AMORTIZED EXPLOITABILITY CERTIFICATES FOR LEARNED RESPONSE EQUILIBRIA

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper proposes an exploitability certificate for learned response models in games with compact convex product action sets and convex costs in each player's action. The certificate audits a prespecified subset of the joint action space. For each player, the audit bounds the norm of the player's cost gradient with respect to its own action and the error between that gradient and the learned response operator, uniformly over the subset. It then certifies any profile in that subset by combining these bounds with the distance between the profile and the model-predicted response. The uniform bounds can be established from finitely many noisy cost evaluations when independently validated smoothness and noise bounds are available. The covering cost can grow exponentially with the subset's intrinsic dimension, so the audit is intended for low-dimensional subsets whose cost can be amortized over many profiles. Experiments illustrate this tradeoff: auditing a region enables subsequent model-only evaluations, whereas a standard model-free bound is tighter for freshly queried profiles. We also show that contraction does not characterize low exploitability and that playerwise convexity is essential.

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. (Generalized) Nash equilibrium (Facchinei & Kanzow, 2010; Pavel & Hong, 2022; Li et al., 2022) is a 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 motivated learning equilibria from players' interaction data using structured learning methods (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 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 consists of an optional feature map, a response head, and a Euclidean projection onto X_i; when no feature transformation is needed, the feature map is the identity on x_{-i}. The projection enforces feasibility. 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 inform 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).

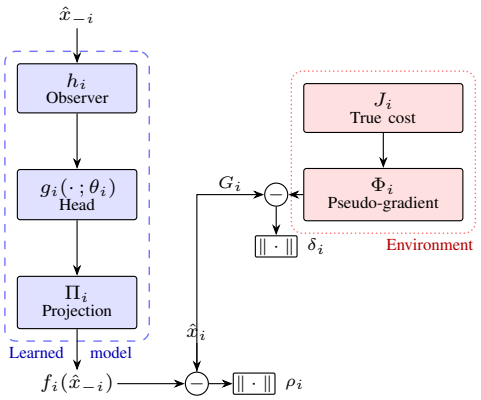

Fig. 1: **A learned response map, and the game it is fitted to.** At a profile $\hat{x}$, $G_i(\hat{x}) = \hat{x}_i - f_i(\hat{x}_{-i})$ is the learned response mismatch and $\rho_i(\hat{x}) = \|G_i(\hat{x})\|$ requires only a model evaluation. The consistency error $\delta_i(\hat{x}) = \|G_i(\hat{x}) - \nabla_{x_i} J_i(\hat{x})\|$ additionally compares that mismatch with the true own-action cost gradient.

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) proved uniqueness and linear convergence of repeated retraining to their decision-dependent equilibrium 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 and δ_i for the discrepancy between the learned response operator and the true own-action cost gradient. For a region Ω, let $\bar{M}_i$ and $\bar{\delta}_i$ upper-bound the true-gradient magnitude and this discrepancy throughout Ω. We call the pair $(\bar{M}, \bar{\delta})$ a *regional envelope*. Figure 1 illustrates the profile-specific quantities. 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 learned-residual candidate earns its cost only under amortization: a finite covering establishes $\bar{\delta}_i$ once for a certified region, after which each residual costs one forward pass. (4) Across two product-action games, seven response families, and 20 random seeds, the certificate upper-bounds exact exploitability in all 280 in-scope configurations. In a separate sweep of 1,000 configurations, a small-gain contraction test is neither necessary nor sufficient for low exploitability, showing that contraction and equilibrium quality measure different properties.

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 states the assumptions, fixes notation, and introduces the objects used in 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,

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

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

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

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

where each bracket is nonnegative, since $x_i \in X_i$ is feasible for the inner minimization. Hence $x \in X$ is a Nash equilibrium 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*, $\mathrm{VI}(\mathcal{K}, T)$, finds a vector $x^\star \in \mathcal{K}$ such that

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

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 $x \mapsto x - \Pi_{\mathcal{K}}\big(x - T(x)\big)$, which vanishes at the solutions of (3). This is the *natural residual* (Facchinei & Pang, 2003, §1.5).

Definition 2.2 (Learned Response Map)**.** Given a dataset of players' observed actions, player i's response map may first apply an *observer* h_i that summarizes the other players' actions in a feature vector $h_i(x_{-i})$. When no feature transformation is needed, h_i is the identity on the concatenated rival actions. A *response head* $g_i(\cdot\,;\theta_i)$, with parameters θ_i, maps these features to an action, and projection onto X_i enforces feasibility:

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

Writing $F(x) = \operatorname{col}\big(f_1(x_{-1}),\dots,f_N(x_{-N})\big)$ 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. The *induced operator* of the projected response map is

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

so that $x - G(x) = F(x) \in X$. Thus, $F(x) = \Pi_X\big(x - G(x)\big)$. The fixed-point residual $x - F(x) = G(x)$ is thus the natural map 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; (ii) the operator-consistency error between (6) and (1); 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.

Small prediction loss can make the operator-consistency error uncontrollable: 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 an analogous point for multi-agent imitation learning and recover imitation-to-Nash bounds only when the demonstrated expert has additional structure, such as the dominant-strategy behavior or continuity of best responses around that expert.

3 An Amortized Exploitability Certificate

This section constructs the amortized exploitability certificate. We first bound each player's regret using the true projected-gradient residual and local, player-specific constants. We then relate that residual to the mismatch of the learned response and its discrepancy from the true game. Finally, we replace the profile-specific game quantities by upper bounds that hold throughout a region and can be reused at later profiles. All proofs are in Appendix C;

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

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

where $\operatorname{diam} X_i = \sup_{u,v\in X_i}\|u - v\|$ is finite by Assum. 2.1. Block i's $r_i(x) := \|\dot{x}\|$, where $\dot{x}$ is projected pseudo-gradient flow $\dot{x} = \Pi_X\big(x - \Phi(x)\big) - x$. Every $r_i(x)$ vanishes at the flow's equilibrium points.

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). $M_i(x)$ and D_i are local and blockwise constants; we evaluate M_i at x. Error bounds of this kind 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. The model-free candidate is $M_i(x)D_i$ so that a learned-residual certificate must improve on the benchmark $\max_i M_i(x)D_i$.

3.1 TRANSFER TO A LEARNED MAP

Write $G_i(x) = x_i - f_i(x_{-i})$ for the i-th block of the induced operator (6), and define the blockwise residual ρ_i and the consistency error δ_i by

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

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.*

3.2 AMORTIZING GAME QUERIES OVER A REGION

At a profile x, the residual $\rho_i(x) = \|x_i - f_i(x_{-i})\|$ requires only an evaluation of the learned model. In contrast, evaluating $\delta_i(x)$ requires the true game gradient $\Phi_i(x)$, obtained either from a gradient query or estimated from cost queries. A regional audit makes those game queries in advance to establish upper bounds that hold throughout a chosen region. Later profiles in that region reuse the same bounds and require only new evaluations of ρ_i.

Definition 3.1 (Regional envelope)**.** Let $\Omega \subseteq X$ be closed. A *regional envelope* of a trained model on Ω is a pair of vectors $(\bar{M}, \bar{\delta})$ that uniformly upper-bound, for every player i, the true-gradient magnitude and the model–game discrepancy:

$$\bar{M}_i \ge \sup_{x\in\Omega} M_i(x), \qquad \bar{\delta}_i \ge \sup_{x\in\Omega} \delta_i(x). \tag{10}$$

Theorem 3.1 (Amortized exploitability certificate)**.** *Let Assumption 2.1 and 2.2 hold, let $(\bar{M}, \bar{\delta})$ be a regional envelope of a trained model on $\Omega \subseteq X$, and let F be of the form (4). Furthermore, let $\big(\bar{M}_i + D_i\big)\big(\rho_i(\hat{x}) + \bar{\delta}_i\big)$ be the* learned-residual candidate bound *and $\bar{M}_i D_i$ the* model-free candidate bound*. Then* every *profile $\hat{x} \in \Omega$ satisfies*

$$\varepsilon(\hat{x}) \;\le\; \max_{i\in\mathcal{N}} \min\Big\{\big(\bar{M}_i + D_i\big)\big(\rho_i(\hat{x}) + \bar{\delta}_i\big),\;\; \bar{M}_i D_i\Big\}, \tag{11}$$

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

Let $\big(\bar{M}_i + D_i\big)\big(\rho_i(\hat{x}) + \bar{\delta}_i\big)$ be the *learned-residual candidate bound* and $\bar{M}_i D_i$ the *model-free candidate bound*.

Proposition 3.1 (The learned map cannot improve a pointwise certificate)**.** *Let Assum. 2.1 hold and specialize the regional envelope in (11) to the single profile $\hat{x}$: $\Omega = \{\hat{x}\}$, $\bar{M}_i = M_i(\hat{x})$, and $\bar{\delta}_i = \delta_i(\hat{x})$. If $M_i(\hat{x}) > 0$, then*

$$\big(M_i(\hat{x}) + D_i\big)\big(\rho_i(\hat{x}) + \delta_i(\hat{x})\big) \;\ge\; M_i(\hat{x})^2 + M_i(\hat{x}) D_i \;>\; M_i(\hat{x}) D_i. \tag{12}$$

That is, player i's minimum is attained by the model-free candidate. If $M_i(\hat{x}) > 0$ for every player, the pointwise certificate reduces to $\max_i M_i(\hat{x}) D_i$.

Proposition 3.1 explains useful regional reuse regime rather than a fresh audit at every profile. Computing $\delta_i(\hat{x})$ at a single profile requires the same true gradient that supplies the model-free candidate; thus, the learned-residual candidate cannot improve the pointwise certificate.

Corollary 3.1 (Non-vacuity threshold)**.** *Fix a player i and suppose $\bar{M}_i + D_i > 0$ holds whenever X_i has more than one point. Then, the learned-residual candidate in (11) improves on the model-free candidate if and only if*

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

Player i's learned-residual candidate must improve on the threshold (13). Because (11) 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. Note: the threshold contains no learned quantity; therefore, training cannot relax it. Since $\rho_i \ge 0$, the regional discrepancy bound $\bar{\delta}_i$ must fall below the threshold before the learned-residual candidate can be selected.

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.

4.2 From probes to a regional envelope

Finite probes control a continuum only through supplied moduli of continuity. For every player, let $L_{\delta,i}$ and $L_{M,i}$ be valid Lipschitz upper bounds for δ_i and M_i on Ω. Because $G_i(x) = x_i - f_i(x_{-i})$, a valid bound for $L_{\delta,i}$ combines the identity block, a certified Lipschitz bound for the projected response f_i, and a valid bound for Φ_i; learned-model weights alone do not bound the environment term. Proposition 4.1 takes these and the other regularity constants as inputs rather than estimating them from the covering data.

