AMORTIZED EXPLOITABILITY CERTIFICATES FOR LEARNED RESPONSE EQUILIBRIA

Anonymous authors
Paper under double-blind review

ABSTRACT

We develop an amortized exploitability certificate for fixed points of learned response maps in games with compact convex product action sets and costs convex in each player's own action. We certify the exploitability of arbitrary strategy profiles with playerwise envelopes without requiring contraction, monotonicity, or uniqueness of equilibrium. Its two game-dependent quantities are each player's own-gradient norm and the consistency error between the learned response operator and the pseudo-gradient. Both are enveloped once by cost-oracle probes on a finite η-covering of the certified region Ω. Subsequent profiles require only a forward pass. Two results delimit the scheme. Amortization is necessary: at a profile where every player's projection is inactive, a single probe cannot improve the model-free bound. Amortization is also costly: coverings at fixed accuracy grow exponentially with the dimension of Ω. Across the two product-set games and seven response families, the certificate is valid in all 280 theorem-supported cells; we report 140 coupled-constraint cells separately as an out-of-scope stress diagnostic. Small-gain contraction certification is neither necessary nor sufficient for low exploitability in the models tested. Finally, on a nonconvex game the natural residual vanishes analytically at a profile of exploitability $1/4$, so that playerwise convexity cannot be dropped.

1 INTRODUCTION

Decision-making in multi-agent settings can be characterized as *non-cooperative games*, where heterogeneous and selfish entities, i.e. *agents or players*, act with independent decisions to optimize their own objective, irrespective of other players' goals. With applications that traverse socio-technical systems (Ekpo et al., 2026), regional science and transportation planning (Facchinei & Pang, 2003), multi-agent/robot path-finding (Zhang et al., 2024; Wang et al., 2024; Ren et al., 2021), and multi-agent reinforcement learning (Zhang et al., 2021), these games are often encumbered with incomplete or conflicting information and multiple conflicting goals. The (generalized) Nash equilibrium (Facchinei & Kanzow, 2010; Pavel & Hong, 2022; Li et al., 2022), is the solution concept used to characterize the strategic interactions and possible outcomes of such games. Modeling the players' interactions and analyzing the resulting equilibria beyond a small number of players proves intractable: the normal-form representation grows exponentially in the number of players, and even computing an approximate equilibrium is PPAD-hard (Rubinstein, 2015). Beyond tractability, for systems where cost functions are unavailable or players are boundedly rational, the Nash solution concept is practically ineffective for characterizing best responses.

This has stimulated the computation of equilibria from players' interaction data rather using structured learning principles (Scarabaggio et al., 2024). These data-driven formulations extend fixed-point theorems that analyze solution existence under convexity (Kakutani, 1941) to response analysis from data. Relaxations such as the first-order Nash equilibrium notion of Scarabaggio et al. (2026) retain the players' objectives while weakening the equilibrium concept, so that each player's action is a fixed point of its own projected gradient, with existence following from algebraic topology rather than from convex analysis. Scarabaggio et al. (2024) replace the objectives, characterizing best responses as a composition of differentiable neural network layers fitted to players' observed actions. Feasibility is enforced by a projection of the network output onto each player's strategy set. Related formulations learn a parametric equilibrium map trained against the Nikaido–Isoda gap instead of regressing on

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

observed play (Bemporad & Tatarenko, 2026a), or take the reacting object to be the data distribution rather than the response (Narang et al., 2023). This paper works in the second setting and certifies against the first: we measure the residual of the learned response map, which costs one forward pass, and bound through it the projected-gradient residual whose zeros define the first-order notion.

Suppose that $\mathcal{N} := \{1, \dots, N\}$ indexes the players and player i chooses an action x_i from a nonempty compact convex set $X_i \subseteq \mathbb{R}^{n_i}$. Learning-based Nash equilibrium (LBNE) (Scarabaggio et al., 2024) fits a map $f_i(x_{-i})$ predicting player i's response to the other players' actions, x_{-i}. The *behavioral equilibrium* is a profile $x^\star$ for which $x_i^\star = f_i(x_{-i}^\star)$ for every $i \in \mathcal{N}$. Each f_i composes a learned observer, a response head, and a Euclidean projection onto X_i, so that feasibility of the response is enforced by the projection. Given the continuity of each f_i and the convexity and compactness of the product X, Scarabaggio et al. (2024) establish that the joint map $F = (f_1, \dots, f_N)$ admits a fixed point by Brouwer's theorem (Brouwer, 1911). However, convergence of the prescribed Picard–Banach iteration $x^{k+1} = F(x^k)$ is secured only under an aggregate contraction condition $\left(\sum_{i \in \mathcal{N}} \gamma_i^2\right)^{1/2} < 1$ on the playerwise Lipschitz constants γ_i, which excludes the nonexpansive case at which the iteration can fail to converge at all. Contraction hypotheses of this kind are replete across response-learning and decision-dependent formulations, where they buy uniqueness, a linear rate, and a perturbation bound $\varsigma/(1-q)$ under a uniform map error ς (Mignoni, 2025; Narang et al., 2023; Bauschke & Combettes, 2017).

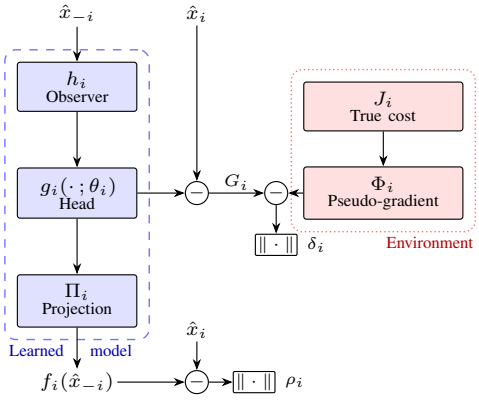

Fig. 1: **A learned response map, and the game it is fitted to.** Both G_i (6) and ρ_i (9) are differences from the same $\hat{x}_i$; δ_i is the consistency error, the one discrepancy that cannot be formed from the model alone.

These guarantees constrain the learned map rather than the game. A contractive F has a unique fixed point of the *learned* game, which need not be an equilibrium of the game that generated the data: the certificate bounds the sensitivity of a fitted network, whereas the quantity of interest is the regret a player would incur at the predicted profile. For arbitrary strategies $\hat{x} \in X$, a constant map $F \equiv \hat{x}$ is nonexpansive and has zero fixed-point residual at $\hat{x}$, while its exploitability is determined by the environment alone and is unconstrained by that residual. Enforcing contraction restricts the hypothesis class where the restriction is least affordable i.e. an interaction whose true best-response map is not contractive admits no contractive model since multi-agent retraining dynamics admit chaotic orbits (Piliouras & Yu, 2023).

