LEARNING IN CONTINUOUS GAMES FROM PAIRWISE PREFERENCE FEEDBACK

Anonymous authors
Paper under double-blind review

ABSTRACT

We study learning in continuous games when players receive only pairwise preference feedback, revealing which of two actions is preferred but neither payoff values nor preference magnitudes. We first show that standard external regret and coarse correlated equilibria are not identifiable from this ordinal information: the same sequence of play can incur zero and linear regret in two ordinally equivalent games, while the distributions that remain CCE across all cardinal representations consistent with the same preferences are exactly those supported on pure Nash equilibria. Motivated by this gap, we develop a first-order ordinal theory based on normalized unilateral preference directions, introducing an ordinal directional regret benchmark and corresponding equilibrium notions. We show that block-normalized pseudogradient dynamics achieve sublinear ordinal regret and, under additional structure, Nash-convergence guarantees. We then use a one-bit comparison estimator to implement these dynamics from finite pairwise comparisons. With one comparison per player and round, the resulting algorithm achieves sublinear finite-resolution ordinal regret against arbitrary opponent behavior and, in ordinal potential games, almost-sure last-iterate convergence to the Nash set. Our results provide regret, dynamics, and equilibrium guarantees directly from preference feedback without reconstructing cardinal utilities.

1 INTRODUCTION

Preference feedback is especially useful for learning when numerical rewards are difficult to specify or observe, but alternatives can still be compared reliably. For instance, humans can compare large language model responses without assigning calibrated scores in reinforcement learning from human feedback (Christiano et al., 2017; Ouyang et al., 2022); users reveal relative preferences over items in recommendation systems (Bobadilla et al., 2013; Jannach et al., 2016); and preference-based robot learning uses comparisons between candidate behaviors or trajectories when numerical objectives are difficult to specify (Hejna & Sadigh, 2022). In continuous optimization, such pairwise comparisons can suffice to optimize an objective without observing its function values (Saha et al., 2021).

In games, this distinction between ordinal preferences and numerical utilities becomes especially consequential. A strategic game can be specified through the players' preference relations over joint action profiles, with payoff functions providing numerical representations of these preferences (Osborne & Rubinstein, 1994). In particular, best responses and pure Nash equilibria depend only on which actions each player prefers, holding the other players' actions fixed. However, much of existing theory on learning in games is formulated in cardinal terms: external regret aggregates numerical payoff differences and its vanishing is related to coarse correlated equilibria (Fudenberg & Levine, 1998; Cesa-Bianchi & Lugosi, 2006; Nisan et al., 2007), while first-order and payoff-based learning rules use utility gradients or numerical payoff observations (see, e.g., Bravo et al. (2018); Mertikopoulos & Zhou (2019)). This creates a structural mismatch: strategic incentives are determined by ordinal preferences, while standard learning objectives and dynamics may depend on the particular numerical utilities used to represent them. This raises a basic question: what remains of learning-in-games theory when players observe only their preferences, rather than a particular numerical representation of them?

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

To isolate what can be learned from preferences alone, we study continuous games under *ordinal pairwise comparison feedback*. At a fixed opponents' profile, each player may query which of two unilateral actions is preferred, but receives no information about the magnitude of the corresponding utility difference. This raises a question beyond payoff estimation: are the learning objects themselves identified by the feedback? Indeed, replacing player i's cost J_i by $\phi_i \circ J_i$ for any strictly increasing ϕ_i preserves every unilateral comparison. A learner observing only pairwise preferences therefore cannot distinguish between these numerical representations. Yet central objects of standard learning theory need not be preserved: such transformations can alter the pseudogradient field, the external regret assigned to the same sequence of play, and the set of coarse correlated equilibria. Thus two games that are indistinguishable from preference feedback can induce different standard learning objects. This leads to our central question:

Which learning objectives, dynamics, and equilibrium guarantees are intrinsic to players' unilateral preferences, and can they be realized efficiently using only binary pairwise comparisons?

Our starting point is that smooth unilateral preferences retain a natural first-order geometry. Infinitesimal comparisons reveal the signs of directional derivatives and therefore identify, whenever $\nabla_i J_i(x) \neq 0$, its direction but not its magnitude (Saha et al., 2021). This suggests taking normalized unilateral preference directions as the primitive first-order objects. They retain the directional information preserved under smooth monotone reparameterizations while discarding precisely the player-specific gradient magnitudes that comparison feedback does not identify. This leads to representation-invariant regret benchmarks, equilibrium notions, and learning dynamics, and provides learning targets that we subsequently estimate directly from finite comparisons.

A key distinction from stochastic ranking models is that our comparison oracle is purely ordinal: conditional on which alternative is preferred, the feedback contains no information about the magnitude of the utility difference. Gap-dependent models such as Plackett–Luce may instead reveal cardinal information through repeated observations (Liu et al., 2026). Likewise, unlike finite context-ordinal games defined directly from rankings (Gemp et al., 2026b), our smooth latent-utility model exposes a local first-order geometry. We use this geometry to define learning objectives, dynamics, and equilibrium guarantees directly from unilateral preferences. See Section 1.2 for a discussion.

1.1 Contributions

We develop a first-order framework for learning in continuous games from *ordinal pairwise comparison feedback*: comparisons reveal only the ordering of two unilateral actions and carry no information about the magnitude of their cost difference. Our performance benchmarks, learning dynamics, and equilibrium guarantees are designed directly from the first-order information that this preference feedback identifies.

1. Identifiability limits of classical learning objectives. We show that ordinally equivalent games—i.e., games inducing exactly the same unilateral comparison oracle—can assign the same sequence of play zero external regret in one cardinal representation of the game and linear regret in another (Proposition 3.1), and can have different coarse correlated equilibrium (CCE) sets (Proposition 3.2). In contrast, ordinally equivalent games have identical best-response correspondences and hence the same pure Nash equilibria. More sharply, for every continuous game $\mathcal{G}$ with compact action spaces, $\bigcap_{\widetilde{\mathcal{G}} \sim_{\mathrm{ord}} \mathcal{G}} \mathrm{CCE}(\widetilde{\mathcal{G}}) = \Delta(\mathrm{NE}(\mathcal{G}))$, where $\sim_{\mathrm{ord}}$ denotes ordinal game equivalence and $\Delta(\mathrm{NE}(\mathcal{G}))$ denotes the distributions supported on the set of Nash equilibria. Thus, the CCE distributions that remain valid under every cardinal realization consistent with the same unilateral preferences are exactly those supported on pure Nash equilibria (Theorem 3.1). Pairwise preference feedback therefore creates not only an estimation problem, but an identifiability problem: classical regret and CCE are not, in general, determined by the information available to the players.

2. A first-order ordinal regret-to-equilibrium theory. Motivated by this identifiability gap, we formulate learning directly on the first-order preference geometry that remains invariant under regular smooth monotone reparameterizations of the cost functions. We introduce *ordinal directional regret*, a first-order benchmark based on normalized unilateral preference directions rather than cardinal payoff differences or gradient magnitudes. This benchmark induces an *ordinal coarse variational equilibrium* residual and an exact no-regret-to-equilibrium correspondence: sublinear ordinal

regret for every player implies vanishing ordinal equilibrium violation of the empirical distribution of play (Section 3.3). We further introduce an ordinal Nash residual whose zero set coincides with the pure Nash equilibria under blockwise pseudoconvexity. Projected block-normalized pseudogradient dynamics achieve $\mathcal{O}(\sqrt{T})$ ordinal directional regret for every player against arbitrary opponent behavior (Theorem 4.1). Consequently, the empirical distribution of play achieves an $\mathcal{O}(T^{-1/2})$ ordinal coarse variational equilibrium residual (Corollary 4.1). Block-normalized pseudogradient dynamics retain the local preference directions preserved under regular monotone reparameterizations while discarding player-specific gradient magnitudes, thereby matching the geometry of the learning dynamics to the information revealed by pairwise preferences.

3. Direct learning from one-bit comparisons and last-iterate Nash convergence. For every comparison radius $\delta > 0$, we identify an exact finite-resolution vector field D^δ determined solely by the pairwise comparison oracle. This vector field need not coincide with the infinitesimal normalized preference field D, and its convergence to D need not be uniform near stationary points. We construct an unbiased one-bit estimator of D^δ and a comparison-normalized learning rule that does not reconstruct a cardinal utility representation. With one comparison per player and round, the resulting algorithm achieves $\mathcal{O}\left(\frac{R_i\sqrt{d_i T}}{\nu_i}\right)$ expected finite-resolution ordinal regret against arbitrary opponent behavior, where R_i, d_i, and ν_i denote the action-set diameter, dimension, and comparison reliability, respectively (Theorem 5.1). In ordinal potential games, where players' comparisons coincide exactly with comparisons of a latent potential, the *last iterate* converges almost surely to the Nash set; if the Nash set is finite, it converges almost surely to a single Nash equilibrium (Theorem 5.2).

1.2 Related Work

We discuss the most closely related work here and defer a broader review to Appendix A.

Dueling bandits and comparison-based optimization. Preference-based feedback has been studied extensively in dueling bandits; see Bengs et al. (2021) for a survey. In continuous optimization, pairwise comparisons can be used to optimize objectives without observing function values (Yue & Joachims, 2009; Jamieson et al., 2012; Kumagai, 2017; Saha et al., 2021), with extensions to more general preferences, adversarial settings, and richer ranking feedback (Saha et al., 2025; Blum et al., 2024; Saha et al., 2024; Tang et al., 2024; Ye, 2025). Most of this literature concerns a single objective, often under convexity or *smooth* stochastic preference models. Our setting is instead a nonconvex multi-player game with a purely ordinal sign oracle, which reveals preference but no cost-gap magnitude. Recent work on online convex optimization with dueling feedback recovers standard regret guarantees under smooth gap-dependent comparison models (Lu et al., 2026); our oracle deliberately excludes such cardinal gap information.

Learning from preference feedback in games. Several recent works study which aspects of games are determined by ordinal rather than cardinal information. In finite games, Biggar & Shames (2025) encode unilateral rankings through preference graphs, while Conitzer (2024) studies equilibria robust across compatible cardinal representations. Gemp et al. (2026b) instead study finite context-ordinal games specified directly by rankings, using social-choice aggregation across opponents' contexts to define mixed equilibrium and learning notions. Our smooth setting assumes latent utility functions observed only ordinally, which leaves accessible a local first-order preference geometry. Abbadi et al. (2026) show that the combinatorial preference structure of a finite game constrains, but does not generally determine, the long-run behavior of no-regret dynamics.

Closer to our continuous game setting, Dindoš & Mezzetti (2006) study stochastic better-reply learning in aggregative games, while Cheung (2014) study pairwise-comparison evolutionary dynamics in continuous-strategy population games. Bervoets et al. (2020) establish almost-sure Nash convergence for payoff-based learning in several classes of continuous games, but their updates use numerical payoff variations. Krupa & Bemporad (2026) propose to learn generalized Nash equilibria (GNE) from pairwise preferences by fitting a numerical surrogate game without providing convergence guarantees. In contrast, we work directly with the first-order field identified by unilateral comparisons, without reconstructing cardinal utilities. Liu et al. (2026) ask when ranking feedback suffices to recover standard external regret. Their Plackett–Luce model is gap-dependent, so repeated rankings can contain cardinal information which is used to estimate numerical utilities; our purely ordinal oracle does not. We therefore formulate regret, equilibrium, and learning dynamics directly from the information preserved by unilateral preferences.

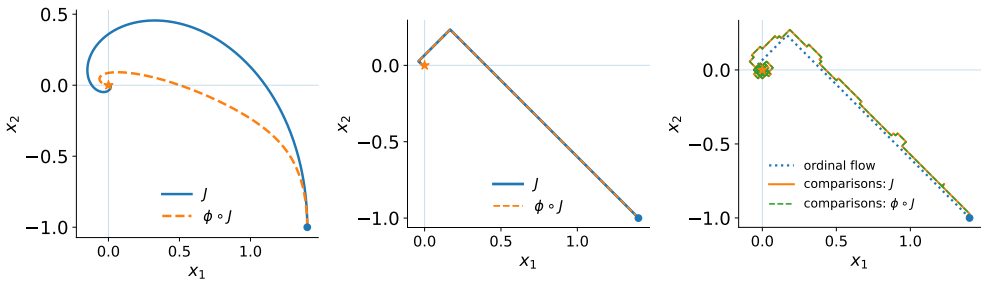

Figure 1: **Same preferences, different numerical representations** (J and $\phi \circ J$). The two games have identical unilateral preferences, best responses, and Nash equilibrium, but different numerical utility representations. **(Left)** Standard gradient play follows different trajectories. **(Center)** Player-wise normalization makes the trajectories coincide. **(Right)** Pairwise-comparison learning is likewise invariant under common oracle randomness. The star denotes the common Nash.

Normalized first-order dynamics. In single-objective optimization, comparison and sign feedback can recover directional information without observing function values (Saha et al., 2021; Cheng et al., 2020; Cai et al., 2022). While our comparison-based implementation in Section 5 builds on this, normalization here is not merely a device for gradient estimation: the block-normalized field is the first-order object identified by unilateral preferences, and induces the corresponding regret, equilibrium, and learning dynamics in the game. Related to our invariance viewpoint, Gupta et al. (2025) use normalized utility gradient directions to obtain bargaining solutions invariant to monotone utility transformations; our setting is decentralized and noncooperative, uses unilateral preference feedback, studies different block-normalized dynamics and provides non-asymptotic guarantees. Poveda et al. (2022) study magnitude-dependent rescalings of gradient play for Nash seeking; our block normalization is instead dictated by ordinal invariance and admits a direct one-bit comparison implementation.

2 Continuous Games with Pairwise Preference Feedback

Notation. For an integer $N \geq 1$, let $[N] := \{1, \dots, N\}$, and let $[a]_+ := \max\{a, 0\}$ denote the positive part of $a \in \mathbb{R}$. For a nonempty closed convex set $\mathcal{C} \subseteq \mathbb{R}^d$, $\Pi_{\mathcal{C}}$ denotes the Euclidean projection onto $\mathcal{C}$ and $\mathrm{cl}(\mathcal{C})$ its closure. At $x \in \mathcal{C}$, its tangent and normal cones are $T_{\mathcal{C}}(x) := \mathrm{cl}\{\lambda(y-x) : y \in \mathcal{C},\ \lambda \geq 0\}$, $N_{\mathcal{C}}(x) := \{v : \langle v, y-x\rangle \leq 0,\ \forall y \in \mathcal{C}\}$. We write $\mathrm{diam}(\mathcal{C}) := \sup_{x,y\in\mathcal{C}} \|x-y\|_2$, $\mathrm{dist}(x, \mathcal{C}) := \inf_{y\in\mathcal{C}} \|x-y\|_2$ where $\|\cdot\|_2$ is the Euclidean 2-norm. For a block vector $v = (v_1, \dots, v_N)$, with $v_i \in \mathbb{R}^{d_i}$, we use $\|v\|_{2,1} := \sum_{i=1}^N \|v_i\|_2$, $\|v\|_{2,\infty} := \max_{i\in[N]} \|v_i\|_2$. Finally, $\mathbb{S}^{d-1} := \{u \in \mathbb{R}^d : \|u\|_2 = 1\}$ and $\mathbb{B}_d$ denote the Euclidean unit sphere and ball, $\Delta(\mathcal{X})$ the set of probability distributions over $\mathcal{X}$, and δ_x the Dirac measure at x.

Continuous games. We consider an N-player continuous game $\mathcal{G} = \big((\mathcal{X}_i)_{i=1}^N, (J_i)_{i=1}^N\big)$, where the action sets $\mathcal{X}_i \subseteq \mathbb{R}^{d_i}$ are nonempty, closed, and convex sets and we define the joint action set $\mathcal{X} := \prod_{i=1}^N \mathcal{X}_i$. The cost functions $J_i : \mathcal{X} \to \mathbb{R}$ are supposed to be continuously differentiable on an open neighborhood of $\mathcal{X}$. Each player i chooses an action $x_i \in \mathcal{X}_i$ and incurs a cost $J_i(x_i, x_{-i})$ where $x_{-i} \in \mathcal{X}_{-i} := \prod_{j\neq i} \mathcal{X}_j$ stands for the actions of all players but i using the standard notation convention. When $\mathcal{X}_i$ is bounded, we write $R_i := \mathrm{diam}(\mathcal{X}_i)$. The best-response correspondence of player i is $\mathrm{BR}_i^{\mathcal{G}}(x_{-i}) := \arg\min_{y_i\in\mathcal{X}_i} J_i(y_i, x_{-i})$. A profile $x^\star \in \mathcal{X}$ is a *Nash equilibrium* if $x_i^\star \in \mathrm{BR}_i^{\mathcal{G}}(x_{-i}^\star), \forall i \in [N]$, and we denote the set of Nash equilibria by $\mathrm{NE}(\mathcal{G})$. A distribution $\mu \in \Delta(\mathcal{X})$ is a *coarse correlated equilibrium* (CCE) if, for every $i \in [N]$ and every fixed deviation $y_i \in \mathcal{X}_i$, $\mathbb{E}_{x\sim\mu}[J_i(x)] \leq \mathbb{E}_{x\sim\mu}[J_i(y_i, x_{-i})]$. Equivalently, $\sup_{y_i\in\mathcal{X}_i} \mathbb{E}_{x\sim\mu}\big[J_i(x) - J_i(y_i, x_{-i})\big] \leq 0, \forall i \in [N]$. We denote the corresponding set by $\mathrm{CCE}(\mathcal{G})$. For any joint strategy $x \in \mathcal{X}$, we denote the standard pseudogradient field of the game by $F(x) = (\nabla_1 J_1(x), \cdots, \nabla_N J_N(x))$, where for any $i \in [N], \nabla_i J_i(x)$ stands for the gradient of J_i w.r.t. its i-th variable x_i.

Pairwise comparison feedback. Player i observes unilateral preferences via the comparison oracle:

$$\mathsf{C}_i(a_i, b_i; x_{-i}) := \mathrm{sgn}\left(J_i(a_i, x_{-i}) - J_i(b_i, x_{-i})\right) \in \{-1, 0, +1\}, \tag{1}$$

when querying the oracle with actions $a_i, b_i \in \mathcal{X}_i$ given fixed opponents' actions $x_{-i} \in \mathcal{X}_{-i}$. Thus $\mathsf{C}_i(a_i, b_i; x_{-i}) = -1$ means that a_i is strictly preferred to b_i under the cost convention. We will

also consider a noisy comparison oracle $Y_i = \xi_i \mathsf{C}_i(a_i, b_i; x_{-i})$, where $\xi_i \in \{-1, +1\}$ satisfies $\mathbb{P}(\xi_i = 1) = \frac{1}{2} + \nu_i, \mathbb{P}(\xi_i = -1) = \frac{1}{2} - \nu_i, \nu_i \in (0, 1/2]$. For a strict comparison, the reported sign is therefore correct with probability $1/2 + \nu_i$; exact comparisons correspond to $\nu_i = 1/2$. The noise distribution depends only on the preference ordering and not on the numerical cost values. In the continuous optimization setting ($N = 1$), this model has been considered, e.g., in Saha et al. (2021).

Unilateral ordinal equivalence. Two continuous games $\mathcal{G} = ((\mathcal{X}_i)_{i\in[N]}, (J_i)_{i\in[N]})$ and $\widetilde{\mathcal{G}} = ((\mathcal{X}_i)_{i\in[N]}, (\widetilde{J}_i)_{i\in[N]})$ with the same action spaces are *ordinally equivalent*, written $\mathcal{G} \sim_{\text{ord}} \widetilde{\mathcal{G}}$, if they induce the same comparisons: $\operatorname{sgn}(J_i(a_i, x_{-i}) - J_i(b_i, x_{-i})) = \operatorname{sgn}\left(\widetilde{J}_i(a_i, x_{-i}) - \widetilde{J}_i(b_i, x_{-i})\right)$ for any $i \in [N]$, $a_i, b_i \in \mathcal{X}_i$ and $x_{-i} \in \mathcal{X}_{-i}$. Equivalently, the two games generate exactly the same unilateral comparison oracles. We denote the corresponding equivalence class by $[\mathcal{G}]_{\text{ord}} := \{\widetilde{\mathcal{G}} : \widetilde{\mathcal{G}} \sim_{\text{ord}} \mathcal{G}\}$. We call $\widetilde{\mathcal{G}}$ a *regular monotone reparameterization* of $\mathcal{G}$ if $\widetilde{J}_i = \phi_i \circ J_i$ (i.e., $\widetilde{J}_i(x) = \phi_i(J_i(x)), \forall x \in \mathcal{X}$), where $\phi_i : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and $\phi_i'(s) > 0$ on a neighborhood of the range of J_i. Every such reparameterization is ordinally equivalent to $\mathcal{G}$.

Ordinal potential games. A continuously differentiable function $\Phi : \mathcal{X} \to \mathbb{R}$ is a *smooth ordinal potential* for $\mathcal{G}$ if $\operatorname{sgn}(J_i(y_i, x_{-i}) - J_i(x_i, x_{-i})) = \operatorname{sgn}(\Phi(y_i, x_{-i}) - \Phi(x_i, x_{-i}))$ for every player $i \in [N]$, every $x_{-i} \in \mathcal{X}_{-i}$, and every $x_i, y_i \in \mathcal{X}_i$. A game admitting such a Φ is called a *smooth ordinal potential game*. Thus, for every fixed opponents' profile, the player's cost and the potential induce exactly the same ordering over unilateral actions. The potential Φ need not be observed.

3 Identifiability from Preference Feedback

3.1 Classical Regret and CCE Are Not Identifiable

For a sequence of joint actions $(x_t)_{t=1}^T$, the standard external regret of player i is defined as:

$$\operatorname{Reg}_{i,T}^J = \sup_{y_i \in \mathcal{X}_i} \sum_{t=1}^{T} J_i(x_t) - J_i(y_i, x_{-i,t}). \tag{2}$$

This performance measure depends on cardinal utility differences. Monotone reparameterizations can change not only its scale, but whether a trajectory is no-regret at all, as we show next.

Proposition 3.1 (External regret is not ordinally identifiable)**.** *There exist ordinally equivalent smooth continuous games $\mathcal{G} \sim_{\text{ord}} \widetilde{\mathcal{G}}$ and a common sequence of joint actions $(x_t)_{t\geq 1}$ such that, for every even T, $\operatorname{Reg}_{i,T}^J = 0, \operatorname{Reg}_{i,T}^{\widetilde{J}} = \Omega(T)$.*

Hence external regret is not determined by the ordinal equivalence class identified by comparison-only feedback. The same is true of CCE.

Proposition 3.2 (CCE is not ordinally invariant)**.** *There exist ordinally equivalent continuous games $\mathcal{G} \sim_{\text{ord}} \widetilde{\mathcal{G}}$ and a distribution $\mu \in \Delta(\mathcal{X})$ such that $\mu \in \operatorname{CCE}(\mathcal{G}), \mu \notin \operatorname{CCE}(\widetilde{\mathcal{G}})$.*

In contrast, ordinally equivalent games have identical best-response correspondences and therefore the same pure Nash equilibria. Thus pure Nash equilibrium is ordinal, whereas external regret and CCE generally are not. Because CCE averages utility differences across different opponents' profiles, it is generally not preserved by nonlinear monotone transformations (Hammond, 2004; 2005; Conitzer, 2024; Gemp et al., 2026b). We sharpen this observation by characterizing the CCE distributions that are robust over the entire unilateral ordinal equivalence class.

Theorem 3.1 (Ordinally robust CCE)**.** *If $\mathcal{X}$ is compact, then $\bigcap_{\tilde{\mathcal{G}} \sim_{\text{ord}} \mathcal{G}} \operatorname{CCE}(\tilde{\mathcal{G}}) = \Delta(\operatorname{NE}(\mathcal{G}))$.*

Classical CCE permits positive conditional regret at some opponents' profiles to be offset by negative regret elsewhere. Requiring the CCE inequalities to hold for every cardinal realization consistent with the same unilateral preferences eliminates this cross-context compensation and forces best-response play almost surely. See Appendix B.4 for a proof and a more detailed discussion. This motivates equilibrium notions defined directly from the ordinal information available to players.

3.2 PREFERENCE-INVARIANT FIRST-ORDER BENCHMARKS

Our starting point is to ask which first-order information is likewise intrinsic to the comparison oracle. Under a smooth monotone reparameterization $\widetilde{J}_i = \phi_i \circ J_i$, $\nabla_i \widetilde{J}_i(x) = \phi_i'(J_i(x))\nabla_i J_i(x)$. Hence the magnitude $\|\nabla_i J_i(x)\|$ depends on the cardinal representation, whereas the ray $\{\lambda \nabla_i J_i(x) : \lambda > 0\}$ does not. This motivates to consider the *preference-direction field*, defined for every $i \in [N]$ and every joint strategy $x \in \mathcal{X}$ as follows:

$$D_i(x) = \begin{cases} \dfrac{\nabla_i J_i(x)}{\|\nabla_i J_i(x)\|}, & \nabla_i J_i(x) \neq 0, \\ 0, & \nabla_i J_i(x) = 0, \end{cases} \qquad D(x) = \begin{pmatrix} D_1(x) \\ \vdots \\ D_N(x) \end{pmatrix}. \tag{3}$$

Lemma 3.1 (Invariance of the preference-direction field)**.** *For every smooth player-wise monotone reparameterization $\widetilde{J}_i = \phi_i \circ J_i$ with $\phi_i' > 0$, $\widetilde{D}_i(x) = D_i(x)$ for all $i \in [N], x \in \mathcal{X}$.*

Player-wise normalization is essential: a global normalization $F/\|F\|$ is not invariant because monotone reparameterizations rescale players' gradient blocks differently.