Proposition 4.1 (Envelope from a finite covering). *Let $\Omega \subseteq X$ be nonempty and let $\mathcal{Z} \subseteq \Omega$ be a supplied finite Euclidean η-cover: every $x \in \Omega$ lies within distance η of some $z \in \mathcal{Z}$. For every player i, suppose supplied constants $L_{\delta,i}$ and $L_{M,i}$ are valid Lipschitz bounds for δ_i and M_i on Ω. Let $\tau > 0$, suppose every required probe segment lies in a domain on which J_i has a C^3 extension satisfying $|\partial^3 J_i/\partial x_{i,k}^3| \le B_3$, and let cost-query errors be independent, mean-zero, and σ-sub-Gaussian. Define*

$$E_i(\tau, \alpha) \triangleq \tfrac{1}{6}B_3\tau^2\sqrt{n_i} + \tfrac{\sigma}{\tau}\sqrt{n_i \log(2d|\mathcal{Z}|/\alpha)}. \tag{16}$$

After the $2d|\mathcal{Z}|$ scalar cost evaluations in (15), the same event of probability at least $1-\alpha$ simultaneously yields, for every player,

$$\bar{M}_i = \max_{z\in\mathcal{Z}} \big\|\widehat{\Phi}_{\tau,i}(z)\big\| + L_{M,i}\eta + E_i(\tau, \alpha), \tag{17}$$

$$\bar{\delta}_i = \max_{z\in\mathcal{Z}} \big\|G_i(z) - \widehat{\Phi}_{\tau,i}(z)\big\| + L_{\delta,i}\eta + E_i(\tau, \alpha), \tag{18}$$

and these quantities satisfy both inequalities in (10) for all players.

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. Proposition 4.1 uses a C^3 extension to a τ-neighborhood of X, so the central Taylor expansion remains valid along the complete probe segment. When such an extension is unavailable, one may instead use a one-sided difference on coordinates within τ of the boundary. Its bias is $O(B_2\tau)$ when B_2 bounds the second derivative. The balanced radius then scales as $\tau \asymp (\sigma/B_2)^{1/2}$, and the corresponding penalty is $O(\sigma^{1/2}B_2^{1/2})$, up to the $\sqrt{n_i}$ and logarithmic factors in (17).

Balancing the bias and noise terms in (17) at $\tau \asymp \left(\sigma\sqrt{\log(2d|\mathcal{Z}|/\alpha)}/B_3\right)^{1/3}$ leaves a penalty of order $\sqrt{n_i}\left(\log(2d|\mathcal{Z}|/\alpha)\right)^{1/3}\sigma^{2/3}B_3^{1/3}$, which vanishes with the oracle noise. The covering term $L_{\delta,i}\eta$ is a modeling cost, and at fixed relative resolution the covering number of a box-like m-dimensional region is $O((\operatorname{diam}\Omega/\eta)^m)$. A rigorous envelope is therefore affordable only for regions with small covering numbers, such as a line segment between two candidate profiles, the trajectory of a response iteration, or a neighborhood of a single deployment point.

Remark 4.2 (The finite-query construction is conditional on supplied regularity). Proposition 4.1 requires valid upper bounds $L_{\delta,i}$, B_3, and $L_{M,i}$, together with a justified independent mean-zero sub-Gaussian noise model. Finite cost values cannot establish these uniform regional quantities because they do not control behavior between queried points. A sampled violation can refute a proposed bound, but the absence of such a violation cannot validate it. A rigorous application must therefore obtain the constants from an analytic simulator, physical specifications, symbolic or interval bounds, a certified environment class, a jointly certified environment–model class, or a separately justified noise model. Weight-derived bounds can control the learned response f_i through the composite $g_i \circ h_i$ and the nonexpansive projection; the environment terms remain external assumptions. If these inputs are unavailable, the finite-cover construction is a diagnostic rather than a uniform certificate. Random sampling without an additional quantile or tolerance argument is also an empirical diagnostic, not a bound on the regional supremum.

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. The bound-and-tightness and coupling-sweep experiments 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. The player-count scaling experiment 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 (Efron, 1979), Wilcoxon signed-rank tests (Wilcoxon, 1945) and Holm's step-down correction (Holm, 1979). The full protocol appears in Appendix F.

The learner and the auditor use different information. For each seed and player, training uses 240 pairs (x_{-i}, x_i) obtained from sixty four-step trajectories of noisy simultaneous best-response play. In these experiments h_i is the identity on the concatenated rival action x_{-i}, so each response head directly learns the map $x_{-i} \mapsto x_i$. Each trajectory starts from an independent uniform draw from the game's bounding box, projected onto the feasible set; these restarts broaden the input support relative to one long trajectory. The multiple-equilibrium experiment uses its game-specific multimodal response process. The learner sees actions but no costs or gradients. The theorem is otherwise training-agnostic. Training loss does not determine either regional envelope component: $\bar{M}_i$ concerns the true game, while $\bar{\delta}_i$ measures uniform agreement between that game and the trained response map. Cost-aware training is compatible with the certificate and may reduce $\bar{\delta}_i$; we study the complementary problem of independently auditing an already trained model.

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.

For each profile we compare three scalar upper bounds. The *model-free candidate* is $B_{\mathrm{MF}} = \max_i \bar{M}_i D_i$. The *learned-residual candidate* is $B_{\mathrm{LR}} = \max_i (\bar{M}_i + D_i)(\rho_i + \bar{\delta}_i)$. The reported blockwise certificate in (11) takes the smaller candidate separately for each player before maximizing. A *direct-probe candidate* queries the true gradient at the evaluated profile, computes the projected-gradient residual r_i, and reports $B_{\mathrm{DP}} = \max_i (M_i + D_i) r_i$. We measure tightness as an upper bound divided by exact exploitability; one is exact and larger values are looser.

Across all 420 game–model–seed configurations in the bound-and-tightness experiment, including the 140 shared-constraint diagnostics outside the theorem's scope, the model-free candidate determined the reported blockwise certificate. Its median tightness was 108.7, compared with 297.9 for the learned-residual candidate and 134.7 for the direct-probe candidate. Comparing the reported/model-free certificate with the direct-probe candidate gave a median paired difference of -0.001165 in $\log_{10}$ tightness ($p_{\mathrm{Holm}} = 0.1519$), so the test did not detect a difference. The direct-probe candidate was, however, tighter than the learned-residual candidate in all 420 paired runs. The model-free candidate is the standard convexity/Cauchy–Schwarz bound and is not a contribution of this paper. Thus this experiment does not show that learning tightens a freshly evaluated bound. Its role is to test validity and motivate the separate regional-audit experiment, where one audit is reused without querying the game at each later profile.

Playerwise scaling. The scaling study varies the player count N 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 candidate. The fixed 240-transition budget makes the largest fits underdetermined, which can inflate learned-map residuals but not the model-free candidate.

Regional auditing creates the amortization regime. At $\eta / \operatorname{diam} \Omega = 0.025$, the median cover contains 20 points on the one-dimensional segment and 8000 points on the three-dimensional box. The learned-residual candidate is selected for 0.5742 and 0.7266 of held-out profiles, respectively. The regional certificate is 1.788 and 1.725 times the pointwise model-free certificate obtained from a fresh gradient query. Every tested profile satisfies the audited envelope, while the covering cost grows rapidly with region dimension. Fig. 3 specifies the game, response model, regions, and evaluation samples. Three-player quadratic product-set game with two-dimensional actions and a two-layer 32–32 tanh response model; medians over 10 seeds and 256 profiles drawn uniformly from each region. The segment joins the learned fixed point to the projection of $x^\star + 0.2Du$, where D is the diagonal length of the joint bounding box and u is a seeded random unit direction. An intrinsic-dimension-m coordinate box varies its first $m \in \{1, 2, 3\}$ coordinates within plus or minus 5% of the corresponding coordinate range around the learned fixed point, clipped to the action bounds; all regions are embedded in the six-dimensional joint action space. (a) Median Euclidean η-cover size, equal to the gradient-block-query break-even against repeated direct probes under the stated accounting. (b) Regional certificate divided by the pointwise model-free certificate obtained from a fresh gradient query; values above one quantify the tightness paid for reuse.

The regional-amortization experiment is a gradient-oracle audit, not an implementation of the noisy scalar-cost oracle in Section 4. The quadratic pseudo-gradient is affine, giving the exact modulus $L_{M,i} = \|A_{i,:}\|_2$. For $G_i = x_i - f_i(x_{-i})$, the audit uses the valid bound $L_{\delta,i} = \sqrt{1 + L_{f_i}^2} + \|A_{i,:}\|_2$, where L_{f_i} is the responder's rigorous weight-derived Lipschitz bound. The midpoint grid uses

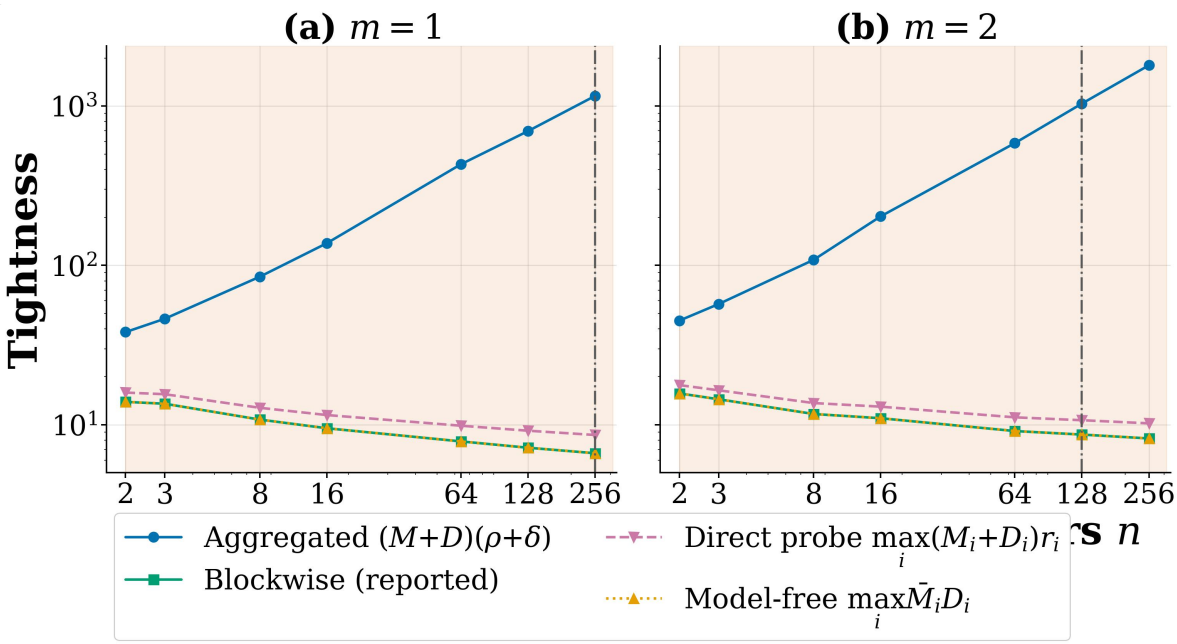

Fig. 2: **Joint constants degrade with the player count; playerwise ones do not.** Coupled quadratic game, (a) scalar and (b) two-dimensional actions, medians over 20 seeds, log–log, lower is tighter.