While Scarabaggio et al. (2024) introduced the projected observer–head response map and showed that its fixed point solves a variational inequality of the learned game, Narang et al. (2023) obtained uniqueness and a linear rate for performatively stable equilibria under an analogous ratio condition. In a similar vein, Bemporad & Tatarenko (2026a) trained a network against the Nikaido–Isoda gap (Nikaidô & Isoda, 1955) of a surrogate game, while Bemporad & Tatarenko (2026b) instead enforce monotonicity of the learned game by construction. In contrast, Gürkan & Pang (2009) characterized when equilibria of approximating games converge to equilibria of the true game; however, the results are asymptotic with no explicit constant, and Bergerault et al. (2026) proved that the Nash gap, the exploitability of a profile, is *not* determined uniformly over games by a measure-matching error, establishing a divergence between prediction loss and any equilibrium guarantee. Neither approach yields a number that can be evaluated at a given profile *in-silico*: the first certifies the model under a hypothesis that may fail on the interaction of interest, and the second delimits what fitting can achieve without quantifying the shortfall anywhere. A certificate should therefore bound regret in the true game of the prediction model's profile, and should (ideally) be finite when the learned map does not contract.

Write ρ_i for player i's fixed-point residual, δ_i for the consistency error between the learned response operator and the true pseudo-gradient, and $\bar{\delta}_i$ for a bound on δ_i over the certified region (see Fig. 1). Our **contributions** are as follows: (1) We certify the exploitability of every profile $\hat{x} \in \Omega$ with a

playerwise envelope, without requiring contraction, monotonicity, or equilibrium uniqueness. (2) We derive a quadratic exploitability bound in the distance to an interior best response, while the general envelope certificate remains linear in $\rho_i + \bar{\delta}_i$. (3) We show that the residual entry earns its cost only under amortization: $\bar{\delta}_i$ is enveloped once on a certified region using a finite covering, after which each residual costs one forward pass. (4) Across the two product-set games with analytic or numerically verified ground truth and seven response families, the bound holds in all 280 theorem-supported cells. The remaining 140 cells use a coupled feasibility constraint and are reported only as a stress diagnostic. Pooling the certifiable product-set cells with a coupling sweep, a small-gain contraction test admits a fraction 0.09 of the 1000 theorem-supported cells, against a fraction 0.524 whose exploitability is at or below 0.01, so that certification and equilibrium quality come apart.

The rest of this paper is organized as follows: Section 2 formulates the game and its learned response map, Section 3 states the certificate, Section 4 gives the envelope and its cost, and Section 5 reports the measurements. Proofs of our main results, discussions and limitations, as well as additional supporting empirical results are available in the appendices.

2 Problem Formulation

This section describes the problem formulation: we state assumptions, form this paper's notations and introduce conceptual machineries that enable our results.

2.1 Notations and Preliminaries

Let $\mathcal{N} = \{1, \dots, N\}$ denote players' indices. For each player i, let x_i denote its action in a nonempty compact convex set $X_i \subseteq \mathbb{R}^{n_i}$. Player i observes other players' actions $x_{-i} = (x_j)_{j \neq i}$ and incurs a local cost $J_i : X \to \mathbb{R}$ with $J = (J_1, \dots, J_N)$ as a game's cost profile. The joint action set is the product $X = \prod_{i \in \mathcal{N}} X_i \subseteq \mathbb{R}^d$, $d = \sum_{i \in \mathcal{N}} n_i$, and $X_{-i} = \prod_{j \neq i} X_j$. The action profile $x = (x_1, \dots, x_N) \in X$ is also written as $x = (x_i, x_{-i})$ when player i's action is singled out. *We reserve $x^\star$ for an equilibrium and $\hat{x}$ for an arbitrary strategy profile at which a certificate is evaluated.* For vectors $v_1, \dots, v_N$ we write $\mathrm{col}(v_1, \dots, v_N)$ for the column vector obtained by stacking them. For a closed convex set S, the Euclidean projection $\Pi_S(x) := \arg\min_{y \in S} \|x - y\|$ is the unique element of S nearest x; since X is a product, $\Pi_X = (\Pi_{X_1}, \dots, \Pi_{X_N})$ acts blockwise. For differentiable J_i we write $\nabla_{x_i} J_i(x)$ for its gradient in player i's own action.

We now state the assumptions that underlie our theoretical machinery.

Assumption 2.1 (Compact convex product actions)**.** Each action set X_i is nonempty, compact, and convex, and the joint action set is the product $X = \prod_{i \in \mathcal{N}} X_i$.

Assumption 2.2 (Playerwise convexity)**.** For each player i and each $x_{-i} \in X_{-i}$, the cost $J_i(\cdot, x_{-i})$ is convex and continuously differentiable on X_i. The *pseudo-gradient* stacks the players' own-action gradients,

$$\Phi(x) \triangleq \mathrm{col}\left(\Phi_1(x), \dots, \Phi_N(x)\right) \in \mathbb{R}^d, \qquad \Phi_i(x) \triangleq \nabla_{x_i} J_i(x), \tag{1}$$

so that $\Phi_i(x) \in \mathbb{R}^{n_i}$ is player i's block of the pseudo-gradient.

We assume no monotonicity of the pseudo-gradient (1) and no map contraction. The product structure in Assum. 2.1 makes Π_X act blockwise and allows variation of a player's action under other players' fixed actions. Games whose feasibility sets couple across players fall outside the development.

All action coordinates and costs are nondimensionalized before applying the maps below: the unit of each action block is the chosen reference action scale, and each cost is divided by a fixed reference cost so that x_i and $\Phi_i(x)$ have the same numerical units. Thus $x - \Phi(x)$ uses a unit proximal step. With physical units, every occurrence of Φ can equivalently be replaced by $\gamma\Phi$ for a stated step size γ; the paper fixes this normalization to $\gamma = 1$.

Players' regret measures a profile's quality. The *exploitability* of $x \in X$, equivalently the max-over-players variant of the Nikaido–Isoda gap (Nikaidô & Isoda, 1955), whose original form sums over players rather than maximizing, is

$$\varepsilon(x) \triangleq \max_{i \in \mathcal{N}} \Big[J_i(x) - \min_{y_i \in X_i} J_i(y_i, x_{-i}) \Big], \tag{2}$$

in which each bracket is nonnegative, since $x_i \in X_i$ is feasible for the inner minimization. Hence $x \in X$ is a Nash equilibrium precisely when $\varepsilon(x) = 0$, and $\varepsilon(x)$ is the smallest tolerance for which x is an ε-Nash equilibrium (Radner, 1980).

