d^3\sqrt{2dL},\end{aligned}$$

where first inequality comes from $\omega \leq 1$, the second from $\eta \leq d^{1/4}$, the third from the upper bound on ω and $d \geq 1$, while the last comes from $\eta \leq \frac{\epsilon}{24d^3\sqrt{2dL}}$. This concludes the proof of the lemma. □

D.4 Proof of Proposition B.5

Proposition B.5. *For any $(x^\star, y^\star)$ which is an (ϵ, δ, ν)-solution to* LocalMinMax*, if the function $f(\cdot, y)$ is convex for any $y \in [0,1]^d$, then $f(x^\star, y^\star) \leq f(x', y^\star) + \frac{\epsilon\sqrt{d}}{\delta} \quad \forall x' \in K_1^\nu(y^\star)$.*

Proof. The proof is immediate by convexity. Take any $x' \in K_1^\nu(y)$ and consider $x'' = x^\star + \frac{\delta}{\sqrt{d}}(x' - x^\star)$. Note that $\|x'' - x^\star\| = \frac{\delta}{\sqrt{d}}\|x' - x^\star\| \leq \delta$. Since $(x^\star, y^\star)$ is a (ϵ, δ, ν) solution to LocalMinMax, and since $f(\cdot, y^\star)$ is convex, it holds:

$$f(x^\star, y^\star) \leq f(x'', y^\star) + \epsilon \leq \left(1 - \frac{\delta}{\sqrt{d}}\right) f(x^\star, y^\star) + \frac{\delta}{\sqrt{d}} f(x', y^\star) + \epsilon.$$

Rearranging we obtain $f(x^\star, y^\star) \le f(x', y^\star) + \frac{\epsilon\sqrt{d}}{\delta}$, as desired. □

E Omitted Proofs for Section 3

Theorem 3.4. *There exists a constant ρ^* such that finding an ρ^*-approximation to* LinVI *is* PPAD*-complete. Moreover, the same result holds even when restricting to deviations along a single component and restricting to $\|D\|_1 \le 1$ and $\|D\|_\infty \le 1$, i.e., finding a z such that*

$$(Dz + c)_i \cdot (z'_i - z_i) \ge -\rho^* \quad \forall i \in [d],\ z'_i \in [0, 1].$$

Proof. First, we show the membership of LinVI in PPAD by reduction to Brouwer. Then we show hardenss by reducing from PolyMatrix.

▷ ***Membership.*** We reduce the problem to Brouwer. Let ρ, D and c be the inputs defining the LinVI instance. Let $K = [0, 1]^d$. Consider the function $F(z) := \Pi_K(z - (Dz + c))$. Since the set K is a polytope with rational coefficients, we can compute the $F(z)$ efficiently. Also note that $F : K \to K$ and it is $(1 + d)$-Lipschitz. We feed F to Brouwer with approximation $\alpha := \frac{\rho}{2\sqrt{d}}$. Thus it admits a fixed point $\|z - F(z)\| \le \alpha$. By Theorem C.5 we know that for all $z' \in K$ we have

$$(Dz + c)^\top (z' - z) \ge -2\alpha\sqrt{d} = -\rho,$$

thus proving it is a solution to LinVI.

▷ ***Hardness.*** We reduce from PolyMatrix with two actions, graphs with degree 3, and approximation $\epsilon^\star$. Recall that we formally define PolyMatrix in Problem 14. Polymatrix games with binary actions are defined by an undirected graph (V, E), where each vertex represents a player. We use $n = |V|$ to denote the number of players in the game. Each player $i \in V$ has two actions a_1 and a_2. For each edge $(i, j) \in E$, there is an 2×2 matrix $A_{i,j} \in [0, 1]^{2\times 2}$ giving the payoffs that player i obtains from their interaction with player j. Formally, we let

$$A_{i,j} = \begin{bmatrix} a_{i,j} & b_{i,j} \\ c_{i,j} & d_{i,j} \end{bmatrix}.$$

Similarly, a 2×2 matrix $A_{j,i}$ gives payoffs for player j's interaction with player i.

We represent a mixed strategy of player i as $x_i \in [0, 1]$, where x_i is the probability of playing action a_1 and $1 - x_i$ the probability of playing action a_2. Then, the utility of player i is given by

$$\begin{aligned} U_i(x_i, x_{-i}) &= \sum_{j:(i,j)\in E} (x_i x_j a_{i,j} + x_i(1 - x_j)b_{i,j} + (1 - x_i)x_j b_{i,j} + d_{i,j}) \\ &= x_i \cdot \left(\sum_{j:(i,j)\in E} (x_j \alpha_{i,j} + \beta_{i,j}) \right) + \sum_{j:(i,j)\in E} (x_j \gamma_{i,j} + \delta_{i,j}), \end{aligned}$$

where $\alpha_{i,j} = a_{i,j} - b_{i,j} + c_{i,j} - d_{i,j}$, $\beta_{i,j} = d_{i,j} - c_{i,j}$, $\gamma_{i,j} = b_{i,j} - c_{i,j}$ and $\delta_{i,j} = c_{i,j}$.