$n_j = \lceil s_j\sqrt{m}/(2\eta) \rceil$ points on an edge of length s_j; its cell half-diagonal is at most η, so it is a Euclidean cover of the embedded m-dimensional region. No finite-difference B_3, σ, or τ enters this experiment. The conversion to $2d|\mathcal{Z}|$ scalar cost evaluations is an extrapolation under Proposition 4.1, not a measured cost-oracle experiment.

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 70, 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 learned fixed point establishes model self-consistency, not low regret in the true game. Our certificate connects the two by combining a model-only residual with independently established regional bounds on true gradients and environment–model discrepancy. Any valid envelope yields profilewise exploitability guarantees without contraction, monotonicity, or equilibrium uniqueness. The finite cost-query construction is narrower: it requires externally certified regularity and noise assumptions and has a covering cost exponential in the audited region's intrinsic dimension. Its practical regime is therefore a structured region that is audited once and reused many times. In the

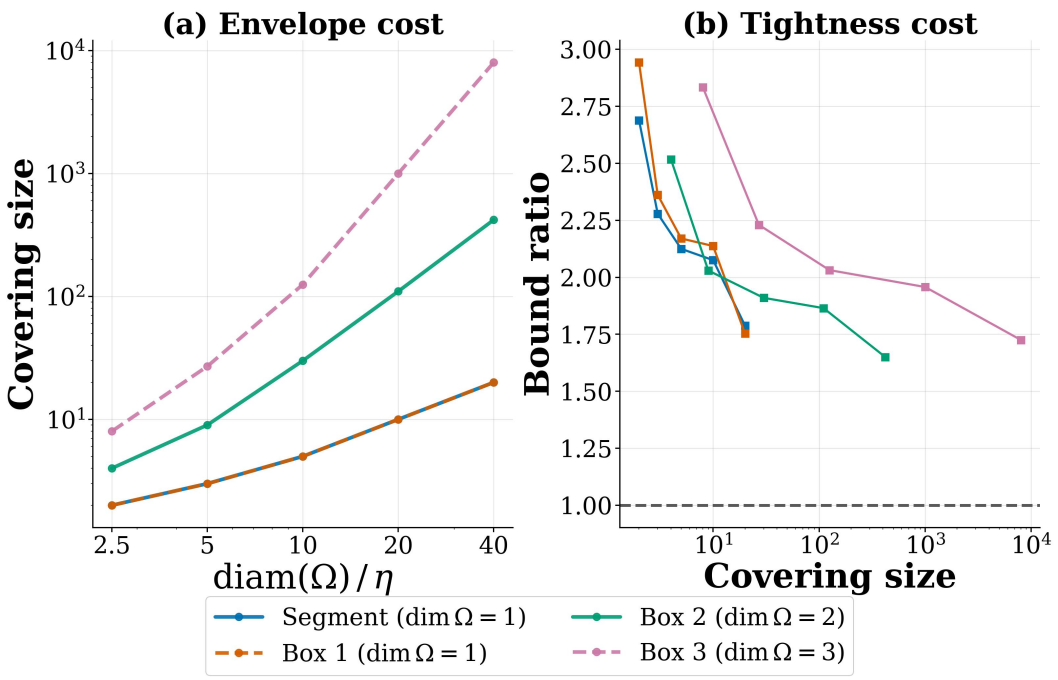

Fig. 3: Amortizing a regional envelope.

pointwise experiments, the standard model-free candidate bound all 280 configurations; the learned map's value appeared only in the regional regime, where it exchanged a looser bound and substantial upfront cost for later environment-query-free evaluation. Without the external assumptions, the finite-query procedure is diagnostic rather than a uniform certificate. Without playerwise convexity, a vanishing projected-gradient residual can coexist with exploitability $\frac{1}{4}$.

USE OF AI TOOLS

In this work, generative AI assistance was used to augment our literature review through search and retrieval of related work, edit prose, and edit and audit the experiment code. We did not use generative AI tools to produce the theoretical contributions or empirical claims. We 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, and artifacts produced with the aid of generative AI.

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.

B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979. doi: 10.1214/aos/1176344552.

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.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. doi: 10.2307/4615733.

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.

Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945. doi: 10.2307/3001968.

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.

APPENDICES

CONTENTS

A Related Works and Distinctions **17**

B Discussion and Limitations **20**

C A Catalog of (Deferred) Proofs **23**

D Constants and Rates of the Blockwise Bound **25**

D.1 Average prediction error does not control the consistency error 25

D.2 Sharpness of the blockwise constants 25

D.3 The quadratic regime 26

E Response Models and Recovering Multiple Equilibria **26**

F Experimental Setup **26**

G Robustness and Scope **29**

G.1 Calibration of the central-difference estimator 29

G.2 Convexity failure 29

A Related Works and Distinctions

The Case for Certification No uncoupled dynamic converges to Nash equilibrium across a family of games containing a neighborhood of a given game (Hart & Mas-Colell, 2003), and their witness lies inside the hypotheses adopted here: two players on unit disks, costs smooth and strongly convex in each player's own action, a unique strict equilibrium. As stated in Hart & Mas-Colell (2003), converging to equilibrium is a more stringent requirement than being in equilibrium. Certificates address being in equilibrium.

Learned response maps. Response-based models of interaction replace a player's optimization with a map fitted to its observed actions. Scarabaggio et al. (2024) introduced the projected observer–head architecture, established via Brouwer's theorem that a fixed point exists, and showed that such a fixed point solves a variational inequality; uniqueness and convergence of the response iteration follow from an aggregate contraction condition $\left(\sum_i \gamma_i^2\right)^{1/2} < 1$, relaxed to a Krasnosel'skiĭ iteration at the nonexpansive boundary. Their guarantees concern the learned game. The distinction drawn in this paper is that Theorem 3.1 bounds exploitability in the game that generated the data, at a profile that need not be an exact fixed point, and needs no condition on γ_i. A subsequent treatment from the same group combines projected maps, existence, and aggregate contraction in the same way (Mignoni, 2025, Chapter 6). Their variational inequality is stated in the learned map and is equivalent to the fixed-point condition itself, so that it constrains the game that generated the data nowhere. Discarding the learned component, (Scarabaggio et al., 2026) characterizes stationary profiles of the true costs as fixed points of the blockwise projected gradient and establish existence and homotopy persistence by Lefschetz–Nielsen theory. We measure the blockwise residual r_i of the map.

A different ε. In the purification of Harsanyi (1973) the parameter scales a perturbation of the payoffs and the equilibrium is exact in the perturbed game; in quantal response (McKelvey & Palfrey, 1995) and in perturbed best-response learning (Hofbauer & Hopkins, 2005) it scales a smoothing and the equilibrium is an exact fixed point of a *modified* operator, whose exploitability in the original game is in general bounded away from zero. The certificate here bounds regret in the unmodified game, so convergence to a perturbed equilibrium implies nothing about (2), and conversely.

Decision-dependent and performative games. When the data distribution reacts to the deployed model, the fixed point of repeated retraining becomes the solution concept. Perdomo et al. (2020) introduced performative prediction and the repeated-risk-minimization fixed point, and Narang et al. (2023) extended it to several players, proving existence, uniqueness, and a linear rate for performatively stable equilibria when a sensitivity modulus divided by a strong-convexity modulus falls below one. That ratio plays the role q plays here, and the resulting sensitivity bound carries the same $1/(1-\cdot)$ constant that Remark B.3 dispenses with. Outside the contractive regime the dynamics are not merely unanalyzed but genuinely irregular: Piliouras & Yu (2023) exhibit chaotic multi-agent retraining, and Dong et al. (2023) study regions of attraction where contraction holds only locally. These results motivate a certificate that does not degrade as the modulus approaches one.

Certified operators and Lipschitz networks. A parallel literature constrains what a network can do rather than enveloping what it did. Monotone operator equilibrium networks obtain well-posed implicit layers by construction (Winston & Kolter, 2020), and Jafarpour et al. (2021) certify implicit networks through a non-Euclidean, weighted one-sided Lipschitz condition that is strictly weaker than a Euclidean contraction; Davydov et al. (2022) develop the contraction theory behind it. Lipschitz bounds are enforced architecturally by direct parameterization (Wang & Manchester, 2023) and unified across several such constructions by Havens et al. (2023), or estimated post hoc by semidefinite relaxation (Fazlyab et al., 2019), exact computation being NP-hard (Virmaux & Scaman, 2018). Two cautions from this literature shape our treatment. Products of layerwise spectral norms bound the joint operator norm only loosely, and the Clarke chain rule for a composition holds as an inclusion rather than an equality (Clarke, 1983), so a layerwise regularity certificate is unsound. Automatic differentiation compounds the problem: it returns an element of a conservative field rather than of the Clarke Jacobian, and the two differ on activation and projection boundaries (Bolte & Pauwels, 2021). We therefore report sampled Jacobian quantities as lower-bound diagnostics and never as certificates.

Error bounds and gap functions. The inequality behind Lemma 3.1 is part of the error-bound theory for finite-dimensional variational inequalities (Facchinei & Pang, 2003, Ch. 6)(Pang, 1997). Gap and merit functions convert a variational inequality into an optimization problem and supply computable residuals: the regularized gap of Fukushima (1992), the gap-function class of Larsson & Patriksson

(1994), the D-gap of Peng (1997), the global error bounds of Yamashita & Fukushima (1997), and the Nikaido–Isoda reformulations for generalized games of von Heusinger & Kanzow (2009). These results run predominantly from a residual to *distance-from-solution* and require strong monotonicity for the constant to exist; our use runs instead to *exploitability*, which is why compactness suffices. Sensitivity of solutions under perturbation of the defining map is classical, through Robinson's strong regularity (Robinson, 1980), its characterization on polyhedral sets (Dontchev & Rockafellar, 1996), and the bound of Dafermos (1988) in which the contraction modulus is replaced by a strong-monotonicity modulus of the environment.