Definition 2.1 (Variational inequality)**.** Let $\mathcal{K} \subseteq \mathbb{R}^d$ be nonempty and let T map $\mathbb{R}^d$ into itself. The *variational inequality problem*, denoted $\mathrm{VI}(\mathcal{K}, T)$, is to find a vector $x^\star \in \mathcal{K}$ such that

$$T(x^\star)^\top (y - x^\star) \geq 0 \quad \text{for all } y \in \mathcal{K}. \tag{3}$$

Geometrically, (3) asks that $T(x^\star)$ make an acute angle with every feasible direction emanating from $x^\star$ (Harker & Pang, 1990). Under Assumption 2.1 and 2.2 the equilibria coincide with the solutions of $\mathrm{VI}(X, \Phi)$ (Rosen, 1965; Facchinei & Pang, 2010). The *natural map* of $\mathrm{VI}(\mathcal{K}, T)$ is $x \mapsto x - \Pi_{\mathcal{K}}\big(x - T(x)\big)$; it vanishes exactly at the solutions of (3), and we call its norm the *natural residual* (Facchinei & Pang, 2003, §1.5).

Definition 2.2 (Learned Response Map)**.** Given a dataset of players' observed actions, the response map of player i composes an *observer* h_i, which summarizes the other players' actions in a feature vector $h_i(x_{-i})$; a *response head* $g_i(\cdot\,; \theta_i)$, a function approximator with parameters θ_i; and a projection onto player i's action set,

$$f_i(x_{-i}; \theta_i) = \Pi_{X_i}\big(g_i(h_i(x_{-i}); \theta_i)\big). \tag{4}$$

Writing $F(x) = \mathrm{col}\,\big(f_1(x_{-1}), \ldots, f_N(x_{-N})\big)$ for the joint response of all players, a *behavioral equilibrium* is a profile $x^\star \in X$ for which every player's prediction reproduces its own input,

$$x_i^\star = f_i(x_{-i}^\star; \theta_i) \quad \text{for all } i \in \mathcal{N}, \tag{5}$$

equivalently $x^\star = F(x^\star)$ (Scarabaggio et al., 2024).

Because Π_{X_i} maps into X_i, we have $F(\mathbb{R}^d) \subseteq X$ and in particular $F : X \to X$, and because the projection onto a closed convex set is nonexpansive (Facchinei & Pang, 2003, §1.5), continuity of f_i reduces to continuity of the learned part $g_i \circ h_i$. Under Assum. 2.1 the joint action set is nonempty, compact and convex, so that Brouwer's theorem (Brouwer, 1911) supplies a profile satisfying (5) whenever each $g_i \circ h_i$ is continuous.

Suppose we write $h(x) = \mathrm{col}\,\big(h_1(x_{-1}), \ldots, h_N(x_{-N})\big)$ and $g = \mathrm{col}\,\big(g_1(\cdot\,; \theta_1), \ldots, g_N(\cdot\,; \theta_N)\big)$, so that $g(h(x))$ is the blockwise pre-projection output of (4). The *induced operator*

$$G(x) \triangleq x - g(h(x)), \tag{6}$$

satisfies $x - G(x) = g(h(x))$, and since Π_X acts blockwise under Assum. 2.1 this gives $F(x) = \Pi_X\big(x - G(x)\big)$. The fixed-point residual $x - F(x)$ of a projected learned map is therefore the natural-map residual of $\mathrm{VI}(X, G)$ (Facchinei & Pang, 2003, §1.5). The learned map is connected to the game via (i) the fixed-point residual of F, which one forward pass of the model computes; (ii) the operator-consistency error between (6) and (1), which cannot be formed without evaluating the true costs; and (iii) the natural residual of the variational inequality $\mathrm{VI}(X, \Phi)$. Only (i) can be computed without access to the game; any certificate that uses (ii) or (iii) must query the environment.

Remark 2.1 (A training loss does not bound δ). Small prediction loss leaves the operator-consistency error uncontrolled: a response map fitted to observed actions constrains G where data was observed, while Φ enters the exploitability (2) through its values on the region being certified. Bergerault et al. (2026) make the analogous point for multi-agent imitation learning: they exhibit games in which exact measure matching still leaves a large Nash gap, and prove that characterizing the Nash gap at a fixed measure-matching error is hard, recovering bounds only under additional structure on the expert equilibrium.

3 An Amortized Exploitability Certificate

Player i's regret is controlled by its blockwise residual and its gradient magnitude, both defined in (7). Lemma 3.1 establishes the pointwise residuals with local constants for the true pseudo-gradient. Lemma 3.2 produces a learned map, trading the variational residual for the fixed-point residual ρ_i (see Fig. 1) and the consistency error δ_i of (9). Every result in this section is proved in Appendix C, and Appendix D collects what the constants and the rate do and do not give: what an average prediction error fails to control, whether the coefficients can be shrunk, and when the rate improves from linear to quadratic.

3.1 A BLOCKWISE GAP BOUND

For a profile $x \in X$ write the gradient magnitude, blockwise residual, and the diameter of each player's block as

$$M_i(x) \triangleq \|\Phi_i(x)\|, \;\; D_i \triangleq \operatorname{diam} X_i, \;\; r_i(x) \triangleq \big\|x_i - \Pi_{X_i}\big(x_i - \Phi_i(x)\big)\big\|. \tag{7}$$

Here $\operatorname{diam} X_i = \sup_{u,v \in X_i} \|u - v\|$, finite by Assum. 2.1. Each $r_i(x)$ is the norm of block i of the right-hand side of the projected pseudo-gradient flow $\dot{x} = \Pi_X\big(x - \Phi(x)\big) - x$, so that every $r_i(x)$ vanishes exactly at the rest points of that flow: where the flow moves slowly, no player gains much by deviating.

Lemma 3.1 (Blockwise gap bound)**.** *Suppose that Assumption 2.1 and 2.2 hold. Then for every $x \in X$,*

$$\varepsilon(x) \;\le\; \max_{i \in \mathcal{N}} \; \min\big\{\big(M_i(x) + D_i\big)\, r_i(x), \;\; M_i(x)\, D_i\big\}. \tag{8}$$

Lemma 3.1 follows from the projection inequality (Facchinei & Pang, 2003, §1.5). Its constants $M_i(x)$ and D_i are local and blockwise: M_i is evaluated at x rather than maximized over X, and each refers only to player i's own gradient and own action set. Error bounds of this kind ordinarily control the distance from x to the solution set, for which the constant exists only under strong monotonicity (Yamashita & Fukushima, 1997). Equation (8) controls exploitability instead, for which compactness suffices. The second entry of the minimum, $M_i(x)D_i$, refers to no learned quantity and is available without a model, so that a certificate is informative only when it falls below $\max_i M_i(x)D_i$ — the comparison Corollary 3.1 makes precise.