Then, the goal is to find an $\epsilon^\star$-approximate solution to PolyMatrix, *i.e.*, an $x \in [0, 1]^d$ that for each $i \in V$ satisfies

$$U_i(x_i, x_{-i}) \ge U_i(x'_i, x_{-i}) - \epsilon^\star \quad \forall x'_i \in [0, 1].$$

As we recalled in Proposition A.1, there exists a constant $\epsilon^\star$ such that polymatrix is PPAD-hard even for graphs with degree 3 [Rub15, Theorem 1], [Del+24, Theorem 4.4].[21]

Consider a polymatrix game with graph (V, E) and matrices $A_{i,j}$. Without loss of generality, we identify the set of vertices V with $[n]$. Then, $x \in [0, 1]^n$ represents the joint strategy for all players.

[21] Notice that we do not try to optimize constants. For example, we could exploit that the payoffs of polymatrix games we reduce from are normalized.

Notice that $U_i(x_i, x_{-i}) \geq U_i(x_i', x_{-i}) - \epsilon^\star$ if and only if

$$x_i \cdot \left(\sum_{j:(i,j)\in E} (x_j \alpha_{i,j} + \beta_{i,j}) \right) \geq x_i' \cdot \left(\sum_{j:(i,j)\in E} (x_j \alpha_{i,j} + \beta_{i,j}) \right) - \epsilon^\star.$$

Now we define the instance of LinVI. We define $d = n$ variables z_i for each $i \in [n]$. The approximation factor is $\rho^\star = \frac{\epsilon^\star}{6}$. The linear operator is defined by a matrix $D \in [-1,1]^{d\times d}$

$$D_{i,j} = -\frac{\alpha_{i,j}\mathbb{I}((i,j)\in E)}{6}$$

and a vector $c \in [-1,1]^d$

$$c_i = -\frac{\sum_{j:(i,j)\in E} \beta_{i,j}}{6}.$$

Notice that since the graph has degree 3 and $\beta_{i,j} \in [-1,1]$, D and c have every entry bounded in $[-1,1]$. Also, the bounded degree of the graph implies that the sum of the absolute values on each row and each column is bounded by 1, *i.e.* $\|D\|_\infty, \|D\|_1 \leq 1$.

Note that the i-th component of $Dz + c$ is equal to

$$\frac{1}{2d}\left[\sum_{j:(i,j)\in E} z_j \alpha_{i,j} + \sum_{j:(i,j)\in E} \beta_{i,j} \right]. \tag{18}$$

Now any solution $z \in K$ to LinVI guarantees that

$$(Dz + c)^\top (z' - z) \geq -\rho^\star \quad \forall z' \in [0,1]^d.$$

Our reduction holds even for a weaker requirement. In particular, it is sufficient to consider deviation on a single component. Formally, we show that any $z \in K$ such that

$$(Dz + c)_i \cdot (z_i' - z_i) \geq -\rho^\star \quad \forall i \in [d], z_i' \in [0,1]$$

Indeed, this clearly implies that for each $i \in [d]$

$$z_i'(Dz + c)_i \geq z_i(Dz + c)_i - \rho^\star \quad \forall z_i' \in [0,1].$$

Using Equation (18) we get that the previous inequality implies that for all $i \in [d]$

$$z_i' \left(\sum_{j:(i,j)\in E} z_j \alpha_{i,j} + \sum_{j:(i,j)\in E} \beta_{i,j} \right) \leq z_i \left(\sum_{j:(i,j)\in E} z_j \alpha_{i,j} + \sum_{j:(i,j)\in E} \beta_{i,j} \right) + 6\rho^\star \quad \forall z_i' \in [0,1],$$

showing that z is a solutions to polymatrix instance since $\rho^\star = \frac{\epsilon^\star}{6}$. □

Lemma 3.5. *Assume that for all $y \in [0,1]^d$ there exist d intervals $\{K_{1,j}(y_j)\}_{j\in[d]}$ such that $K_1(y) = \bigtimes_{j\in[d]} K_{1,j}(y_j)$ and d intervals $\{K_{2,j}(x_j)\}_{j\in[d]}$ such that $K_2(x) = \bigtimes_{j\in[d]} K_{2,j}(x_j)$. Let $(x^\star, y^\star) \in K_1(y^\star) \times K_2(x^\star)$ satisfy*

$$\|(x^\star, y^\star) - F^\nu_{\mathrm{GDA}}(x^\star, y^\star)\| \leq \alpha$$

Then, for all $i \in [d]$ and all single-component deviation $x' = (x_i', x^\star_{-i}), y' = (y_i', y^\star)$ it holds

$$f(x^\star, y^\star) \leq f(x', y^\star) + \alpha(1 + G_\infty) + \frac{L}{2}\delta^2 \quad \forall x' \in K_1^\nu(y^\star) \quad \textit{and} \quad \|x' - x^\star\| \leq \delta,$$
$$f(x^\star, y^\star) \geq f(x^\star, y') - \alpha(1 + G_\infty) - \frac{L}{2}\delta^2 \quad \forall y' \in K_2^\nu(x^\star) \quad \textit{and} \quad \|y' - y^\star\| \leq \delta,$$

where $G_\infty = \sup_{x,y\in[0,1]^d} \|\nabla f(x,y)\|_\infty$.