Certifying a limit vs. a profile. Mertikopoulos & Zhou (2018) work in our setting, with compact convex product action sets and payoffs only C^1 in each player's own action. However, what they certify is a limit point of dual averaging which is Nash only if the iterates converge and the game is pseudo-concave. Their global results require that the Nash set be variationally stable, the pseudo-gradient be monotone, or the game to admit a concave potential. Their one quantitative gap estimate is an $\mathcal{O}(1/\sqrt{n})$ rate for an ergodic average that they note may behave differently from the played sequence. Variational stability is adoptable and would furnish isolation and asymptotics. However, it cannot furnish a computable certificate because it relates $\Phi(x)$ to an unknown equilibrium. Thus, with that equilibrium unknown it yields no computable number, and verifying it requires information about the game's Hessian at the equilibrium one is trying to locate. For C^2 costs it is moreover equivalent to local asymptotic stability (Mazumdar et al., 2020), so that it excludes precisely the equilibria that gradient-based learners avoid from almost every initialization — the profiles most in need of an external certificate.

Continuous-time equilibrium seeking. A parallel literature drives dynamics to an equilibrium rather than certifying a profile. The survey of Hu et al. (2022) collects the consensus-based, passivity-based and gradient-free designs, and Belgioioso et al. (2022) give the operator-splitting counterpart for coupled constraints. Two features bear on this work. Their convergence certificates are Lyapunov functions of the form $\frac{1}{2}\|x - x^\star\|^2$, evaluable only once $x^\star$ is known, which is unavailable when the game is not; and their hypotheses begin at monotonicity of Φ and strengthen from there. Gap and merit functions for variational inequalities are cited in neither survey, and no ε-equilibrium notion appears in either. The single quantitative rate in Belgioioso et al. (2022) is an ergodic $\mathcal{O}(1/k)$ bound on a fixed-point residual, and their operator-to-game link holds at an exact zero of that residual, so that a positive residual is converted into a statement about the game nowhere in this line.

Identifying a game from observed play. Whether observed actions pin down a game is the question our consistency error confronts. Gürkan & Pang (2009) give conditions under which equilibria of approximating games converge to equilibria of the true game, via a multi-epiconvergence notion; the statements are asymptotic and contain no explicit constant, and the authors note the connection between their regularized Nikaido–Isoda function and variational gap functions without developing it quantitatively. Bemporad & Tatarenko (2026a) learn solution maps of parametric generalized games by training against the Nikaido–Isoda gap rather than regressing on precomputed equilibria, which requires only best-response data, and Bemporad & Tatarenko (2026b) instead secure monotonicity of the learned game by construction. Bergerault et al. (2026) prove the sharpest negative result in this direction: characterizing the Nash gap at a fixed measure-matching error is hard, and exact matching of observed play can still leave a large Nash gap. Their positive results impose structure on the expert, such as dominant-strategy behavior or continuity of best responses around the expert, to transfer distributional imitation error into a Nash gap. Our assumptions constrain different objects: playerwise convexity constrains the true one-shot game, and a regional envelope directly controls the learned operator's discrepancy from its pseudo-gradient. Neither assumption set implies the other. This distinction is why **??** measures operator consistency rather than distributional fit, and why the environment access of Section 4 is stated as a cost rather than assumed away. Payoff-based equilibrium learning supplies the access model we adopt, in which players evaluate costs but do not differentiate them (Tatarenko & Kamgarpour, 2024).

Generative responders. Our set-valued responders draw on conditional generative models, which represent multimodal conditional distributions rather than conditional means. Proposition 3.2 shows why that distinction is not cosmetic for equilibrium computation where playerwise convexity fails. Where it holds, our measurements in Section 5 attribute recovery of multiple equilibria to model capacity rather than to a multimodal likelihood; we therefore do not argue that generativity is required. We are not aware of prior work analyzing diffusion policies, flow matching, or vision–language–action models as response correspondences in a game, though we have not surveyed those

literatures systematically and state this as an impression rather than a priority claim. Two scope limits apply when the certificate is read alongside them. Theorem 3.1 constrains the *environment* through Assum. 2.2, so that continuity of the learned map is not by itself sufficient; and under playerwise convexity each best-response set is convex, so the multimodality that motivates a generative responder must come from multiple equilibria rather than from a multimodal best response. Certifying a genuinely set-valued response, where the residual is defined for a sampler rather than for a map, remains open, as does certifying a generalist policy deployed simultaneously as several players, whose playerwise responses share parameters.

B Discussion and Limitations

The regional audit has two distinct burdens. First, Proposition 4.1 requires externally justified regularity and noise bounds; finite cost queries do not infer them. Without those inputs, Theorem 3.1 remains valid for an envelope established by another method, but Proposition 4.1 does not construct one. Second, even with valid inputs, a uniform audit requires a finite covering. For an m-dimensional box, its size grows as $O((\operatorname{diam}\Omega/\eta)^m)$. The construction is therefore limited to regions with small covering numbers, such as a segment or a low-dimensional deployment neighborhood with controlled geometry.

In the amortization study, the tanh-MLP responder uses a three-player quadratic game with two-dimensional actions, ten seeds, and 256 held-out profiles in each audited region. The segment is one-dimensional. The three-dimensional box is embedded in the six-dimensional joint action space by varying only its first three coordinates, with distances and the η-cover computed in the induced Euclidean geometry. At $\eta/\operatorname{diam}\Omega = 0.025$, the median covering sizes are 20 on the segment and 8,000 on the embedded box. E9 actually uses an exact gradient oracle: one query returns one player's own-gradient block, so a pointwise audit costs N block queries and a regional audit costs $N|\mathcal{Z}|$. Under the scalar cost-oracle model of Section 4, central differences would instead cost $2d$ evaluations per point and $2d|\mathcal{Z}|$ per cover. Either accounting gives an oracle-call break-even of $|\mathcal{Z}|$ later profiles when the same stencil is used. This comparison is conditional on already having the regularity constants and excludes cover construction, storage, training, model evaluation, and any rollout replication used to produce a cost value. The ratio of the regional bound to the fresh pointwise bound is computed first as a median over held-out profiles and then as a median over seeds; it is 1.788 on the segment and 1.725 on the embedded box. The construction buys reuse without new environment queries; it does not beat a fresh query in tightness. Replacing the covering by samples provides an empirical diagnostic unless a separate distributional guarantee is proved.

Coupled feasibility sets are excluded. Assum. 2.1 requires a product action set, and the requirement is load-bearing twice: it makes the projection act blockwise, and it permits one player's action to be varied while the others are held fixed. Generalized games with shared constraints therefore fall outside every result here. Our coupled-constraint experiment shows what the certificate does when its own hypotheses are violated, and is not a claim of coverage.

Playerwise convexity is necessary, and it makes the set-valued question unanswerable here. Proposition 3.2 shows that the blockwise residual certifies nothing once Assum. 2.2 fails. The assumption also removes the motivation for generative responders: under playerwise convexity each best-response set is convex, so that responses of the bimodal kind in Proposition 3.2 are excluded by hypothesis. The multimodality that survives arises from multiple *equilibria* of a playerwise-convex game, which non-monotone coupling produces readily. Our evaluation on those games (Section 5) reports what a *conditional-mean* responder recovers, since the residual ρ_i of (9) requires a single-valued evaluation and the generative families expose one only through their mean map. It therefore speaks to neither generativity nor capacity: the comparison is between mean maps, and the generative heads carry roughly four times the parameters and still do worse. What it does show is that a linear regressor fitted to play pooled across basins lands between them on a fraction 0.724 of instances, while nonlinear responders do not. Whether a certificate can be given for a genuinely set-valued response we leave open, and the open half is the metric one rather than the topological one. Existence for a response correspondence with nonempty closed convex values is classical (Kakutani, 1941; Facchinei & Kanzow, 2010); what is missing is a *residual*, since no notion of approximate fixed-pointness of a correspondence appears in the equilibrium-seeking literature we surveyed, and ρ_i as defined here requires a single-valued evaluation.

The quadratic rate is not a computable certificate. Proposition D.2 governs the interior regime but requires distances to an $\arg\min$, hence N constrained convex solves in the unknown costs. We report it as a rate and measure it only where ground truth is available. A computable surrogate would need interiority certified from probe data together with a curvature modulus, which we have not established.

Sampled Jacobians are diagnostics, not certificates. Where the experiments report operator norms of learned maps, rigorous bounds are computed from weights and sampled quantities are labeled as lower-bound diagnostics. The distinction is not pedantic: the Clarke chain rule for a composition holds only as an inclusion (Clarke, 1983), and automatic differentiation returns an element of a

conservative field rather than of the Clarke Jacobian, the two differing on activation and projection boundaries (Bolte & Pauwels, 2021). No certificate in this paper reads a sampled Jacobian.

Remark B.1 (Why certify rather than solve). The same oracle that supplies an envelope would also drive projected gradient play, so that the certificate must earn its place against simply solving the game. It does so on three counts. (i) An envelope is paid once and then certifies every profile in Ω for the price of a forward pass, whereas solving pays along every trajectory and yields one profile. (ii) Projected gradient play carries no convergence guarantee at our hypotheses, and the failure arrives one rung earlier than the absence of monotonicity: under monotone Φ the Lyapunov derivative is still nonpositive, but the set on which it vanishes gives only $\langle \Phi(x) - \Phi(x^\star), x - x^\star \rangle = 0$, which does not force $x = x^\star$, so the invariance argument does not close (Hu et al., 2022). Below monotonicity the dynamics are irregular rather than merely unanalyzed, and multi-player retraining outside the contraction regime admits chaotic orbits (Piliouras & Yu, 2023). (iii) The learned map is often the deliverable for reasons unrelated to equilibrium computation, such as simulating counterfactual play, so the question is whether the model one already holds is trustworthy where it is evaluated.

Remark B.2 (What the certificate does not claim). The bound (11) concerns profiles in Ω. It does not assert that $\mathrm{Fix}(F)$ is a singleton, nor that a certified profile approximates any particular equilibrium. For a nonexpansive F the fixed-point set is nonempty, closed, and convex, but need not be a singleton, as $F = \mathrm{Id}$ shows. We therefore write *a* behavioral equilibrium throughout, never *the* behavioral equilibrium.

Remark B.3 (Behavior at the contraction boundary). Classical fixed-point perturbation controls equilibrium displacement by $\varsigma/(1-q)$, for a uniform map error ς and a map of contraction modulus $q < 1$ (Bauschke & Combettes, 2017), and that constant diverges as $q \to 1$. The certificate (11) contains no q. Certifying a nonexpansive or expansive learned map is thus a regime the bound already covers, which is the operative difference from the aggregate-contraction condition of Scarabaggio et al. (2024) and from the sensitivity-to-curvature ratio governing performatively stable equilibria in Narang et al. (2023).