3.2 TRANSFER TO A LEARNED MAP

Write $G_i(x) = x_i - g_i(h_i(x_{-i}))$ for the i-th block of the induced operator (6), and define the blockwise residual, ρ_i, and consistency error between the environment and learning scheme, δ_i as

$$\rho_i(x) \triangleq \big\|x_i - f_i(x_{-i})\big\|, \qquad \delta_i(x) \triangleq \big\|G_i(x) - \Phi_i(x)\big\|, \tag{9}$$

respectively.

Lemma 3.2 (Blockwise residual transfer)**.** *Let Assum. 2.1 hold and let Φ_i be defined on X. Then $r_i(x) \le \rho_i(x) + \delta_i(x)$ for every $x \in X$ and every player i.*

Proposition 3.1 (The learned map cannot improve a pointwise certificate where the projection is inactive)**.** *Let Assum. 2.1 hold and let the envelope of (12) be evaluated pointwise at $\hat{x}$, that is $\Omega = \{\hat{x}\}$, so that $\bar{M}_i = M_i(\hat{x})$ and $\bar{\delta}_i = \delta_i(\hat{x})$. Suppose that for every player i the projection in r_i is inactive at $\hat{x}$, that is $\hat{x}_i - \Phi_i(\hat{x}) \in X_i$, so that $r_i(\hat{x}) = M_i(\hat{x})$, and that $M_i(\hat{x}) > 0$. Then for every player i,*

$$\big(M_i(\hat{x}) + D_i\big)\big(\rho_i(\hat{x}) + \delta_i(\hat{x})\big) \;\ge\; M_i(\hat{x})^2 + M_i(\hat{x})D_i \;>\; M_i(\hat{x})D_i, \tag{10}$$

so at such a profile, each player's minimum in (8) is attained by the model-free entry and the certificate reduces to $\max_i M_i(\hat{x})D_i$.

Since forming $\delta_i(\hat{x})$ requires $\Phi_i(\hat{x})$, which already yields the model-free entry, the residual entry is worth its cost only when the consistency error is bounded once over a region rather than at the profile.

Scope of the pointwise limitation. Proposition 3.1 applies only when every player's projection is inactive and $M_i(\hat{x}) > 0$. It does not say that a learned response can never improve a pointwise certificate: active projections can make the learned residual strictly smaller, as the capacitated Cournot experiment illustrates. Its conclusion is instead that, in the interior regime, a single environment query already supplies the model-free branch and the learned branch cannot improve it.

3.3 AMORTIZING THE ENVELOPE OVER A REGION

The consistency error is a property of the model against the environment, so it can be bounded once on a region and reused. The residual is a property of the model alone, so it is free wherever it is evaluated. Separating the two is what makes the certificate worth computing.

Definition 3.1 (Regional envelope)**.** Let $\Omega \subseteq X$ be closed. An *envelope* of a trained model on Ω is a pair of vectors $(\bar{M}, \bar{\delta})$ with, for every player i,

$$\bar{M}_i \geq \sup_{x \in \Omega} M_i(x), \qquad \bar{\delta}_i \geq \sup_{x \in \Omega} \delta_i(x). \tag{11}$$

Theorem 3.1 (Amortized exploitability certificate)**.** *Let Assumption 2.1 and 2.2 hold, let $(\bar{M}, \bar{\delta})$ envelope a trained model on $\Omega \subseteq X$, and let F be of the form (4). Then* every *profile $\hat{x} \in \Omega$ satisfies*

$$\varepsilon(\hat{x}) \;\leq\; \max_{i \in \mathcal{N}} \; \min\big\{ \big(\bar{M}_i + D_i\big)\big(\rho_i(\hat{x}) + \bar{\delta}_i\big),\;\; \bar{M}_i D_i \big\}, \tag{12}$$

and the residuals $\rho_i(\hat{x})$ require no access to the environment.

When $\bar{\delta}$ is measured by probes rather than supplied by an oracle, (12) inherits the confidence of that measurement and holds with probability at least $1 - \alpha$ (Proposition 4.1); the same qualification attaches to Corollary 3.1.

Corollary 3.1 (Non-vacuity threshold)**.** *Fix a player i and suppose $\bar{M}_i + D_i > 0$, which holds whenever X_i has more than one point. Then the residual entry of (12) improves on the model-free entry if and only if*

$$\rho_i(\hat{x}) + \bar{\delta}_i \;<\; \frac{\bar{M}_i D_i}{\bar{M}_i + D_i} \;\leq\; \min\{\bar{M}_i,\, D_i\}. \tag{13}$$

The threshold (13) is what player i's residual entry must improve on over the model-free one; Theorem 3.1 holds either way. Because (12) reports a maximum over players, the certificate improves on $\max_i \bar{M}_i D_i$ only when (13) holds at every player attaining that maximum; clearing it for some other player buys nothing. It is built from $\bar{M}_i$ and D_i, which carry no learned quantity, hence training cannot relax it, and a lower training loss need not lower $\bar{\delta}_i$ either (Proposition D.1). Since $\rho_i \geq 0$, the envelope alone must clear the threshold, at an accuracy fixed by the game rather than by the model; through (16) that accuracy in turn fixes the covering radius η, and the covering size grows as $(\operatorname{diam}\Omega/\eta)^{\dim \Omega}$.

Proposition 3.2 (A spurious stationary profile)**.** *Let $X_i = [-c, c]$ with $c > (1-\kappa)^{-1/2}$ and let*

$$J_i(x_i, x_{-i}) = \tfrac{1}{4}x_i^4 - \tfrac{1}{2}w_i(x_{-i})\, x_i^2, \qquad w_i(x_{-i}) = 1 + \tfrac{\kappa}{N-1}\sum_{j \neq i} x_j^2, \tag{14}$$

with $\kappa \in [0, 1)$. Then $\Phi(0) = 0$, hence $r_i(0) = 0$ for every player, while $\varepsilon(0) = \frac{1}{4}$.

The origin is a strict local maximum of each $J_i(\cdot, x_{-i})$ in its own argument, and it is the conditional mean of play whenever the data-generating process is symmetric across the two minimizers $\pm\sqrt{w_i(x_{-i})}$. A regressor fitted to conditional means therefore lands where the blockwise residual vanishes and exploitability is nonetheless $\frac{1}{4}$: without playerwise convexity, a vanishing residual certifies nothing. Under Assum. 2.2, multimodal play arises from multiple equilibria rather than from nonconvexity.

4 Envelope Costs

This section states the costs of the envelope in the sense of Definition 3.1 and the weaker guarantees available when those costs are omitted.