When each $J_i(\cdot, x_{-i})$ is convex, standard external regret is bounded by its linearized counterpart $\sup_{y_i \in \mathcal{X}_i} \sum_{t=1}^{T} \langle \nabla_i J_i(x_t), x_{i,t} - y_i \rangle$. Replacing $\nabla_i J_i$ by the preference direction D_i yields a representation-invariant first-order benchmark, motivating the following definition.

Definition 3.1 (Ordinal directional regret)**.** The *ordinal directional regret* of player $i \in [N]$ is

$$\mathrm{Reg}_{i,T}^{D} := \sup_{y_i \in \mathcal{X}_i} \sum_{t=1}^{T} \langle D_i(x_t), x_{i,t} - y_i \rangle \,. \tag{4}$$

We say that player i has no ordinal directional regret if $\limsup_{T\to\infty} \frac{\mathrm{Reg}_{i,T}^{D}}{T} \leq 0$.

Since D is independent of the arbitrary magnitude of the utility gradient, so is $\mathrm{Reg}_{i,T}^{D}$.

An ordinal Nash residual. The usual unilateral-deviation gap $\sum_{i=1}^{N} [J_i(x) - \inf_{y_i \in \mathcal{X}_i} J_i(y_i, x_{-i})]$ has an ordinally invariant zero set but a representation-dependent numerical value. For compact convex action sets, define instead

$$\mathrm{Gap}_D(x) = \sup_{y \in \mathcal{X}} \langle D(x), x - y \rangle = \sum_{i=1}^{N} \sup_{y_i \in \mathcal{X}_i} \langle D_i(x), x_i - y_i \rangle \,. \tag{5}$$

This is the standard Stampacchia variational inequality gap associated with the ordinal pseudogradient D. The gap is invariant under all player-wise smooth monotone utility reparameterizations with strictly positive derivative and satisfies $0 \leq \mathrm{Gap}_D(x) \leq \sum_{i=1}^{N} R_i$ since $\|D_i(x)\| \leq 1$. To relate its zero set to Nash equilibria, we impose the standard first-order sufficiency condition below.

Assumption 3.1 (Blockwise pseudoconvexity)**.** For every $i \in [N]$ and every $x_{-i} \in \mathcal{X}_{-i}$, the function $z_i \longmapsto J_i(z_i, x_{-i})$ is pseudoconvex on $\mathcal{X}_i$, i.e., satisfies: For any $y_i \in \mathcal{X}_i, x = (x_i, x_{-i}) \in \mathcal{X}$, if $\langle \nabla_i J_i(x), y_i - x_i \rangle \geq 0$ then $J_i(y_i, x_{-i}) \geq J_i(x_i, x_{-i})$.

Pseudoconvexity is preserved by regular increasing reparameterizations and ensures that unilateral first-order stationarity is globally optimal. Under this assumption, we obtain the following characterization of Nash equilibria.

Proposition 3.3 (Ordinal Nash characterization)**.** *Suppose the action sets are convex and Assumption 3.1 holds. Then $\mathrm{Gap}_D(x) = 0$ if and only if $x \in \mathrm{NE}(\mathcal{G})$.*

Hence Gap_D provides a representation-invariant first-order certificate of pure Nash equilibrium.

3.3 FROM ORDINAL REGRET TO ORDINAL COARSE EQUILIBRIUM

Ordinal directional regret induces the following distributional analogue of the classical no-regret-to-CCE correspondence.

Definition 3.2 (Ordinal coarse variational equilibrium)**.** For $\mu \in \Delta(\mathcal{X})$, define

$$\mathcal{E}_D(\mu) := \sum_{i=1}^{N} \left[\sup_{y_i \in \mathcal{X}_i} \mathbb{E}_{x\sim\mu} \langle D_i(x), x_i - y_i \rangle \right]_+ . \tag{6}$$

We call μ an *ordinal coarse variational equilibrium* if $\mathcal{E}_D(\mu) = 0$.

For a point mass, $\mathcal{E}_D(\delta_x) = \mathrm{Gap}_D(x)$, so under Assumption 3.1, $\mathcal{E}_D(\delta_x) = 0$ iff $x \in \mathrm{NE}(\mathcal{G})$ (see Proposition 3.3). For the empirical distribution of play $\widehat{\mu}_T = T^{-1} \sum_{t=1}^{T} \delta_{x_t}$, it follows directly from the definitions of directional regret $\mathrm{Reg}_{i,T}^D$ in (4) and $\mathcal{E}_D(\mu)$ in (6) that:

$$\mathcal{E}_D(\widehat{\mu}_T) = \frac{1}{T} \sum_{i=1}^{N} [\mathrm{Reg}_{i,T}^D]_+. \tag{7}$$

Hence sublinear ordinal regret for each player implies $\mathcal{E}_D(\widehat{\mu}_T) \to 0$, i.e., the empirical distribution has asymptotically vanishing ordinal coarse equilibrium violation.

4 Learning on First-Order Preference Geometry

We now study learning dynamics induced by the preference-direction field D.

4.1 Continuous-Time Block-Normalized Dynamics

For compact convex action sets, consider the projected *block-normalized pseudogradient dynamics*:

$$\dot{x}_i(t) = \Pi_{T_{\mathcal{X}_i}(x_i(t))}(-D_i(x(t))), \qquad i \in [N]. \tag{BNPGD}$$

At interior points this reduces to $\dot{x}_i = -D_i(x)$. By Lemma 3.1, the dynamics are invariant under regular player-wise monotone reparameterizations. Because D may be discontinuous at stationary blocks, our statements apply to absolutely continuous solutions of (BNPGD) satisfying the dynamics almost everywhere. The continuous-time regret and ordinal-potential analyses are in Appendix C.

4.2 Discrete-Time No-Ordinal-Regret Learning

Inspired by (BNPGD) descent, we introduce the corresponding projected discrete-time dynamics:

$$x_{i,t+1} = \Pi_{\mathcal{X}_i}(x_{i,t} - \eta_t D_i(x_t)), \qquad i \in [N], \tag{8}$$

starting from an initial point $x_0 \in \mathcal{X}$. We assume exact access to $D_i(x_t)$ for now. Section 5 shows how these directions can be learned from finite pairwise comparisons.

Theorem 4.1 (Directional regret of discrete block-normalized learning)**.** *Suppose $\eta_t = \eta > 0$. Then, the algorithm (8) incurs $\mathrm{Reg}_{i,T}^D \le \frac{R_i^2}{2\eta} + \frac{\eta T}{2}$. With $\eta = R_i/\sqrt{T}$, $\mathrm{Reg}_{i,T}^D \le R_i\sqrt{T}$.*

Combining Theorem 4.1 with (7) gives:

Corollary 4.1 (Convergence to coarse variational equilibria)**.** *Let $\widehat{\mu}_T = T^{-1} \sum_{t=1}^{T} \delta_{x_t}$. If every player runs (8) with $\eta_i = R_i/\sqrt{T}$, then $\mathcal{E}_D(\widehat{\mu}_T) \le \frac{1}{\sqrt{T}} \sum_{i=1}^{N} R_i$.*

This shows that the ordinal coarse equilibrium residual of empirical play decays as $\mathcal{O}(T^{-1/2})$.

5 One-Bit Learning from Pairwise Comparisons

We now replace direct access to the preference-direction field D by pairwise comparisons.

Feasible comparison queries. Throughout this section, we assume that each $\mathcal{X}_i$ is compact, convex, and has nonempty interior. To ensure feasible symmetric queries, fix $c_i \in \mathrm{int}(\mathcal{X}_i)$ and $\rho_i > 0$ such that $c_i + \rho_i \mathbb{B}_{d_i} \subseteq \mathcal{X}_i$, and let $\rho := \min_{i\in[N]} \rho_i$. For $0 < \delta < \rho_i$, define for any $z_i \in \mathcal{X}_i$, $\mathsf{Q}_{i,\delta}(z_i) := c_i + \left(1 - \frac{\delta}{\rho_i}\right)(z_i - c_i)$. Then, $\mathsf{Q}_{i,\delta}(z_i) \pm \delta u \in \mathcal{X}_i$ and $\|\mathsf{Q}_{i,\delta}(z_i) - z_i\| = \frac{\delta}{\rho_i}\|z_i - c_i\| \le \frac{R_i}{\rho_i}\delta$ for any $u \in \mathbb{S}^{d_i-1}$ We write $\mathcal{X}_i^\delta := \mathsf{Q}_{i,\delta}(\mathcal{X}_i)$ and $\mathcal{X}^\delta := \prod_{i=1}^{N} \mathcal{X}_i^\delta$.

5.1 The Finite-Resolution Comparison Vector Field

Let $U_i \sim \mathrm{Unif}(\mathbb{S}^{d_i-1})$. For any radius $\delta > 0$ and any $x \in \mathcal{X}^\delta$, define:

$$S_i^\delta(x, U_i) := \mathsf{C}_i\left(x_i + \delta U_i,\, x_i - \delta U_i;\, x_{-i}\right) = \mathrm{sgn}(J_i(x_i + \delta U_i, x_{-i}) - J_i(x_i - \delta U_i, x_{-i}))\,. \quad (9)$$

Differentiability gives $J_i(x_i+\delta U_i, x_{-i}) - J_i(x_i - \delta U_i, x_{-i}) = 2\delta\langle \nabla_i J_i(x), U_i\rangle + o(\delta)$, so infinitesimal comparisons reveal the sign of $\langle D_i(x), U_i\rangle$, but not the magnitude of the gradient. Define the *δ-comparison vector field* induced by the comparison oracle at finite resolution $\delta > 0$ as follows:

$$D_i^\delta(x) := \frac{1}{\kappa_{d_i}} \mathbb{E}_{U_i}\left[S_i^\delta(x, U_i) U_i\right], \quad \forall x \in \mathcal{X}^\delta\,, \quad (10)$$

where $\kappa_d := \mathbb{E}_{U\sim\mathrm{Unif}(\mathbb{S}^{d-1})}|U_1|$ and U_1 is the first coordinate of the random variable U. We collect some of the properties of D_i^δ in the following result.

Proposition 5.1 (Properties of the comparison field). *For any $i \in [N]$ and every $\delta > 0$, (i) D_i^δ is invariant under every strictly increasing transformation $J_i \mapsto \phi_i \circ J_i$; (ii) $\|D_i^\delta(x)\| \le 1$ for every $x \in \mathcal{X}^\delta$; (iii) for any $x \in \mathcal{X}^\delta$ s.t. $x_i \in \mathrm{int}(\mathcal{X}_i)$, if $\nabla_i J_i(x) \neq 0$, then $\lim_{\delta\downarrow 0} D_i^\delta(x) = D_i(x)$.*

Hence D^δ is comparison-invariant and converges pointwise to D away from stationary blocks.

Comparison-normalized gradient play. We now use the finite-resolution comparison vector field D^δ and its expectation expression (10) to motivate the stochastic algorithm we consider, which uses only pairwise stochastic comparisons. At round t, each player i samples $U_{i,t} \sim \mathrm{Unif}(\mathbb{S}^{d_i-1})$ and queries the symmetric comparison (9) at radius $\delta_t > 0$. Under the noisy comparison model of Section 2, the observed stochastic sign is $Y_{i,t} = \xi_{i,t} S_i^{\delta_t}(x_t, U_{i,t})$, with exact comparisons corresponding to $\nu_i = 1/2$. We define the associated one-bit estimator $\widehat{D}_{i,t} := \frac{Y_{i,t}}{2\nu_i\kappa_{d_i}} U_{i,t}$. This random-direction construction is standard in comparison-based optimization (Saha et al., 2021; Cai et al., 2022); here it is applied blockwise to a coupled game and evaluated through ordinal regret.

Lemma 5.1 (One-bit estimator). *Let $\mathcal{F}_t = \sigma(\xi_0, \dots, \xi_{t-1}, U_0, \dots, U_{t-1})$ be the σ-algebra induced by the random uniform directions and comparison noise up to round $t-1$ for $t \ge 1$. In particular, z_t and $x_t = \mathsf{Q}_{\delta_t}(z_t)$ are $\mathcal{F}_t$-measurable, and $x_t \in \mathcal{X}^{\delta_t}$ by construction. Conditionally on $\mathcal{F}_t$ and the current action profile x_t, $\mathbb{E}[\widehat{D}_{i,t} \mid \mathcal{F}_t] = D_i^{\delta_t}(x_t)$, and $\mathbb{E}[\|\widehat{D}_{i,t}\|^2 \mid \mathcal{F}_t] \le G_i^2$, where $G_i = \frac{1}{2\nu_i\kappa_{d_i}} = \Theta\left(\frac{\sqrt{d_i}}{\nu_i}\right)$.*

To keep all comparison queries feasible, each player maintains a reference iterate $z_{i,t} \in \mathcal{X}_i$ and plays the contracted action $x_{i,t} = \mathsf{Q}_{i,\delta_t}(z_{i,t}) \in \mathcal{X}_i^{\delta_t}$. Player i then samples $U_{i,t} \sim \mathrm{Unif}(\mathbb{S}^{d_i-1})$ and queries $(x_{i,t} + \delta_t U_{i,t}, x_{-i,t})$ versus $(x_{i,t} - \delta_t U_{i,t}, x_{-i,t})$, which are feasible by construction. From the observed sign $Y_{i,t}$, the player updates the reference iterate as follows:

$$z_{i,t+1} = \Pi_{\mathcal{X}_i}\left(z_{i,t} - \eta_{i,t}\widehat{D}_{i,t}\right), \quad \widehat{D}_{i,t} = \frac{Y_{i,t}}{2\nu_i\kappa_{d_i}} U_{i,t}, \quad \text{(CNGP)}$$

We call this procedure comparison-normalized gradient play (CNGP). The profile x_t is the action profile actually played and used as the center of the comparison queries, while z_t is the internal reference iterate. Note that $\|x_{i,t} - z_{i,t}\| \le \frac{R_i}{\rho_i}\delta_t$, so the two sequences become asymptotically indistinguishable when $\delta_t \to 0$. Each round still uses one pairwise comparison per player, and all players' queries can be performed in parallel. The full pseudocode is provided in Appendix E.1.

5.2 One-Bit Ordinal Regret Guarantees

Define the finite-resolution regret and expected pseudo-regret respectively defined as follows:

$$\mathrm{Reg}_{i,T}^\delta := \sup_{y_i\in\mathcal{X}_i} \sum_{t=1}^T \left\langle D_i^{\delta_t}(x_t), x_{i,t} - y_i\right\rangle, \; \overline{\mathrm{Reg}}_{i,T}^\delta := \sup_{y_i\in\mathcal{X}_i} \mathbb{E}\left[\sum_{t=1}^T \left\langle D_i^{\delta_t}(x_t), x_{i,t} - y_i\right\rangle\right]. \quad (11)$$

The benchmark is evaluated at the played actions x_t; the reference iterates z_t enter only the implementation of (CNGP). Both quantities are invariant under arbitrary strictly increasing transformations of the players' utilities since D_i^δ is invariant (by Proposition 5.1).

Theorem 5.1 (Directional regret from one comparison per round)**.** *Let player $i \in [N]$ run (CNGP) with constant stepsize $\eta_i > 0$ and radii $0 < \delta_t < \rho$. Then $\overline{\mathrm{Reg}}_{i,T}^{\delta} \le \frac{R_i^2}{2\eta_i} + \frac{\eta_i G_i^2 T}{2} + \frac{R_i}{\rho_i}\sum_{t=1}^{T}\delta_t$ for any $T \ge 1$. If in addition $\eta_i = \frac{R_i}{G_i\sqrt{T}}, \delta_t = \frac{\delta}{\sqrt{T}}, 0 < \delta < \rho$, then*

$$\overline{\mathrm{Reg}}_{i,T}^{\delta} \le R_i(G_i+1)\sqrt{T} = O\left(\frac{R_i\sqrt{d_i T}}{\nu_i}\right), \mathbb{E}[\mathrm{Reg}_{i,T}^{\delta}] \le 3R_iG_i\sqrt{T} = O\left(\frac{R_i\sqrt{d_i T}}{\nu_i}\right). \tag{12}$$

Furthermore, $\mathbb{E}\big[[\mathrm{Reg}_{i,T}^{\delta}]_+\big] \le 4R_iG_i\sqrt{T}$.

The guarantee requires no monotonicity, potential structure, or assumptions on opponents; the $\sqrt{d_i}$ factor is the cost of one random comparison per round (minibatching is discussed in Appendix E.2.5). Importantly, $D^\delta(x) \to D(x)$ only pointwise away from stationary blocks, so vanishing resolution does not in general imply sublinear infinitesimal regret; see Proposition E.1. This motivates evaluating finite-query learning using the exact oracle-identified field D^δ.

5.3 Last-Iterate Nash Convergence in Ordinal Potential Games

In this section, we will suppose that the game has an additional ordinal potential structure to prove convergence to Nash equilibria. Under this structure, although the potential Φ is never observed, every unilateral comparison generated by a player coincides with the corresponding comparison of Φ: for every $i \in [N]$, every $\delta \in (0, \rho)$, $x \in \mathcal{X}^\delta$, and $u \in \mathbb{S}^{d_i-1}$,

$$\mathrm{sgn}\left(J_i(x_i + \delta u, x_{-i}) - J_i(x_i - \delta u, x_{-i})\right) = \mathrm{sgn}\left(\Phi(x_i + \delta u, x_{-i}) - \Phi(x_i - \delta u, x_{-i})\right).$$

Consequently, $D_{i,J_i}^{\delta}(x) = D_{i,\Phi}^{\delta}(x)$ for any $x \in \mathcal{X}^\delta$. Hence the comparison signal used by CNGP is exactly the finite-resolution comparison field of a common ordinal potential, even though the potential itself is never observed.

Assumption 5.1 (Nondegeneracy)**.** For every $i \in [N]$ and $x \in \mathcal{X}$, $\nabla_i J_i(x) = 0 \iff \nabla_i \Phi(x) = 0$.

This condition rules out ordinal representations that preserve all unilateral comparisons while introducing or removing stationary points. Together with the ordinal-potential property, it ensures that the players' and potential's first-order preference directions are positively aligned whenever nonzero.

Assumption 5.2 (Robbins-Monro stepsizes and vanishing resolution)**.** The stepsizes (η_t) and the resolution (δ_t) satisfy: $\sum_{t=1}^{\infty}\eta_t = \infty, \sum_{t=1}^{\infty}\eta_t^2 < \infty$ with $\eta_t > 0, \delta_t \in (0, \rho), \forall t \in \mathbb{N}$ and $\delta_t \to 0$.

Theorem 5.2 (Almost-sure last-iterate convergence to Nash equilibria)**.** *Let $\mathcal{G}$ be an ordinal potential game with ordinal potential Φ, and suppose that each $\mathcal{X}_i$ is compact and convex with nonempty interior. Assume that the costs J_i and the potential Φ are continuously differentiable on an open neighborhood of $\mathcal{X}$ on which Φ is L-smooth. Let Assumptions 3.1 and 5.1 hold, and suppose that $\mathrm{int}_{\mathbb{R}}\big(\Phi(\mathrm{NE}(\mathcal{G}))\big) = \varnothing$. Let (x_t) be generated by (CNGP), where $\widehat{D}_{i,t}$ is defined in (CNGP) under the gap-independent noisy comparison model with $\nu_i \in (0, 1/2]$. Suppose that the stepsizes and comparison radii satisfy Assumption 5.2. Then, with probability one, $\mathrm{dist}\big(x_t, \mathrm{NE}(\mathcal{G})\big) \longrightarrow 0$ and $\Phi(x_t)$ converges. If $\mathrm{NE}(\mathcal{G})$ is finite, then there exists an $\mathrm{NE}(\mathcal{G})$-valued random variable X_∞ such that $x_t \to X_\infty$ almost surely. In particular, if the Nash equilibrium is unique, the iterates converge to it almost surely.*

A convenient admissible choice is $\eta_t = \frac{\eta}{t^p}, \delta_t = \frac{\delta}{t^q}$, with $1/2 < p \le 1$, $q > 1 - p$ and $\delta \in (0, \rho)$.

6 Conclusion and Future Work

We developed an ordinal framework for learning in continuous games from pairwise preference feedback, aligning the information available to players with the regret benchmark, learning dynamics, and equilibrium guarantees. This perspective opens several directions, including richer preference-feedback models and extensions to dynamical games.

AI USE STATEMENT

In this work, ChatGPT was used to assist with literature search and identification of relevant references, drafting and editing portions of the manuscript, code generation for the illustrative figure, and technical proof development. In particular, ChatGPT was used interactively to explore proof strategies, refine technical arguments, and assist in the development of proofs for some of the main results.

The research questions, conceptual framing, main mathematical definitions, theorem statements, and decisions about which results to pursue were developed by the authors, including through brainstorming with AI. AI-generated proof suggestions were subsequently checked, revised, and integrated by the authors, and references identified with AI assistance were verified against the original sources. The authors take full responsibility for the final content of the manuscript, including all claims, proofs, citations, and artifacts produced with the assistance of generative AI.