Remark B.4 (Regularity of the learned map certifies nothing). Let $F \equiv \hat{x}$ be constant for an arbitrary $\hat{x} \in X$. That map is nonexpansive and $\rho_i(\hat{x}) = 0$ for every player, while $\varepsilon(\hat{x})$ is as large as the game permits. Its induced operator satisfies $G_i(\hat{x}) = 0$, so $\delta_i(\hat{x}) = M_i(\hat{x})$ and the learned-residual candidate of (11) reads $(\bar{M}_i + D_i)\bar{M}_i$, which exceeds the model-free candidate $\bar{M}_i D_i$. The constant map is therefore excluded by the model-free bound rather than by the envelope, and the certificate reports nothing better than the trivial bound for it. Regularity of a learned map carries no information about the game.

Remark B.5 (Remark on Proposition 4.1.). A strict contraction F with modulus $q < 1$ gives geometric convergence of the Picard iteration $x^{k+1} = F(x^k) = \Pi_X\big(x^k - g(h(x^k))\big)$. To reduce the fixed-point residual below ϵ, the number of value queries scales as

$$\mathcal{O}\left(\frac{\log(1/\epsilon)}{1-q}\right)$$

when q is close to one. Equivalently, if $q = 1 - \varepsilon$ with $0 < \varepsilon \ll 1$, the cost is $\mathcal{O}(\varepsilon^{-1}\log(1/\epsilon))$. At the nonexpansive boundary $q = 1$, this contraction argument disappears. General value-oracle lower bounds for Brouwer fixed-point computation are exponential in the dimension and in the number of digits of accuracy, under the corresponding Lipschitz assumptions (Hirsch et al., 1989); related reductions yield query lower bounds for equilibrium computation (Babichenko, 2016). These results provide complexity context, rather than a bound on the exploitability certificate (11): our result certifies a profile without assuming contraction, whereas Picard iteration seeks a fixed point under contraction.

Remark B.6 (Sampling changes the guarantee). Replacing the covering by independent draws from a distribution ν on Ω removes the covering cost but also removes deterministic uniform coverage. Under a fixed deployment distribution and exchangeability, a separate order-statistic or tolerance theorem can provide a fresh-draw or quantile guarantee; such a result still does not bound the regional supremum. Because we prove no such distributional theorem here, we treat sample maxima only as empirical diagnostics.

The boundary-stencil count is not a measurement. The fraction of coordinates requiring the one-sided stencil of Remark 4.1 is 0 *by construction* rather than by measurement: the probe used in our runs is an unconstrained central difference that never switches stencil, so the number could not have come out otherwise. The companion quantity we can measure is the fraction of probe steps

that leave the action set, which is 0 at every radius on the grid. A feasibility-respecting probe would therefore not have switched either, but that is a property of these arenas and not evidence about the remedy.

C A CATALOG OF (DEFERRED) PROOFS

This appendix catalogs the proofs of the results of Section 3 and Section 4, in order.

Proof of Lemma 3.1. Fix a player i and let $y_i^\star \in \arg\min_{y_i \in X_i} J_i(y_i, x_{-i})$, which exists because X_i is compact by Assum. 2.1 and $J_i(\cdot, x_{-i})$ is continuous by Assum. 2.2. Convexity of $J_i(\cdot, x_{-i})$ implies that $J_i(y_i^\star, x_{-i}) \geq J_i(x) + \langle \Phi_i(x), y_i^\star - x_i \rangle$. Hence, player i's regret obeys

$$J_i(x) - J_i(y_i^\star, x_{-i}) \;\leq\; \big\langle \Phi_i(x),\, x_i - y_i^\star \big\rangle. \tag{19}$$

Set $p_i = \Pi_{X_i}\big(x_i - \Phi_i(x)\big)$ and $w_i = x_i - p_i$, so that $\|w_i\| = r_i(x)$ by (7). Projecting $x_i - \Phi_i(x)$ onto X_i implies $\langle x_i - \Phi_i(x) - p_i,\, y_i - p_i \rangle \leq 0$ for every $y_i \in X_i$, the variational characterization of the projection (Facchinei & Pang, 2003, §1.5); equivalently, p_i solves $\mathrm{VI}\big(X_i,\, u \mapsto u - (x_i - \Phi_i(x))\big)$ in the sense of Definition 2.1. It is easy to verify that $\langle \Phi_i(x),\, p_i - y_i \rangle \leq \langle w_i,\, p_i - y_i \rangle$; applying the Cauchy–Schwarz inequality with $\|p_i - y_i\| \leq D_i := \operatorname{diam} X_i$, since $p_i, y_i \in X_i$, gives

$$\langle \Phi_i(x),\, p_i - y_i \rangle \;\leq\; r_i(x)\, D_i \qquad \text{for every } y_i \in X_i. \tag{20}$$

Splitting (19) at p_i as $\langle \Phi_i(x), x_i - y_i^\star \rangle = \langle \Phi_i(x), w_i \rangle + \langle \Phi_i(x), p_i - y_i^\star \rangle$, then bounding the first term by $M_i(x) r_i(x)$ with the Cauchy–Schwarz inequality and the second by (20), we find that the bound on the player i's regret is $\big(M_i(x) + D_i\big) r_i(x)$, which is the first entry of (8). Separately, (19) and Cauchy–Schwarz give $J_i(x) - J_i(y_i^\star, x_{-i}) \leq M_i(x) \|x_i - y_i^\star\| \leq M_i(x) D_i$. Both bounds apply to the same player, hence player i's regret is at most their minimum, and (8) follows by taking the maximum over players. Writing $a_i = \big(M_i(x) + D_i\big) r_i(x)$ and $b_i = M_i(x) D_i$, the inequality $\max_i \min\{a_i, b_i\} \leq \min\{\max_i a_i, \max_i b_i\}$ shows that taking the minimum after the maximum is valid but never tighter, and strict exactly when no single player has both $a_i \geq m$ and $b_i \geq m$ for $m = \min\{\max_j a_j, \max_j b_j\}$. □

Proof of Lemma 3.2. Since $x_i - G_i(x) = f_i(x_{-i}) \in X_i$, projection fixes this point. Inserting it into $r_i(x) = \|x_i - \Pi_{X_i}(x_i - \Phi_i(x))\|$ and using the triangle inequality gives $r_i(x) \leq \rho_i(x) + \|\Pi_{X_i}(x_i - G_i(x)) - \Pi_{X_i}(x_i - \Phi_i(x))\|$. Nonexpansiveness of Π_{X_i} bounds the second term by $\|G_i(x) - \Phi_i(x)\| = \delta_i(x)$, proving the claim. □

Proof of Proposition 3.1. The reverse triangle inequality and $\rho_i(\hat{x}) = \|G_i(\hat{x})\|$ give $\rho_i(\hat{x}) + \delta_i(\hat{x}) \geq \|\Phi_i(\hat{x})\| = M_i(\hat{x})$. Multiplying by $M_i(\hat{x}) + D_i \geq 0$ yields the first inequality of (12), and the second is strict whenever $M_i(\hat{x}) > 0$, since then $M_i(\hat{x})^2 > 0$. □

Proof of Theorem 3.1. Fix $\hat{x} \in \Omega$. By (10), $M_i(\hat{x}) \leq \bar{M}_i$ and $\delta_i(\hat{x}) \leq \bar{\delta}_i$, and Lemma 3.2 gives $r_i(\hat{x}) \leq \rho_i(\hat{x}) + \bar{\delta}_i$. Both entries of (8) are nondecreasing in M_i and in r_i on the nonnegative orthant $\mathbb{R}_+$ since $D_i \geq 0$. Hence, substituting the enveloped upper bounds preserves each inequality. Applying Lemma 3.1 playerwise gives (11). The envelope is fixed before $\hat{x}$ is chosen, so the conclusion holds simultaneously for every $\hat{x} \in \Omega$. □

Proof of Corollary 3.1. Dividing $(\bar{M}_i + D_i)(\rho_i + \bar{\delta}_i) < \bar{M}_i D_i$ by it gives the first inequality of (13). For the second, $\bar{M}_i D_i / (\bar{M}_i + D_i) \leq \bar{M}_i$ because $D_i \leq \bar{M}_i + D_i$, and $\bar{M}_i D_i / (\bar{M}_i + D_i) \leq D_i$ because $\bar{M}_i \leq \bar{M}_i + D_i$; hence it is at most $\min\{\bar{M}_i, D_i\}$. □

Proof of Proposition 3.2. Differentiating (14) gives $\Phi_i = x_i\big(x_i^2 - w_i(x_{-i})\big)$, which vanishes at $x = 0$ for every x_{-i}. Since $0 \in X_i = [-c, c]$, the projection fixes it, so $r_i(0) = \|\Pi_{X_i}(0)\| = 0$ for every player. Each $J_i(\cdot, x_{-i})$ is even in x_i with minimizers $\pm\sqrt{w_i(x_{-i})}$ and minimum value $-\frac{1}{4} w_i(x_{-i})^2$, while $J_i(0, x_{-i}) = 0$. At the origin $w_i(0) = 1$, hence every player's regret is $\frac{1}{4}$. The condition on c places $\pm\sqrt{w_i}$ inside X_i for every $x_{-i} \in X_{-i}$. □

Proof of Proposition 4.1. Fix a player i, a covering point $z \in \mathcal{Z}$ and a coordinate k of player i's block, and write $\varphi(t) = J_i(z + t e_{i,k})$, so that $\varphi'(0) = [\Phi_i(z)]_k$ is the quantity (15) estimates. The assumed extension makes φ three times continuously differentiable on $[-\tau, \tau]$, hence Taylor's

theorem with Lagrange remainder gives $\varphi(\pm\tau) = \varphi(0) \pm \tau\varphi'(0) + \frac{1}{2}\tau^2\varphi''(0) \pm \frac{1}{6}\tau^3\varphi'''(\xi_\pm)$ for some $|\xi_\pm| \le \tau$. Subtracting the two expansions cancels the even orders, so that

$$\frac{\varphi(\tau) - \varphi(-\tau)}{2\tau} \; - \; \varphi'(0) \; = \; \frac{\tau^2}{12}\big(\varphi'''(\xi_+) + \varphi'''(\xi_-)\big). \tag{21}$$

Since $\varphi''' = \partial^3 J_i / \partial x_{i,k}^3$ is bounded by B_3 on the probe segment, the right side of (21) is at most $\frac{1}{6}B_3\tau^2$ in magnitude. Stacking the n_i coordinates, player i's bias vector has norm at most $\frac{1}{6}B_3\tau^2\sqrt{n_i}$.

For the noise, a difference of two independent σ-sub-Gaussian variables is $\sqrt{2}\sigma$-sub-Gaussian, so dividing by 2τ makes each difference quotient s-sub-Gaussian about its mean with $s = \sigma/(\sqrt{2}\tau)$. There are $d|\mathcal{Z}|$ scalar estimates in all, one per coordinate and covering point, and each has two-sided tail $2\exp(-t^2/2s^2)$. Equating $2d|\mathcal{Z}|\exp(-t^2/2s^2)$ with α and solving for t, a union bound gives that every coordinatewise deviation is at most $s\sqrt{2\log(2d|\mathcal{Z}|/\alpha)} = \frac{\sigma}{\tau}\sqrt{\log(2d|\mathcal{Z}|/\alpha)}$ simultaneously, with probability at least $1-\alpha$; the $\sqrt{2}$ in s cancels the one from the tail. On that event player i's noise vector has norm at most $\frac{\sigma}{\tau}\sqrt{n_i\log(2d|\mathcal{Z}|/\alpha)}$.