4.1 Probing at a profile

Two access models bracket what a probe can learn. A *gradient oracle* returns $\Phi_i(x)$ directly; a *cost oracle* returns only the values $J_i(x)$, so that Φ_i must be estimated. We assume the cost oracle, the access model of payoff-based equilibrium learning (Tatarenko & Kamgarpour, 2024). Differencing player i's own coordinates at radius $\tau > 0$ gives

$$\big[\widehat{\Phi}_\tau(x)\big]_{i,k} = \frac{J_i(x + \tau e_{i,k}) - J_i(x - \tau e_{i,k})}{2\tau}, \tag{15}$$

in which $e_{i,k}$ is the unit vector along coordinate k of player i's block. Only own coordinates are differenced, since $\Phi_i = \nabla_{x_i} J_i$ involves no cross-partial, and the estimator costs $2d$ evaluations per profile, with $d = \sum_{i \in \mathcal{N}} n_i$. Costs returned by a simulator, or averaged over finitely many rollouts of a stochastic environment, are not exact, so we take a query at x to return $J_i(x) + Z$ with Z independent across queries, mean-zero and σ-sub-Gaussian. Equation (15) then carries both a smoothing bias and a noise term, and τ trades one against the other.

Remark 4.1 (Probes at the boundary). Since every f_i ends in a projection onto X_i, certified profiles frequently lie on ∂X, and then $x \pm \tau e_{i,k}$ leaves X where the costs are neither defined nor assumed smooth. Two remedies are available. Proposition 4.1 adopts the first, assuming each J_i extends three times continuously differentiably to a τ-neighborhood of X, letting its Taylor expansion run along the whole probe segment. When unavailable, we replace (15) by a one-sided difference on coordinates within τ of the boundary. This has bias $O(B_2\tau)$, with B_2 bounding the second derivative. Thus, the balanced radius on the coordinates becomes $\tau \asymp (\sigma/B_2)^{1/2}$ and their contribution to the penalty is $O\big(\sigma^{1/2} B_2^{1/2}\big)$, in both cases up to the $\sqrt{n_i}$ and logarithmic factors that (16) carries.

4.2 From probes to a regional envelope

A bound on $\sup_{x \in \Omega} \delta_i(x)$ from finitely many probes needs a modulus of continuity. Let δ_i be L_δ-Lipschitz on Ω, which holds whenever both Φ_i and the learned head are Lipschitz there.

Proposition 4.1 (Envelope from a finite covering)**.** *Let $\mathcal{Z} \subseteq \Omega$ be a finite η-covering of the certified region Ω, let δ_i be L_δ-Lipschitz on Ω, let each $J_i(\cdot, x_{-i})$ extend three times continuously differentiably to a τ-neighborhood of X with $|\partial^3 J_i / \partial x_{i,k}^3| \le B_3$ there, and let each cost evaluation be corrupted by independent, mean-zero, σ-sub-Gaussian noise. Then for any $\alpha \in (0,1)$, with probability at least $1 - \alpha$, the quantities*

$$\bar{\delta}_i \;=\; \max_{z \in \mathcal{Z}} \big\| G_i(z) - \widehat{\Phi}_{\tau,i}(z) \big\| \;+\; L_\delta \eta \;+\; \tfrac{1}{6} B_3 \tau^2 \sqrt{n_i} \;+\; \tfrac{\sigma}{\tau} \sqrt{n_i \log(2d|\mathcal{Z}|/\alpha)} \tag{16}$$

satisfy the second half of (11). If in addition M_i is L_M-Lipschitz on Ω, the same construction with L_M in place of L_δ bounds $\bar{M}_i$.

Balancing the bias and noise terms in (16) at $\tau \asymp \big(\sigma \sqrt{\log(2d|\mathcal{Z}|/\alpha)}/B_3\big)^{1/3}$ leaves a penalty of order $\sqrt{n_i}\,\big(\log(2d|\mathcal{Z}|/\alpha)\big)^{1/3} \sigma^{2/3} B_3^{1/3}$, which vanishes with the oracle noise. The covering term $L_\delta \eta$ is a modeling cost, and $|\mathcal{Z}|$ grows by $(\operatorname{diam}\Omega/\eta)^{\dim \Omega}$, so that a rigorous envelope is affordable only when Ω is low-dimensional. The regions for which this is practical are therefore structured ones: a line segment between two candidate profiles, the trajectory of a response iteration, or a neighborhood of a single deployment point.

Remark 4.2 (What the envelope assumes about the environment). Proposition 4.1 needs a Lipschitz modulus L_δ for the consistency error, a third-derivative bound B_3 on costs that are by assumption unknown, a noise scale σ, and L_M for the gradient envelope. None is playerwise and none is measurable from cost-oracle queries alone, so that a global hypothesis on the *model*, that it contracts, has been exchanged for global hypotheses on the *environment*. The exchange is worth making because on a structured low-dimensional Ω these constants are falsifiable: a probe that violates a posited L_δ or B_3 exhibits the violation, whereas a contraction modulus asserted of a trained network is not checkable at all.

5 Experiments

We test soundness, the pointwise limitation, and regional amortization. Exploitability is computed from the analytic game, not the learned response. The seven responders are an affine baseline, two-layer 32–32 tanh/ReLU, spectral-normalized and Almost-Orthogonal-Layer (AOL) Lipschitz-certified networks (Prach & Lampert, 2022), and 64–64 conditional diffusion and flow models. Standard E1–E2 runs supply 240 noisy observational transitions; neural responders use Adam with batch size 128 for 200 epochs at learning rate 3×10^{-3} (AOL: 5×10^{-3}), with least-squares, denoising, and flow-matching losses as appropriate. E8 uses its stated reduced training budget. A cell is one game, response family, and seed; comparisons are paired within cells, with 10,000-resample bootstrap intervals, Wilcoxon signed-rank tests, and Holm correction. The full protocol appears in Appendix F.

Bound and Tightness of Learned Map. The tightness study fits seven response families over 20 seeds on three games: a coupled quadratic game whose coupling parameter fixes the monotonicity modulus exactly, a capacitated Cournot game whose capacities bind at equilibrium, and a Rosen game carrying one constraint shared across players. The quadratic and Cournot games have product action sets and test Theorem 3.1 directly: $2 \times 7 \times 20 = 280$ cells, all valid. The Rosen game violates the product requirement of Assum. 2.1; its 140 cells are reported only as a stress diagnostic and are excluded from theorem coverage. The shared game has an analytically specified best-response oracle, but its variational equilibrium is computed to the documented numerical tolerance rather than in closed form.

The learned-map-free arm is active in every cell, with median tightness 108.7; the residual arm has median 297.9(about 2.7 times larger). Its paired difference from the direct-probe comparison is -0.001165in $\log_{10}$ tightness ($p_{\text{Holm}} = 0.1519$); the test does not detect a difference. The learned response helps only after the consistency error is audited over a region and reused.