REFERENCES

Omar Abbadi, Rida Laraki, and Panayotis Mertikopoulos. What preferences can—and cannot—predict in multi-agent online learning. In *Forty-third International Conference on Machine Learning*, 2026. 3, 16

Akhil Agnihotri, Rahul Jain, Deepak Ramachandran, and Zheng Wen. Best policy learning from trajectory preference feedback. In *International Conference on Artificial Intelligence and Statistics*, pp. 2116–2124, 2026. 16

Mete Şeref Ahunbay. First-order (coarse) correlated equilibria in non-concave games. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, pp. 1800–1811, 2026. 16

Viktor Bengs, Róbert Busa-Fekete, Adil El Mesaoudi-Paul, and Eyke Hüllermeier. Preference-based online learning with dueling bandits: A survey. *Journal of Machine Learning Research*, 22(7): 1–108, 2021. 3, 16

Sebastian Bervoets, Mario Bravo, and Mathieu Faure. Learning with minimal information in continuous games. *Theoretical Economics*, 15(4):1471–1508, 2020. 3, 16

Oliver Biggar and Iman Shames. Preference graphs: a combinatorial tool for game theory. *arXiv preprint arXiv:2502.03546*, 2025. 3, 16

Avrim Blum, Meghal Gupta, Gene Li, Naren Sarayu Manoj, Aadirupa Saha, and Yuanyuan Yang. Dueling optimization with a monotone adversary. In *International Conference on Algorithmic Learning Theory*, pp. 221–243. PMLR, 2024. 3, 16

Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013. 1

Mario Bravo, David Leslie, and Panayotis Mertikopoulos. Bandit learning in concave n-person games. *Advances in Neural Information Processing Systems*, 31, 2018. 1, 16

HanQin Cai, Daniel McKenzie, Wotao Yin, and Zhenliang Zhang. A one-bit, comparison-based gradient estimator. *Applied and Computational Harmonic Analysis*, 60:242–266, 2022. 4, 8, 16

Yang Cai, Constantinos Daskalakis, Haipeng Luo, Chen-Yu Wei, and Weiqiang Zheng. Proximal regret and proximal correlated equilibria: A new tractable solution concept for online learning and games. In *Proceedings of the 58th Annual ACM Symposium on Theory of Computing*, pp. 1639–1650, 2026. 16

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press Cambridge, 2006. 1

Peter Chen, Xi Chen, Wotao Yin, and Tianyi Lin. ComPO: Preference alignment via comparison oracles. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 16

Minhao Cheng, Simranjit Singh, Patrick H. Chen, Pin-Yu Chen, Sijia Liu, and Cho-Jui Hsieh. Sign-opt: A query-efficient hard-label adversarial attack. In *International Conference on Learning Representations*, 2020. 4

Man-Wah Cheung. Pairwise comparison dynamics for games with continuous strategy space. *Journal of Economic Theory*, 153:344–375, 2014. 3, 16

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017. 1, 16

Vincent Conitzer. The complexity of computing robust mediated equilibria in ordinal games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 9607–9615, 2024. 3, 5, 16

Jorge Cortés. Finite-time convergent gradient flows with applications to network consensus. *Automatica*, 42(11):1993–2000, 2006. 16

Jose B Cruz Jr and Marwan A Simaan. Ordinal games and generalized nash and stackelberg solutions. *Journal of optimization theory and applications*, 107(2):205–222, 2000. 16

Ashok Cutkosky and Harsh Mehta. Momentum improves normalized sgd. In *International conference on machine learning*, pp. 2260–2268. PMLR, 2020. 16

Martin Dindoš and Claudio Mezzetti. Better-reply dynamics and global convergence to nash equilibrium in aggregative games. *Games and Economic Behavior*, 54(2):261–292, 2006. 3

Jacques Durieu, Hans Haller, Nicolas Quérou, and Philippe Solal. Ordinal games. *International Game Theory Review*, 10(02):177–194, 2008. 16

Christian Ewerhart. Ordinal potentials in smooth games: C. ewerhart. *Economic Theory*, 70(4): 1069–1100, 2020. 16

Xuening Feng, Zhaohui Jiang, Timo Kaufmann, Eyke Hüllermeier, Paul Weng, and Yifei Zhu. Comparing comparisons: Informative and easy human feedback with distinguishability queries. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 16727–16743. PMLR, 13–19 Jul 2025. 16

Weibo Fu, Guanyang Wang, and Jun Yan. On the minimax spherical designs. *Random Structures & Algorithms*, 62(1):131–154, 2023. 30, 43

Drew Fudenberg and David K Levine. *The theory of learning in games*, volume 2. MIT press, 1998. 1

Xiand Gao, Xiaobo Li, and Shuzhong Zhang. Online learning with non-convex losses and non-stationary regret. In Amos Storkey and Fernando Perez-Cruz (eds.), *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pp. 235–243. PMLR, 09–11 Apr 2018. 16

Ian Gemp, Daniel Halpern, and Kate Larson. Rl without rewards: A social choice approach to intermediate preferences in context-ordinal mdps. In *Third Reinforcement Learning Beyond Rewards Workshop at the Reinforcement Learning Conference*, 2026a. 16

Ian Gemp, Crystal Qian, Marc Lanctot, and Kate Larson. Nash without numbers: A social choice approach to mixed equilibria in context-ordinal games. *arXiv preprint arXiv:2605.07996*, 2026b. 2, 3, 5, 16

Kushagra Gupta, Surya Murthy, Mustafa Karabag, Ufuk Topcu, and David Fridovich-Keil. Cooperative bargaining games without utilities: Mediated solutions from direction oracles. *Advances in Neural Information Processing Systems*, 38:3780–3805, 2025. 4

Peter J Hammond. Expected utility in non-cooperative game theory. In *Handbook of Utility Theory: Volume 2 Extensions*, pp. 979–1063. Springer, 2004. 5

Peter J Hammond. Utility invariance in non-cooperative games. In *Advances in Public Economics: Utility, Choice and Welfare: A Festschrift for Christian Seidl*, pp. 31–50. Springer, 2005. 5, 16

Elad Hazan, Kfir Levy, and Shai Shalev-Shwartz. Beyond convexity: Stochastic quasi-convex optimization. *Advances in neural information processing systems*, 28, 2015. 16

Donald Joseph Hejna and Dorsa Sadigh. Few-shot preference learning for human-in-the-loop RL. In *6th Annual Conference on Robot Learning*, 2022. 1

Amélie Heliou, Johanne Cohen, and Panayotis Mertikopoulos. Learning with bandit feedback in potential games. *Advances in Neural Information Processing Systems*, 30, 2017. 16

Kevin G Jamieson, Robert Nowak, and Ben Recht. Query complexity of derivative-free optimization. *Advances in neural information processing systems*, 25, 2012. 3, 16

Dietmar Jannach, Paul Resnick, Alexander Tuzhilin, and Markus Zanker. Recommender systems—beyond matrix completion. *Communications of the ACM*, 59(11):94–102, 2016. 1

CÃŠdric Josz. Implicit regularization of normalized gradient descent. *arXiv preprint arXiv:2602.08177*, 2026. 16

Pablo Krupa and Alberto Bemporad. Learning generalized nash equilibria from pairwise preferences. *IEEE Control Systems Letters*, 2026. 3

Wataru Kumagai. Regret analysis for continuous dueling bandit. *Advances in Neural Information Processing Systems*, 30, 2017. 3, 16

Mingyang Liu, Yongshan Chen, Zhiyuan Fan, Gabriele Farina, Asuman E. Ozdaglar, and Kaiqing Zhang. Online learning and equilibrium computation with ranking feedback. In *The Fourteenth International Conference on Learning Representations*, 2026. 2, 3, 16, 17

Yiyang Lu, Hareshkumar Jadav, Mohammad Pedramfar, Ranveer Singh, and Vaneet Aggarwal. Online convex optimization with dueling feedback. *arXiv preprint arXiv:2608.15050*, 2026. 3, 16

Szymon Majewski, Błażej Miasojedow, and Eric Moulines. Analysis of nonsmooth stochastic approximation: the differential inclusion approach. *arXiv preprint arXiv:1805.01916*, 2018. 34, 37

Panayotis Mertikopoulos and Zhengyuan Zhou. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1):465–507, 2019. 1, 16

Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996. 16

Remi Munos, Michal Valko, Daniele Calandriello, Mohammad Gheshlaghi Azar, Mark Rowland, Zhaohan Daniel Guo, Yunhao Tang, Matthieu Geist, Thomas Mesnard, Côme Fiegel, Andrea Michi, Marco Selvi, Sertan Girgin, Nikola Momchev, Olivier Bachem, Daniel J Mankowitz, Doina Precup, and Bilal Piot. Nash learning from human feedback. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 36743–36768, 21–27 Jul 2024. 16

Ryan Murray, Brian Swenson, and Soummya Kar. Revisiting normalized gradient descent: Fast evasion of saddle points. *IEEE Transactions on Automatic Control*, 64(11):4818–4824, 2019. doi: 10.1109/TAC.2019.2914998. 16

Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani (eds.). *Algorithmic Game Theory*. Cambridge University Press, 2007. 1

Martin J Osborne and Ariel Rubinstein. *A course in game theory*. MIT press, 1994. 1

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35: 27730–27744, 2022. 1, 16

Jorge I Poveda, Miroslav Krstić, and Tamer Başar. Fixed-time nash equilibrium seeking in time-varying networks. *IEEE Transactions on Automatic Control*, 68(4):1954–1969, 2022. 4, 16

H. Robbins and D. Siegmund. A convergence theorem for nonnegative almost supermartingales and some applications. In *Optimizing Methods in Statistics*, pp. 233–257. Academic Press, 1971. 44

Aadirupa Saha, Tomer Koren, and Yishay Mansour. Dueling convex optimization. In *International Conference on Machine Learning*, 2021. 1, 2, 3, 4, 5, 8, 16

Aadirupa Saha, Vitaly Feldman, Yishay Mansour, and Tomer Koren. Faster convergence with multiway preferences. In *International Conference on Artificial Intelligence and Statistics*, pp. 433–441. PMLR, 2024. 3, 16

Aadirupa Saha, Tomer Koren, and Yishay Mansour. Dueling convex optimization with general preferences. In *International Conference on Machine Learning*, 2025. 3, 16

Katya Scheinberg and Zikai Xiong. Function-free optimization via comparison oracles. *arXiv preprint arXiv:2604.26867*, 2026. 16

Zhiwei Tang, Dmitry Rybin, and Tsung-Hui Chang. Zeroth-order optimization meets human feedback: Provable learning via ranking oracles. In *The Twelfth International Conference on Learning Representations*, 2024. 3, 16

Emanuel Tewolde and Vincent Conitzer. Game transformations that preserve nash equilibria or best-response sets. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pp. 2984–2993. International Joint Conferences on Artificial Intelligence Organization, 8 2024. Main Track. 16

Helin Wang, Chenyi Zhang, Xiwen Tao, Yexin Zhang, and Tongyang Li. Finding stationary points by comparisons. In *Forty-third International Conference on Machine Learning*, 2026. 16

Haishan Ye. Explicit and non-asymptotic query complexities of rank-based zeroth-order algorithms on smooth functions. *arXiv preprint arXiv:2512.16200*, 2025. 3, 16

Yisong Yue and Thorsten Joachims. Interactively optimizing information retrieval systems as a dueling bandits problem. In *Proceedings of the 26th annual international conference on machine learning*, pp. 1201–1208, 2009. 3

Shen-Yi Zhao, Yin-Peng Xie, and Wu-Jun Li. On the convergence and improvement of stochastic normalized gradient descent. *Science China Information Sciences*, 64(3):132103, 2021. 16

Shen-Yi Zhao, Chang-Wei Shi, Yin-Peng Xie, and Wu-Jun Li. Stochastic normalized gradient descent with momentum for large-batch training. *Science China Information Sciences*, 67(11): 212101, 2024. 16

Banghua Zhu, Michael Jordan, and Jiantao Jiao. Principled reinforcement learning with human feedback from pairwise or k-wise comparisons. In *International Conference on Machine Learning*, pp. 43037–43067. PMLR, 2023. 16

Tianchen Zhu, Yue Qiu, Haoyi Zhou, and Jianxin Li. Decoding global preferences: Temporal and cooperative dependency modeling in multi-agent preference-based reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 17202–17210, 2024. 16

CONTENTS

1 Introduction **1**

1.1 Contributions . . . 2

1.2 Related Work . . . 3

2 Continuous Games with Pairwise Preference Feedback **4**

3 Identifiability from Preference Feedback **5**

3.1 Classical Regret and CCE Are Not Identifiable . . . 5

3.2 Preference-Invariant First-Order Benchmarks . . . 6

3.3 From Ordinal Regret to Ordinal Coarse Equilibrium . . . 6

4 Learning on First-Order Preference Geometry **7**

4.1 Continuous-Time Block-Normalized Dynamics . . . 7

4.2 Discrete-Time No-Ordinal-Regret Learning . . . 7

5 One-Bit Learning from Pairwise Comparisons **7**

5.1 The Finite-Resolution Comparison Vector Field . . . 8

5.2 One-Bit Ordinal Regret Guarantees . . . 8

5.3 Last-Iterate Nash Convergence in Ordinal Potential Games . . . 9

6 Conclusion and Future Work **9**

A Extended Related Work Discussion **16**

B Proofs for Section 3: Ordinal Feedback, Invariance and Identification **17**

B.1 Proof of Lemma 3.1 . . . 17

B.2 Proof of Proposition 3.1 . . . 17

B.3 Proof of Proposition 3.2 . . . 17

B.4 Proof of Theorem 3.1 . . . 18

B.5 Proof of Proposition 3.3 . . . 21

C Continuous-Time Analysis of (BNPGD) **22**

C.1 Discontinuity example of the preference-direction vector field . . . 22

C.2 Continuous-time regret and equilibrium . . . 22

C.3 Nash convergence in ordinal potential games . . . 23

D Proofs for Section 4: Ordinal Learning Dynamics **23**

D.1 Proofs for Section 4.1: Continuous-time ordinal dynamics . . . 23

D.1.1 Proof of Theorem C.1 . . . 23

D.1.2 Proof of Lemma C.1 . . . 24

D.1.3 Proof of Theorem C.2 25
D.2 Proofs for Section 4.2: Discrete-time ordinal gradient play 27
D.2.1 Proof of Theorem 4.1 27

E Proofs for Section 5: Learning from Pairwise Comparisons **27**
E.1 Pseudocode for Comparison-Normalized Gradient Play 27
E.2 Proofs for Section 5.1: Directional regret bound from comparisons 27
E.2.1 Finite-resolution regret does not control infinitesimal ordinal regret 27
E.2.2 Proof of Proposition 5.1 29
E.2.3 Proof of Lemma 5.1 30
E.2.4 Proof of Theorem 5.1 30
E.2.5 Trading comparisons for dimension dependence using minibatch 33
E.3 Proofs for Section 5.3 34
E.3.1 Proof of Theorem 5.2: Almost sure convergence to Nash in ordinal potential games 34
E.3.2 Almost sure convergence to Nash in ordinal potential games in the unconstrained case 41
E.4 Finite-time first-order stationarity 45

A Extended Related Work Discussion

Comparison-based optimization and ranking feedback. Preference-based online learning has a long history in dueling bandits; see Bengs et al. (2021) for a survey. In continuous single-objective optimization, comparison queries can replace numerical function evaluations and still provide useful directional information (Jamieson et al., 2012; Kumagai, 2017; Saha et al., 2021; Cai et al., 2022). Subsequent work considers richer preference models, adversarial comparisons, multiway feedback, and ranking oracles (Blum et al., 2024; Saha et al., 2024; 2025; Tang et al., 2024; Ye, 2025; Chen et al., 2025; Lu et al., 2026). Recent work also studies nonconvex optimization from comparisons: Wang et al. (2026) give comparison-query guarantees for reaching first-order stationary points of a smooth objective, while Scheinberg & Xiong (2026) take a representation-free perspective in which the preference relation itself is primitive and formulate optimality and stationarity certificates through the geometry of preference level sets. This distinction is relevant here because quantitative gradient-norm criteria depend on the chosen cardinal representation, whereas our goal is to define learning objects determined by comparisons themselves. Relatedly, Liu et al. (2026) study standard external regret from ranking observations. For instantaneous-utility rankings they show that sublinear external regret is impossible in general, and obtain positive results under additional temporal structure; their stochastic Plackett–Luce model is gap-dependent, so repeated rankings can be used to estimate numerical utilities. In contrast, our comparison noise is gap-independent: conditional on the underlying ordering, repeated observations contain no information about the magnitude of the utility gap.

Learning from human preferences. Pairwise and ranking feedback is central to reinforcement learning from human feedback (Christiano et al., 2017; Ouyang et al., 2022; Zhu et al., 2023). Beyond the standard reward-modeling pipeline, Munos et al. (2024) learn a pairwise preference model and formulate policy optimization through a regularized Nash equilibrium; other recent work studies best-policy identification from trajectory preferences (Agnihotri et al., 2026), preference-based multi-agent reinforcement learning (Zhu et al., 2024), and ordinal formulations of reinforcement learning (Gemp et al., 2026a). The information contained in the feedback itself is also an active concern: Feng et al. (2025), for example, augment ordinary trajectory comparisons with queries designed to elicit preference strength. Our setting isolates the opposite extreme—binary feedback containing only the ordering—and asks what learning and equilibrium guarantees remain identifiable without recovering preference magnitudes.

Ordinal and invariant learning in games. Ordinal games and ordinal potential games have a long history in game theory (Monderer & Shapley, 1996; Cruz Jr & Simaan, 2000; Durieu et al., 2008; Ewerhart, 2020), as do questions concerning utility representations and invariance (Hammond, 2005; Tewolde & Conitzer, 2024). Recent work has renewed these questions from an algorithmic perspective. Preference graphs encode the unilateral ordering structure of finite games (Biggar & Shames, 2025); Conitzer (2024) studies equilibrium concepts robust to compatible cardinal representations; and Gemp et al. (2026b) develops mixed equilibrium, regret, and learning notions directly for finite context-ordinal games. Abbadi et al. (2026) investigate which aspects of the long-run behavior of no-regret dynamics are, and are not, determined by the preference structure of a finite game. In smooth games, Ahunbay (2026) develop first-order notions of coarse correlated equilibrium based on regret against local vector-field deviations. Concurrently, Cai et al. (2026) develop similar notions.

Normalized and limited-feedback dynamics. Normalized gradient and subgradient methods are classical in optimization, with deterministic, stochastic, and online variants analyzed under a range of convex and nonconvex assumptions (Cortés, 2006; Hazan et al., 2015; Gao et al., 2018; Murray et al., 2019; Cutkosky & Mehta, 2020; Zhao et al., 2021; 2024; Josz, 2026). Our use of normalization has a different purpose: player-wise normalization removes precisely the gradient magnitudes that unilateral preference comparisons do not identify. Accordingly, our stochastic algorithm does not normalize a noisy gradient; a one-bit comparison instead gives an unbiased estimator of the exact finite-resolution ordinal field D^δ. Learning in continuous games with limited numerical feedback has also been studied under bandit and payoff-based information (Heliou et al., 2017; Bravo et al., 2018; Mertikopoulos & Zhou, 2019; Bervoets et al., 2020), while Cheung (2014) studies deterministic pairwise-comparison dynamics in continuous-strategy population games and Poveda et al. (2022) considers rescaled dynamics for Nash seeking. In contrast, the dynamics studied here are dictated

by ordinal identifiability, and their finite-comparison implementation yields both no-ordinal-regret and last-iterate Nash-convergence guarantees.

Additional details on the comparison to Liu et al. (2026). The Plackett–Luce (PL) model is not invariant under arbitrary strictly increasing transformations of utilities: its ranking probabilities depend on utility differences and therefore encode cardinal utility-gap information. As the temperature $\tau \to 0^+$, the PL model becomes increasingly deterministic and approaches purely ordinal ranking feedback, since the ordering remains observable while information about the magnitude of utility gaps vanishes. This limiting regime is therefore closely related to the information structure considered here. For instantaneous-utility rankings, Liu et al. (2026) show that standard sublinear external regret is impossible in general. Their positive result in this setting assumes that the utility sequence has sublinear temporal variation, allowing recent ranking observations to be pooled in a sliding window to estimate an approximately constant cardinal utility vector. Their estimation bound contains a factor $\tau(e^{1/\tau}+1)^2$, which diverges as $\tau \to 0^+$. Thus, precisely as the PL observations approach purely ordinal rankings, recovering the cardinal information needed for standard external regret becomes increasingly ill-conditioned.

B Proofs for Section 3: Ordinal Feedback, Invariance and Identification

B.1 Proof of Lemma 3.1

Proof. If $F_i(x) \neq 0$, then $\frac{\widetilde{F}_i(x)}{\|\widetilde{F}_i(x)\|} = \frac{\phi_i'(J_i(x))F_i(x)}{\phi_i'(J_i(x))\|F_i(x)\|} = D_i(x)$, and $\widetilde{F}_i(x) = 0$ if and only if $F_i(x) = 0$. □

B.2 Proof of Proposition 3.1

Proof. Consider two players with $\mathcal{X}_i = \mathcal{X}_j = [0,1]$, and let player i have the smooth cost

$$J_i(x_i, x_j) = 2 - x_j + x_i(1 - 2x_j).$$

The cost of the second player is immaterial. Suppose player i always selects $x_{i,t} = 1$, while player j alternates between $x_{j,t} = 0$ and $x_{j,t} = 1$. Over every pair of rounds, player i incurs total cost $3 + 0 = 3$. Any fixed comparator $y \in [0,1]$ incurs

$$J_i(y, 0) + J_i(y, 1) = (2 + y) + (1 - y) = 3.$$

Hence $\mathrm{Reg}_{i,T}^{J} = 0$ for every even T.

Now consider the ordinally equivalent representation $\widetilde{J}_i = \exp(J_i)$. The learner incurs $e^3 + 1$ over every pair of rounds, while a fixed comparator y incurs $e^{2+y} + e^{1-y}$. Since $y \mapsto e^{2+y} + e^{1-y}$ is strictly increasing on $[0,1]$, its minimum is attained at $y = 0$. Therefore

$$\mathrm{Reg}_{i,T}^{\widetilde{J}} = \frac{T}{2}\left(e^3 + 1 - e^2 - e\right) = \Omega(T),$$

while $\widetilde{J}_i = \exp(J_i)$ induces exactly the same preference comparisons as J_i. □

B.3 Proof of Proposition 3.2

Proof. Consider a two-player game with $\mathcal{X}_1 = \mathcal{X}_2 = [0,1]$, and costs

$$J_1(x_1, x_2) = 2 - x_2 + x_1(1 - 2x_2), \qquad J_2(x_1, x_2) = 0.$$

Let $\mu = \frac{1}{2}\delta_{(1,0)} + \frac{1}{2}\delta_{(1,1)}$. We first show that μ is a coarse correlated equilibrium of $\mathcal{G}$. Under μ, player 1 incurs expected cost

$$\mathbb{E}_{x\sim\mu}[J_1(x)] = \frac{1}{2}J_1(1,0) + \frac{1}{2}J_1(1,1) = \frac{1}{2}(3+0) = \frac{3}{2}.$$

For any fixed deviation $y_1 \in [0,1]$,

$$\mathbb{E}_{x\sim\mu}\big[J_1(y_1, x_2)\big] = \frac{1}{2}J_1(y_1,0) + \frac{1}{2}J_1(y_1,1) = \frac{1}{2}(2+y_1) + \frac{1}{2}(1-y_1) = \frac{3}{2}.$$

Thus player 1 cannot improve its expected cost by any fixed deviation. Player 2 is indifferent between all actions since $J_2 \equiv 0$. Therefore $\mu \in \mathrm{CCE}(\mathcal{G})$.

Now define the ordinally equivalent game $\widetilde{\mathcal{G}}$ by $\widetilde{J}_1 = e^{J_1}, \widetilde{J}_2 = J_2$. Since $s \mapsto e^s$ is strictly increasing, the transformation preserves all unilateral preference comparisons and best responses. Under the same distribution μ, player 1 now incurs expected cost

$$\mathbb{E}_{x\sim\mu}[\widetilde{J}_1(x)] = \frac{1}{2}e^{J_1(1,0)} + \frac{1}{2}e^{J_1(1,1)} = \frac{1}{2}(e^3+1).$$

On the other hand, the fixed deviation $y_1 = 0$ yields

$$\mathbb{E}_{x\sim\mu}\big[\widetilde{J}_1(0,x_2)\big] = \frac{1}{2}e^{J_1(0,0)} + \frac{1}{2}e^{J_1(0,1)} = \frac{1}{2}(e^2+e).$$

Since $e^3 + 1 - e^2 - e = (e-1)^2(e+1) > 0$, we have

$$\mathbb{E}_{x\sim\mu}[\widetilde{J}_1(x)] > \mathbb{E}_{x\sim\mu}\big[\widetilde{J}_1(0,x_2)\big].$$

Hence player 1 has a profitable fixed deviation, so $\mu \notin \mathrm{CCE}(\widetilde{\mathcal{G}})$. This proves that the CCE correspondence is not preserved under arbitrary strictly increasing transformations. □

B.4 Proof of Theorem 3.1

The inclusion $\Delta(\mathrm{NE}(\mathcal{G})) \subseteq \bigcap_{\widetilde{\mathcal{G}}\sim_{\mathrm{ord}}\mathcal{G}} \mathrm{CCE}(\widetilde{\mathcal{G}})$ is immediate from the invariance of best responses under unilateral ordinal transformations. For the reverse inclusion, recall first that the usual CCE condition only controls *average* regret, so positive regret at some opponents' profiles may be offset by negative regret at others. Robustness over the ordinal equivalence class rules out precisely this cross-context compensation: if, for some fixed deviation y_i, the conditional regret were positive on a set of opponents' profiles of positive measure, one could choose α_i large on that set and small elsewhere, without changing any unilateral comparison, thereby making the overall expected regret positive. Hence ordinal robustness forces every fixed-deviation conditional regret to be nonpositive almost everywhere. Taking the infimum over deviations then implies that the player's conditional expected cost equals the pointwise minimum cost, which is possible only if the player is best-responding almost surely. Applying this argument to every player yields support on $\mathrm{NE}(\mathcal{G})$. We provide a complete proof below in this section.

Proof idea. We start by highlighting the intuition behind the proof of the more challenging and interesting inclusion $\Big(\bigcap_{\widetilde{\mathcal{G}}\in[\mathcal{G}]_{\mathrm{ord}}} \mathrm{CCE}(\widetilde{\mathcal{G}}) \subseteq \Delta(\mathrm{NE}(\mathcal{G}))\Big)$.

For a fixed deviation y_i, ordinary CCE only requires

$$\int q_{i,y_i}(z)\,\lambda(dz) \le 0, \qquad q_{i,y_i}(z) = \mathbb{E}[J_i(X_i,z) \mid X_{-i} = z] - J_i(y_i,z).$$

Hence positive conditional regret at some opponents' profiles may be offset by negative conditional regret at others. Unilateral comparison feedback, however, does not identify the relative cardinal scale of regret across different z: replacing $J_i(x_i,z)$ by $\alpha_i(z)J_i(x_i,z)$ for any positive $\alpha_i(z)$ leaves all unilateral comparisons unchanged. Requiring the CCE inequality to hold under every such reweighting eliminates cross-context cancellation and forces

$$q_{i,y_i}(z) \le 0 \qquad \text{a.e.}$$

for every deviation y_i. Taking the infimum over y_i then gives

$$\mathbb{E}[J_i(X_i,z) \mid X_{-i} = z] = \min_{y_i} J_i(y_i,z),$$

which is possible only if X_i is a best response almost surely. Thus robustness to all cardinal representations consistent with the comparison oracle turns the global expected-regret inequalities defining CCE into pointwise best-response conditions, yielding Nash support.

In what follows, we provide a complete and detailed proof of Theorem 3.1.

Proof. We prove the two inclusions separately.