Hence, on the same event and for every $z \in \mathcal{Z}$ at once, the triangle inequality bounds $\delta_i(z) = \|G_i(z) - \Phi_i(z)\|$ by $\|G_i(z) - \widehat{\Phi}_{\tau,i}(z)\| + E_i(\tau,\alpha)$ and bounds $M_i(z)$ by $\|\widehat{\Phi}_{\tau,i}(z)\| + E_i(\tau,\alpha)$. Finally, given $x \in \Omega$, choose $z \in \mathcal{Z}$ with $\|x - z\| \le \eta$. Lipschitzness gives $\delta_i(x) \le \delta_i(z) + L_{\delta,i}\eta$ and $M_i(x) \le M_i(z) + L_{M,i}\eta$. Taking the two maxima over the covering yields (17). The coordinatewise union bound already ranges over all d coordinates and all covering points, so both envelope inequalities hold for every player on this one event. □

D CONSTANTS AND RATES OF THE BLOCKWISE BOUND

Three analytic supplements to Lemma 3.1 and Theorem 3.1: what an average prediction error fails to control, whether the coefficients can be shrunk, and when the rate in $\rho_i + \delta_i$ improves from linear to quadratic.

D.1 AVERAGE PREDICTION ERROR DOES NOT CONTROL THE CONSISTENCY ERROR

A small training loss leaves the operator-consistency error of Lemma 3.2 uncontrolled. The following one-player construction makes the failure explicit and elementary, and shows that the obstruction is a gap between an average and a supremum rather than anything about multi-player coupling.

Proposition D.1 (Small mean-squared error, displaced fixed point)**.** *Let $X = [0,1]$, let the reference response be $F_0 \equiv 0$, and for $0 < h < \frac{1}{4}$ let*

$$T_h(x) = \tfrac{1}{2} \max\big\{1 - |x - \tfrac{1}{2}|/h,\ 0\big\}. \tag{22}$$

Then under observations uniform on X, $\mathbb{E}\big|T_h(x) - F_0(x)\big|^2 = h/6 \to 0$ as $h \to 0$, while $\sup_{x\in X}\big|T_h(x) - F_0(x)\big| = \frac{1}{2}$ for every h. The map T_h retains the reference fixed point at 0 and acquires a second fixed point at $\frac{1}{2}$, whose fixed-point residual is zero and whose distance from the reference fixed point is $\frac{1}{2}$.

Proof. The triangle T_h is supported on $[\frac{1}{2} - h, \frac{1}{2} + h]$ with peak $\frac{1}{2}$ at $x = \frac{1}{2}$, so its supremum is $\frac{1}{2}$ and $\int_0^1 T_h^2 = h/6$ by direct integration of the squared triangular profile. Evaluating (22) gives $T_h(0) = 0$ and $T_h(\frac{1}{2}) = \frac{1}{2}$, hence both points are fixed. The residual at $\frac{1}{2}$ vanishes because $\frac{1}{2}$ is a fixed point. □

To read this as a statement about the consistency error of Lemma 3.2, use a two-player decoupled game on $X_1 = X_2 = [0,1]$ with $J_i(x) = \frac{1}{2}x_i^2$, hence $\Phi_i(x) = x_i$. Define the admissible response map $F(x_1, x_2) = (T_h(x_2), T_h(x_1))$. Then $G_i(x) = x_i - T_h(x_{-i})$ and $\delta_i(x) = T_h(x_{-i})$, so $\sup_X \delta_i = \frac{1}{2}$ while its squared expectation under the uniform distribution on X is $h/6$. Both $(0,0)$ and $(\frac{1}{2}, \frac{1}{2})$ are fixed points. On the full region X, $\bar{M}_i = D_i = 1$ and $\bar{\delta}_i = \frac{1}{2}$, so at the displaced fixed point the two certificate entries tie at 1. The bound itself still holds, since each player's regret there is $\frac{1}{8}$; what a small average error is compatible with is a displaced fixed point together with a regional learned-residual candidate that improves on nothing.

Three readings follow. An average prediction error of $h/6$ is compatible with a uniform map error of $\frac{1}{2}$, so that no bound on the former controls the latter without a uniform regularity hypothesis; the spurious fixed point at $\frac{1}{2}$ has zero residual, which is Remark B.4 in miniature and for a different reason; and the construction does not contradict any contraction result, since T_h has slope magnitude $1/(2h) > 1$ on its support. What is required to connect an average loss to the consistency error is a uniform statement, which is what the regional envelope of Proposition 4.1 supplies by probing a finite covering rather than by appealing to a training loss.

D.2 SHARPNESS OF THE BLOCKWISE CONSTANTS

Neither coefficient in (8) can be rescaled downward. One limitation bounds what follows and is worth stating before it rather than after: the family below has $\varepsilon = MD$ at every parameter value, so the model-free candidate binds throughout and the learned-residual candidate never does. What is established is therefore sharpness of the coefficients at profiles where a practitioner would report the model-free bound instead.

Take $N = 1$, $X = [0,1]$, $J(x) = ax$ and evaluate at $x = 1$, so that $M = a$, $D = 1$ and $r = \min\{1, a\}$. The attainment ratio $\varepsilon/[(M + D)r]$ equals $1/(1 + a)$ for $a \le 1$, which tends to one as $a \to 0$ and so forbids replacing D by cD with $c < 1$; and it equals $a/(1 + a)$ for $a \ge 1$, which tends to one as $a \to \infty$ and so forbids replacing M by cM. The argument excludes rescaling either coefficient and not a differently shaped constant.

D.3 THE QUADRATIC REGIME

Whether the rate in (11) is linear or quadratic turns on whether best responses lie in the relative interior of the action sets. Write $\mathrm{BR}_i(x_{-i}) = \arg\min_{y_i \in X_i} J_i(y_i, x_{-i})$.

Proposition D.2 (Quadratic certificate under interior best responses)**.** *Let Assum. 2.1 hold, let $J_i(\cdot, x_{-i})$ be differentiable with L_i-Lipschitz gradient $\Phi_i(\cdot, x_{-i})$ on X_i, and let some minimizer $b_i \in \mathrm{BR}_i(\hat{x}_{-i})$ lie in the relative interior of X_i for every player i. With $\rho_i = \rho_i(\hat{x})$ and $\delta_i^{\mathrm{BR}} \triangleq \|g_i(h_i(\hat{x}_{-i})) - b_i\|$,*

$$\varepsilon(\hat{x}) \;\le\; \max_{i \in \mathcal{N}} \tfrac{L_i}{2} \left(\rho_i + \delta_i^{\mathrm{BR}}\right)^2. \tag{23}$$

Proof. Fix a player i. Since $b_i \in X_i$ we have $\Pi_{X_i} b_i = b_i$, hence nonexpansiveness of Π_{X_i} gives $\|\hat{x}_i - b_i\| \le \rho_i + \delta_i^{\mathrm{BR}}$ as in Lemma 3.2. Minimality of b_i gives $\langle \Phi_i(b_i, \hat{x}_{-i}), y_i - b_i \rangle \ge 0$ for every $y_i \in X_i$. Because b_i lies in the relative interior of X_i, for each such y_i the point $b_i - t(y_i - b_i)$ also lies in X_i for all small $t > 0$, which supplies the reverse inequality; hence $\langle \Phi_i(b_i, \hat{x}_{-i}), y_i - b_i \rangle = 0$ for every $y_i \in X_i$, and in particular at $y_i = \hat{x}_i$. The descent lemma along the segment $[b_i, \hat{x}_i] \subseteq X_i$ then bounds player i's regret by $\frac{L_i}{2}\|\hat{x}_i - b_i\|^2$. Taking the maximum over players gives (23). □

Two scope statements belong with Proposition D.2. Relative interiority, rather than topological interiority, is the sharp hypothesis: at a relative-interior point of a lower-dimensional X_i the gradient need not vanish, but its inner product against every feasible direction does, which is all the proof uses. Playerwise convexity is not needed, hence (23) holds outside Assum. 2.2. What the refinement does require is a strictly stronger oracle than the envelope of Definition 3.1: δ_i^{BR} is a distance to an $\arg\min$, hence obtaining it means solving N constrained convex programs in the unknown costs. We therefore report (23) as the rate that governs the interior regime, not as a computable certificate, and Section 5 measures it only where ground truth is available.

E RESPONSE MODELS AND RECOVERING MULTIPLE EQUILIBRIA

Multiple equilibria and what recovers them. Non-monotone coupling makes multiple equilibria the common case even under Assum. 2.2: of 120 playerwise-convex instances, 113 carry at least two distinct Nash equilibria, verified by active-set enumeration to a worst residual of 1.42×10^{-14}. A linear regressor fitted to play pooled across both basins lands between them rather than at either, recovering both equilibria on a fraction 0.431 of instances. A unimodal nonlinear network recovers both on 0.862, against 0.224 for flow matching and 0.0517 for diffusion.

These numbers do not separate capacity from generativity, and we do not read them as doing so. Every family is evaluated through a single-valued map, which for the two generative families is a conditional mean rather than a sample, so the comparison is between mean maps and the question of set-valuedness is not posed. The parameter counts run the other way besides: the generative heads carry roughly four times the parameters per player of the unimodal network and still recover fewer equilibria. What the measurement does establish is narrower and still useful: a linear responder fitted across basins lands between them, and sufficient capacity avoids that failure. We record it because it severs any empirical benefit claim from the observation that set-valued responses sit inside the hypotheses of Theorem 3.1; the latter is a scope statement, not a performance one.

Set-valuedness where convexity fails. On the nonconvex witness of Proposition 3.2 the conditional-mean profile has exploitability 0.25 analytically while its variational residual is 0, so the residual alone certifies nothing there. Mode-selected samples reduce exploitability from 0.256 to 0.0286 for diffusion ($p_{\mathrm{Holm}} = 1.91 \times 10^{-6}$) and from 0.252 to 0.0175 for flow matching ($p_{\mathrm{Holm}} = 1.91 \times 10^{-6}$), over 20 paired seeds. The certificate's hypotheses fail on this game and we never evaluate the bound as a claim about it.