Playerwise scaling. The scaling study varies the player count while holding the observational-data budget fixed. At a fixed displacement from the reference equilibrium, the aggregate ratio of certificate to true exploitability increases from 38.15 at $n = 2$ to 1157 at $n = 256$. The corresponding blockwise ratio changes from 13.92 to 6.628. The aggregate ratio therefore grows by a factor of about thirty, whereas the blockwise ratio remains on the same order, which Fig. 2 shows at both action dimensions and Table 1 lists at scalar actions. Shading marks where the aggregated bound exceeds the model-free bound; the dash-dot rule marks where the responder fit becomes underdetermined.

The aggregate form uses one global constant and a joint residual, whereas the playerwise form uses (7); here the blockwise series equals the model-free entry. The fixed 240-transition budget makes the largest fits underdetermined, which can inflate learned-map residuals but not the model-free arm.

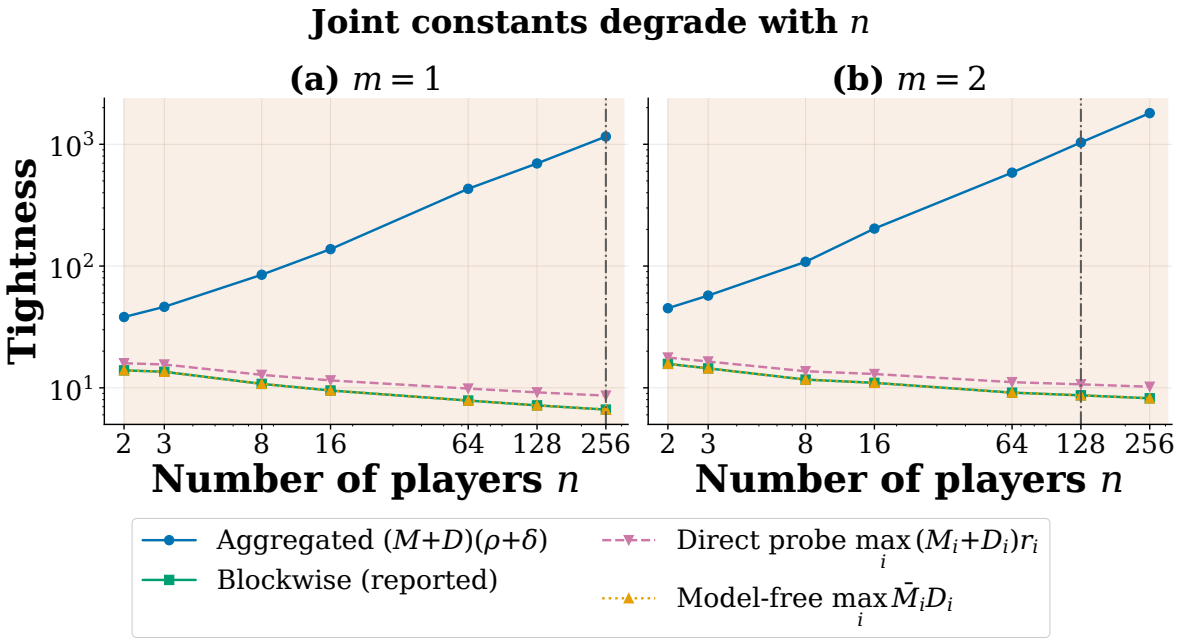

Fig. 2: **Joint constants degrade with the player count; playerwise ones do not.** Coupled quadratic game, (a) scalar and (b) two-dimensional actions, medians over 20 seeds, log–log, lower is tighter.

Regional auditing creates the amortization regime. At $\eta/\operatorname{diam}\Omega = 0.025$, the one-dimensional segment requires 20 covering points and the three-dimensional box requires 8000. Their residual arms bind on 0.5742 and 0.7266 of held-out profiles, with regional-to-pointwise ratios 1.788 and 1.725. Every tested profile satisfies the audited envelope; covering cost grows with region dimension. Fig. 3 shows the tradeoff.

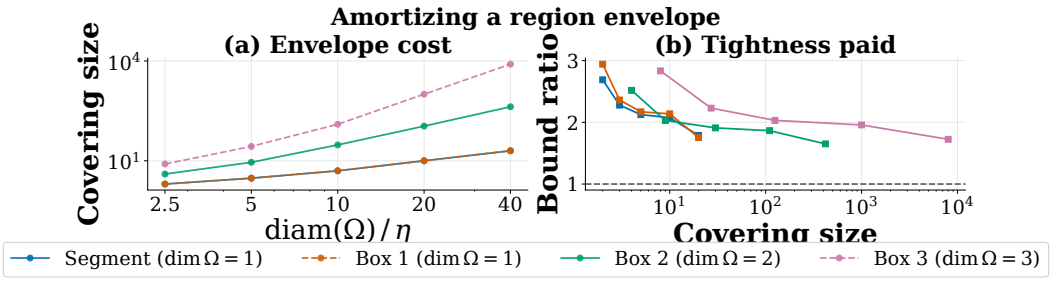

Fig. 3: **Regional envelope costs and tightness.** Medians over 10 seeds and 256 held-out profiles per region. (a) Covering size=break-even certified profile numbers. (b) Amortized bound over pointwise bound; the dashed rule is the value a single-profile probe attains.

Contraction and exploitability measure different properties. Pooling the theorem-supported product-set cells from the tightness study with the coupling sweep gives 1000 model cells; the shared-constraint diagnostics are excluded. Small-gain certification accepts 0.09, while 0.524 have exploitability at most 0.01. The four contingency counts are 66, 20, 454, and 456; contraction is therefore neither necessary nor sufficient here. Generative families are excluded because they have no rigorous global Lipschitz bound.

Multiplicity, solvers, and probe diagnostics. On the playerwise-convex non-monotone game, generative responders are evaluated through conditional-mean maps, so the comparison is silent about generativity and capacity; their heads have roughly four times as many parameters yet recover fewer equilibria (Appendix E). The nonconvex witness remains a scope diagnostic: zero variational residual can coexist with nonzero exploitability. On the nonexpansive game, Picard reaches tolerance 10^{-8} in none of 20 runs, while Krasnosel'skiĭ–Mann and Halpern require medians of 126 and 5 iterations. Probe-radius sweeps reproduce the bias–noise tradeoff of Proposition 4.1; both are diagnostics, not replacements for its regularity assumptions.

6 Conclusion

A contraction certificate establishes the regularity of a learned response map without resolving its fidelity to the game that produced the data. We address this with a blockwise bound on exploitability in the true game, whose hypotheses constrain the environment rather than the model. Two limits bound what it delivers: Proposition 3.1 shows the learned map cannot improve a certificate at a single probed profile where the projection is inactive, and Corollary 3.1 shows the regional envelope that escapes that argument must reach an accuracy set by the game, at a cost exponential in $\dim \Omega$. Proposition 3.2 isolates the indispensable hypothesis: without playerwise convexity a vanishing residual coexists with exploitability $\frac{1}{4}$. Empirically, contraction certification and low exploitability come apart. Appendix Bdiscusses our limitations.