(i) Proof of $\Delta(\mathrm{NE}(\mathcal{G})) \subseteq \bigcap_{\widetilde{\mathcal{G}}\in[\mathcal{G}]_{\mathrm{ord}}} \mathrm{CCE}(\widetilde{\mathcal{G}})$. Let $\mu \in \Delta(\mathrm{NE}(\mathcal{G}))$, and consider an arbitrary $\widetilde{\mathcal{G}} \in [\mathcal{G}]_{\mathrm{ord}}$. By ordinal equivalence, for every $i \in [N]$, every $x_{-i} \in \mathcal{X}_{-i}$, and every $a_i, b_i \in \mathcal{X}_i$,

$$\mathrm{sgn}\big(J_i(a_i, x_{-i}) - J_i(b_i, x_{-i})\big) = \mathrm{sgn}\big(\widetilde{J}_i(a_i, x_{-i}) - \widetilde{J}_i(b_i, x_{-i})\big).$$

Hence the two cost functions induce exactly the same minimizers on every unilateral slice: $\mathrm{BR}_i^{\mathcal{G}}(x_{-i}) = \mathrm{BR}_i^{\widetilde{\mathcal{G}}}(x_{-i})$. Therefore, $\mathrm{NE}(\mathcal{G}) = \mathrm{NE}(\widetilde{\mathcal{G}})$.

Since μ is supported on $\mathrm{NE}(\mathcal{G})$, for every player $i \in [N]$ and every fixed deviation $y_i \in \mathcal{X}_i$, $\widetilde{J}_i(x_i, x_{-i}) \le \widetilde{J}_i(y_i, x_{-i})$ for μ-a.e. x. Integrating gives $\mathbb{E}_{x\sim\mu}\big[\widetilde{J}_i(x)\big] \le \mathbb{E}_{x\sim\mu}\big[\widetilde{J}_i(y_i, x_{-i})\big]$. Thus $\mu \in \mathrm{CCE}(\widetilde{\mathcal{G}})$. Since $\widetilde{\mathcal{G}} \in [\mathcal{G}]_{\mathrm{ord}}$ was arbitrary, $\mu \in \bigcap_{\widetilde{\mathcal{G}}\in[\mathcal{G}]_{\mathrm{ord}}} \mathrm{CCE}(\widetilde{\mathcal{G}})$.

(ii) Proof of $\bigcap_{\widetilde{\mathcal{G}}\in[\mathcal{G}]_{\mathrm{ord}}} \mathrm{CCE}(\widetilde{\mathcal{G}}) \subseteq \Delta(\mathrm{NE}(\mathcal{G}))$. Let $\mu \in \bigcap_{\widetilde{\mathcal{G}}\in[\mathcal{G}]_{\mathrm{ord}}} \mathrm{CCE}(\widetilde{\mathcal{G}})$.

We show that μ is supported on $\mathrm{NE}(\mathcal{G})$.

Proof strategy. Fix a player $i \in [N]$. Conditional on an opponents' profile $z = x_{-i} \in \mathcal{X}_{-i}$, let

$$A_i(z) := \mathbb{E}_\mu[J_i(X_i, z) \mid X_{-i} = z]$$

denote player i's conditional expected cost. For a fixed deviation $y_i \in \mathcal{X}_i$, define its conditional regret for any $z \in \mathcal{X}_{-i}$ as follows:

$$q_{i,y_i}(z) := A_i(z) - J_i(y_i, z).$$

The key step is to show that ordinal robustness forces

$$q_{i,y_i}(z) \le 0 \qquad \text{for a.e. } z$$

for every fixed y_i. Otherwise, an ordinally invisible positive rescaling of costs could amplify the contexts in which $q_{i,y_i} > 0$ and violate the CCE inequality. We then remove the dependence on y_i of the exceptional null set using a countable dense set of deviations. This yields

$$A_i(z) \le \min_{y_i\in\mathcal{X}_i} J_i(y_i, z) \qquad \text{a.e.}$$

The reverse inequality holds automatically because $A_i(z)$ is an average of the values $J_i(\cdot, z)$. Hence the conditional average equals the minimum, which is possible only if the conditional distribution of X_i is supported on best responses. Repeating this argument for every player implies $\mu(\mathrm{NE}(\mathcal{G})) = 1$.

Step 1: Ordinal robustness implies nonpositive conditional regret. Fix $i \in [N]$, write $z := x_{-i}$. Denote by $\pi_{-i} : \mathcal{X} \to \mathcal{X}_{-i}, \pi_{-i}(x) = x_{-i}$, the canonical projection onto the opponents' action space. Let $\lambda := \mu_{-i} := (\pi_{-i})_{\#}\mu$ be the marginal distribution of X_{-i} induced by μ; equivalently, for every Borel set $B \subseteq \mathcal{X}_{-i}$, $\mu_{-i}(B) = \mu\big(\mathcal{X}_i \times B\big)$. Since $\mathcal{X}_i$ and $\mathcal{X}_{-i}$ are compact metric spaces, there exists a regular conditional distribution $\kappa_i(\cdot \mid z)$ of X_i given $X_{-i} = z$ such that the disintegration $\mu(dx_i, dz) = \kappa_i(dx_i \mid z)\,\mu_{-i}(dz) = \kappa_i(dx_i \mid z)\lambda(dz)$ holds.

Define for any $z \in \mathcal{X}_{-i}$ the average cost of player $i \in [N]$ as follows:

$$A_i(z) := \int_{\mathcal{X}_i} J_i(x_i, z)\,\kappa_i(dx_i|z).$$

Since J_i is continuous on the compact set $\mathcal{X}$, it is bounded, and therefore A_i is bounded and measurable. Fix an arbitrary deviation $y_i \in \mathcal{X}_i$ and define the conditional regret

$$q_{i,y_i}(z) := A_i(z) - J_i(y_i, z).$$

We claim that

$$q_{i,y_i}(z) \le 0 \qquad \text{for } \lambda\text{-a.e. } z. \tag{13}$$

To see this, let h be any nonnegative smooth function defined on an open neighborhood of $\mathcal{X}_{-i}$, and let $t \geq 0$. Consider the game $\mathcal{G}^{t,h}$ obtained by replacing player i's cost by

$$J_i^{t,h}(x_i, z) := \big(1 + th(z)\big) J_i(x_i, z)\,, \quad \forall x_i \in \mathcal{X}_i, \forall z \in \mathcal{X}_{-i}\,,$$

and leaving every other player's cost unchanged. Since $1 + th(z) > 0$, this transformation preserves every unilateral comparison: for any $a_i, b_i \in \mathcal{X}_i, z \in \mathcal{X}_{-i}$,

$$\operatorname{sgn}\Big(J_i^{t,h}(a_i, z) - J_i^{t,h}(b_i, z)\Big) = \operatorname{sgn}(J_i(a_i, z) - J_i(b_i, z))\,.$$

Therefore, $\mathcal{G}^{t,h} \sim_{\text{ord}} \mathcal{G}$. Since $\mu \in \bigcap_{\widetilde{\mathcal{G}} \in [\mathcal{G}]_{\text{ord}}} \text{CCE}(\widetilde{\mathcal{G}})$, it follows that $\mu \in \text{CCE}(\mathcal{G}^{t,h})$. Then by definition of $\text{CCE}(\mathcal{G}^{t,h})$, we have for the fixed deviation $y_i \in \mathcal{X}_i$,

$$\mathbb{E}_\mu\big[\big(1 + th(X_{-i})\big)\big(J_i(X_i, X_{-i}) - J_i(y_i, X_{-i})\big)\big] \leq 0.$$

Disintegrating with respect to X_{-i} (recall that $\mu(dx_i, dz) = \kappa_i(dx_i|z)\lambda(dz)$) yields:

$$\int_{\mathcal{X}_{-i}} \big(1 + th(z)\big) q_{i,y_i}(z) \lambda(dz) \leq 0\,, \qquad \forall t \geq 0. \tag{14}$$

Therefore,

$$\int_{\mathcal{X}_{-i}} h(z) q_{i,y_i}(z)\, \lambda(dz) \leq -\frac{1}{t} \int_{\mathcal{X}_{-i}} q_{i,y_i}(z)\, \lambda(dz)\,, \qquad \forall t > 0.$$

Taking the limit $t \to +\infty$ implies that necessarily:

$$\int_{\mathcal{X}_{-i}} h(z) q_{i,y_i}(z) \lambda(dz) \leq 0 \tag{15}$$

for every nonnegative smooth h.

We now localize (15), i.e., we use the freedom to choose the test function h so that it concentrates on an arbitrary region of $\mathcal{X}_{-i}$, to pass from the global integral inequality 15 to the local/almost-everywhere statement (13) by using test functions with localized support. Suppose, toward a contradiction, that

$$\lambda\big(\{z \in \mathcal{X}_{-i} : q_{i,y_i}(z) > 0\}\big) > 0.$$

Then, for some $\varepsilon > 0$, $E_\varepsilon := \{z : q_{i,y_i}(z) \geq \varepsilon\}$ has positive λ-measure. By inner regularity of λ, choose a compact set $K \subseteq E_\varepsilon$, $\lambda(K) > 0$.

Since q_{i,y_i} is bounded by boundedness of J_i over the compact set $\mathcal{X}$, let $M := \|q_{i,y_i}\|_\infty < \infty$. By outer regularity, choose an open neighborhood $U \supseteq K$ such that

$$\lambda(U \setminus K) < \frac{\varepsilon \lambda(K)}{2M},$$

where the claim is immediate if $M = 0$. Choose a smooth bump function h satisfying:

$$0 \leq h \leq 1, \qquad h = 1 \text{ on } K, \qquad \operatorname{supp}(h) \cap \mathcal{X}_{-i} \subseteq U.$$

Then

$$\begin{aligned}
\int h(z) q_{i,y_i}(z)\, \lambda(dz) &\geq \int_K q_{i,y_i}(z)\, \lambda(dz) - \int_{U \setminus K} |q_{i,y_i}(z)|\, \lambda(dz) \\
&\geq \varepsilon \lambda(K) - M \lambda(U \setminus K) \\
&> \frac{\varepsilon \lambda(K)}{2} > 0,
\end{aligned}$$

contradicting (15). Hence $q_{i,y_i}(z) \leq 0, \lambda$-a.e. for every fixed $y_i \in \mathcal{X}_i$. Equivalently,

$$A_i(z) \leq J_i(y_i, z)\,, \qquad \lambda\text{-a.e.} \tag{16}$$

Step 2: Obtain a common full-measure set for all deviations. The exceptional null set in (16) may depend on y_i. To remove this dependence, let $D_i \subseteq \mathcal{X}_i$ be a countable dense subset. For each

$y_i \in D_i$, let Z_{i,y_i} be a set of full λ-measure on which $A_i(z) \le J_i(y_i, z)$. Because D_i is countable, $Z_i := \bigcap_{y_i \in D_i} Z_{i,y_i}$ still satisfies $\lambda(Z_i) = 1$. Thus, for every $z \in Z_i$,

$$A_i(z) \le J_i(y_i, z) \qquad \forall y_i \in D_i.$$

By continuity of $J_i(\cdot, z)$ and density of D_i,

$$A_i(z) \le \inf_{y_i \in D_i} J_i(y_i, z) = \min_{y_i \in \mathcal{X}_i} J_i(y_i, z)\,, \qquad \forall z \in Z_i.$$

We have therefore shown that

$$A_i(z) \le \min_{y_i \in \mathcal{X}_i} J_i(y_i, z) \qquad \lambda\text{-a.e.} \tag{17}$$

Step 3: Nonpositive conditional regret forces best-response play. Note first that for any $z \in \mathcal{X}_{-i}$,

$$J_i(x_i, z) \ge \min_{y_i \in \mathcal{X}_i} J_i(y_i, z)\,, \qquad \forall x_i \in \mathcal{X}_i.$$

Averaging with respect to $\kappa_i(\cdot|z)$ gives

$$A_i(z) = \int_{\mathcal{X}_i} J_i(x_i, z)\, \kappa_i(dx_i \mid z) \ge \min_{y_i \in \mathcal{X}_i} J_i(y_i, z).$$

Together with (17), this yields

$$A_i(z) = \min_{y_i \in \mathcal{X}_i} J_i(y_i, z) \qquad \lambda\text{-a.e.} \tag{18}$$

Now define for any $x_i \in \mathcal{X}_i, x_{-i} \in \mathcal{X}_{-i}$, $r_i(x_i, z) := J_i(x_i, z) - \min_{y_i \in \mathcal{X}_i} J_i(y_i, z) \ge 0$.

For λ-a.e. z, (18) implies

$$\int_{\mathcal{X}_i} r_i(x_i, z)\, \kappa_i(dx_i \mid z) = A_i(z) - \min_{y_i \in \mathcal{X}_i} J_i(y_i, z) = 0.$$

Since $r_i(\cdot, z)$ is nonnegative, it must vanish $\kappa_i(\cdot \mid z)$-almost surely. Therefore

$$\kappa_i\big(\mathrm{BR}_i^{\mathcal{G}}(z) \mid z\big) = 1 \qquad \lambda\text{-a.e. } z.$$

Integrating over z gives

$$\mu\left(\left\{x \in \mathcal{X} : x_i \in \mathrm{BR}_i^{\mathcal{G}}(x_{-i})\right\}\right) = 1\,. \tag{19}$$

Step 4: Intersect the best-response conditions across players. Since (19) holds for every $i \in [N]$, and the number of players is finite,

$$\mu\left(\bigcap_{i=1}^{N} \left\{x \in \mathcal{X} : x_i \in \mathrm{BR}_i^{\mathcal{G}}(x_{-i})\right\}\right) = 1.$$

The set inside the probability is exactly $\mathrm{NE}(\mathcal{G})$. Hence $\mu(\mathrm{NE}(\mathcal{G})) = 1$, or equivalently, $\mu \in \Delta(\mathrm{NE}(\mathcal{G}))$, which concludes the proof of the desired inclusion. □

B.5 Proof of Proposition 3.3

Proof. Since every summand in (5) is nonnegative, $\mathrm{Gap}_D(x) = 0$ if and only if,

$$\langle D_i(x), y_i - x_i \rangle \ge 0 \quad \forall i \in [N], \forall\, y_i \in \mathcal{X}_i. \tag{20}$$

If $\nabla_i J_i(x) \neq 0$, multiplying the previous inequality (20) by the positive scalar $\|\nabla_i J_i(x)\|$ yields $\langle \nabla_i J_i(x), y_i - x_i \rangle \ge 0$ for all y_i. If $\nabla_i J_i(x) = 0$, the same condition holds trivially.

By pseudoconvexity, this first-order condition is sufficient for x_i to globally minimize $J_i(\cdot, x_{-i})$ and hence x is a Nash equilibrium.

Conversely, if x is a NE, since every global minimizer of a differentiable function over a convex set satisfies the corresponding first-order variational inequality, it follows that (20) holds and hence by the first step in this proof we obtain that $\mathrm{Gap}_D(x) = 0$. □

C CONTINUOUS-TIME ANALYSIS OF (BNPGD)

In the unconstrained case (and at interior points of a full-dimensional action set), (BNPGD) reduces to $\dot{x}_i = -\frac{\nabla_i J_i(x)}{\|\nabla_i J_i(x)\|}$ whenever $\nabla_i J_i(x) \neq 0$, and $\dot{x}_i = 0$ otherwise. Because normalization is singular at stationary blocks, the field D_i is continuous on $\{x \in \mathcal{X} : \nabla_i J_i(x) \neq 0\}$ but may be discontinuous where $\nabla_i J_i(x) = 0$ (see Appendix C.1 for a simple example), despite the convention $D_i = 0$ at such points. Accordingly, our continuous-time statements apply to any absolutely continuous trajectory satisfying (BNPGD) almost everywhere. The derivative identities are likewise understood almost everywhere.

Proposition C.1 (Invariance of the dynamics). *Suppose $\widetilde{J}_i = \phi_i \circ J_i$ for every i, where $\phi_i' > 0$. Then the block-normalized projected vector fields generated by $\mathcal{G} = ((\mathcal{X})_{i\in[N]}, (J_i)_{i\in[N]})$ and $\widetilde{\mathcal{G}} = ((\mathcal{X})_{i\in[N]}, (\tilde{J}_i)_{i\in[N]})$ coincide.*

In contrast, ordinary gradient play is transformed blockwise as $-\nabla_i J_i(x) \longmapsto -\phi_i'(J_i(x))\nabla_i J_i(x)$, so ordinally equivalent games may induce different relative adaptation speeds and different joint trajectories due to the player-wise factor $\phi_i'(J_i(x))$. This is specific to the multi-player setting: for a single objective, normalization only reparametrizes the speed along a gradient-flow trajectory, whereas player-wise normalization generally changes the geometry of the coupled dynamics.

In this section, we show that (BNPGD) enjoys a $\mathcal{O}(1/T)$ continuous-time average ordinal regret (see Theorem C.1), and in ordinal potential games the potential is a Lyapunov function yielding convergence to Nash equilibria under suitable conditions (see Theorem C.2).

C.1 DISCONTINUITY EXAMPLE OF THE PREFERENCE-DIRECTION VECTOR FIELD

In one dimension and for a single player ($N = 1$), let $J(x) = \frac{1}{2}x^2$ for any $x \in \mathbb{R}$. Then,

$$D(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases}$$

Hence D is well defined everywhere under the convention $D(0) = 0$, but it is discontinuous at the stationary point $x = 0$.

C.2 CONTINUOUS-TIME REGRET AND EQUILIBRIUM

For continuous play, define

$$\mathrm{Reg}_i^D(T) := \sup_{y_i \in \mathcal{X}_i} \int_0^T \langle D_i(x(t)), x_i(t) - y_i \rangle \, dt. \tag{21}$$

Theorem C.1 (Constant continuous-time directional regret). *Every trajectory of (BNPGD) satisfies $\mathrm{Reg}_i^D(T) \leq \frac{\mathrm{diam}(\mathcal{X}_i)^2}{2}$ for every $T \geq 0$. Consequently, $\frac{\mathrm{Reg}_i^D(T)}{T} = O(T^{-1})$.*

Thus block-normalized play is no-regret with respect to the ordinal benchmark against arbitrary opponent behavior, without monotonicity or potential assumptions.

Let $\mu_T = \frac{1}{T}\int_0^T \delta_{x(t)}\, dt$ denote the occupation measure. Combining Theorem C.1 with (7) gives:

Corollary C.1 (Ordinal coarse equilibrium at rate $O(1/T)$). *The occupation measure of block-normalized play satisfies $\mathcal{E}_D(\mu_T) \leq \frac{1}{2T}\sum_{i=1}^N R_i^2$. Hence time-average play converges to the set of ordinal coarse variational equilibria at rate $O(T^{-1})$.*

This is the ordinal counterpart of the classical no-regret-to-CCE correspondence.

C.3 Nash convergence in ordinal potential games

No-regret alone controls time-average behavior. We next identify a class of games in which the same dynamics also admits a Lyapunov function and converges in the last iterate.

Lemma C.1 (First-order alignment of costs and the ordinal potential)**.** *Suppose that $\mathcal{G}$ is an ordinal potential game with costs J_i and ordinal potential Φ that are C^1 on an open neighborhood of $\mathcal{X}$, and suppose that each $\mathcal{X}_i$ has nonempty interior. Under Assumption 5.1, for every $i \in [N]$ and every $x \in \mathcal{X}$, either $\nabla_i J_i(x) = \nabla_i \Phi(x) = 0$, or there exists a scalar $\lambda_i(x) > 0$ s.t.*

$$\nabla_i J_i(x) = \lambda_i(x) \nabla_i \Phi(x). \tag{22}$$

In particular, whenever the gradients are nonzero, $D_i(x) = \frac{\nabla_i J_i(x)}{\|\nabla_i J_i(x)\|} = \frac{\nabla_i \Phi(x)}{\|\nabla_i \Phi(x)\|}$.

Remark C.1 (Role of Assumption 5.1). The nondegeneracy condition is needed because ordinal equivalence alone does not preserve first-order stationarity. For example, the functions $s \mapsto s$ and $s \mapsto s^3$ induce the same ordering on $\mathbb{R}$, although their derivatives behave differently at the origin. Thus, without nondegeneracy, an ordinal representation could create or remove stationary blocks. Assumption 5.1 rules out precisely this pathology and allows first-order stationarity of the potential to be identified with first-order stationarity of the players' costs.

Thus block normalization removes the player-specific factor $\lambda_i(x)$: within an ordinal potential game, (BNPGD) follows the normalized block gradients of a common potential.

Theorem C.2 (Convergence of (BNPGD) in ordinal potential games)**.** *Let $\mathcal{G}$ be an ordinal potential game with potential Φ such that Assumption 5.1 holds. Then the following statements hold:*

(i) (potential dissipation) $\frac{d}{dt}\Phi(x(t)) = -\sum_{i=1}^N \|\nabla_i \Phi(x(t))\| \, \|\Pi_{T_{\mathcal{X}_i}(x_i)}(-D_i(x))\|^2 \le 0$ for almost every t along every absolutely continuous solution of (BNPGD).

(ii) (block-$\ell_{2,1}$ dissipation) $\frac{d}{dt}\Phi(x(t)) = -\sum_{i=1}^N \|\nabla_i \Phi(x(t))\| = -\|\nabla \Phi(x(t))\|_{2,1}$ for almost every t, if the trajectory remains in $\operatorname{int}(\mathcal{X})$.

(iii) (first-order stationarity) For any $T > 0$, $\min_{0 \le t \le T} \|\nabla \Phi(x(t))\|_{2,1} \le \frac{\Phi(x(0)) - \Phi_{\inf}}{T}$, where $\Phi_{\inf} := \inf_{x \in \mathcal{X}} \Phi(x) > -\infty$, if again the trajectory remains interior.

(iv) (convergence) Assuming a solution of (BNPGD) remains in a compact set $K \subset \operatorname{int}(\mathcal{X})$ and $\nabla \Phi$ is Lipschitz on K, then for every $i \in [N]$, $\|\nabla_i \Phi(x(t))\| \longrightarrow 0$. If in addition Assumption 3.1 holds, then the rest points of (BNPGD) coincide exactly with the set of Nash equilibria and $\operatorname{dist}\big(x(t), \mathrm{NE}\big) \longrightarrow 0$.

The interior dissipation identity admits a useful geometric interpretation. Since $\|\cdot\|_{2,1}$ is dual to $\|v\|_{2,\infty} = \max_i \|v_i\|_2$, $\min_{\|v_i\| \le 1, \, \forall i} \langle \nabla \Phi(x), v \rangle = -\|\nabla \Phi(x)\|_{2,1}$. Hence block-normalized play is the steepest-descent direction of the ordinal potential when each player's instantaneous speed is bounded by one.

D Proofs for Section 4: Ordinal Learning Dynamics

D.1 Proofs for Section 4.1: Continuous-time ordinal dynamics

D.1.1 Proof of Theorem C.1

Proof. Fix $y_i \in \mathcal{X}_i$ and define $V_i(t) = \frac{1}{2}\|x_i(t) - y_i\|^2$. The characterization of the projection in the (BNPGD) dynamics yields $-D_i(x) = \dot{x}_i + n_i$ for some $n_i \in N_{\mathcal{X}_i}(x_i)$, where $N_{\mathcal{X}_i}(x_i)$ denotes the normal cone to the set $\mathcal{X}_i$. Since $\langle n_i, y_i - x_i \rangle \le 0$ by definition of the normal cone,

$$\dot{V}_i(t) = \langle \dot{x}_i, x_i - y_i \rangle = -\langle D_i(x), x_i - y_i \rangle + \langle n_i, y_i - x_i \rangle \le -\langle D_i(x), x_i - y_i \rangle.$$

Integrating the above inequality from 0 to T and rearranging gives:

$$\int_0^T \langle D_i(x(t)), x_i(t) - y_i \rangle \, dt \le \frac{1}{2}\|x_i(0) - y_i\|^2 - \frac{1}{2}\|x_i(T) - y_i\|^2 \le \frac{\operatorname{diam}(\mathcal{X}_i)^2}{2}.$$

Taking the supremum over y_i proves the claim. □

D.1.2 PROOF OF LEMMA C.1

Proof. Let $i \in [N]$ and $x \in \mathcal{X}$. By Assumption 5.1, the two block gradients vanish simultaneously. Thus it remains only to consider the case $\nabla_i J_i(x) \neq 0, \nabla_i \Phi(x) \neq 0$.

Suppose first that $x_i \in \mathrm{int}(\mathcal{X}_i)$. Define the two C^1 functions

$$f_i(y_i) := J_i(y_i, x_{-i}), \qquad \phi_i(y_i) := \Phi(y_i, x_{-i}).$$

The ordinal-potential property implies that, for all $y_i, z_i \in \mathcal{X}_i$,

$$\mathrm{sgn}\big(f_i(y_i) - f_i(z_i)\big) = \mathrm{sgn}\big(\phi_i(y_i) - \phi_i(z_i)\big).$$

We claim that the two gradients $\nabla f_i(x_i)$ and $\nabla \phi_i(x_i)$ are positively collinear.