F EXPERIMENTAL SETUP

The studies named in Section 5 correspond to the modules of the released package as follows: the *tightness study* to `e1_tightness`, the *coupling sweep* to `e2_coupling`, the *dissociation study*

to `e3_incomparability`, the *set-valued witness* to `e4_set_valued`, the *multiplicity study* to `e5_multiplicity`, the *solver comparison* to `e6_solvers`, the *probe study* to `e7_probe`, the *scaling study* to `e8_scaling`, and the *amortization study* to `e9_amortization`.

Each game is chosen so that it isolates one hypothesis of Theorem 3.1 and supplies either analytic ground truth or a documented numerical oracle, so that a discrepancy is attributable to the bound rather than to an undocumented solver. We describe the five in turn, in each case stating what makes the instance hard and which claim it exercises.

Coupled quadratic: monotonicity on a knob. Set $Q_i = I$ and $A = I + \kappa(W + \frac{1}{2}Z)$ with W symmetric and Z antisymmetric, both with zero diagonal blocks, W rescaled so that $\lambda_{\min}(W) = -1$ and Z so that $\|Z\|_2 = 1$. Then

$$\tfrac{1}{2}\big(A + A^\top\big) = I + \kappa W, \qquad \text{hence} \qquad \lambda_{\min}\big(\tfrac{1}{2}(A + A^\top)\big) = 1 - \kappa, \tag{24}$$

so the monotonicity modulus is known analytically and the sweep crosses monotonicity exactly at $\kappa = 1$. The skew part keeps $A \neq A^\top$, so that the pseudo-gradient is not the gradient of any potential and the instance is not a potential game. This is the only family in which instances can be placed arbitrarily far on either side of the monotone boundary with the modulus known in closed form, which is what lets the coupling sweep attribute an outcome to non-monotonicity rather than to conditioning. It carries the claim that the certificate requires no monotonicity.

Capacitated Cournot: an active projection. Firms choose scalar quantities on $[0, \mathrm{cap}_i]$ with capacities alternating between 0.6 and 5.0, so that the tight capacities bind at equilibrium. This tests the projection-dependent part of the blockwise gap bound. A fresh true-gradient query can make the direct projected-gradient candidate $(M_i + D_i)r_i$ sharper than the simpler model-free candidate $M_i D_i$. Proposition 3.1 still rules out an improvement by the learned-residual candidate $(M_i + D_i)(\rho_i + \delta_i)$ when the model–game discrepancy is evaluated at that same profile.

Rosen shared constraint: outside the product assumption. Boxes intersected with a single coupled row $w^\top x \leq h$, so that player i optimizes over $\mathcal{X}_i(x_{-i}) = \{x_i \in [\mathrm{lo}_i, \mathrm{hi}_i] : w_i^\top x_i \leq h - \sum_{j \neq i} w_j^\top x_j\}$. This violates the product requirement of Assum. 2.1, so every result here is formally silent about it. We include it to record what the certificate does when its own hypothesis fails, and we report those cells as a stress diagnostic and never as coverage. Two consequences are worth naming: the best response is a projection onto a box intersected with a halfspace, which is computed by exact bisection on the single multiplier and so needs no inner solver; and D_i is conservative here, since it is the diameter of the box rather than of the coupled set.

Bimodal coordination: convexity removed. With J_i and w_i as in (14), $\partial^2 J_i / \partial x_i^2(0) = -w_i < 0$, so playerwise convexity fails. All 2^N sign patterns of $s = (1 - \kappa)^{-1/2}$ are equilibria in closed form, while the origin is a stationary point of Φ at which the variational residual vanishes and the exploitability is exactly $\frac{1}{4}$. This is the witness of Proposition 3.2. It tests the *necessity* of Assum. 2.2 and not the validity of the bound, and we never evaluate (11) as a claim about it.

Nonexpansive boundary: monotone yet not contractive. Let $J_i(x) = \frac{1}{2}(x_i - \sigma_i x_{i+1})^2$ cyclically with $\sigma_i = +1$ for $i < N - 1$ and $\sigma_{N-1} = -1$. Then $F(x) = Px$ with P a signed cyclic permutation, so that $P^\top P = I$ and the best-response modulus is exactly one: nonexpansive, and not a contraction. The equilibrium is nevertheless unique and interior at the origin, because $P^N = (\prod_i \sigma_i) I = -I$ makes $I - P$ invertible. Picard iteration is $2N$-periodic and therefore never converges from any $x^0 \neq 0$, its residual staying exactly constant. This is the sharpest separation we can construct between monotonicity of the variational inequality and convergence of naive fixed-point iteration: the game is strongly monotone with a unique equilibrium, and the iteration still fails. It is the regime in which contraction-based guarantees say nothing while Theorem 3.1 still applies, and it is where the solver comparison is run.

Ground truth and evaluation protocol. Every game supplies a closed-form or verified-global best response, so that no inner minimization is approximated and the exploitability against which each bound is checked is exact rather than a solver output. Equilibria are verified to a residual below 10^{-10} when an equilibrium is required: the capacitated Cournot equilibrium by active-set enumeration, the

coupled-constraint variational equilibrium by projected extragradient to its documented tolerance, and the bimodal game's 2^N equilibria in closed form.

The nine experiments use the following fixed protocol. E1 tests bound validity and tightness on the three games above (420 cells, of which 280 satisfy the theorem assumptions); E2 sweeps eight coupling values on the product quadratic game (1,120 cells); E3 pools only the five families with rigorous Lipschitz bounds to compare joint and small-gain contraction certificates; E4 tests the nonconvex set-valued witness; E5 enumerates equilibria and compares responder families; E6 compares Picard, Krasnosel'ski–Mann, and Halpern on the analytic nonexpansive map; E7 varies finite-difference radius and oracle noise; E8 varies player count at a fixed observational-data budget; and E9 varies covering resolution and region dimension. Unless stated otherwise, E1–E2 use seeds 0–19, seven response families, 200 training epochs, and 240 observational transitions per seed. E3–E9 use the seed counts, tolerances, held-out profile counts, and correction procedures reported with their respective results files. All comparisons pair the same seed, game, family, and coupling wherever those factors are present; confidence intervals use $10{,}000$ paired bootstrap resamples, and Wilcoxon tests use Holm correction within the declared experiment family.

Response families. Seven response models receive the concatenated rival profile x_{-i} and are fitted on observational play under a shared budget, so that no family is advantaged by a longer fit: a least-squares affine map, two-layer 32–32 tanh and ReLU networks, a spectral-normalized network, an Almost-Orthogonal-Layer (AOL) Lipschitz-certified network (Prach & Lampert, 2022), a conditional diffusion sampler, and conditional flow matching. The first five expose a rigorous Lipschitz bound computed from weights; the two samplers expose none, so that their certificate status is unknown and is recorded as not-certified rather than as passing. Capacity is not matched across the two groups. The deterministic heads use two hidden layers of width 32 and the generative ones two of width 64 together with a time embedding, so that the generative group carries roughly four times the parameters per player — a mismatch that runs against, rather than for, any reading in which capacity explains the outcomes of Appendix E. All neural responders use Adam with batch size 128 for 200 epochs on the same 240 noisy observational transitions per seed; the learning rate is 3×10^{-3} except for AOL, which uses 5×10^{-3}. Affine and feedforward responders minimize mean-squared prediction error, diffusion uses denoising-score matching, and flow matching uses its conditional flow objective.

Table 1: **Certificate divided by true exploitability as players are added.** Coupled quadratic game, scalar actions, medians over 20 seeds; lower is tighter and one is exact. The aggregated column is the single-constant form $(M + D)(\rho + \delta)$, the blockwise column is Theorem 3.1. The aggregated form grows with N because exploitability is a maximum over players whereas an aggregated residual is a joint norm. In this game the blockwise column coincides with the model-free bound $\max_i \bar{M}_i D_i$, hence it reports tightness of the model-free candidate, not of the learned-residual candidate.

players N	2	8	64	128	256
aggregated	38.15	84.85	431.1	696.1	1157
blockwise	13.92	10.78	7.862	7.176	6.628

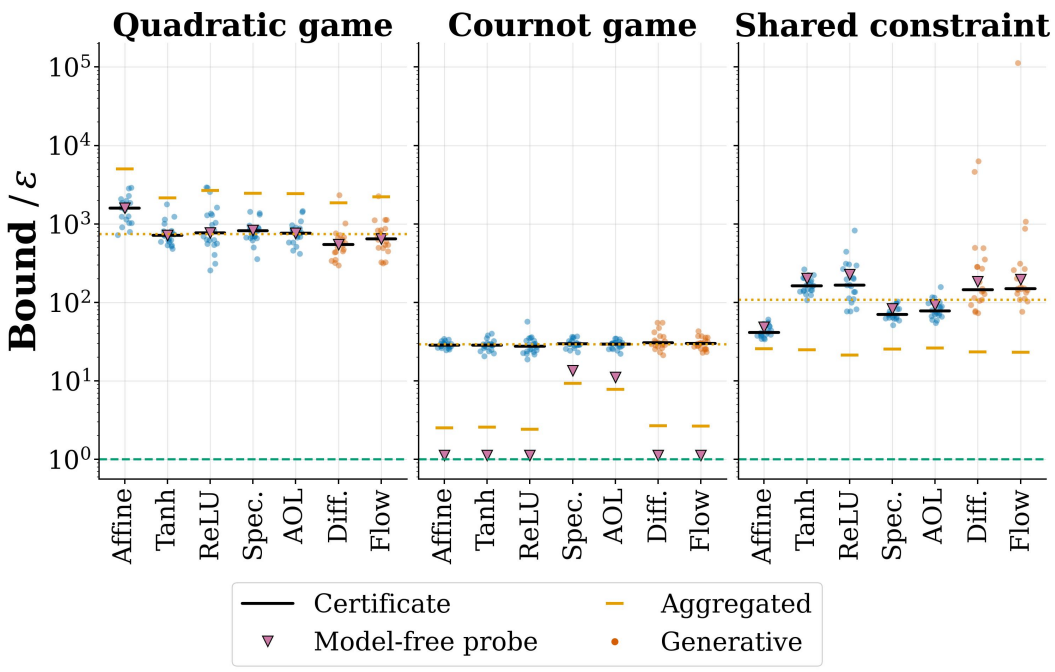

Fig. 4: **Tightness by response family.** Certified bound over true exploitability. Blue marks families with a rigorous Lipschitz bound, orange those without.

G Robustness and Scope

We report support claims in the form of visual charts to support our main results.

Fig. 4 illustrates the certified bound of (11) over true exploitability, testing the pointwise result against architecture changes and game parameters. Here, the dots denote seeds and the bars denote medians. We compare seven response families across three games. We find that the responses are close within each game type.

Fig. 5 sweeps the coupling parameter, which fixes $\lambda_{\min}$ of the symmetrical pseudo-gradient Jacobian, so that the grid crosses monotonicity at one. The bound holds in all grid cells including the non-monotone half. The median tightness falls monotonically from 3.01×10^3 to 9.86. The sweep establishes robustness of validity. The bound holds at every coupling; the median tightness ratio, on the right axis, falls monotonically by a factor of about 300 across the grid. Thus, the sweep shows stability of the validity.