The statement holds even if we only have $\|(x^\star, y^\star) - F^\nu_{\mathrm{GDA}}(x^\star, y^\star)\|_\infty \leq \alpha$.

Proof. We prove the result for the x player as symmetric arguments hold for the y player. Consider the projection $\Pi_{K_1^\nu(y^\star)}(x^\star - \nabla_x f(x^\star, y^\star))$ and any $i \in [d]$. Since $K_1(y) = \bigtimes_{j\in[d]} K_{1,j}(y_j)$ we have that the projection is independent among components [BV04, Section 8.1.1]. In particular, its i-th component is equal to

$$\Pi_{K_{1,i}^\nu(y_i^\star)}(x_i^\star - \partial_{x_i} f(x^\star, y^\star)).$$

Thus, since $\|(x^\star, y^\star) - F_{\mathrm{GDA}}^\nu(x^\star, y^\star)\| \le \alpha$ we have that

$$|x_i^\star - \Pi_{K_{1,i}^\nu(y_i^\star)}(x_i^\star - \partial_{x_i} f(x^\star, y^\star))| \le \alpha.$$

Now only have to consider single-dimensional problems and it is easy to see that this implies that either $|\partial_{x_i} f(x^\star, y^\star)| \le \alpha$ (when no projection incurs) or $d(x_i^\star, \partial K_{1,i}^\nu(y_i^\star)) \le \alpha$, where we denote with $\partial K_{1,i}^\nu(y_i^\star)$ the boundary of the set $K_{1,i}^\nu(y_i^\star)$. Notice that in the second case $x_i^\star - \partial_{x_i} f(x^\star, y^\star)$ gets projected on the boundary of the interval $K_{1,j}^\nu(y_j))$. Hence, the distance of $x_i^\star$ from the boundary of $K_{1,i}^\nu(y_i^\star)$ must be at most α. Now we bound $\partial_{x_i} f(x^\star, y^\star)(x_i^\star - x_i')$ with different arguments depending on the case.

▷ ***Case 1:*** $|\partial_{x_i} f(x^\star, y^\star)| \le \alpha$**.** In this case it is trivial to observe that $\partial_{x_i} f(x^\star, y^\star)(x_i^\star - x_i') \le |\partial_{x_i} f(x^\star, y^\star)| \cdot |x_i^\star - x_i'| \le \alpha$, since by assumption $|\partial_{x_i} f(x^\star, y^\star)| \le \alpha$ and $x_i, x_i' \in [0, 1]$.

▷ ***Case 2:*** $d(x_i^\star, \partial K_{1,i}^\nu(y_i^\star)) \le \alpha$**.** We denote the interval $K_{1,i}^\nu(y_i^\star))$ with $[a, b]$. We now consider two sub-cases

- $\partial_{x_i} f(x^\star, y^\star) \ge 0$. Then this means that $a \le x_i^\star \le a+\alpha$. If $x_i' \in [a, x_i^\star]$ then $\partial_{x_i} f(x^\star, y^\star)(x_i^\star - x_i') \le \alpha \cdot G_\infty$. If $x_i' > x_i^\star$ then $\partial_{x_i} f(x^\star, y^\star)(x_i^\star - x_i') \le 0$.
- $\partial_{x_i} f(x^\star, y^\star) < 0$. This means that $b - \alpha \le x_i^\star \le b$. If $x_i' \in [x_i^\star, b]$ then $\partial_{x_i} f(x^\star, y^\star)(x_i^\star - x_i') \le \alpha \cdot G_\infty$. if $x_i' < x_i^\star$ then $\partial_{x_i} f(x^\star, y^\star)(x_i^\star - x_i') \le 0$.

In conclusion we have that $\partial_{x_i} f(x^\star, y^\star)(x_i^\star - x_i') \le \max(\alpha, \alpha G_\infty) \le \alpha(1 + G_\infty)$.

Now a simple smoothness argument finishes the proof. Indeed L-smoothness of f implies that

$$f(x^\star, y^\star) \le f(x', y^\star) \le \partial_{x_i} f(x^\star, y^\star)(x_i^\star - x_i') + \frac{L}{2}\|x^\star - x'\|^2 \le \alpha(1 + G_\infty) + \frac{L}{2}\delta^2,$$

concluding the proof. □

E.1 Proof of Theorem 3.1

Construction of the hard instance Take any LinVI instance defined by the matrix $D \in [-1, 1]^{d\times d}$, a vector $c \in [-1, 1]^d$ and approximation $\alpha_{\mathrm{LinVI}} = \rho^*$, where ρ^* is the constant in Theorem 3.4. We showed that this problem is PPAD-hard in Theorem 3.4.

Objective function The objective function $f : \mathbb{R}^d \times \mathbb{R}^d \to [-B, B]$ is given by

$$f(x, y) = (x - y)^\top (Dx + c).$$

We observe a few facts about the function f. This function is linear in y and nonconvex quadratic in x, and hence it is a nonconvex-concave function. Moreover, the function f is a second-degree polynomial in x and y. Finally, since in Theorem 3.4 we can restrict to D such that $\|D\|_1, \|D\|_\infty \le 1$, the objective satisfies the following smoothness conditions.