Proof strategy via orthogonal complements. Set $a := \nabla f_i(x_i), b := \nabla \phi_i(x_i)$. By assumption, $a \neq 0$ and $b \neq 0$. To show that a and b are collinear, it suffices to show that $\mathrm{span}\{a\} = \mathrm{span}\{b\}$. In the following, we equivalently show by taking orthogonal complements that $a^\perp := \{d : \langle a, d\rangle = 0\}$ and $b^\perp := \{d : \langle b, d\rangle = 0\}$ coincide, i.e., $a^\perp = b^\perp$. Finally, since a and b are both nonzero, $a^\perp$ and $b^\perp$ are both $(d_i - 1)$-dimensional subspaces, it suffices to show that $a^\perp \subseteq b^\perp$ to conclude the equality $a^\perp = b^\perp$.

Proof of the inclusion $a^\perp \subseteq b^\perp$. Fix any direction $d \in \mathbb{R}^{d_i}$ satisfying $\langle a, d\rangle = 0$. We prove that $\langle b, d\rangle = 0$ as well. For $\varepsilon > 0$, define $d_\varepsilon^+ := d + \varepsilon a, d_\varepsilon^- := d - \varepsilon a$. Since $x_i \in \mathrm{int}(\mathcal{X}_i)$, for every fixed $\varepsilon > 0$ and all sufficiently small $h > 0$, $x_i + h d_\varepsilon^\pm \in \mathcal{X}_i$.

By differentiability of f_i (which follows from differentiability of J_i),

$$f_i(x_i + h d_\varepsilon^+) - f_i(x_i) = h\langle a, d_\varepsilon^+\rangle + o(h) = h\varepsilon\|a\|^2 + o(h) > 0$$

for all sufficiently small $h > 0$. By the ordinal-potential property,

$$\phi_i(x_i + h d_\varepsilon^+) - \phi_i(x_i) > 0$$

for the same sufficiently small h. Dividing by h and letting $h \downarrow 0$ gives $\langle b, d + \varepsilon a\rangle \geq 0$. Similarly,

$$f_i(x_i + h d_\varepsilon^-) - f_i(x_i) = h\langle a, d_\varepsilon^-\rangle + o(h) = -h\varepsilon\|a\|^2 + o(h) < 0$$

for sufficiently small $h > 0$. Ordinal equivalence therefore gives

$$\phi_i(x_i + h d_\varepsilon^-) - \phi_i(x_i) < 0,$$

and hence $\langle b, d - \varepsilon a\rangle \leq 0$. Letting $\varepsilon \downarrow 0$ in the preceding two inequalities yields $\langle b, d\rangle \geq 0$ and $\langle b, d\rangle \leq 0$, so $\langle b, d\rangle = 0$. This proves that $a^\perp \subseteq b^\perp$.

Proof of positive collinearity. We have proved that a and b are collinear: there exists $\mu_i(x) \neq 0$ such that $b = \mu_i(x)a$. It remains to show that $\mu_i(x) > 0$. Consider the direction a. Since $\langle a, a\rangle = \|a\|^2 > 0$, differentiability gives

$$f_i(x_i + ha) - f_i(x_i) > 0$$

for every sufficiently small $h > 0$. The ordinal-potential property therefore implies

$$\phi_i(x_i + ha) - \phi_i(x_i) > 0.$$

Dividing by h and letting $h \downarrow 0$ yields $\langle b, a\rangle \geq 0$. Since $b = \mu_i(x)a$, $\langle b, a\rangle = \mu_i(x)\|a\|^2$. Thus $\mu_i(x) \geq 0$. Because $b \neq 0$, we have $\mu_i(x) \neq 0$, and consequently $\mu_i(x) > 0$. Equivalently, there exists $\lambda_i(x) > 0$ such that

$$\nabla f_i(x_i) = \lambda_i(x)\nabla \phi_i(x_i).$$

Proof for the case $x_i \in \partial\mathcal{X}_i$. It remains to treat $x_i \in \partial\mathcal{X}_i$. We will use the result for the case $x_i \in \mathrm{int}(\mathcal{X}_i)$ proved above. Fix $c_i \in \mathrm{int}(\mathcal{X}_i)$ which always exists since $\mathcal{X}_i$ has a nonempty interior by assumption, and let $\varepsilon_n \downarrow 0$ with $\varepsilon_n \in (0, 1)$. Define $x_i^n := (1 - \varepsilon_n)x_i + \varepsilon_n c_i$. By convexity and since $c_i \in \mathrm{int}(\mathcal{X}_i)$, we have $x_i^n \in \mathrm{int}(\mathcal{X}_i)$ for every n, and $x_i^n \to x_i$.

Set $x^n := (x_i^n, x_{-i})$. Since $\nabla_i J_i(x) \neq 0, \nabla_i \Phi(x) \neq 0$, continuity implies that both gradients remain nonzero at x^n for all sufficiently large n. Applying the interior argument gives

$$\frac{\nabla_i J_i(x^n)}{\|\nabla_i J_i(x^n)\|} = \frac{\nabla_i \Phi(x^n)}{\|\nabla_i \Phi(x^n)\|}.$$

Passing to the limit and using continuity yields

$$\frac{\nabla_i J_i(x)}{\|\nabla_i J_i(x)\|} = \frac{\nabla_i \Phi(x)}{\|\nabla_i \Phi(x)\|}.$$

Since both gradients are nonzero,

$$\nabla_i J_i(x) = \frac{\|\nabla_i J_i(x)\|}{\|\nabla_i \Phi(x)\|} \nabla_i \Phi(x).$$

Thus (22) holds with $\lambda_i(x) := \frac{\|\nabla_i J_i(x)\|}{\|\nabla_i \Phi(x)\|} > 0$. This completes the proof. □

D.1.3 Proof of Theorem C.2

Proof. **Proof of (i).** Let $x(t)$ be an absolutely continuous solution of (BNPGD). By the differentiability of Φ and using the chain rule, we have for any $t \geq 0$,

$$\frac{d}{dt}\Phi(x(t)) = \langle \dot{x}(t), \nabla\Phi(x(t))\rangle = \sum_{i=1}^{N} \langle \dot{x}_i(t), \nabla_i\Phi(x(t))\rangle = \sum_{i=1}^{N} \|\nabla_i\Phi(x(t))\| \cdot \langle \dot{x}_i(t), D_i(x(t))\rangle\,, \tag{23}$$

where the last inequality follows from using $D_i(x(t)) = \frac{\nabla_i\Phi(x(t))}{\|\nabla_i\Phi(x(t))\|}$ as proved in Lemma C.1. When $\nabla_i\Phi(x) = 0$, Assumption 5.1 gives $\nabla_i J_i(x) = 0$, hence $D_i(x) = 0$ and that block contributes zero to the chain-rule identity.

Now by Moreau's decomposition theorem applied, we can write for every $i \in [N]$ and any $t \geq 0$, $-D_i(x(t)) = \dot{x}_i(t) + n_i(t)$ where $\dot{x}_i(t) \in T_{\mathcal{X}_i}(x_i(t))$ and $n_i(t) \in N_{\mathcal{X}_i}(x_i(t))$. In particular, we have $\langle \dot{x}_i(t), n_i(t)\rangle = 0$. (BNPGD) gives $\dot{x}_i = \Pi_{T_{\mathcal{X}_i}(x_i)}(-D_i(x))$ and using this in (23) gives:

$$\frac{d}{dt}\Phi(x(t)) = \sum_{i=1}^{N} \|\nabla_i\Phi(x(t))\| \cdot \langle \dot{x}_i(t), -\dot{x}_i(t) - n_i(t)\rangle = -\sum_{i=1}^{N} \|\nabla_i\Phi(x(t))\| \cdot \|\dot{x}_i(t)\|^2\,, \tag{24}$$

which concludes the proof of the first item using (BNPGD), i.e., $\dot{x}_i(t) = \Pi_{T_{\mathcal{X}_i}(x_i)}(-D_i(x))$.

Proof of (ii). If the trajectory $x(t)$ remains in $\mathrm{int}(\mathcal{X})$, it follows that $\dot{x}_i(t) = \Pi_{T_{\mathcal{X}_i}(x_i)}(-D_i(x)) = -D_i(x(t))$ as the projection is not active. Since $\|D_i(x(t))\| = 1$ when $\nabla_i\Phi(x(t)) \neq 0$, we obtain that for any t s.t. $\nabla_i\Phi(x(t)) \neq 0$,

$$\frac{d}{dt}\Phi(x(t)) = -\sum_{i=1}^{N} \|\nabla_i\Phi(x(t))\| = -\|\nabla\Phi(x(t))\|_{2,1}\,. \tag{25}$$

Proof of (iii). Let $T > 0$. Integrating the previous identity between 0 and T, and rearranging the resulting identity yields:

$$\int_0^T \|\nabla\Phi(x(t))\|_{2,1} dt = \Phi(x(0)) - \Phi(x(T)) \leq \Phi(x(0)) - \Phi_{\inf}\,. \tag{26}$$

As a consequence,

$$\min_{0\leq t\leq T} \|\nabla\Phi(x(t))\|_{2,1} \leq \frac{1}{T}\int_0^T \|\nabla\Phi(x(t))\|_{2,1} dt \leq \frac{\Phi(x(0)) - \Phi_{\inf}}{T}\,. \tag{27}$$

Proof of (iv). The rest points of (BNPGD) correspond to points $x^\star$ s.t. $\Pi_{T_{\mathcal{X}_i}(x_i^\star)}(-D_i(x^\star)) = 0$ for any $i \in [N]$. Hence $-D_i(x^\star) \in N_{\mathcal{X}_i}(x_i^\star)$ and by definition of the normal cone this means that

$$\langle -D_i(x^\star), x_i - x_i^\star\rangle \leq 0\,, \quad \forall x_i \in \mathcal{X}_i, \forall i \in [N],\,. \tag{28}$$

Now if $\nabla J_i(x^\star) \neq 0$, multiplying the above inequality by $-\|\nabla J_i(x^\star)\|$ yields:

$$\langle \nabla_i J_i(x^\star), x_i - x_i^\star\rangle \geq 0\,, \quad \forall x_i \in \mathcal{X}_i, \forall i \in [N],\,. \tag{29}$$

Invoking block pseudo-convexity (Assumption 3.1) yields:

$$J_i(x_i, x^*_{-i}) \geq J_i(x^\star_i, x^\star_{-i}), \quad \forall x_i \in \mathcal{X}_i, \tag{30}$$

which shows that $x^\star$ is a Nash equilibrium.

Now it follows from (26) by taking $T \to +\infty$ that:

$$\int_0^{+\infty} \|\nabla\Phi(x(t))\|_{2,1} dt < +\infty. \tag{31}$$

We now prove that if $x(t)$ remains in a compact set $K \subset \operatorname{int}(\mathcal{X})$ and $\nabla\Phi$ is Lipschitz on K, then $\|\nabla\Phi(x(t))\|_{2,1} \longrightarrow 0$. The proof of this statement follows from proving the following two claims:

(i) For all $i \in [N]$, $\int_0^\infty \|\nabla_i\Phi(x(t))\|\, dt < \infty$.

(ii) For any $i \in [N]$, $g_i : t \mapsto \|\nabla_i\Phi(x(t))\|$ is uniformly continuous.

We start by proving each one of the statements before concluding.

Along the trajectory, we have, for almost every $t \geq 0$,

$$\frac{d}{dt}\Phi(x(t)) = \sum_{i=1}^N \langle \nabla_i\Phi(x(t)), \dot{x}_i(t)\rangle = -\sum_{i=1}^N \|\nabla_i\Phi(x(t))\|. \tag{32}$$

Hence, for every $T > 0$,

$$\int_0^T \sum_{i=1}^N \|\nabla_i\Phi(x(t))\|\, dt = \Phi(x(0)) - \Phi(x(T)). \tag{33}$$

Since Φ is bounded from below, it follows that for every $i \in [N]$,

$$\int_0^\infty \|\nabla_i\Phi(x(t))\|\, dt < \infty, \tag{34}$$

which proves the first claim above.

As for the second claim, observe first that $\|\dot{x}_i(t)\| \leq 1$ by definition of the normalized flow. It follows then that $\|\dot{x}(t)\| \leq \sqrt{N}$ for the Euclidean norm and as a consequence, for all $s, t \geq 0$, we have $\|x(t) - x(s)\| \leq \sqrt{N}\,|t - s|$. Since $\nabla\Phi$ is L-Lipschitz on the compact set K in which the trajectory lies, we obtain for all $s, t \geq 0$:

$$\begin{aligned}|g_i(t) - g_i(s)| \leq \|\nabla_i\Phi(x(t)) - \nabla_i\Phi(x(s))\| \leq \|\nabla\Phi(x(t)) - \nabla\Phi(x(s))\| &\leq L\|x(t) - x(s)\| \\ &\leq L\sqrt{N}\,|t - s|,\end{aligned}$$

which proves that g_i is Lipschitz, and in particular uniformly continuous. This concludes the proof of the second claim above.

We now show that $g_i(t) \to 0$ using a standard argument, we provide a detailed proof for completeness. Suppose otherwise for contradiction. Then there exist $\varepsilon > 0$ and a sequence $t_n \to \infty$ such that $g_i(t_n) \geq \varepsilon$ for all n. By uniform continuity, there exists $\delta > 0$, independent of n, such that

$$|t - t_n| \leq \delta \quad \Longrightarrow \quad g_i(t) \geq \frac{\varepsilon}{2}.$$

Passing to a subsequence if necessary, we may assume that the intervals $[t_n - \delta, t_n + \delta]$ are pairwise disjoint. It follows that

$$\int_0^\infty g_i(t)\, dt \geq \sum_n \int_{t_n-\delta}^{t_n+\delta} g_i(t)\, dt \geq \sum_n \varepsilon\delta = \infty,$$

contradicting the integrability result proved above (34). Therefore $\lim_{t\to\infty} \|\nabla_i\Phi(x(t))\| = 0$.

We conclude by proving that $\operatorname{dist}\big(x(t), \mathrm{NE}\big) \longrightarrow 0$. Suppose, by contradiction, that this limit statement does not hold. Then there exist $\varepsilon > 0$ and a sequence $t_k \to \infty$ such that

$$\operatorname{dist}(x(t_k), \mathrm{NE}) \geq \varepsilon, \qquad \forall k \geq 0. \tag{35}$$

Since K is compact, after passing to a subsequence we may assume that $x(t_k) \to \bar{x}$ for some $\bar{x} \in K$. By continuity of $\nabla_i \Phi$ and using the limit statement we already established, we obtain that for every player $i \in [N]$,

$$\nabla_i \Phi(\bar{x}) = \lim_{k \to \infty} \nabla_i \Phi(x(t_k)) = 0.$$

For any $i \in [N]$, since the action constraints are inactive at $\bar{x}$, the point $\bar{x}_i$ is an unconstrained stationary point of the function $z_i \mapsto \Phi(z_i, \bar{x}_{-i})$. By pseudoconvexity of this function, every stationary point is a global minimizer. Hence

$$\Phi(\bar{x}_i, \bar{x}_{-i}) \leq \Phi(y_i, \bar{x}_{-i}), \quad y_i \in \mathcal{X}_i. \tag{36}$$

Since Φ is an ordinal potential, unilateral improvements in a player's objective are accompanied by strict decreases of the potential. Therefore (36) rules out any profitable unilateral deviation by player i. Since this holds for every $i \in [N]$, we conclude that $\bar{x}$ is a Nash equilibrium. But then $\mathrm{dist}(x(t_k), \mathrm{NE}) \leq \|x(t_k) - \bar{x}\| \longrightarrow 0$ because $\bar{x}$ is a Nash equilibrium and $x(t_k) \to \bar{x}$. This contradicts $\mathrm{dist}(x(t_k), \mathrm{NE}) \geq \varepsilon$. Therefore $\mathrm{dist}(x(t), \mathrm{NE}) \longrightarrow 0$, as claimed. □

D.2 Proofs for Section 4.2: Discrete-time ordinal gradient play

D.2.1 Proof of Theorem 4.1

Proof. Nonexpansiveness of projection gives, for every $y_i \in \mathcal{X}_i$,

$$\|x_{i,t+1} - y_i\|^2 \leq \|x_{i,t} - y_i\|^2 - 2\eta \langle D_i(x_t), x_{i,t} - y_i \rangle + \eta^2 \|D_i(x_t)\|^2.$$

Since $\|D_i(x_t)\| \leq 1$, summing over t and telescoping gives the result. □

Remark D.1. On unbounded action sets (e.g., $\mathcal{X}_i = \mathbb{R}^{d_i}$), the uniform comparator supremum in $\mathrm{Reg}_{i,T}^{D}$ is generally infinite. Indeed, it is infinite whenever $\sum_t D_i(x_t) \neq 0$. This differs from standard external regret, where nonlinear utilities may make the cumulative comparator objective coercive and hence yield a finite infimum. Such quantitative growth away from an optimum is not encoded by the normalized preference field. On unbounded domains, we may use a comparator-dependent regret (i.e. for a fixed comparator y_i rather than supremum in $\mathrm{Reg}_{i,T}^{D}$), or restrict the comparator class to a bounded neighborhood.

E Proofs for Section 5: Learning from Pairwise Comparisons

E.1 Pseudocode for Comparison-Normalized Gradient Play

Algorithm 1 Comparison-Normalized Gradient Play (CNGP)

Initialize $z_{i,1} \in \mathcal{X}_i$ for every player $i \in [N]$.
for $t = 1, \ldots, T$ **do**
  for each player i, in parallel **do**
    Set $x_{i,t} = \mathsf{Q}_{i,\delta_t}(z_{i,t})$.
    Sample $U_{i,t} \sim \mathrm{Unif}(\mathbb{S}^{d_i - 1})$.
    Compare $(x_{i,t} + \delta_t U_{i,t}, x_{-i,t})$ and $(x_{i,t} - \delta_t U_{i,t}, x_{-i,t})$, and observe $Y_{i,t}$.
    Set $\widehat{D}_{i,t} = \frac{Y_{i,t}}{2\nu_i \kappa_{d_i}} U_{i,t}$.
    Update $z_{i,t+1} = \Pi_{\mathcal{X}_i}\left(z_{i,t} - \eta_{i,t} \widehat{D}_{i,t}\right)$.
  end for
end for

E.2 Proofs for Section 5.1: Directional regret bound from comparisons

E.2.1 Finite-resolution regret does not control infinitesimal ordinal regret

The following result shows that even with exact comparisons and vanishing comparison radii, sublinear finite-resolution ordinal regret does not in general imply sublinear infinitesimal ordinal directional regret.

Proposition E.1 (Finite-resolution regret does not control infinitesimal ordinal regret)**.** *There exist a compact convex action set $\mathcal{X} \subset \mathbb{R}$, a C^∞ strongly convex cost function $J : \mathcal{X} \to \mathbb{R}$, a sequence of comparison radii δ_t converging to zero, and a sequence of actions $(x_t)_{t\geq 1}$ such that all symmetric comparison queries $x_t \pm \delta_t$ are feasible and, for every $K \geq 1$,*

$$\mathrm{Reg}^{\delta}_{2K} \leq \frac{\pi^2}{512}, \qquad \textit{while} \qquad \mathrm{Reg}^{D}_{2K} \geq 2K.$$

Proof sketch. We provide an overview of the proof to highlight the intuition. The construction exploits the fact that $D^\delta(x) \to D(x)$ only pointwise away from stationary points. Consider $J(x) = x^2 - \frac{1}{4}x^3$ on $\mathcal{X} = [-1, 1]$. For every small $x > 0$, $J'(x) > 0$, so the infinitesimal direction is $D(x) = 1$. On the other hand, the symmetric finite difference is

$$J(x+\delta) - J(x-\delta) = 2\delta\left(2x - \frac{3}{4}x^2 - \frac{1}{4}\delta^2\right).$$

Thus the first-order contribution is of size δx, while the finite-resolution correction is of size δ^3. Choosing $x = \Theta(\delta^2)$ makes these terms comparable, so the sign of the finite comparison can differ from the infinitesimal gradient direction even as $\delta \downarrow 0$.

For each resolution ε_k, we choose two positive actions on opposite sides of this sign-change threshold:

$$x_{2k-1} = \frac{\varepsilon_k^2}{16}, \qquad x_{2k} = \frac{\varepsilon_k^2}{4}, \qquad \delta_{2k-1} = \delta_{2k} = \varepsilon_k.$$

Then $D(x_{2k-1}) = D(x_{2k}) = 1$, whereas the corresponding finite-resolution directions satisfy

$$D^{\delta_{2k-1}}(x_{2k-1}) = -1, \qquad D^{\delta_{2k}}(x_{2k}) = 1.$$

Hence the finite-resolution directions cancel pairwise, so the comparator term in Reg^δ cancels as well; the remaining contribution is of order $\sum_k \varepsilon_k^2$. Taking $\varepsilon_k \asymp 1/k$ makes this sum finite. By contrast, the infinitesimal directions are equal to $+1$ at every round, so the fixed comparator $y = -1$ yields a unit-order advantage per round and therefore linear Reg^D. The example shows that vanishing comparison radii alone do not ensure that the finite-resolution field tracks the infinitesimal one along trajectories approaching the stationary set.

Proof. Consider the one-dimensional action set $\mathcal{X} = [-1, 1]$ and the cost function $J : \mathcal{X} \to \mathbb{R}$ defined for any $x \in \mathcal{X}$ by:

$$J(x) = x^2 - \frac{1}{4}x^3. \tag{37}$$

Since the first and second derivatives of J are respectively given by $J'(x) = 2x - \frac{3}{4}x^2$, $J''(x) = 2 - \frac{3}{2}x$, we have $J''(x) \geq 1/2$ for every $x \in [-1, 1]$. Hence J is $1/2$-strongly convex on $\mathcal{X}$. Moreover, for every $x \in (0, 1]$, $J'(x) = x\left(2 - \frac{3}{4}x\right) > 0$, so the infinitesimal preference direction satisfies $D(x) = 1$ for any $x \in (0, 1]$.

In dimension one, $\mathbb{S}^0 = \{-1, +1\}$ and $\kappa_1 = 1$. Therefore, for every feasible $x \pm \delta$,

$$D^\delta(x) = \mathrm{sgn}(J(x+\delta) - J(x-\delta)).$$

Indeed, if $s = \mathrm{sgn}(J(x+\delta) - J(x-\delta))$, then the two possible values of $S^\delta(x, U)U$, corresponding to $U = 1$ and $U = -1$, are both equal to s.

A direct calculation yields

$$J(x+\delta) - J(x-\delta) = 4x\delta - \frac{1}{4}\left(6x^2\delta + 2\delta^3\right) = 2\delta\left(2x - \frac{3}{4}x^2 - \frac{1}{4}\delta^2\right). \tag{38}$$

For every $k \geq 1$, let $\varepsilon_k := \frac{1}{4k}$, $\delta_{2k-1} = \delta_{2k} := \varepsilon_k$, and define

$$x_{2k-1} := \frac{\varepsilon_k^2}{16}, \qquad x_{2k} := \frac{\varepsilon_k^2}{4}.$$

Clearly, $\delta_t \to 0$, and all query points $x_t \pm \delta_t$ belong to $\mathcal{X}$. Since $x_t > 0$ for every t,

$$D(x_t) = 1.$$

For the first action in each pair, using (38),

$$2x_{2k-1} - \frac{3}{4}x_{2k-1}^2 - \frac{1}{4}\varepsilon_k^2 = \frac{\varepsilon_k^2}{8} - \frac{3\varepsilon_k^4}{1024} - \frac{\varepsilon_k^2}{4} = -\frac{\varepsilon_k^2}{8} - \frac{3\varepsilon_k^4}{1024} < 0, \tag{39}$$

and hence

$$D^{\delta_{2k-1}}(x_{2k-1}) = -1.$$

For the second action,

$$2x_{2k} - \frac{3}{4}x_{2k}^2 - \frac{1}{4}\varepsilon_k^2 = \frac{\varepsilon_k^2}{2} - \frac{3\varepsilon_k^4}{64} - \frac{\varepsilon_k^2}{4} = \frac{\varepsilon_k^2}{4} - \frac{3\varepsilon_k^4}{64} > 0, \tag{40}$$

where the last inequality follows from $\varepsilon_k \le 1/4$. Thus

$$D^{\delta_{2k}}(x_{2k}) = 1.$$

The finite-resolution directions therefore cancel pairwise:

$$D^{\delta_{2k-1}}(x_{2k-1}) + D^{\delta_{2k}}(x_{2k}) = 0.$$

For $T = 2K$, the finite-resolution regret is consequently

$$\begin{aligned}
\mathrm{Reg}_{2K}^{\delta} &= \sup_{y\in[-1,1]} \sum_{t=1}^{2K} D^{\delta_t}(x_t)(x_t - y) \\
&= \sum_{k=1}^{K} (-x_{2k-1} + x_{2k}) \\
&= \frac{3}{16}\sum_{k=1}^{K} \varepsilon_k^2 \\
&= \frac{3}{256}\sum_{k=1}^{K} \frac{1}{k^2} \\
&\le \frac{3}{256}\frac{\pi^2}{6} = \frac{\pi^2}{512}.
\end{aligned}$$

On the other hand, since $D(x_t) = 1$ for every t,

$$\mathrm{Reg}_{2K}^{D} = \sup_{y\in[-1,1]} \sum_{t=1}^{2K} (x_t - y) = \sum_{t=1}^{2K} x_t + 2K,$$

where the supremum is attained at $y = -1$. Therefore

$$\mathrm{Reg}_{2K}^{D} \ge 2K.$$

This proves the claim. □

E.2.2 Proof of Proposition 5.1

Proof. **Proof of (i).** Observe that strictly increasing transformations preserve the sign, i.e., for any strictly increasing function $\phi_i : \mathbb{R} \to \mathbb{R}$, any $x \in \mathcal{X}, U_i$ s.t. $x_i \pm \delta U_i \in \mathcal{X}_i$,

$$\mathrm{sgn}(J_i(x_i + \delta U_i, x_{-i}) - J_i(x_i - \delta U_i, x_{-i})) = \mathrm{sgn}(\phi_i(J_i(x_i + \delta U_i, x_{-i})) - \phi_i(J_i(x_i - \delta U_i, x_{-i}))).$$

This implies that $S_i^{\delta}(x, U_i)$ is invariant under transformations of the cost functions J_i by strictly increasing functions. As a consequence, the *δ-comparison field* $D_i^{\delta}(x) := \frac{1}{\kappa_{d_i}} \mathbb{E}_{U_i}\left[S_i^{\delta}(x, U_i) U_i\right]$ defined in (10) is also invariant.