USE OF AI TOOLS

In this work, generative AI assistance was used to augment our literature review via its search and retrieval of related works, for editing prose, and for editing and auditing the experiment code. We have not used generative AI tools for our theoretical contributions, nor empirical claims. We have reviewed all AI-assisted work, including literature-survey and prose edits. The authors take full responsibility for the final content of the paper, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY.

The article's main text introduces and states the problem, assumptions, certificate, and evaluation measures. Complete proofs appear in Appendix C. The experimental appendix specifies the five games, seven response families, training and solver protocols, random seeds, oracle evaluations, covering construction, and statistical tests used in E1–E9.

REFERENCES

Yakov Babichenko. Query Complexity of Approximate Nash Equilibria. *Journal of the ACM*, 63(4): 1–24, 2016. doi: 10.1145/2908734.

Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, Cham, 2nd edition, 2017. doi: 10.1007/978-3-319-48311-5.

Giuseppe Belgioioso, Peng Yi, Sergio Grammatico, and Lacra Pavel. Distributed Generalized Nash Equilibrium Seeking: An Operator-Theoretic Perspective. *IEEE Control Systems Magazine*, 42(4): 87–102, 2022. doi: 10.1109/MCS.2022.3171480.

Alberto Bemporad and Tatiana Tatarenko. Learning Approximate Solutions to Multiparametric Generalized Nash Equilibrium Problems, 2026a. arXiv preprint arXiv:2605.28757.

Alberto Bemporad and Tatiana Tatarenko. Learning Parametric Monotone Games, 2026b. arXiv preprint arXiv:2609.02494.

Antoine Bergerault, Volkan Cevher, and Negar Mehr. Matching Multiple Experts: On the Exploitability of Multi-Agent Imitation Learning, 2026. arXiv preprint arXiv:2602.21020.

Jérôme Bolte and Edouard Pauwels. Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning. *Mathematical Programming*, 188(1):19–51, 2021. doi: 10.1007/s10107-020-01501-5.

L. E. J. Brouwer. Über Abbildung von Mannigfaltigkeiten. *Mathematische Annalen*, 71(1):97–115, 1911. doi: 10.1007/BF01456931.

Frank H. Clarke. *Optimization and Nonsmooth Analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons, New York, 1983. Reprinted as SIAM Classics in Applied Mathematics 5, 1990, doi:10.1137/1.9781611971309.

Stella Dafermos. Sensitivity analysis in variational inequalities. *Mathematics of Operations Research*, 13(3):421–434, 1988. doi: 10.1287/moor.13.3.421.

Alexander Davydov, Saber Jafarpour, and Francesco Bullo. Non-Euclidean Contraction Theory for Robust Nonlinear Stability. *IEEE Transactions on Automatic Control*, 67(12):6667–6681, 2022. doi: 10.1109/TAC.2022.3183966.

Roy Dong, Heling Zhang, and Lillian J. Ratliff. Approximate Regions of Attraction in Learning with Decision-Dependent Distributions. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 206 of *Proceedings of Machine Learning Research*, pp. 11172–11184. PMLR, 2023.

Asen L. Dontchev and R. Tyrrell Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM Journal on Optimization*, 6(4):1087–1105, 1996. doi: 10.1137/S1052623495284029.

Promise Ekpo, Angelique Taylor, and Lekan Molu. A Generalized Nash Equilibrium-Seeking Scheme for Trauma Resuscitation. In *Proceedings of the IFAC World Congress*, Busan, Republic of Korea, 2026.

Francisco Facchinei and Christian Kanzow. Generalized Nash Equilibrium Problems. *Annals of Operations Research*, 175(1):177–211, 2010. doi: 10.1007/s10479-009-0653-x.

Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2003. doi: 10.1007/b97544. Two volumes, continuously paginated pp. 1–1234.

Francisco Facchinei and Jong-Shi Pang. Nash Equilibria: The Variational Approach. In Daniel P. Palomar and Yonina C. Eldar (eds.), *Convex Optimization in Signal Processing and Communications*, pp. 443–493. Cambridge University Press, Cambridge, 2010. doi: 10.1017/CBO9780511804458.013. Published online 3 December 2009.

Mahyar Fazlyab, Alexander Robey, Hamed Hassani, Manfred Morari, and George J. Pappas. Efficient and Accurate Estimation of Lipschitz Constants for Deep Neural Networks. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, 2019.

Masao Fukushima. Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Mathematical Programming*, 53(1–3):99–110, 1992. doi: 10.1007/BF01585696.

Gül Gürkan and Jong-Shi Pang. Approximations of Nash Equilibria. *Mathematical Programming*, 117(1–2):223–253, 2009. doi: 10.1007/s10107-007-0156-y. Published online 21 August 2007.

Patrick T. Harker and Jong-Shi Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming*, 48(1–3):161–220, 1990. doi: 10.1007/BF01582255.

John C. Harsanyi. Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory*, 2(1):1–23, 1973. doi: 10.1007/BF01737554.

Sergiu Hart and Andreu Mas-Colell. Uncoupled Dynamics Do Not Lead to Nash Equilibrium. *American Economic Review*, 93(5):1830–1836, 2003. doi: 10.1257/000282803322655581.

Aaron J. Havens, Alexandre Araujo, Siddharth Garg, Farshad Khorrami, and Bin Hu. Exploiting Connections between Lipschitz Structures for Certifiably Robust Deep Equilibrium Models. In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, pp. 21658–21674, 2023.

Michael D. Hirsch, Christos H. Papadimitriou, and Stephen A. Vavasis. Exponential lower bounds for finding Brouwer fix points. *Journal of Complexity*, 5(4):379–416, 1989. doi: 10.1016/0885-064X(89)90017-4.

Josef Hofbauer and Ed Hopkins. Learning in perturbed asymmetric games. *Games and Economic Behavior*, 52(1):133–152, 2005. doi: 10.1016/j.geb.2004.06.006.

Guoqiang Hu, Yipeng Pang, Chao Sun, and Yiguang Hong. Distributed Nash Equilibrium Seeking: Continuous-Time Control-Theoretic Approaches. *IEEE Control Systems Magazine*, 42(4):68–86, 2022. doi: 10.1109/MCS.2022.3171479.

Saber Jafarpour, Alexander Davydov, Anton V. Proskurnikov, and Francesco Bullo. Robust Implicit Networks via Non-Euclidean Contractions. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, pp. 9857–9868, 2021.

Shizuo Kakutani. A generalization of Brouwer's fixed point theorem. *Duke Mathematical Journal*, 8 (3):457–459, 1941. doi: 10.1215/S0012-7094-41-00838-4.