Lemma E.1. *The function f is G-Lipschitz and L-smooth with $G = 5\sqrt{d}$ and $L = 7$.*

Proof. We recall that $\nabla_x f(x, y) = Dx + c + D^\top(x - y)$ and $\nabla_y f(x, y) = -(Dx + c)$. We bound the Lipschitz parameter of f as the supremum of the norm of the gradient. For all $(x, y) \in [0, 1]^{2d}$ we have

$$\begin{aligned}
\|\nabla f(x, y)\|^2 &= \|Dx + c + D^\top(x - y)\|^2 + \|Dx + c\|^2 \\
&\le 6\|Dx\|^2 + 6\|c\|^2 + 2\|D^\top(x - y)\|^2 && \text{(parallelogram inequality)} \\
&\le 6\|x\|^2 + 6d + 2\|x - y\|^2 && \text{(Lemma F.2)} \\
&\le 20d, && \text{(since } \|x\|_2, \|y\|_2 \le \sqrt{d})
\end{aligned}$$

where in the last inequality we exploit that $\|D\| \leq \sqrt{\|D\|_1 \|D\|_\infty}$ (Lemma F.2) and that $\|D\|, \|D^\top\| \leq 1$. Taking the square root on both sides, we conclude that the function f is G-Lipschitz with $L \leq 5\sqrt{d}$. We now focus on the smoothness parameter of the function (*i.e.*, the Lipschitz parameter of the function's gradient).

Then, it holds

$$\begin{aligned}\|\nabla f(x,y) - \nabla f(x',y')\|^2 &= \|D(x-x') + D^\top(x-x'-(y-y'))\|^2 + \|D(x-x')\|^2 \\ &\leq 3\|D(x-x')\|^2 + 2\|D^\top(x-x'-(y-y'))\|^2 \\ &\leq 3\|x-x'\|^2 + 4(\|x-x'\|^2 + \|y-y'\|^2) \\ &\leq 7\|(x,y)-(x',y')\|^2,\end{aligned}$$

where we recall that $\|D\| \leq 1$. This proves that $L \leq 7$ and concludes the proof. □

Constraints We define the constraint function g by the induced set $K = \{(x,y) \in [0,1]^{2d} : \|x-y\|_\infty \leq \Delta\}$, where $\Delta = \frac{\rho^*}{4}$. It is easy to see that there exists a function g that implements such set.

In the following, we first prove the hardness result for LR-LocalMinMax-JC. Then, we exploit the reduction from LR-LocalMinMax-JC to GDAFixedPoint-JC in Lemma 3.5 to extend the hardness result to the problem of finding fixed points of gradient ascent-descent dynamics. For both settings, restricting to single-dimensional deviations will be essential to provide tighter hardness results.

From LR-LocalMinMax-JC to LinVI We provide a reduction for any $0 < \gamma \leq 1$, $\delta = \gamma\frac{\rho^*}{15}$, and $\epsilon = \gamma\frac{\rho^{*,2}}{60}$. It is easy to see that $\delta < \sqrt{\frac{2\epsilon}{L}}$ for any $\gamma \in (0,1]$ and thus we are in the local regime. Now we want to prove that any solution to LR-LocalMinMax-JC $(x^\star, y^\star)$ is a solution to LinVI. To obtain a stronger hardness result, we focus on the specialized (but still PPAD-Hard as proved in Theorem 3.4) problem of finding an $x^\star$ that is robust to single-dimensional deviations, *i.e.*

$$(Dx^\star + c)_i \cdot (x'_i - x^\star_i) \geq -\rho^* \quad \forall i \in [d],\, x'_i \in [0,1]. \tag{19}$$

For each $i \in [d]$ and $x'_i \in [0,1]$, we consider two cases depending on whether or not the deviation is *inside* in the segment $(x^\star_i, y^\star_i)$, *i.e.*, $x'_i \in [\min(x^\star_i, y^\star_i), \max(x^\star_i, y^\star_i)]$. Figure 9 illustrates the different cases.

▷ ***Case* 1.** We start from the case in which $x'_i \in [\min(x^\star_i, y^\star_i), \max(x^\star_i, y^\star_i)]$. Then,

$$(Dx^\star + c)_i \cdot (x'_i - x^\star_i) \geq -|(Dx^\star + c)_i| \cdot |x'_i - x^\star_i| \geq -2\Delta \geq -\rho^*.$$