Proof of (ii). The desired bound follows from:

$$\|D_i^\delta(x)\| = \sup_{\|v\|=1} \frac{1}{\kappa_{d_i}} \mathbb{E}\big[S_i^\delta(x, U_i)\langle U_i, v\rangle\big] \le \frac{1}{\kappa_{d_i}} \sup_{\|v\|=1} \mathbb{E}[|\langle U_i, v\rangle|] = 1\,, \tag{41}$$

where the first equality uses the fact that $\|z\| = \sup_{\|v\|=1}\langle z, v\rangle$ (which follows from the Cauchy-Schwarz inequality) and the definition of $D_i^\delta(x)$ in (10), and the last inequality uses Lemma 3.1 in Fu et al. (2023) (which relies on rotational invariance).

Proof of (iii). Suppose $\nabla_i J_i(x) \neq 0$. Then differentiability of $J_i(\cdot, x_{-i})$ gives, for every u,

$$\frac{J_i(x_i + \delta u, x_{-i}) - J_i(x_i - \delta u, x_{-i})}{2\delta} \xrightarrow[\delta\to 0]{} \langle \nabla_i J_i(x), u\rangle.$$

For all u outside the measure-zero great sphere $\{u : \langle \nabla_i J_i(x), u\rangle = 0\}$,

$$S_i^\delta(x, u) \xrightarrow[\delta\to 0]{} \operatorname{sgn}(\langle \nabla_i J_i(x), u\rangle).$$

Dominated convergence therefore yields

$$D_i^\delta(x) = \frac{1}{\kappa_{d_i}} \mathbb{E}_{U_i}\left[S_i^\delta(x, U_i) U_i\right] \xrightarrow[\delta\to 0]{} \frac{1}{\kappa_{d_i}} \mathbb{E}\left[\operatorname{sgn}(\langle \nabla_i J_i(x), U_i\rangle) U_i\right].$$

By rotational symmetry, the expectation on the right must be collinear with $\nabla_i J_i(x)$. Taking its inner product with $\nabla_i J_i(x)/\|\nabla_i J_i(x)\|$ gives $\mathbb{E}\left|\left\langle U_i, \frac{\nabla_i J_i(x)}{\|\nabla_i J_i(x)\|}\right\rangle\right| = \kappa_{d_i}$, and hence

$$\frac{1}{\kappa_{d_i}} \mathbb{E}\left[\operatorname{sgn}(\langle \nabla_i J_i(x), U_i\rangle) U_i\right] = \frac{\nabla_i J_i(x)}{\|\nabla_i J_i(x)\|} = D_i(x)\,.$$

□

E.2.3 Proof of Lemma 5.1

Proof. Conditional on $U_{i,t}$,

$$\begin{aligned}
\mathbb{E}[Y_{i,t} \mid U_{i,t}, \mathcal{F}_t] &= \mathbb{E}[\xi_{i,t} S_i^{\delta_t}(x_t, U_{i,t}) \mid U_{i,t}, \mathcal{F}_t] \\
&= \left(\frac{1}{2} + \nu_i\right) S_i^{\delta_t}(x_t, U_{i,t}) - \left(1 - \frac{1}{2} - \nu_i\right) S_i^{\delta_t}(x_t, U_{i,t}) \\
&= 2\nu_i S_i^{\delta_t}(x_t, U_{i,t}).
\end{aligned}$$

Hence

$$\mathbb{E}[\widehat{D}_{i,t} \mid \mathcal{F}_t] = \mathbb{E}\left[\frac{Y_{i,t}}{2\nu_i \kappa_{d_i}} U_{i,t} \mid \mathcal{F}_t\right] = \frac{1}{\kappa_{d_i}} \mathbb{E}_{U_{i,t}}[S_i^{\delta_t}(x_t, U_{i,t}) U_{i,t}] = D_i^{\delta_t}(x_t).$$

Moreover, $\|U_{i,t}\| = 1$ and $|Y_{i,t}| \le 1$, so

$$\|\widehat{D}_{i,t}\|^2 = \left\|\frac{Y_{i,t}}{2\nu_i \kappa_{d_i}} U_{i,t}\right\|^2 \le \frac{1}{4\nu_i^2 \kappa_{d_i}^2} = G_i^2\,.$$

Taking conditional expectation concludes the proof. □

Remark E.1. For exact comparisons, $G_i = \kappa_{d_i}^{-1} = \sqrt{\frac{\pi d_i}{2}}\,(1 + o(1))$.

E.2.4 Proof of Theorem 5.1

Proof. Let $y_i \in \mathcal{X}_i$. Projection nonexpansiveness gives

$$\|z_{i,t+1} - y_i\|^2 \le \|z_{i,t} - y_i\|^2 - 2\eta_i \langle \widehat{D}_{i,t}, z_{i,t} - y_i\rangle + \eta_i^2 \|\widehat{D}_{i,t}\|^2. \tag{42}$$

Rearranging and summing,

$$\sum_{t=1}^T \langle \widehat{D}_{i,t}, z_{i,t} - y_i\rangle \le \frac{\|z_{i,1} - y_i\|^2}{2\eta_i} + \frac{\eta_i}{2} \sum_{t=1}^T \|\widehat{D}_{i,t}\|^2 \le \frac{R_i^2}{2\eta_i} + \frac{\eta_i}{2} \sum_{t=1}^T \|\widehat{D}_{i,t}\|^2. \tag{43}$$

Now, introducing the sequence $(x_{i,t})$, we obtain:

$$\sum_{t=1}^{T}\langle \widehat{D}_{i,t}, x_{i,t}-y_i\rangle = \sum_{t=1}^{T}\langle \widehat{D}_{i,t}, z_{i,t}-y_i\rangle + \sum_{t=1}^{T}\langle \widehat{D}_{i,t}, x_{i,t}-z_{i,t}\rangle\,. \tag{44}$$

Taking expectations, using $\mathbb{E}[\widehat{D}_{i,t} \mid \mathcal{F}_t] = D_i^{\delta_t}(x_t)$, $\mathbb{E}[\|\widehat{D}_{i,t}\|^2 \mid \mathcal{F}_t] \le G_i^2$ (Lemma 5.1), (43), and noting that $z_{i,t}-y_i$ and $x_{i,t}-z_{i,t}$ are $\mathcal{F}_t$-measurable yields:

$$\mathbb{E}\left[\sum_{t=1}^{T}\langle D_i^{\delta_t}(x_t), x_{i,t}-y_i\rangle\right] \le \frac{R_i^2}{2\eta_i} + \frac{\eta_i G_i^2 T}{2} + \mathbb{E}\left[\sum_{t=1}^{T}\langle D_i^{\delta_t}(x_t), x_{i,t}-z_{i,t}\rangle\right].$$

Since this is true for every fixed $y_i \in \mathcal{X}_i$, taking the supremum gives:

$$\overline{\mathrm{Reg}}_{i,T}^{\delta} \le \frac{R_i^2}{2\eta_i} + \frac{\eta_i G_i^2 T}{2} + \mathbb{E}\left[\sum_{t=1}^{T}\langle D_i^{\delta_t}(x_t), x_{i,t}-z_{i,t}\rangle\right]. \tag{45}$$

To obtain the first desired inequality, it remains to use Cauchy-Schwarz inequality for the last inner product, with the fact that $\|D_i^{\delta}(x_t)\| \le 1$ (Proposition 5.1) and $\|x_{i,t}-z_{i,t}\| \le \frac{R_i}{\rho_i}\delta_t$. We obtain:

$$\overline{\mathrm{Reg}}_{i,T}^{\delta} \le \frac{R_i^2}{2\eta_i} + \frac{\eta_i G_i^2 T}{2} + \frac{R_i}{\rho_i}\sum_{t=1}^{T}\delta_t\,. \tag{46}$$

Setting $\eta_i = \frac{R_i}{G_i\sqrt{T}}$ and $\delta_t = \frac{\delta}{\sqrt{T}}$ (with $\delta \in (0,\rho)$) gives:

$$\overline{\mathrm{Reg}}_{i,T}^{\delta} \le R_i G_i\sqrt{T} + R_i\frac{\rho}{\rho_i}\sqrt{T} \le R_i(G_i+1)\sqrt{T} = O\left(\frac{R_i\sqrt{d_i T}}{\nu_i}\right). \tag{47}$$

To control the realized regret, define for all $i \in [N]$, the martingale difference sequence $(M_{i,t})$:

$$M_{i,t} := D_i^{\delta_t}(x_t) - \widehat{D}_{i,t}\,, \quad \forall t \ge 0\,.$$

For an arbitrary reference point $y_i^0 \in \mathcal{X}_i$, every $y_i \in \mathcal{X}_i$ and any $T > 0$,

$$\sum_{t=1}^{T}\langle M_{i,t}, x_{i,t}-y_i\rangle = \sum_{t=1}^{T}\langle M_{i,t}, x_{i,t}-y_i^0\rangle + \left\langle \sum_{t=1}^{T} M_{i,t}, y_i^0 - y_i\right\rangle.$$

Hence, for any $i \in [N]$ and any $T > 0$,

$$\begin{aligned}
\mathrm{Reg}_{i,T}^{\delta} &= \sup_{y_i\in\mathcal{X}_i}\sum_{t=1}^{T}\langle D_i^{\delta_t}(x_t), x_{i,t}-y_i\rangle \\
&= \sup_{y_i\in\mathcal{X}_i}\sum_{t=1}^{T}\langle \widehat{D}_{i,t}, x_{i,t}-y_i\rangle + \sum_{t=1}^{T}\langle M_{i,t}, x_{i,t}-y_i\rangle \\
&= \sup_{y_i\in\mathcal{X}_i}\sum_{t=1}^{T}\langle \widehat{D}_{i,t}, x_{i,t}-y_i\rangle + \sum_{t=1}^{T}\langle M_{i,t}, x_{i,t}-y_i^0\rangle + \sum_{t=1}^{T}\langle M_{i,t}, y_i^0-y_i\rangle \\
&\le \sup_{y_i\in\mathcal{X}_i}\sum_{t=1}^{T}\langle \widehat{D}_{i,t}, x_{i,t}-y_i\rangle + A_T + R_i\cdot\left\|\sum_{t=1}^{T} M_{i,t}\right\|,
\end{aligned} \tag{48}$$

where $A_T := \sum_{t=1}^{T}\langle M_{i,t}, x_{i,t}-y_i^0\rangle$.

By the pathwise projection inequality proved above in (43),

$$\sup_{y_i\in\mathcal{X}_i}\sum_{t=1}^{T}\langle \widehat{D}_{i,t}, x_{i,t}-y_i\rangle \le \frac{R_i^2}{2\eta_i} + \frac{\eta_i}{2}\sum_{t=1}^{T}\|\widehat{D}_{i,t}\|^2 + \sum_{t=1}^{T}\langle \widehat{D}_{i,t}, x_{i,t}-z_{i,t}\rangle$$

Moreover, by construction, $\|x_{i,t} - z_{i,t}\| \le \frac{R_i}{\rho_i}\delta_t$, and, from the definition of the one-bit estimator, $\|\widehat{D}_{i,t}\| = \frac{|Y_{i,t}|}{2\nu_i \kappa_{d_i}} \le G_i$ almost surely. Hence, using Cauchy-Schwarz inequality, we have:

$$\sum_{t=1}^{T} \langle \widehat{D}_{i,t}, x_{i,t} - z_{i,t} \rangle \le \sum_{t=1}^{T} \|\widehat{D}_{i,t}\| \cdot \|x_{i,t} - z_{i,t}\| \le \frac{R_i G_i}{\rho_i} \sum_{t=1}^{T} \delta_t. \tag{49}$$

Moreover, since $x_{i,t}$ is $\mathcal{F}_t$-measurable and $\mathbb{E}[M_{i,t} \mid \mathcal{F}_t] = 0$, $\mathbb{E}[A_T] = 0$.

We also have, using $M_{i,t} = D_i^{\delta_t}(x_t) - \widehat{D}_{i,t}$, $D_i^{\delta_t}(x_t) = \mathbb{E}[\widehat{D}_{i,t} \mid \mathcal{F}_t]$, we obtain

$$\begin{aligned}
\mathbb{E}[\|M_{i,t}\|^2 \mid \mathcal{F}_t] &= \mathbb{E}\left[\left\| \widehat{D}_{i,t} - \mathbb{E}[\widehat{D}_{i,t} \mid \mathcal{F}_t] \right\|^2 \,\middle|\, \mathcal{F}_t \right] \\
&= \mathbb{E}[\|\widehat{D}_{i,t}\|^2 \mid \mathcal{F}_t] - \left\| \mathbb{E}[\widehat{D}_{i,t} \mid \mathcal{F}_t] \right\|^2 \\
&\le \mathbb{E}[\|\widehat{D}_{i,t}\|^2 \mid \mathcal{F}_t] \le G_i^2 \,,
\end{aligned}$$

where the last inequality follows from using Lemma 5.1.

By martingale orthogonality,

$$\mathbb{E}\left\| \sum_{t=1}^{T} M_{i,t} \right\|^2 = \sum_{t=1}^{T} \mathbb{E}\|M_{i,t}\|^2 \le G_i^2 T.$$

Therefore, by Cauchy-Schwarz inequality,

$$\mathbb{E}\left\| \sum_{t=1}^{T} M_{i,t} \right\| \le G_i \sqrt{T}.$$

Taking expectations in (48) therefore yields

$$\mathbb{E}[\mathrm{Reg}_{i,T}^{\delta}] \le \frac{R_i^2}{2\eta_i} + \frac{\eta_i G_i^2 T}{2} + R_i G_i \sqrt{T} + \frac{R_i G_i}{\rho_i} \sum_{t=1}^{T} \delta_t.$$

With $\eta_i = \frac{R_i}{G_i \sqrt{T}}$, the first two terms sum to $R_i G_i \sqrt{T}$, and with $\delta_t = \frac{\delta}{\sqrt{T}}$ the last term is bounded by $R_i G_i \sqrt{T}$ as well. Hence,

$$\mathbb{E}[\mathrm{Reg}_{i,T}^{\delta}] \le 3 R_i G_i \sqrt{T} = O\left(\frac{R_i \sqrt{d_i T}}{\nu_i} \right).$$

Finally, we bound the positive part. Since the right-hand side of (48), apart from A_T, consists of nonnegative terms,

$$[\mathrm{Reg}_{i,T}^{\delta}]_+ \le \frac{R_i^2}{2\eta_i} + \frac{\eta_i}{2} \sum_{t=1}^{T} \|\widehat{D}_{i,t}\|^2 + \frac{R_i G_i}{\rho_i} \sum_{t=1}^{T} \delta_t + |A_T| + R_i \left\| \sum_{t=1}^{T} M_{i,t} \right\|.$$

It remains to control A_T. Note first that its increments form a scalar martingale difference sequence. To see this, define $Z_t := \langle M_{i,t}, x_{i,t} - y_i^0 \rangle$, so that $A_T = \sum_{t=1}^{T} Z_t$. Since $x_{i,t}$ is $\mathcal{F}_t$-measurable and $\mathbb{E}[M_{i,t} \mid \mathcal{F}_t] = 0$, we have $\mathbb{E}[Z_t \mid \mathcal{F}_t] = \langle \mathbb{E}[M_{i,t} \mid \mathcal{F}_t], x_{i,t} - y_i^0 \rangle = 0$. Thus (Z_t) is a scalar martingale difference sequence. Expanding the square gives

$$A_T^2 = \sum_{t=1}^{T} Z_t^2 + 2 \sum_{1 \le s < t \le T} Z_s Z_t.$$

For $s < t$, Z_s is $\mathcal{F}_t$-measurable, so by the tower property, $\mathbb{E}[Z_s Z_t] = \mathbb{E}[\mathbb{E}[Z_s Z_t \mid \mathcal{F}_t]] = \mathbb{E}[Z_s \, \mathbb{E}[Z_t \mid \mathcal{F}_t]] = 0$. Therefore the cross terms vanish, and

$$\mathbb{E}[A_T^2] = \sum_{t=1}^{T} \mathbb{E}[Z_t^2].$$

By Cauchy–Schwarz inequality,

$$Z_t^2 = \langle M_{i,t}, x_{i,t} - y_i^0 \rangle^2 \le \|M_{i,t}\|^2 \|x_{i,t} - y_i^0\|^2.$$

Since $x_{i,t}, y_i^0 \in \mathcal{X}_i$, $\|x_{i,t} - y_i^0\| \le R_i$, and hence

$$\mathbb{E}[A_T^2] \le R_i^2 \sum_{t=1}^{T} \mathbb{E}\|M_{i,t}\|^2 \le R_i^2 G_i^2 T.$$

Therefore, by Cauchy–Schwarz inequality,

$$\mathbb{E}[|A_T|] \le \sqrt{\mathbb{E}[A_T^2]} \le R_i G_i \sqrt{T}.$$

Thus, it follows that:

$$\mathbb{E}\big[[\mathrm{Reg}_{i,T}^{\delta}]_+\big] \le \frac{R_i^2}{2\eta_i} + \frac{\eta_i G_i^2 T}{2} + 2R_i G_i \sqrt{T} + \frac{R_i G_i}{\rho_i} \sum_{t=1}^{T} \delta_t.$$

Choosing again $\eta_i = \frac{R_i}{G_i\sqrt{T}}$ and $\delta_t = \frac{\delta}{\sqrt{T}}$ with $\delta \in (0, \rho)$ gives

$$\mathbb{E}\big[[\mathrm{Reg}_{i,T}^{\delta}]_+\big] \le 4R_i G_i \sqrt{T} = O\left(\frac{R_i \sqrt{d_i T}}{\nu_i}\right).$$

This proves the claimed bound. □

E.2.5 Trading comparisons for dimension dependence using minibatch

The $\sqrt{d_i}$ factor in regret comes from using only one random directional comparison per iteration. It can be reduced by batching comparisons.

Suppose player i draws m_i independent directions and averages the corresponding estimators:

$$\widehat{D}_{i,t}^{(m_i)} = \frac{1}{m_i} \sum_{j=1}^{m_i} \widehat{D}_{i,t}^{(j)}.$$

Since $\|D_i^{\delta}(x)\| \le 1$,

$$\mathbb{E}\left[\|\widehat{D}_{i,t}^{(m_i)}\|^2 \mid \mathcal{F}_t\right] = \|D_i^{\delta}(x_t)\|^2 + \frac{1}{m_i}\mathbb{E}\left[\|\widehat{D}_{i,t} - D_i^{\delta}(x_t)\|^2 \mid \mathcal{F}_t\right] \le 1 + \frac{G_i^2}{m_i}.$$

The proof of Theorem 5.1 therefore gives

$$\mathrm{Reg}_{i,T}^{\delta} = O\left(R_i \sqrt{T\left(1 + \frac{d_i}{m_i \nu_i^2}\right)}\right)$$

in expectation, up to universal constants. Thus $m_i = 1$ gives $\mathrm{Reg}_{i,T}^{\delta} = O\left(\frac{R_i\sqrt{d_i T}}{\nu_i}\right)$ whereas $m_i = \Theta(d_i/\nu_i^2)$ gives $\mathrm{Reg}_{i,T}^{\delta} = O(R_i\sqrt{T})$ in terms of iterations, at the cost of $\Theta(d_i/\nu_i^2)$ comparisons per iteration.

Remark E.2. To compare the finite-resolution guarantee with the infinitesimal benchmark, let $b_{i,t} := \|D_i^{\delta_t}(x_t) - D_i(x_t)\|$. Then $\mathrm{Reg}_{i,T}^{D} \le \mathrm{Reg}_{i,T}^{\delta} + R_i \sum_{t=1}^{T} b_{i,t}$. Uniform control of $b_{i,t}$ is impossible in general near vanishing gradients; the normalized infinitesimal field is nonsmooth in general and may change sharply even when the underlying utilities are smooth. This is why D^{δ} is the natural oracle-matched benchmark at finite comparison resolution.

E.3 Proofs for Section 5.3

E.3.1 Proof of Theorem 5.2: Almost sure convergence to Nash in ordinal potential games

The proof relies on the projected differential inclusion stochastic approximation framework of Majewski et al. (2018).

We use the following C^1 specialization of Majewski et al. (2018, Theorem 4.2). The limiting dynamics remain a projected differential inclusion; the specialization only replaces the general set-valued Lie derivative by the ordinary directional derivative of the smooth Lyapunov function used here.

Theorem E.1 (C^1 specialization of Theorem 4.2 in Majewski et al. (2018))**.** *Let $\mathcal{K} \subset \mathbb{R}^d$ be nonempty, compact, and convex, let $\mathcal{U} \supseteq \mathcal{K}$ be open, and let $F : \mathcal{U} \rightrightarrows \mathbb{R}^d$ be a set-valued mapping. Consider a recursion of the form*

$$z_{t+1} = \Pi_{\mathcal{K}}\left[z_t + \eta_t\big(f_t + e_{t+1} + r_t\big)\right], \qquad f_t \in F(y_t), \tag{50}$$

where the set-valued mapping F, the stepsize sequence (η_t), the noise sequence (e_t) and the bias sequence (r_t) satisfy the following assumptions:

- *(A1) F is upper hemicontinuous, locally bounded, and has nonempty compact convex values.*
- *(A2) $\eta_t > 0$, $\eta_t \to 0$, and $\sum_{t=1}^{\infty} \eta_t = \infty$.*
- *(A3) $r_t \to 0$ and $\sum_{t=1}^{\infty} \eta_t e_{t+1}$ converges.*
- *(A4) $y_t \in \mathcal{U}$ and $\|y_t - z_t\| \to 0$.*
- *(A5) $\sup_t \|e_{t+1}\| < \infty$.*

Suppose that $V \in C^1(\mathcal{U})$ and that there exists an upper semicontinuous function $U : \mathcal{U} \to \mathbb{R}$ such that

$$\big\langle \nabla V(x), \Pi_{T_{\mathcal{K}}(x)} u \big\rangle \le U(x) \le 0, \qquad \forall x \in \mathcal{K}, \quad \forall u \in F(x). \tag{51}$$

Let $\mathcal{S} := \{x \in \mathcal{K} : U(x) = 0\}$. Then the image under V of the limit set of (z_t) is a compact interval contained in $V(\mathcal{S})$. If $\operatorname{int}_{\mathbb{R}} V(\mathcal{S}) = \varnothing$, then

$$\operatorname{dist}(z_t, \mathcal{S}) \longrightarrow 0.$$

Under the empty-interior condition, the preceding theorem also implies that $V(z_t)$ converges. Indeed, the image under V of the limit set of (z_t) is an interval contained in $V(\mathcal{S})$. Since $V(\mathcal{S})$ has empty interior, this interval must be a singleton, say $\{V_\infty\}$. Compactness of $\mathcal{K}$ and continuity of V then imply that $V(z_t) \to V_\infty$.

Proof outline. To apply Theorem E.1, we proceed in four steps:

1. **Completion of the limiting normalized field.** We complete the normalized potential field at zero-gradient blocks and show that the resulting set-valued field is upper hemicontinuous, locally bounded, and has nonempty compact convex values (definition in (52) and Lemma E.1).
2. **Graph consistency and stochastic-approximation representation.** We show that the finite-resolution comparison fields approach the completed limiting field in the sense of graphs (Lemma E.2). This allows us to represent our (CNGP) algorithm as a projected perturbed approximate discretization of the projected differential inclusion
$$\dot{x}(t) \in \Pi_{T_{\mathcal{X}}(x(t))} F(x(t)),$$
see Lemma E.3.
3. **Lyapunov analysis and potential stationarity.** We show that the ordinal potential Φ is a Lyapunov function for the projected limiting inclusion and identify the zero set of its dissipation with the constrained stationary points of the potential (Lemma E.4).
4. **From potential stationarity to Nash equilibrium.** We establish first-order alignment between the players' costs and the ordinal potential under Assumption 5.1 and, using blockwise pseudoconvexity, show that the potential stationary set coincides with the Nash equilibrium set (Lemmas C.1 and E.6). Theorem E.1 then yields the desired almost-sure convergence result.