Torbjörn Larsson and Michael Patriksson. A class of gap functions for variational inequalities. *Mathematical Programming*, 64(1–3):53–79, 1994. doi: 10.1007/BF01582565.

Tao Li, Guanze Peng, Quanyan Zhu, and Tamer Başar. The confluence of networks, games, and learning a game-theoretic framework for multiagent decision making over networks. *IEEE Control Systems Magazine*, 42(4):35–67, 2022. doi: 10.1109/MCS.2022.3171478.

Eric Mazumdar, Lillian J. Ratliff, and S. Shankar Sastry. On Gradient-Based Learning in Continuous Games. *SIAM Journal on Mathematics of Data Science*, 2(1):103–131, 2020. doi: 10.1137/18M1231298.

Richard D. McKelvey and Thomas R. Palfrey. Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior*, 10(1):6–38, 1995. doi: 10.1006/game.1995.1023.

Panayotis Mertikopoulos and Zhengyuan Zhou. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1–2):465–507, 2018. doi: 10.1007/s10107-018-1254-8.

Nicola Mignoni. *Non-Cooperative Game Theoretical Control for Green and Efficient Energy Communities*. PhD thesis, Politecnico di Bari, Bari, Italy, 2025. URL `https://hdl.handle.net/11589/284440`. Chapter 6.

Adhyyan Narang, Evan Faulkner, Dmitriy Drusvyatskiy, Maryam Fazel, and Lillian J. Ratliff. Multiplayer Performative Prediction: Learning in Decision-Dependent Games. *Journal of Machine Learning Research*, 24(202):1–56, 2023. URL https://jmlr.org/papers/v24/22-0131.html.

Hukukane Nikaidô and Kazuo Isoda. Note on non-cooperative convex game. *Pacific Journal of Mathematics*, 5(5):807–815, 1955. doi: 10.2140/pjm.1955.5.807.

Jong-Shi Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79(1–3): 299–332, 1997. doi: 10.1007/BF02614322.

Lacra Pavel and Yiguang Hong. Distributed games and nash equilibrium seeking in multiagent systems over networks: An introduction to the special issue. *IEEE Control Systems Magazine*, 42 (4):32–34, 2022. doi: 10.1109/MCS.2022.3171477.

Ji-Ming Peng. Equivalence of variational inequality problems to unconstrained minimization. *Mathematical Programming*, 78(3):347–355, 1997. doi: 10.1007/BF02614360.

Juan C. Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative Prediction. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7599–7609. PMLR, 2020. URL https://proceedings.mlr.press/v119/perdomo20a.html.

Georgios Piliouras and Fang-Yi Yu. Multi-Agent Performative Prediction: From Global Stability and Optimality to Chaos. In *Proceedings of the 24th ACM Conference on Economics and Computation (EC)*, pp. 1047–1074. ACM, 2023. doi: 10.1145/3580507.3597759.

Bernd Prach and Christoph H. Lampert. Almost-orthogonal layers for efficient general-purpose lipschitz networks. In *Computer Vision – ECCV 2022*, volume 13681 of *Lecture Notes in Computer Science*, pp. 350–365. Springer, 2022. doi: 10.1007/978-3-031-19803-8_21.

Roy Radner. Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives. *Journal of Economic Theory*, 22(2):136–154, 1980. doi: 10.1016/0022-0531(80)90037-X.

Zhongqiang Ren, Sivakumar Rathinam, and Howie Choset. Multi-objective conflict-based search for multi-agent path finding. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 8786–8791, 2021. doi: 10.1109/ICRA48506.2021.9560985.

Stephen M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5(1):43–62, 1980. doi: 10.1287/moor.5.1.43.

J. B. Rosen. Existence and uniqueness of equilibrium points for concave N-person games. *Econometrica*, 33(3):520–534, 1965. doi: 10.2307/1911749.

Aviad Rubinstein. Inapproximability of Nash Equilibrium. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 409–418, 2015. doi: 10.1145/2746539.2746578.

Paolo Scarabaggio, Nicola Mignoni, Raffaele Carli, and Mariagrazia Dotoli. Equilibrium Seeking in Learning-Based Noncooperative Nash Games. In *IEEE 63rd Conference on Decision and Control (CDC)*, pp. 210–215, Milan, Italy, December 2024. IEEE. doi: 10.1109/CDC56724.2024.10886614.

Paolo Scarabaggio, Nicola Mignoni, Raffaele Carli, and Mariagrazia Dotoli. Equilibria Existence in Games: A Perspective from the Lefschetz–Nielsen Theory. In *European Control Conference (ECC)*, pp. 2938–2943, Reykjavík, Iceland, July 2026. EUCA. ISBN 978-3-907144-13-8.

Tatiana Tatarenko and Maryam Kamgarpour. Payoff-based Learning of Nash Equilibria in Merely Monotone Games. *IEEE Transactions on Control of Network Systems*, 11(4):1790–1799, 2024. doi: 10.1109/TCNS.2024.3355035.

Aladin Virmaux and Kevin Scaman. Lipschitz Regularity of Deep Neural Networks: Analysis and Efficient Estimation. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, 2018.

Anna von Heusinger and Christian Kanzow. Optimization reformulations of the generalized Nash equilibrium problem using Nikaido–Isoda-type functions. *Computational Optimization and Applications*, 43(3):353–377, 2009. doi: 10.1007/s10589-007-9145-6.

Fangji Wang, Han Zhang, Sven Koenig, and Jiaoyang Li. Efficient approximate search for multi-objective multi-agent path finding. In *International Conference on Automated Planning and Scheduling (ICAPS)*, volume 34, pp. 613–622, 2024. doi: 10.1609/icaps.v34i1.31524.

Ruigang Wang and Ian R. Manchester. Direct Parameterization of Lipschitz-Bounded Deep Networks. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, volume 202 of *Proceedings of Machine Learning Research*, pp. 36093–36110. PMLR, 2023.

Ezra Winston and J. Zico Kolter. Monotone Operator Equilibrium Networks. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, pp. 10718–10728, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/798d1c2813cbdf8bcdb388db0e32d496-Abstract.html.

Nobuo Yamashita and Masao Fukushima. Equivalent unconstrained minimization and global error bounds for variational inequality problems. *SIAM Journal on Control and Optimization*, 35(1): 273–284, 1997. doi: 10.1137/S0363012994277645.

Han Zhang, Oren Salzman, Ariel Felner, Carlos Hernández Ulloa, and Sven Koenig. A-a*pex: Efficient anytime approximate multi-objective search. In *Symposium on Combinatorial Search (SoCS)*, volume 17, pp. 179–187, 2024. doi: 10.1609/socs.v17i1.31556.

Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. In *Handbook of Reinforcement Learning and Control*, pp. 321–384. Springer, 2021. doi: 10.1007/978-3-030-60990-0_12.