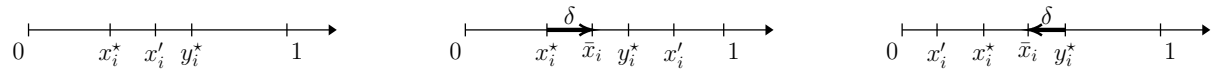
▷ ***Case* 2.** Conversely, we consider the case $x'_i \notin [\min(x^\star_i, y^\star_i), \max(x^\star_i, y^\star_i)]$. We denote by $Q_i = \mathbb{I}(|x'_i - x^\star_i| \geq |x'_i - y^*_i|)$ the indicator of whether $x^\star_i$ is closer to $y^\star_i$ than x'_i ($Q_i = 1$) or not ($Q_i = 0$). Then, we define the deviations

$$\begin{cases} \bar{x}_i = \delta Q_i (x'_i - x^\star_i) + x^\star_i \\ \bar{x}_j = x^\star_j \qquad\qquad\qquad \forall j \neq i, \end{cases} \quad \text{and} \quad \begin{cases} \bar{y}_i = \delta(1-Q_i)(x'_i - x^\star_i) + y^\star_i \\ \bar{y}_j = y^\star_j \qquad\qquad\qquad \forall j \neq i. \end{cases}$$

Intuitively, we move the point farther from x' ($x^\star$ or $y^\star$) closer to it by an amount $\delta(x'_i - x^\star_i)$. Few considerations about $\bar{x}$ and $\bar{y}$ are in order. First, recall that x'_i is outside of the interval $(x^\star_i, y^\star_i)$ by hypothesis. Hence, when $Q_i = 1$, it must be the case that $x^\star_i - y^\star_i$ and $x'_i - x^\star_i$ have opposite sign. Similarly, when $Q_i = 0$, it is the case that $y^\star_i - x^\star_i$ and $x'_i - x^\star_i$ have opposite sign. Finally, it is also easy to see that $\|\bar{x} - x^\star\| \leq \delta$ and $\|\bar{y} - y^\star\| \leq \delta$.

We now show that $\bar{x}$ and $\bar{y}$ are both in $[0,1]^d$ and $\|\bar{y} - x^\star\|_\infty, \|\bar{x} - y^\star\|_\infty \leq \Delta$. We break the analysis into two cases. When $Q_i = 1$, then $\bar{y} = y^\star$ and thus $\bar{y} \in [0,1]^d$ and $\|\bar{y} - x^\star\|_\infty \leq \Delta$ are trivially satisfied. Moreover, $\bar{x}$ is a convex combination of $x^\star$ and x', and thus it is guaranteed that $\bar{x} \in [0,1]^d$. Furthermore, since $\bar{x}$ is obtained by changing the i-th component of $x^\star$, we have

$$\|\bar{x} - y^\star\|_\infty \leq \max(\|x^\star - y^\star\|_\infty, |\bar{x}_i - y^\star_i|).$$



(a) Case 1 (b) Case 2. $Q_i = 0$ (c) Case 2. $Q_i = 1$

Figure 9: Illustration of the case analysis. In Figure 9a we have $x_i' \in [x_i^\star, y_i^\star]$. While in Figure 9b and Figure 9c we are in the setting in which $x_i' \notin [\min(x_i^\star, y_i^\star), \max(x_i^\star, y_i^\star)]$, where $Q_i = 0$ and $Q_i = 1$, respectively.

We argue that $|\bar{x}_i - y_i^\star|$ is smaller then Δ. Indeed we can rewrite $|\bar{x}_i - y_i^\star|$ as $|\delta(x_i' - x_i^\star) + x_i^\star - y_i^\star|$. Then, we notice that $\delta(x_i' - x_i^\star)$ and $x_i^\star - y_i^\star$ have different sign and are both smaller than Δ since $\delta \leq \Delta$. Hence, we can use that $\|a + b\| \leq \max(|a|, |b|)$ for any a and b with different signs. This shows that $|\bar{x}_i - y_i^\star| \leq \Delta$.

Conversely, when $Q_i = 0$, then $\bar{x} = x^\star$ and thus $\bar{x} \in [0,1]^d$ and $\|\bar{x} - y^\star\|_\infty \leq \Delta$ are trivially satisfied. Again we can write

$$\|\bar{y} - x^\star\|_\infty \leq \max(\|x^\star - y^\star\|_\infty, |\bar{y}_i - x_i^\star|),$$

and by a similar argument as before $|\bar{y}_i - x_i^\star| \leq \Delta$. Now we are left to check that $\bar{y} \in [0,1]^d$. If $y_i^\star \geq x_i^\star$ then $x_i^\star \geq x_i'$. Thus

$$\bar{y}_i = \delta(x_i' - x_i^\star) + y_i^\star \geq (1-\delta)x_i^\star + \delta x_i' \geq 0,$$

and $\bar{y}_i \leq y_i^\star \leq 1$. Similarly, if $y_i^\star \leq x_i^\star$ then $x_i^\star \leq x_i'$. This clearly implies that $\bar{y}_i \geq y_i^\star \geq 0$ and $\delta(x_i' - x_i^\star) + y_i^\star \leq (1-\delta)x_i^\star + \delta x_i' \leq 1$. This shows that $\bar{y} \in [0,1]^d$. Thus we can conclude that both $\bar{y}$ and $\bar{x}$ are valid deviations.

Since $\bar{y}$ and $\bar{x}$ are valid local deviations and $(x^\star, y^\star)$ satisfies the local optimality condition, it holds

$$(x^\star - y^\star)^\top (Dx^\star + c) \leq (\bar{x} - y^\star)^\top (D\bar{x} + c) + \epsilon. \tag{20}$$

and

$$(x^\star - y^\star)^\top (Dx^\star + c) \geq (x^\star - \bar{y})^\top (Dx^\star + c) - \epsilon \tag{21}$$

In the next step of the proof, we "globalize" such optimality condition, showing that the deviation to x' only slightly increases the value of the VI.