Let $\mathcal{U}$ be the open neighborhood of $\mathcal{X}$ on which Φ is C^1. For every $i \in [N]$, let $g_i(x) := \nabla_i \Phi(x)$, and define $\mathcal{D}_i : \mathcal{U} \rightrightarrows \mathbb{R}^{d_i}$ by

$$\mathcal{D}_i(x) := \begin{cases} \left\{ \dfrac{g_i(x)}{\|g_i(x)\|} \right\}, & g_i(x) \neq 0, \\ \overline{\mathbb{B}}_{d_i}, & g_i(x) = 0. \end{cases} \tag{52}$$

Set $\mathcal{D}(x) := \prod_{i=1}^{N} \mathcal{D}_i(x)\,, F(x) := -\mathcal{D}(x)$.

At zero-gradient blocks, the unit ball provides a closed convex completion of the normalized field, while away from such blocks the completion reduces to the unique normalized gradient direction. The graph-consistency result below (Lemma E.2) shows that this completion captures all asymptotic cluster behavior of the finite-resolution comparison fields as the comparison radius vanishes.

We start by checking Assumption (A1) regarding the regularity of the set-valued mapping F.

Lemma E.1 (Regularity of the completed field). *The map $F = -\mathcal{D} : \mathcal{U} \rightrightarrows \mathbb{R}^d$ defined above in (52) is upper hemicontinuous, locally bounded, and has nonempty compact convex values. Moreover,* $\sup_{v \in \mathcal{D}(x)} \|v\| \leq \sqrt{N}$.

Proof. Nonemptiness, compactness, convexity, and uniform boundedness are immediate from the definition.

It remains to establish upper hemicontinuity. Since the values are compact and the map is locally bounded, it suffices in finite dimension to prove that its graph is closed.

Let $x_n \to x$ and $v_n \to v$ with $v_n \in \mathcal{D}(x_n)$. Let $i \in [N]$. If $g_i(x) \neq 0$, continuity of g_i gives $g_i(x_n) \neq 0$ for sufficiently large n, and hence

$$v_{i,n} = \frac{g_i(x_n)}{\|g_i(x_n)\|} \longrightarrow \frac{g_i(x)}{\|g_i(x)\|}.$$

Thus $v_i \in \mathcal{D}_i(x)$. If $g_i(x) = 0$, then $\|v_{i,n}\| \leq 1$, so $\|v_i\| \leq 1$ and $v_i \in \overline{\mathbb{B}}_{d_i} = \mathcal{D}_i(x)$. Thus $v \in \mathcal{D}(x)$ and the graph is closed. □

Lemma E.2 (Graph consistency). *As $\delta \downarrow 0$,*

$$\sup_{x \in \mathcal{X}^\delta} \operatorname{dist}\left((x, D_\Phi^\delta(x)), \operatorname{gph}(\mathcal{D})\right) \longrightarrow 0\,,$$

where $\operatorname{gph}(\mathcal{D}) := \{(x, v) \in \mathcal{X} \times \mathbb{R}^d : \ v \in \mathcal{D}(x)\}$ *is the graph of the set-valued mapping $\mathcal{D}$.*

Proof. Suppose, towards a contradiction, that the graph-consistency statement fails. Then there exist $\varepsilon > 0$, a sequence $\delta_n \downarrow 0$, and $x_n \in \mathcal{X}^{\delta_n}$ such that

$$\operatorname{dist}\left((x_n, D_\Phi^{\delta_n}(x_n)), \operatorname{gph}(\mathcal{D})\right) \geq \varepsilon, \qquad \forall n. \tag{53}$$

Since $\mathcal{X}$ is compact, after passing to a subsequence we may assume that $x_n \to \bar{x} \in \mathcal{X}$. Moreover, since $\|D_\Phi^{\delta_n}(x_n)\| \leq \sqrt{N}$, the sequence $(D_\Phi^{\delta_n}(x_n))$ is bounded. After extracting a further subsequence, we may therefore assume that $D_\Phi^{\delta_n}(x_n) \to v$ for some $v = (v_1, \ldots, v_N) \in \mathbb{R}^d$.

We claim that $v \in \mathcal{D}(\bar{x})$. To prove this, let $i \in [N]$.

If $\nabla_i \Phi(\bar{x}) = 0$, then $\|D_i^{\delta_n}(x_n)\| \leq 1$ for all n. Passing to the limit gives $\|v_i\| \leq 1$. Hence $v_i \in \overline{\mathbb{B}}_{d_i} = \mathcal{D}_i(\bar{x})$.

Suppose now that $\nabla_i \Phi(\bar{x}) \neq 0$. By continuity of $\nabla_i \Phi$, $\nabla_i \Phi(x_n) \neq 0$ for all sufficiently large n. Define

$$q_n := \frac{\nabla_i \Phi(x_n)}{\|\nabla_i \Phi(x_n)\|}.$$

By Lemma E.7,

$$\left\langle q_n, D_i^{\delta_n}(x_n) \right\rangle \geq 1 - \frac{L\delta_n}{\kappa_{d_i}\|\nabla_i \Phi(x_n)\|}.$$

Since $\|D_i^{\delta_n}(x_n)\| \leq 1$, we have

$$\begin{aligned}\left\| D_i^{\delta_n}(x_n) - q_n \right\|^2 &= \|D_i^{\delta_n}(x_n)\|^2 + 1 - 2\left\langle D_i^{\delta_n}(x_n), q_n \right\rangle \\ &\leq 2\left(1 - \left\langle D_i^{\delta_n}(x_n), q_n \right\rangle\right) \\ &\leq \frac{2L\delta_n}{\kappa_{d_i}\|\nabla_i \Phi(x_n)\|} \longrightarrow 0\,,\end{aligned}$$

where the above limit follows from noting that $\delta_n \to 0$ while $\|\nabla_i \Phi(x_n)\| \longrightarrow \|\nabla_i \Phi(\bar{x})\| > 0$. Therefore, $D_i^{\delta_n}(x_n) - q_n \to 0$. Since $q_n \longrightarrow \frac{\nabla_i \Phi(\bar{x})}{\|\nabla_i \Phi(\bar{x})\|}$, and $D_i^{\delta_n}(x_n) \to v_i$, we conclude that

$$v_i = \frac{\nabla_i \Phi(\bar{x})}{\|\nabla_i \Phi(\bar{x})\|} \in \mathcal{D}_i(\bar{x}).$$

Since $i \in [N]$ was arbitrary, we have shown that $v_i \in \mathcal{D}_i(\bar{x})$ for all $i \in [N]$. Hence $v \in \prod_{i=1}^N \mathcal{D}_i(\bar{x}) = \mathcal{D}(\bar{x})$, so $(\bar{x}, v) \in \mathrm{gph}(\mathcal{D})$.

Consequently,

$$\mathrm{dist}\left((x_n, D_\Phi^{\delta_n}(x_n)), \mathrm{gph}(\mathcal{D})\right) \leq \left\|(x_n, D_\Phi^{\delta_n}(x_n)) - (\bar{x}, v)\right\| \longrightarrow 0,$$

because $x_n \to \bar{x}, D_\Phi^{\delta_n}(x_n) \to v$. This contradicts (53). Therefore, as $\delta \downarrow 0$,

$$\sup_{x \in \mathcal{X}^\delta} \mathrm{dist}\left((x, D_\Phi^\delta(x)), \mathrm{gph}(\mathcal{D})\right) \longrightarrow 0\,.$$

□

Lemma E.3 (CNGP tracks the completed normalized inclusion)**.** *Under the assumptions of Theorem 5.2, there exist sequences $y_t \in \mathcal{X}$, $v_t \in \mathcal{D}(y_t)$, $r_t \to 0$, and a bounded martingale-difference sequence e_{t+1} such that, almost surely,*

$$z_{t+1} = \Pi_{\mathcal{X}}\left[z_t + \eta_t(-v_t + e_{t+1} + r_t)\right], \qquad \|y_t - z_t\| \to 0,$$

and $\sum_{t=1}^\infty \eta_t e_{t+1}$ converges. Consequently, (z_t) is an $\mathcal{X}$-projected perturbed approximate discretization of $F = -\mathcal{D}$ (in the sense of Theorem E.1).

Proof. Recall first that (CNGP) can be written as follows:

$$z_{t+1} = \Pi_{\mathcal{X}}(z_t - \eta_t \hat{D}_t)\,. \tag{54}$$

Define $e_{t+1} := -M_{t+1} = -(\widehat{D}_t - D_\Phi^{\delta_t}(x_t))$. We can then write:

$$-\hat{D}_t = -(\hat{D}_t - D_\Phi^{\delta_t}(x_t)) - D_\Phi^{\delta_t}(x_t) = -D_\Phi^{\delta_t}(x_t) + e_{t+1}\,. \tag{55}$$

Our next step is to decompose the term $-D_\Phi^{\delta_t}(x_t)$ using the graph consistency result (Lemma E.2) relating it to our completed set-valued mapping $-\mathcal{D}$ defined in (52).

Define

$$\varepsilon_t := \sup_{x \in \mathcal{X}^{\delta_t}} \mathrm{dist}\left((x, D_\Phi^{\delta_t}(x)), \mathrm{gph}(\mathcal{D})\right).$$

By Lemma E.2 and $\delta_t \to 0$, we have $\varepsilon_t \to 0$. Since $x_t \in \mathcal{X}^{\delta_t}$,

$$\mathrm{dist}\left((x_t, D_\Phi^{\delta_t}(x_t)), \mathrm{gph}(\mathcal{D})\right) \leq \varepsilon_t. \tag{56}$$

Since $\mathcal{X}$ is compact and $\sup_{v\in\mathcal{D}(x)}\|v\| \le \sqrt{N}$, the closed graph $\mathrm{gph}(\mathcal{D})$ is a compact subset of $\mathcal{X}\times\mathbb{R}^d$. Hence, the distance in (56) is attained and there exist $y_t \in \mathcal{X}$ and $v_t \in \mathcal{D}(y_t)$ such that

$$\Big\|(x_t, D_\Phi^{\delta_t}(x_t)) - (y_t, v_t)\Big\| = \mathrm{dist}\left((x_t, D_\Phi^{\delta_t}(x_t)), \mathrm{gph}(\mathcal{D})\right) \le \varepsilon_t.$$

In particular,

$$\|y_t - x_t\| \le \varepsilon_t \longrightarrow 0, \qquad \|v_t - D_\Phi^{\delta_t}(x_t)\| \le \varepsilon_t \longrightarrow 0. \tag{57}$$

We define now $r_t := v_t - D_\Phi^{\delta_t}(x_t)$ and we use (55) to obtain:

$$-\hat{D}_t = -D_\Phi^{\delta_t}(x_t) + e_{t+1} = -v_t + e_{t+1} + r_t\,, \quad v_t \in \mathcal{D}(y_t)\,. \tag{58}$$

Therefore, using (54) and the above decomposition (58), we obtain:

$$z_{t+1} = \Pi_{\mathcal{X}}\left[z_t + \eta_t\big(f_t + e_{t+1} + r_t\big)\right]\,,$$

where $f_t := -v_t \in F(y_t) = -\mathcal{D}(y_t), e_{t+1} = -M_{t+1} = -(\widehat{D}_t - D_\Phi^{\delta_t}(x_t)), r_t = v_t - D_\Phi^{\delta_t}(x_t)$.

We now check that the assumptions (A1) to (A5) in Theorem E.1 are satisfied.

Assumption (A1) is satisfied by Lemma E.1.

Assumption (A2) follows from Assumption 5.2; indeed $\sum_t \eta_t^2 < \infty$ implies $\eta_t \to 0$.

Assumption (A3): Since the comparison field of the game and that of the potential coincide exactly at finite resolution, Lemma 5.1 gives $\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] = 0$.

Moreover, the one-bit estimator is uniformly bounded: $\|\widehat{D}_t\| \le G := \left(\sum_{i=1}^N G_i^2\right)^{1/2}$ almost surely. Consequently, $\|M_{t+1}\| \le \|\hat{D}_t\| + \|D_\Phi^{\delta_t}(x_t)\| \le G + \sqrt{N}$ almost surely. Therefore,

$$\sum_{t=1}^{+\infty} \mathbb{E}[\|\eta_t M_{t+1}\|^2 | \mathcal{F}_t] \le (G+\sqrt{N})^2 \sum_{t=1}^{+\infty} \eta_t^2 < \infty.$$

Since $\sum_{t=1}^\infty \eta_t^2 < \infty$ by Assumption 5.2, the vector-valued martingale series $\sum_{t=1}^T \eta_t M_{t+1}$ converges almost surely. Noting that $e_{t+1} = -M_{t+1}$, it follows that $\sum_{t=1}^\infty \eta_t e_{t+1}$ converges almost surely, as required by Assumption (A3). We henceforth fix a sample path in the probability-one event on which the preceding martingale series converges and the uniform bounds hold.

Recall from (57) that $r_t \to 0$. Furthermore, recall that by construction, $x_{i,t} = \mathsf{Q}_{i,\delta_t}(z_{i,t})$, and hence $\|x_{i,t} - z_{i,t}\| \le \frac{R_i}{\rho_i}\delta_t$. Therefore, with $C_{\mathcal{X}} := \left(\sum_{i=1}^N \frac{R_i^2}{\rho_i^2}\right)^{1/2}$, we have

$$\|x_t - z_t\| \le C_{\mathcal{X}}\delta_t \longrightarrow 0. \tag{59}$$

Hence, we obtain using (57) and (59),

$$\|y_t - z_t\| \le \|y_t - x_t\| + \|x_t - z_t\| \le \varepsilon_t + C_{\mathcal{X}}\delta_t \longrightarrow 0\,, \tag{60}$$

where the limit follows from using $\varepsilon_t \to 0, \delta_t \to 0$. Thus Assumptions (A3) and (A4) are satisfied.

Assumption (A5) also holds as we have shown above that $\sup_t \|M_{t+1}\| < \infty$ almost surely. □

Lyapunov analysis of the projected limiting inclusion. We now verify the Lyapunov condition required by Majewski et al. (2018, Theorem 4.2). Recall that $F(x) = -\mathcal{D}(x)$ where $\mathcal{D}$ is defined in (52).

The projected limiting field on $\mathcal{X}$ is

$$F_{\mathcal{X}}(x) := \Pi_{T_{\mathcal{X}}(x)} F(x).$$

Since $\mathcal{X} = \prod_{i=1}^N \mathcal{X}_i$, we have $T_{\mathcal{X}}(x) = \prod_{i=1}^N T_{\mathcal{X}_i}(x_i)$, and Euclidean projection onto the tangent cone acts blockwise.

We next show that the ordinal potential Φ acts as a Lyapunov function for the projected differential inclusion induced by the (CNGP) stochastic approximation algorithm, i.e. we verify the second part

in Theorem E.1, namely (51) and the identification of the zero set of the upper-bounding function U.

Lemma E.4 (Lyapunov inequality for the projected normalized field)**.** *For every $i \in [N]$ define*

$$\Psi_i(x) := \begin{cases} \|g_i(x)\| \cdot \left\| \Pi_{T_{\mathcal{X}_i}(x_i)} \left(-\frac{g_i(x)}{\|g_i(x)\|} \right) \right\|^2, & \text{if } g_i(x) \neq 0, \\ 0, & \text{if } g_i(x) = 0, \end{cases} \tag{61}$$

and let $U(x) := -\sum_{i=1}^N \Psi_i(x)$. Then $U : \mathcal{X} \to \mathbb{R}$ is upper semicontinuous on $\mathcal{X}$, $U \leq 0$, and for every $x \in \mathcal{X}$ and every $u \in F(x)$,

$$\left\langle \nabla\Phi(x), \Pi_{T_{\mathcal{X}}(x)} u \right\rangle = U(x). \tag{62}$$

Moreover,

$$\{x \in \mathcal{X} : U(x) = 0\} = \{x \in \mathcal{X} : -\nabla\Phi(x) \in N_{\mathcal{X}}(x)\}. \tag{63}$$

Proof. **Proof of (62).** Let $x \in \mathcal{X}$ and $w \in F_{\mathcal{X}}(x)$. By definition of the projected field, there exists $u \in F(x) = -\mathcal{D}(x)$ such that $w = \Pi_{T_{\mathcal{X}}(x)}(u)$. Since both the action set and tangent cone are Cartesian products, this means blockwise that $w_i = \Pi_{T_{\mathcal{X}_i}(x_i)}(u_i), i \in [N]$.

Observe first that

$$\langle \nabla\Phi(x), w \rangle = \sum_{i=1}^N \langle g_i(x), w_i \rangle .$$

To prove (62), we will show in what follows that for every $i \in [N]$,

$$\langle g_i(x), w_i \rangle = -\Psi_i(x) . \tag{64}$$

Consider first a block $i \in [N]$ such that $g_i(x) \neq 0$. Then $F_i(x)$ is a singleton and necessarily $u_i = -\frac{g_i(x)}{\|g_i(x)\|}$. For a closed convex cone C, Moreau's decomposition gives

$$u_i = \Pi_C(u_i) + \Pi_{C^\circ}(u_i), \qquad \langle \Pi_C(u_i), \Pi_{C^\circ}(u_i) \rangle = 0.$$

Taking $C = T_{\mathcal{X}_i}(x_i), C^\circ = N_{\mathcal{X}_i}(x_i)$, and recalling that $w_i = \Pi_C(u_i)$, we obtain

$$\langle u_i, w_i \rangle = \langle \Pi_C(u_i) + \Pi_{C^\circ}(u_i), \Pi_C(u_i) \rangle = \|\Pi_C(u_i)\|^2 = \|w_i\|^2 .$$

Since $g_i(x) = -\|g_i(x)\| u_i$, it follows that

$$\begin{aligned} \langle g_i(x), w_i \rangle &= -\|g_i(x)\| \langle u_i, w_i \rangle \\ &= -\|g_i(x)\| \cdot \|w_i\|^2 \\ &= -\|g_i(x)\| \cdot \left\| \Pi_{T_{\mathcal{X}_i}(x_i)} \left(-\frac{g_i(x)}{\|g_i(x)\|} \right) \right\|^2 \\ &= -\Psi_i(x). \end{aligned} \tag{65}$$

If instead $g_i(x) = 0$, the block u_i may be any element of $-\overline{\mathbb{B}}_{d_i}$, but independently of this choice,

$$\langle g_i(x), w_i \rangle = 0 = -\Psi_i(x).$$

Summing over the blocks gives

$$\langle \nabla\Phi(x), w \rangle = -\sum_{i=1}^N \Psi_i(x) = U(x).$$

The right-hand side is independent of the particular selection $u \in F(x)$ at zero-gradient blocks. Consequently, $\left\langle \nabla\Phi(x), \Pi_{T_{\mathcal{X}}(x)} u \right\rangle = U(x)$. Since $\Psi_i(x) \geq 0$ for every i, we also have $U(x) \leq 0$.

Proof of upper semicontinuity of U. Since $U = -\sum_{i=1}^{N} \Psi_i$, it suffices to show that each Ψ_i is lower semicontinuous to conclude that their negative sum is upper semicontinuous, i.e., for every $x^n \to x$,

$$\Psi_i(x) \leq \liminf_n \Psi_i(x^n)\,. \tag{66}$$

First, by definition, $\Psi_i(y) \geq 0$ for every $y \in \mathcal{X}$. If $g_i(x) = 0$, then $\Psi_i(x) = 0$ by definition (see (61)). Therefore for any such x for which $g_i(x) = 0$, (66) holds since $0 = \Psi_i(x) \leq \liminf_n \Psi_i(x^n)$.

It remains to show lower semicontinuity of Ψ_i whenever $g_i(x) \neq 0$. Positive homogeneity of the projection onto a cone yields:

$$\Psi_i(x) = \|g_i(x)\| \cdot \left\| \Pi_{T_{\mathcal{X}_i}(x_i)} \left(-\frac{g_i(x)}{\|g_i(x)\|} \right) \right\|^2 = \frac{\left\| \Pi_{T_{\mathcal{X}_i}(x_i)} \left(-g_i(x) \right) \right\|^2}{\|g_i(x)\|}\,. \tag{67}$$

Using Moreau's decomposition gives:

$$\operatorname{dist}\left(-g_i(x), N_{\mathcal{X}_i}(x_i)\right) = \left\| -g_i(x) - \Pi_{\mathcal{N}_{\mathcal{X}_i}(x_i)}(-g_i(x)) \right\| = \left\| \Pi_{T_{\mathcal{X}_i}(x_i)} \left(-g_i(x) \right) \right\|\,. \tag{68}$$

Combining (67) and (68) yields:

$$\Psi_i(x) = \frac{\operatorname{dist}^2\left(-g_i(x), N_{\mathcal{X}_i}(x_i)\right)}{\|g_i(x)\|}. \tag{69}$$

We now use this expression to show that Ψ_i is lower semicontinuous in the case where $g_i(x) \neq 0$. Noting that g_i is continuous, our main task is to show that the function $d_i : \mathcal{X} \to \mathbb{R}_+$ defined for any $x \in \mathcal{X}$ by:

$$d_i(x) := \operatorname{dist}\left(-g_i(x), N_{\mathcal{X}_i}(x_i)\right)$$

is lower semicontinuous. Note here, that this is not a simple distance function to a fixed convex set which is known to be continuous. Here the set $\mathcal{N}_{\mathcal{X}_i}(x_i)$ also depends on the variable x_i. We prove the lower semicontinuity claim below.

Let $x^n \to x$ and, after passing to a subsequence realizing $\liminf_n d_i(x^n)$, let

$$v_i^n := \Pi_{N_{\mathcal{X}_i}(x_i^n)}\big(-g_i(x^n)\big).$$

Since $0 \in N_{\mathcal{X}_i}(x_i^n)$, $\|v_i^n\| \leq \|g_i(x^n)\|$ by nonexpansiveness of the projection operator. As a consequence, since $(g(x^n))$ converges by continuity of g, it follows that it is bounded and hence that (v_i^n) is also bounded. We can then pass to a further subsequence and let $v_i^n \to v_i$. The normal-cone mapping of a closed convex set has closed graph, and since $v_i^n \in N_{\mathcal{X}_i}(x_i^n)$, we obtain $v_i \in N_{\mathcal{X}_i}(x_i)$. Therefore,

$$d_i(x) = \inf_{y_i \in N_{\mathcal{X}_i}(x_i)} \| - g_i(x) - y_i\| \leq \| - g_i(x) - v_i\| = \lim_{n\to\infty} \| - g_i(x^n) - v_i^n\| = \liminf_{n\to\infty} d_i(x^n)\,,$$

where the penultimate identity follows from continuity of g_i and $x^n \to x, v_i^n \to v_i$, and the last identity follows from recalling that x^n is a subsequence realizing $\liminf_n d_i(x^n)$. We have proved that d_i is lower semicontinuous.

We now conclude to show that Ψ_i as expressed in (69) is lower semicontinuous. We use the following elementary fact:

Lemma E.5. *If $a_n \geq 0$ and $b_n \to b > 0$, then $\liminf_{n\to\infty} \frac{a_n}{b_n} = \frac{\liminf_{n\to\infty} a_n}{b}$.*

Proof. If $\liminf_n a_n = +\infty$, the claim follows immediately from $b_n \to b > 0$. Hence assume that $\liminf_n a_n < +\infty$. We prove the two inequalities separately.

For the first inequality, let (a_{n_k}) be a subsequence such that $a_{n_k} \to \liminf_{n\to\infty} a_n$. Since $b_{n_k} \to b > 0$, we have $\frac{a_{n_k}}{b_{n_k}} \longrightarrow \frac{\liminf_{n\to\infty} a_n}{b}$. Since the liminf of a sequence is no larger than any of its subsequential limits,

$$\liminf_{n\to\infty} \frac{a_n}{b_n} \leq \frac{\liminf_{n\to\infty} a_n}{b}.$$

For the reverse inequality, let $\left(a_{n_k}/b_{n_k}\right)$ be a subsequence such that $\frac{a_{n_k}}{b_{n_k}} \longrightarrow \liminf_{n\to\infty} \frac{a_n}{b_n}$. Since $b_{n_k} \to b > 0$,

$$a_{n_k} = b_{n_k} \frac{a_{n_k}}{b_{n_k}} \longrightarrow b \liminf_{n\to\infty} \frac{a_n}{b_n}.$$

Again using that the liminf is no larger than any subsequential limit,

$$\liminf_{n\to\infty} a_n \leq b \liminf_{n\to\infty} \frac{a_n}{b_n}.$$

Since $b > 0$, $\frac{\liminf_{n\to\infty} a_n}{b} \leq \liminf_{n\to\infty} \frac{a_n}{b_n}$. Combining the two inequalities proves the claim. □

Take now $a_n = d_i(x^n)^2, b_n = \|g_i(x^n)\|$. Since $g_i(x) \neq 0$ and g_i is continuous, it follows that $b_n \to \|g_i(x)\| > 0$. By lower semicontinuity of d_i, $d_i(x)^2 \leq (\liminf_{n\to\infty} d_i(x^n))^2 = \liminf_{n\to\infty} d_i(x^n)^2$. Therefore, we can conclude using Lemma E.5:

$$\Psi_i(x) = \frac{d_i(x)^2}{\|g_i(x)\|} \leq \frac{\liminf_{n\to\infty} d_i(x^n)^2}{\|g_i(x)\|} = \liminf_{n\to\infty} \frac{d_i(x^n)^2}{\|g_i(x^n)\|} = \liminf_{n\to\infty} \Psi_i(x^n)\,, \tag{70}$$

which shows that Ψ_i is lower semicontinuous.