G.1 Calibration of the central-difference estimator

Fig. 6 reports the central-difference error in estimating Φ_i against the probe radius τ at five noise scales, with medians over twelve probe points at each radius. The first panel has bias $B_3 > 0$ and exhibits the interior minimum of the bias–noise prediction (17); the second panel sets $B_3 = 0$, where no interior minimum can arise and none does. Dotted verticals mark the balance radius of the prediction of Proposition 4.1. The minimizer over the radius is 2.14 against a multiplicative grid step 2.15. The first panel is the multimodal arena, where $B_3 > 0$. A bias order $B_3\tau^2$ dominates at large τ while noise order σ/τ dominated at small τ. The second panel is a control with $B_3 = 0$, wherein no interior minimum can occur.

G.2 Convexity failure

Fig. 7 compares the exploitability reached from a conditional-mean response with that reached from mode-selected samples, over 20 paired seeds. At the origin, the variational residual vanishes

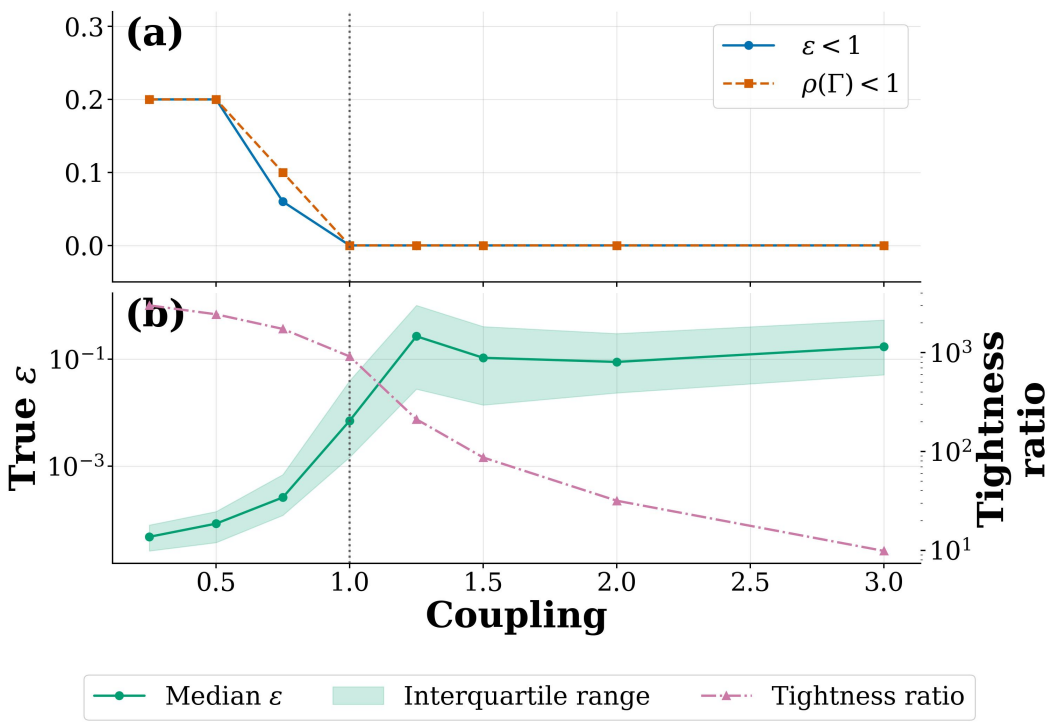

Fig. 5: **Coupling sweep across the monotone boundary.** Coupled quadratic game; the coupling parameter sets $\lambda_{\min}$ of the symmetrized pseudo-gradient Jacobian exactly, so the grid crosses monotonicity at one.

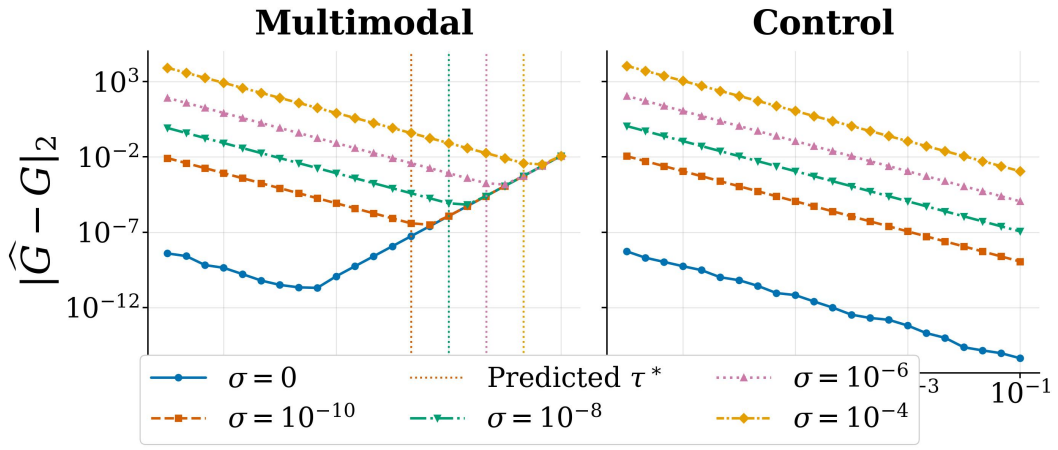

Fig. 6: **Probe error vs. probe radius.** Central-difference error in estimating Φ_i versus τ, at several noise scales σ; medians over twelve probe points and twenty-four noise replicates each.

analytically for an exploitability that is 0.25. Mode selection lowers exploitability to 0.0286 and 0.0175. The comparison is paired and between groups: a fraction 0.825 of mode-selected profiles carry both a smaller exploitability and a larger residual than the conditional-mean profile.

Fig. 8 cross-classifies the joint-spectral against the small-gain condition. The empty cell illustrates the fits, since $\rho(\Gamma) \le \|J_F\|_2$ held in all 1100 of them with median slack 0.03969. The direction

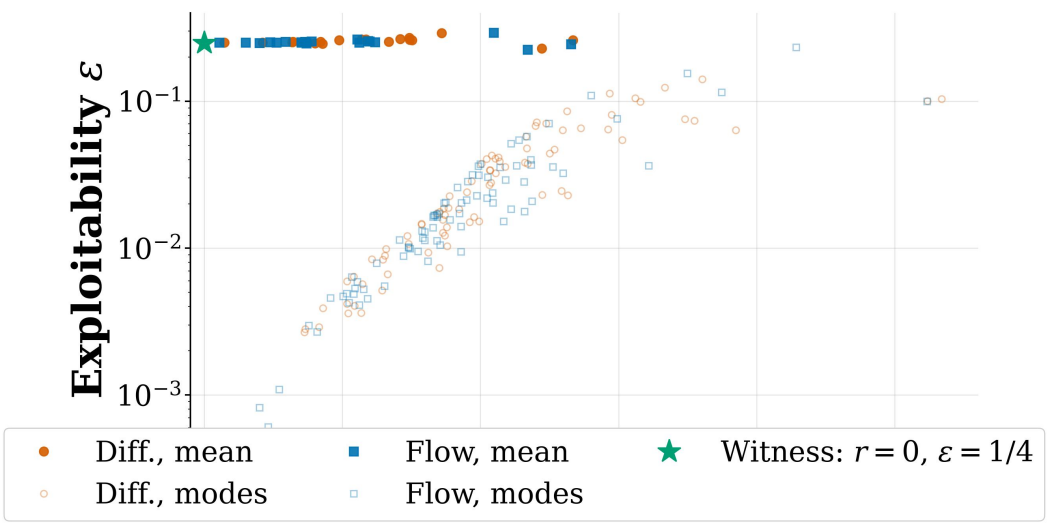

Fig. 7: **A vanishing residual is no evidence of a small gap.** Exploitability against variational residual. Filled markers: conditional-mean fixed points; hollow markers: four profiles per seed; star: analytic result at zero residual and $1/4$ exploitability.

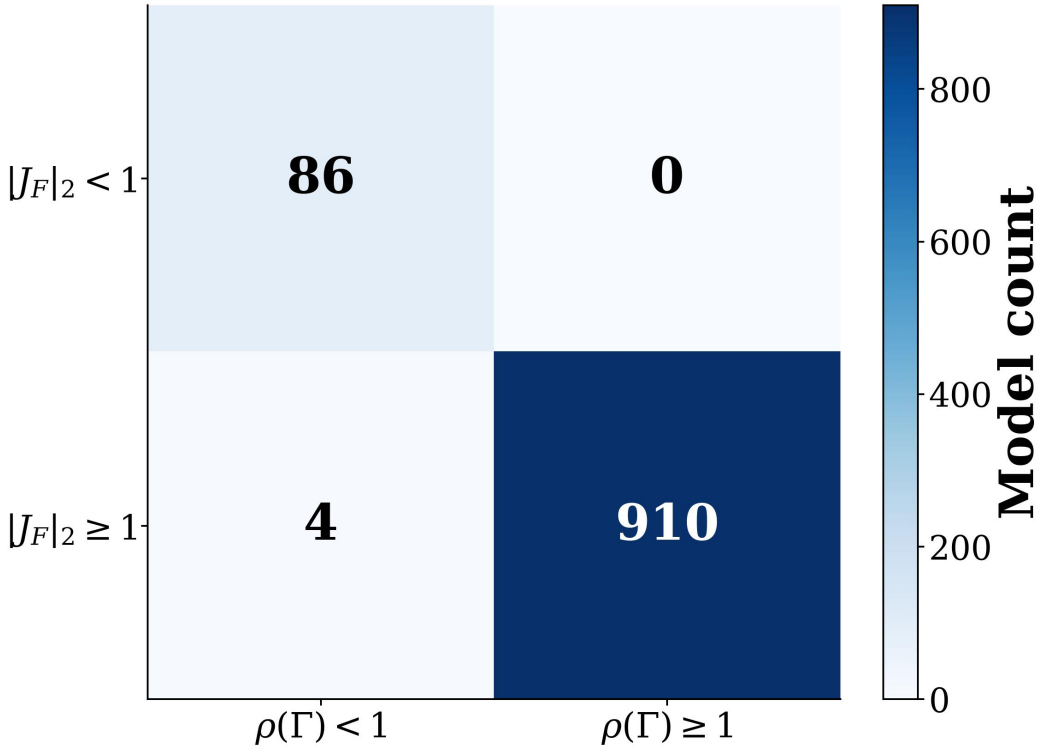

Fig. 8: **Joint-spectral against small-gain certification on trained models.** .

separating the two conditions analytically is realized nowhere in this population, with all 106 certified cells are affine.