We start from Equation (20). Using the fact that, by construction, $\bar{x} = x^\star + \kappa \cdot e_i$, where $\kappa = \delta Q_i(x_i' - x_i^\star)$, and using the fact that $|\kappa| \leq \delta$, we can therefore write

$$\begin{aligned}
&(\bar{x} - y^\star)^\top (D\bar{x} + c) \\
&\qquad = (x^\star + \kappa \cdot e_i - y^\star)^\top (D(x^\star + \kappa \cdot e_i) + c) \\
&\qquad = (x^\star - y^\star)^\top (Dx^\star + c) + \kappa e_i^\top (Dx^\star + c) + \kappa (x^\star + \kappa e_i - y^\star)^\top (De_i) \\
&\qquad \leq (x^\star - y^\star)^\top (Dx^\star + c) + \kappa e_i^\top (Dx^\star + c) + \delta \|x^\star - y^\star\|_\infty \|D\|_1 + \delta^2 |D_{i,i}| \\
&\qquad \leq (x^\star - y^\star)^\top (Dx^\star + c) + \kappa e_i^\top (Dx^\star + c) + \delta\Delta + \delta^2 \\
&\qquad \leq (x^\star - y^\star)^\top (Dx^\star + c) + \kappa e_i^\top (Dx^\star + c) + 2\delta\Delta. && (\delta \leq \Delta)
\end{aligned}$$

Hence, combined with Equation (20) we obtain

$$\kappa e_i^\top (Dx^\star + c) = (Dx^\star + c)^\top (\bar{x} - x^\star) \geq -\epsilon - 2\delta\Delta. \tag{22}$$

Then, we globalize the equilibrium condition of player y. Equation (21) directly implies that

$$(Dx^\star + c)^\top (\bar{y} - y^\star) \geq -\epsilon. \tag{23}$$

Summing Equation (22) and Equation (23) we obtain

$$(Dx^\star + c)^\top (\bar{x} + \bar{y} - x^\star - y^\star) \geq -2\epsilon - 2\delta\Delta. \tag{24}$$

Notice that either $\bar{x}_i \neq x_i^\star$ or $\bar{y}_i \neq y_i^\star$ and hence we could consider only one of the inequalities depending on Q_i. However, considering the two together simplifies the proof.

Finally, we observe that by combining $\bar{x} - x^\star$ and $\bar{y} - y^\star$ we implement the vector $\delta(x'_i - x_i^\star)e_i$. Formally,

$$\bar{x} + \bar{y} = x^\star + y^\star + \delta(x'_i - x_i^\star)e_i, \tag{25}$$

where e_i is the i-th standard basis vector. Hence, the joint deviation implements a deviation from $x^\star$ toward x'.

Thus, combining Equation (25) and Equation (24) we obtain that

$$(Dx^\star + c)_i(x' - x_i^\star) \geq -2\left(\frac{\epsilon}{\delta} + \Delta\right) \geq -\rho^*,$$

concluding the proof.

From GDAFixedPoint-JC to LR-LocalMinMax-JC The relationship between fixed point of gradient descent-ascent dynamics and local min-max of Daskalakis et al. [DSZ21] (Theorem 2.1) is not enough to obtain constant inapproximability of GDAFixedPoint-JC in ℓ_∞. This is because we could not design a function f with constant Lipschitzness G. However, since we proved that LR-LocalMinMax-JC is hard even with single component deviations (or equivalently "norm 1" deviations), we can use the specialized reduction that we show in Lemma 3.5.

Now, with this result, we can prove hardness for GDAFixedPoint-JC by a reduction from LR-LocalMinMax-JC. Take the same instance of LR-LocalMinMax-JC built at the beginning of Section E.1. It is easy to check that K is defined such that $K_1(y)$ and $K_2(x)$ satisfy the conditions of Lemma 3.5 and that $G_\infty \leq 3$. Thus, we can use Lemma 3.5 with $\nu = 0$, and $\alpha = \rho^{*,2}/120$. This implies that any α fixed point to GDAFixedPoint-JC, is a (δ, ϵ)-solution to LR-LocalMinMax-JC, where $\delta = \rho^*/15$ and $\epsilon = \rho^{*,2}/60$. Thus, GDAFixedPoint-JC is PPAD-hard for any $\alpha < \rho^{*,2}/120$, where the function f is $O(1)$-smooth, $G = O(\sqrt{d})$-Lipschitz, and nonconvex-concave (quadratic-linear). In Section 3.2 we discuss how this result can be exploited to prove constant inapproximability of GDAFixedPoint-JC in ℓ_∞-norm.

E.2 Proof of Theorem 3.2

Take any LinVI instance defined by the matrix $D \in [-1, 1]^{d \times d}$, a vector $c \in [-1, 1]^d$ and approximation $\alpha_{\text{LinVI}} = \rho^*$, where ρ^* is the constant in Theorem 3.4. We showed that this problem is PPAD-hard in Theorem 3.4.

Objective function We define the objective function $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ as $f(x, y) = x^\top(Dy + c)$.

Similarly to Theorem 3.1, we first prove some properties about the Lipschitzness and smoothness of the function. Recall that we can restrict to D such that $\|D\|_1, \|D\|_\infty \leq 1$.

Lemma E.2. *The function f is G-Lipschitz and L-smooth with $G = 3\sqrt{d}$ and $L = 1$.*

Proof. We recall that $\nabla_x f(x, y) = Dy + c$ and $\nabla_y f(x, y) = D^\top x$. We bound the Lipschitz parameter of f as the supremum of the norm of the gradient. For all $(x, y) \in [0, 1]^{2d}$ we have

$$\begin{aligned}
\|\nabla f(x, y)\|^2 &= \|Dy + c + \|^2 + \|D^\top x\|^2 \\
&\leq 2\|Dx\|^2 + 2\|c\|^2 + 2\|D^\top x\|^2 && \text{(Parallelogram inequality)} \\
&\leq 2\|x\|^2 + 2d + 2\|x\|^2 && \text{(Lemma F.2)} \\
&\leq 5d. && \text{(Since } \|x\|_2, \|y\|_2 \leq \sqrt{d}\text{)}
\end{aligned}$$

Taking a square root, we conclude that the function f is G-Lipschitz with $L \leq 3\sqrt{d}$. We now focus on the smoothness parameter of the function (*i.e.*, the Lipschitz parameter of the function's gradient). We have

$$\begin{aligned}\|\nabla f(x,y) - \nabla f(x',y')\|^2 &= \|D(y-y')\|^2 + \|D^\top(x-x')\|^2 \\ &\leq \|x-x'\|^2 + \|y-y'\|^2 \qquad \text{(Lemma F.2)} \\ &\leq \|(x,y)-(x',y')\|^2.\end{aligned}$$

This proves that $L \leq 1$ and concludes the proof. □

Constraints The constraint functions are given by $g_2(x,y) := \|x-y\|_\infty \leq 0$, while player 1 is unconstrained. It is easy to see that $K_1(y)$ and $K_2(x)$ are non empty for each $(x,y) \in [0,1]^d$ and that the constraints have the required structure of independent constraints of Equation (4).

In the following, we follow the same blueprint of Theorem 3.1 First, prove the hardness result for LR-LocalMinMax-Indep. Then, we exploit the reduction from LR-LocalMinMax-Indep to GDAFixedPoint-Indep in Lemma 3.5 to extend the hardness result to the problem of finding fixed points of gradient ascent-descent dynamics. Also, in this setting, restricting to single-dimensional deviations will be essential to provide tighter hardness results.

From LR-LocalMinMax-Indep to LinVI We provide a reduction for any $0 < \gamma \leq 1$, $\delta = \gamma\frac{\rho^*}{15}$, $\epsilon = \gamma\frac{\rho^{*,2}}{60}$, and $\nu = \frac{\rho^*\delta}{4d}$. It is easy to check that $\delta < \sqrt{\frac{2\epsilon}{L}}$ for any $\gamma \in (0,1]$.

Now we want to prove that any solution to LR-LocalMinMax-JC $(x^\star, y^\star)$ is a solution to LinVI. To obtain a tighter hardness result, we focus on the specialized (but still PPAD-Hard as proved in Theorem 3.4) problem of finding a z such that it is robust to single-dimensional deviations, *i.e.*

$$(Dx^\star + c)_i \cdot (x'_i - x^\star_i) \geq -\rho^* \quad \forall d \in [d],\, x'_i \in [0,1].$$

Consider any $i \in [d]$, and $x'_i \in [0,1]$. Then define $\bar{x} \in [0,1]^d$ as $\bar{x}_i = \delta x'_i + (1-\delta)x^\star_i$ and $\bar{x}'_j = x^\star_j$ for each $j \neq i$.

We check that $\bar{x}_i$ is indeed a feasible deviation. First, $\|\bar{x} - x^\star\| \leq \delta$. Moreover, $\bar{x}_i$ is in $[0,1]$ as $\bar{x}_i$ is a convex combination of two points in $[0,1]$.

Then by the local optimality condition of $x^\star$ with respect to $\bar{x}$ we have:

$$\begin{aligned}x^{\star,\top}(Dx^\star + c) &= x^{\star,\top}D(x^\star - y^\star) + x^{\star,\top}(Dy^\star + c) \\ &\leq \|D^\top x^\star\|_1 \|x^\star - y^\star\|_\infty + x^{\star,\top}(Dy^\star + c) \\ &\leq d\nu + x^{\star,\top}(Dy^\star + c) \\ &\leq d\nu + \bar{x}^\top(Dy^\star + c) + \epsilon \\ &\leq d\nu + \bar{x}^\top D(y^\star - x^\star) + \bar{x}^\top(Dx^\star + c) + \epsilon \\ &\leq 2d\nu + \bar{x}^\top(Dx^\star + c) + \epsilon \\ &= 2d\nu + \delta(x'_i - x^\star_i)\cdot(Dx^\star + c)_i + x^{\star\top}(Dx^\star + c) + \epsilon,\end{aligned}$$

where the first inequality is the Hölder's inequality, the second holds since and $x \in [0,1]^d$ and $\|D\|_1 \leq 1$, and the third one follows from $(x^\star, y^\star)$ being a solution to LR-LocalMinMax-Indep with parameters (ϵ, δ, ν). The last two inequalities hold for similar reasons to the first two. Rearranging the above inequality and using the definition of ν, ϵ, and δ we obtain

$$(Dx^\star + c)_i \cdot (x'_i - x^\star_i) \geq -\frac{\epsilon + 2d\nu}{\delta} = -\rho,$$

which holds for all $i \in [d]$ and $x'_i \in [0,1]$. This proves that $x^\star$ is a solution to the LinVI instance.

From GDAFixedPoint-Indep to LR-LocalMinMax-Indep Similarly to what we have done in the proof of Theorem 3.1, we reduce from LR-LocalMinMax-Indep to GDAFixedPoint-Indep through Lemma 3.5.

Take the same instance of LR-LocalMinMax-Indep built at the beginning of Section E.2. It is easy to check that K^ν is defined such that $K_1^\nu(y)$ and $K_2^\nu(x)$ satisfy the conditions of Lemma 3.5. Notice that working with the "enlarged" sets simply modifies the sets K_1 and K_2 without changing the properties of the constraints. Moreover, it is easy to check that $G_\infty \leq 1$. Thus, we can apply Lemma 3.5 with $\nu = \frac{\rho^* \delta}{4d}$, and $\alpha = \rho^{*,2}/120$. This implies that any α fixed point to GDAFixedPoint-Indep, is a (δ, ϵ, ν)-solution to LR-LocalMinMax-Indep, where $\delta = \rho^*/15$ and $\epsilon = \rho^{*,2}/60$.

Thus, GDAFixedPoint-Indep is PPAD-hard for any $\alpha < \rho^{*,2}/120$, where the function f is $O(1)$-smooth, $G = O(\sqrt{d})$-Lipschitz, and convex-concave (linear-linear).

F Basic Mathematical Preliminaries

F.1 Matrix Norms

Definition F.1. For a matrix $M \in \mathbb{R}^{m \times n}$ we define the operator norm as

$$\|M\|_{r,s} = \sup_{x \in \mathbb{R}^n : \|x\|_r = 1} \|Mx\|_s.$$

When $r = s$, we simply write $\|M\|_r$. Some special cases that we use in this paper are $\|M\|_\infty = \max_{i \in [m]} \sum_{j \in [n]} |M_{i,j}|$ and $\|M\|_1 = \max_{j \in [n]} \sum_{i \in [m]} |M_{i,j}|$. Clearly $\|M\|_1 = \|M^\top\|_\infty$.

Finally, we use the following lemma.

Lemma F.2. *Given a matrix $M \in \mathbb{R}^{m \times n}$ it holds that*

$$\|M\|_2 \leq \sqrt{\|M\|_1 \|M\|_\infty}.$$

In particular if $\|M\|_1, \|M\|_\infty \leq 1$ then $\|Mx\| \leq \|x\|$ and $\|M^\top x\| \leq \|x\|$ for all $x \in \mathbb{R}^n$.

Proof. Consider λ the greatest eigenvalue of $M^\top M$ and x an eigenvector. Now:

$$\|M\|_2^2 = \lambda = \frac{\|M^\top M x\|_\infty}{\|x\|_\infty} \leq \|M^\top M\|_\infty \leq \|M^\top\|_\infty \|M\|_\infty \leq \|M\|_1 \|M\|_\infty.$$

The proof is concluded by using Cauchy-Schwartz inequality for matrix norms on $\|Mx\|$. □

F.2 Correspondences

Here, we define some basic concepts about semicontinuity of correspondences.

Definition F.3. (Upper/Lower-semicontinuity) A correspondence $Q : X \rightrightarrows Y$ is:

- *upper-semicontinuous* at $x \in X$ if for any open set $V \subseteq Y$ such that $Q(x) \subseteq V$ there is a neighborhood U of x such that $\cup_{\tilde{x} \in U} Q(\tilde{x}) \subseteq V$.
- *lower-semicontinuous* at $x \in X$ if for any open set $V \subseteq Q$ such that $Q(x) \cap V \neq \emptyset$ there is a neighborhood U of x such that $Q(\tilde{x}) \cap V \neq \emptyset$ for all $\tilde{x} \in U$.
- continuous at $x \in X$ if it is both upper- and lower-semicontinuous at x.

We say that a correspondence $Q : X \rightrightarrows Y$ is continuous (resp., upper- or lower-semicontinuous) in a set $\tilde{X} \subseteq X$ if it is continuous (resp., upper- or lower-semicontinuous) at x for all $x \in \tilde{X}$.

The following well-known theorem shows the existence result of fixed points for semicontinuous correspondences.

Theorem F.4 (Kakutani fixed point theorem)**.** *If $Q : X \rightrightarrows X$ is an upper-semicontinuous correspondence which is closed, convex, and non-empty valued, then there exists $x \in X$ such that $x \in Q(x)$.*