Consequently, $U = -\sum_{i=1}^N \Psi_i$ is upper semicontinuous.

Proof of (63): zeros of the Lyapunov upper bound. Since each $\Psi_i(x) \geq 0$, $U(x) = 0$ if and only if $\Psi_i(x) = 0$ for all $i \in [N]$.

If $g_i(x) = 0$, then $-g_i(x) = 0 \in N_{\mathcal{X}_i}(x_i)$ and $\Psi_i(x) = 0$ by definition.

Suppose now that $g_i(x) \neq 0$. Then $\Psi_i(x) = 0$ if and only if $\Pi_{T_{\mathcal{X}_i}(x_i)}\left(-\frac{g_i(x)}{\|g_i(x)\|}\right) = 0$.

For a closed convex cone C, $\Pi_C(v) = 0$ if and only if $v \in C^\circ$. Since $T_{\mathcal{X}_i}(x_i)^\circ = N_{\mathcal{X}_i}(x_i)$, we obtain $\Psi_i(x) = 0$ if and only if $-\frac{g_i(x)}{\|g_i(x)\|} \in N_{\mathcal{X}_i}(x_i)$. Since the normal cone is a cone, this is equivalent to $-g_i(x) \in N_{\mathcal{X}_i}(x_i)$. This proves (63). The product-space characterization follows from noting that $N_{\mathcal{X}}(x) = \prod_{i=1}^N N_{\mathcal{X}_i}(x_i)$. □

For compatibility with Theorem E.1, we extend U from $\mathcal{X}$ to the ambient open set $\mathcal{U}$. Define

$$\widetilde{U}(x) := U\big(\Pi_{\mathcal{X}}(x)\big), \qquad x \in \mathcal{U}. \tag{71}$$

Since $\mathcal{X}$ is nonempty, closed, and convex, $\Pi_{\mathcal{X}}$ is continuous. Hence $\widetilde{U}$ is upper semicontinuous on $\mathcal{U}$ as the composition of the upper semicontinuous function U with the continuous map $\Pi_{\mathcal{X}}$. Moreover, $\widetilde{U}(x) = U(x), \forall x \in \mathcal{X}$, and $\widetilde{U} \leq 0$ on $\mathcal{U}$. With a slight abuse of notation, we henceforth denote this extension again by U.

Lemma E.6 (Potential stationarity is equivalent to Nash equilibrium)**.** *Under Assumption 3.1 and Assumption 5.1,* $\mathcal{S} = \mathrm{NE}(\mathcal{G})$, *where* $\mathcal{S} = \{x \in \mathcal{X} : U(x) = 0\}$.

Proof. By (63) in Lemma E.4 , $x \in \mathcal{S}$ if and only if $-\nabla_i \Phi(x) \in N_{\mathcal{X}_i}(x_i), \forall i \in [N]$.

We first prove that $\mathcal{S} \subseteq \mathrm{NE}(\mathcal{G})$. Let $x \in \mathcal{S}$, $i \in [N]$.

If $\nabla_i \Phi(x) = 0$, Assumption 5.1 gives $\nabla_i J_i(x) = 0$. Suppose for contradiction that x_i is not a best response to x_{-i}. Then there exists $y_i \in \mathcal{X}_i$ such that $J_i(y_i, x_{-i}) < J_i(x_i, x_{-i})$. By blockwise pseudoconvexity, $\langle \nabla_i J_i(x), y_i - x_i \rangle < 0$, contradicting $\nabla_i J_i(x) = 0$. Thus x_i is a best response.

Suppose now that $\nabla_i \Phi(x) \neq 0$. By Lemma C.1, $\nabla_i J_i(x) = \lambda_i(x) \nabla_i \Phi(x)$ for some $\lambda_i(x) > 0$. Since $-\nabla_i \Phi(x) \in N_{\mathcal{X}_i}(x_i)$ and the normal cone is a cone,

$$-\nabla_i J_i(x) = -\lambda_i(x) \nabla_i \Phi(x) \in N_{\mathcal{X}_i}(x_i).$$

Equivalently, $\langle \nabla_i J_i(x), y_i - x_i \rangle \geq 0, \forall y_i \in \mathcal{X}_i$. If there were $y_i \in \mathcal{X}_i$ satisfying

$$J_i(y_i, x_{-i}) < J_i(x_i, x_{-i}),$$

blockwise pseudoconvexity would imply $\langle \nabla_i J_i(x), y_i - x_i \rangle < 0$, a contradiction. Hence x_i is a best response. Since this holds for every $i \in [N]$, $x \in \mathrm{NE}(\mathcal{G})$.

Conversely, suppose $x \in \mathrm{NE}(\mathcal{G})$. For every $i \in [N]$, x_i minimizes the differentiable function $y_i \mapsto J_i(y_i, x_{-i})$ over the convex set $\mathcal{X}_i$. The first-order necessary optimality condition therefore gives $\langle \nabla_i J_i(x), y_i - x_i \rangle \geq 0, \forall y_i \in \mathcal{X}_i$, or equivalently, $-\nabla_i J_i(x) \in N_{\mathcal{X}_i}(x_i)$.

If $\nabla_i J_i(x) = 0$, nondegeneracy gives $\nabla_i \Phi(x) = 0$, and consequently $-\nabla_i \Phi(x) = 0 \in N_{\mathcal{X}_i}(x_i)$.

If $\nabla_i J_i(x) \neq 0$, Lemma C.1 gives $\nabla_i J_i(x) = \lambda_i(x) \nabla_i \Phi(x), \lambda_i(x) > 0$. Hence

$$-\nabla_i \Phi(x) = \frac{1}{\lambda_i(x)} \big(-\nabla_i J_i(x)\big) \in N_{\mathcal{X}_i}(x_i).$$

Thus the potential stationarity condition holds for every player, so $x \in \mathcal{S}$. This proves $\mathcal{S} = \mathrm{NE}(\mathcal{G})$. □

End of the proof of Theorem 5.2. Applying Theorem E.1 with $\mathcal{K} = \mathcal{X}, F = -\mathcal{D}, V = \Phi$, and with the upper semicontinuous extension of U defined in (71), the Lyapunov condition (51) follows from Lemma E.4, since this extension agrees with the original U on $\mathcal{X}$. Using $\mathcal{S} = \mathrm{NE}(\mathcal{G})$ (see Lemma E.6) and $\mathrm{int}_{\mathbb{R}}\, \Phi(\mathcal{S}) = \mathrm{int}_{\mathbb{R}}\, \Phi(\mathrm{NE}(\mathcal{G})) = \varnothing$ by assumption, we obtain, almost surely,

$$\mathrm{dist}\big(z_t, \mathrm{NE}(\mathcal{G})\big) \longrightarrow 0\,, \quad \Phi(z_t) \longrightarrow \Phi_\infty\,, \tag{72}$$

for some finite random variable Φ_∞.

We now transfer this conclusion to the sequence (x_t). By construction, $x_{i,t} = \mathsf{Q}_{i,\delta_t}(z_{i,t})$, and hence $\|x_{i,t} - z_{i,t}\| \leq \frac{R_i}{\rho_i}\delta_t$. Therefore, with $C_{\mathcal{X}} := \Big(\sum_{i=1}^N \frac{R_i^2}{\rho_i^2}\Big)^{1/2}$, we have

$$\|x_t - z_t\| \leq C_{\mathcal{X}} \delta_t \longrightarrow 0. \tag{73}$$

Combining (72) and (73),

$$\mathrm{dist}\big(x_t, \mathrm{NE}(\mathcal{G})\big) \leq \mathrm{dist}\big(z_t, \mathrm{NE}(\mathcal{G})\big) + \|x_t - z_t\| \underset{t\to\infty}{\longrightarrow} 0\,.$$

Moreover, uniform continuity of Φ on the compact set $\mathcal{X}$ yields $|\Phi(x_t) - \Phi(z_t)| \longrightarrow 0$, and hence $\Phi(x_t) \longrightarrow \Phi_\infty$.

Suppose $\mathrm{NE}(\mathcal{G}) = \{x_\star^1, \ldots, x_\star^m\}$ is finite. The rest of the proof is standard. For $m \geq 2$ let $\Delta := \min_{j\neq k} \|x_\star^j - x_\star^k\| > 0$. Since $\mathrm{dist}(z_t, \mathrm{NE}(\mathcal{G})) \to 0$, eventually

$$z_t \in \bigcup_{j=1}^m B(x_\star^j, \Delta/4).$$

On the other hand, nonexpansiveness of projection and the uniform bound $\|\widehat{D}_t\| \leq G$ give

$$\|z_{t+1} - z_t\| \leq G\eta_t \longrightarrow 0.$$

Hence, eventually, $\|z_{t+1} - z_t\| < \Delta/2$. Since distinct balls $B(x_\star^j, \Delta/4)$ are separated by distance at least $\Delta/2$, while eventually $\|z_{t+1} - z_t\| < \Delta/2$, the sequence cannot move between two distinct such balls after some finite time, and remains in only one such ball. Together with $\mathrm{dist}(z_t, \mathrm{NE}(\mathcal{G})) \to 0$, this implies

$$z_t \to X_\infty$$

for some random variable $X_\infty \in \mathrm{NE}(\mathcal{G})$. Finally, $\|x_t - z_t\| \to 0$ gives $x_t \to X_\infty$.

If $\mathrm{NE}(\mathcal{G})$ is a singleton, the same conclusion follows immediately from $\mathrm{dist}\big(x_t, \mathrm{NE}(\mathcal{G})\big) \longrightarrow 0$. This completes the proof.

E.3.2 Almost sure convergence to Nash in ordinal potential games in the unconstrained case

We slightly strengthen Assumption 5.2 with the condition $\sum_{t=1}^\infty \eta_t \delta_t < \infty$ for our alternative simpler proof based on the Robbins-Monro theorem in the unconstrained case ($\mathcal{X}_i = \mathbb{R}^{d_i}$). The stepsize

and radii choices made in the main part for Theorem 5.2, $\eta_t = \frac{\eta}{t^p}, \delta_t = \frac{\delta}{t^q}$, with $1/2 < p \leq 1$, $q > 1 - p$ and $\delta \in (0, \rho)$, remains valid and satisfies the following assumption.

Assumption E.1 (Robbins-Monro stepsizes and vanishing resolution)**.** The stepsizes (η_t) and the resolution (δ_t) satisfy: $\sum_{t=1}^{\infty} \eta_t = \infty, \sum_{t=1}^{\infty} \eta_t^2 < \infty$ and $\sum_{t=1}^{\infty} \eta_t \delta_t < \infty$, with $\eta_t > 0, \delta_t \in (0, \rho), \forall t \in \mathbb{N}$ and $\delta_t \to 0$.

In the unconstrained setting $X_i = \mathbb{R}^{d_i}$, set $x_t = z_t$, query $x_{i,t} \pm \delta_t U_{i,t}$, and update $x_{t+1} = x_t - \eta_t \widehat{D}_t$ instead of (CNGP).

Theorem E.2 (Almost-sure convergence to Nash equilibrium)**.** *Let $\mathcal{G}$ be an ordinal potential game with potential Φ. Let Assumptions 3.1 and 5.1 hold. Suppose that Φ is L-smooth and coercive[a]. Let (x_t) be generated by (CNGP) with the projection inactive, where $\widehat{D}_{i,t}$ is defined by (CNGP) under the gap-independent noisy comparison model with $\nu_i \in (0, 1/2]$. Suppose that the stepsizes and comparison radii satisfy Assumption E.1. Then, with probability one,*

$$\Phi(x_t) \text{ converges}, \qquad \|\nabla \Phi(x_t)\|_{2,1} \to 0\,, \quad \operatorname{dist}(x_t, \mathrm{NE}(\mathcal{G})) \to 0\,.$$

If $\mathrm{NE}(\mathcal{G})$ is finite, then there exists an $\mathrm{NE}(\mathcal{G})$-valued random variable X_∞ such that $x_t \to X_\infty$ almost surely. In particular, if $\mathrm{NE}(\mathcal{G})$ is a singleton, the iterates converge to it almost surely.

[a] i.e., $\Phi(x) \to +\infty$ as $\|x\| \to \infty$.

A convenient admissible choice is $\eta_t = \frac{\eta}{t^p}, \delta_t = \frac{\delta}{t^q}$, with $1/2 < p \leq 1$ and $q > 1 - p$.

The next estimate is useful because it controls the quantity that enters potential descent directly. In contrast, a uniform bound on $\|D_i^\delta - D_i\|$ necessarily becomes ill-conditioned as the gradient vanishes.

Lemma E.7 (Finite-resolution potential alignment)**.** *Under the assumptions of Theorem 5.2, for every player $i \in [N]$, every $\delta > 0$ and every $x \in \mathcal{X}^\delta$,*

$$\left\langle \nabla_i \Phi(x), D_i^\delta(x) \right\rangle \geq \|\nabla_i \Phi(x)\| - \frac{L\delta}{\kappa_{d_i}}. \tag{74}$$

If in addition $\nabla_i \Phi(x) \neq 0$, then $D_i(x) = \frac{\nabla_i \Phi(x)}{\|\nabla_i \Phi(x)\|}$ and

$$\left\langle \nabla_i \Phi(x), D_i^\delta(x) - D_i(x) \right\rangle \geq -\frac{L\delta}{\kappa_{d_i}}. \tag{75}$$

Proof. Let $x \in \mathcal{X}, i \in [N]$. Recalling the definition of the finite resolution vector field D_i^δ, we have

$$\left\langle \nabla_i \Phi(x), D_i^\delta(x) \right\rangle = \frac{1}{\kappa_{d_i}} \langle \nabla_i \Phi(x), \mathbb{E}_{U_i}[S_i^\delta(x, U_i) U_i] \rangle = \frac{1}{\kappa_{d_i}} \mathbb{E}_{U_i}[S_i^\delta(x, U_i) \langle \nabla_i \Phi(x), U_i \rangle]\,. \tag{76}$$

We will prove first that for any $u \in \mathbb{R}^{d_i}$ such that $\|u\|_2 = 1$,

$$S_i^\delta(x, u) \langle \nabla_i \Phi(x), u \rangle \geq |\langle \nabla_i \Phi(x), u \rangle| - L\delta\,. \tag{77}$$

Define for any $\delta > 0$ and any u s.t. $x_i \pm \delta u \in \mathcal{X}_i$ and $\|u\|_2 = 1$ the central difference quotient:

$$\Delta_{\Phi,i}^\delta(u) := \frac{\Phi(x_i + \delta u, x_{-i}) - \Phi(x_i - \delta u, x_{-i})}{2\delta}\,. \tag{78}$$

By L-smoothness of Φ, we have:

$$|\Phi(x_i + \delta u, x_{-i}) - (\Phi(x) + \delta \langle \nabla_i \Phi(x), u \rangle)| \leq \frac{L\delta^2}{2}\,, \tag{79}$$

$$|\Phi(x_i - \delta u, x_{-i}) - (\Phi(x) - \delta \langle \nabla_i \Phi(x), u \rangle)| \leq \frac{L\delta^2}{2}\,. \tag{80}$$

Using both the above inequalities, it follows that:

$$
\begin{aligned}
&|\Delta^{\delta}_{\Phi,i}(u) - \langle \nabla_i \Phi(x), u\rangle| \\
&= \frac{1}{2\delta}\,|\Phi(x_i + \delta u, x_{-i}) - \Phi(x_i - \delta u, x_{-i}) - 2\delta\langle \nabla_i \Phi(x), u\rangle| \\
&= \frac{1}{2\delta}\,|\Phi(x_i + \delta u, x_{-i}) - (\Phi(x) + \delta\langle \nabla_i \Phi(x), u\rangle) - (\Phi(x_i - \delta u, x_{-i}) - (\Phi(x) - \delta\langle \nabla_i \Phi(x), u\rangle))| \\
&\leq \frac{1}{2\delta}\left(\frac{L\delta^2}{2} + \frac{L\delta^2}{2}\right) = \frac{\delta L}{2}\,. \qquad (81)
\end{aligned}
$$

By the ordinal-potential property, $S^{\delta}_i(x,u) = \mathrm{sgn}(\Delta^{\delta}_{\Phi,i}(u))$. If $\mathrm{sgn}(\Delta^{\delta}_{\Phi,i}(u)) = \mathrm{sgn}(\langle \nabla_i \Phi(x), u\rangle)$, then $S^{\delta}_i(x,u)\langle \nabla_i \Phi(x), u\rangle = |\langle \nabla_i \Phi(x), u\rangle|$. Otherwise, the signs are opposite and (81) implies that $|\langle \nabla_i \Phi(x), u\rangle| \leq \frac{\delta L}{2}$, and hence

$$
S^{\delta}_i(x,u)\langle \nabla_i \Phi(x), u\rangle = -|\langle \nabla_i \Phi(x), u\rangle| \geq |\langle \nabla_i \Phi(x), u\rangle| - L\delta. \qquad (82)
$$

Therefore, the above inequality holds for any u, in both cases and even if one of the two quantities is zero, which proves (77). Using now (76) and taking expectation in (77) yields:

$$
\left\langle \nabla_i \Phi(x), D^{\delta}_i(x)\right\rangle = \frac{1}{\kappa_{d_i}}\mathbb{E}_{U_i}[S^{\delta}_i(x,U_i)\langle \nabla_i \Phi(x), U_i\rangle] \geq \frac{1}{\kappa_{d_i}}\mathbb{E}[|\langle \nabla_i \Phi(x), U_i\rangle|] - \frac{L\delta}{\kappa_{d_i}}\,. \qquad (83)
$$

Rotational invariance of the uniformly distributed random variable U_i on the sphere gives $\mathbb{E}[|\langle \nabla_i \Phi(x), U_i\rangle|] = \kappa_{d_i} \cdot \|\nabla_i \Phi(x)\|$ (see Lemma 3.1 in Fu et al. (2023)) and this concludes the proof of (74). Equation (75) follows from recalling that $D_i(x) = \frac{\nabla_i \Phi(x)}{\|\nabla_i \Phi(x)\|}$ when $\nabla_i \Phi(x) \neq 0$ by Lemma C.1, which implies that $\langle \nabla_i \Phi(x), D_i(x)\rangle = \|\nabla_i \Phi(x)\|\,.$ □

We are now ready to prove Theorem 5.2.

Proof. We divide the proof into four steps.

Step 1: Conditional Lyapunov drift. Since $\nabla\Phi$ is L-Lipschitz, we have

$$
\Phi(x_{t+1}) \leq \Phi(x_t) + \langle \nabla\Phi(x_t), x_{t+1} - x_t\rangle + \frac{L}{2}\|x_{t+1} - x_t\|^2.
$$

Using the update rule of the algorithm, it follows that:

$$
\Phi(x_{t+1}) \leq \Phi(x_t) - \eta_t \left\langle \nabla\Phi(x_t), \widehat{D}_t\right\rangle + \frac{L\eta_t^2}{2}\|\widehat{D}_t\|^2. \qquad (84)
$$

Taking conditional expectation with respect to $\mathcal{F}_t$ and using Lemma 5.1 and the definition of D^{δ}_i,

$$
\mathbb{E}[\Phi(x_{t+1}) \mid \mathcal{F}_t] \leq \Phi(x_t) - \eta_t \sum_{i=1}^{N} \left\langle \nabla_i \Phi(x_t), D^{\delta_t}_i(x_t)\right\rangle + \frac{LC_{2,\nu}}{2}\eta_t^2\,, \qquad (85)
$$

where $C_{2,\nu} = \sum_{i=1}^{N} \frac{1}{(2\nu_i \kappa_{d_i})^2}$.

Lemma E.7 yields

$$
\sum_{i=1}^{N} \left\langle \nabla_i \Phi(x_t), D^{\delta_t}_i(x_t)\right\rangle \geq \|\nabla\Phi(x_t)\|_{2,1} - LC_1\delta_t\,, \qquad (86)
$$

where $C_1 = \sum_{i=1}^{N} \frac{1}{\kappa_{d_i}}$. Hence, combining (85) with (86), we obtain:

$$
\mathbb{E}[\Phi(x_{t+1}) \mid \mathcal{F}_t] \leq \Phi(x_t) - \eta_t\|\nabla\Phi(x_t)\|_{2,1} + LC_1\eta_t\delta_t + \frac{LC_{2,\nu}}{2}\eta_t^2. \qquad (87)
$$

Step 2: Application of Robbins-Siegmund's theorem. Since Φ is continuous and coercive, it is bounded below by $\Phi_\star = \inf_{x\in\mathcal{X}} \Phi(x) > -\infty$. Set

$$
V_t := \Phi(x_t) - \Phi_\star \geq 0\,, \quad a_t := LC_1\eta_t\delta_t + \frac{LC_{2,\nu}}{2}\eta_t^2\,, \quad \forall t \geq 0\,. \qquad (88)
$$

We can then rewrite (87) as follows:

$$\mathbb{E}[V_{t+1} \mid \mathcal{F}_t] \leq V_t - \eta_t \|\nabla\Phi(x_t)\|_{2,1} + a_t. \tag{89}$$

By Assumption 5.2, $\sum_{t=1}^{\infty} a_t < \infty$ and we can apply the Robbins-Siegmund almost supermartingale theorem (Robbins & Siegmund, 1971) to obtain that V_t converges almost surely to a finite random variable and that

$$\sum_{t=1}^{\infty} \eta_t \|\nabla\Phi(x_t)\|_{2,1} < \infty \qquad \text{a.s.} \tag{90}$$

Since $\Phi(x_t)$ converges (as V_t does) and Φ is coercive, the sequence of play (x_t) is almost surely bounded.

Step 3: First-order stationarity guarantee. Define for any $x \in \mathcal{X}, t \geq 1$,

$$r(x) := \|\nabla\Phi(x)\|_{2,1} = \sum_{i=1}^{N} \|\nabla_i \Phi(x)\|, \qquad r_t := r(x_t).$$

Since $\sum_t \eta_t = \infty$ by Assumption 5.2 and $\sum_{t=1}^{\infty} \eta_t \|\nabla\Phi(x_t)\|_{2,1} < \infty$ as shown in (90), it follows that $\liminf_{t\to\infty} r_t = 0$.

We show that in fact $r_t \to 0$. The proof is similar to the proof of $\lim_{t\to\infty} \|\nabla_i \Phi(x(t))\| = 0$ in continuous time in the proof of Theorem C.2. However, here, we deal with the discrete time setting and we provide a complete proof using again standard arguments.

First, by the reverse triangle inequality and the Lipschitz continuity of $\nabla\Phi$, for any $x, y \in \mathcal{X}$,

$$|r(x) - r(y)| \leq \sum_{i=1}^{N} \|\nabla_i \Phi(x) - \nabla_i \Phi(y)\| \leq \sqrt{N}\, \|\nabla\Phi(x) - \nabla\Phi(y)\| \leq \sqrt{N} L \|x - y\|. \tag{91}$$

Furthermore, using again the update rule of the algorithm, we have:

$$\|x_{t+1} - x_t\| = \eta_t \|\hat{D}_t\| = \eta_t \left\| \left(\frac{1}{\kappa_{d_i}} S_{i,t} U_{i,t} \right)_{i \in [N]} \right\| \leq \sqrt{C_{2,\nu}}\, \eta_t. \tag{92}$$

Combining the above inequality with (91) yields for any $t \geq 1$,

$$|r_{t+1} - r_t| \leq K \eta_t, \qquad K := L\sqrt{N C_{2,\nu}}. \tag{93}$$

Suppose, towards a contradiction, that $\limsup_{n\to\infty} r_n > 0$. Together with $\liminf_{t\to\infty} r_t = 0$, this implies that there exists $\varepsilon > 0$ and infinitely many disjoint upcrossings from below ε to above 2ε. Consider one such sufficiently late upcrossing. Let $p < q$ be chosen so that $r_p < \varepsilon, r_q \geq 2\varepsilon$, and $r_n \geq \varepsilon$ for $p < n \leq q$. Using (93), we have

$$\varepsilon < r_q - r_p \leq K \sum_{t=p}^{q-1} \eta_t.$$

Since $\eta_t \to 0$, for all sufficiently large p, $K\eta_p \leq \frac{\varepsilon}{2}$ which implies, using the above inequality, that

$$\sum_{t=p+1}^{q-1} \eta_t \geq \frac{\varepsilon}{2K}, \tag{94}$$

up to adjustments of the endpoints in the sum. On this interval, $r_t \geq \varepsilon$, and hence

$$\sum_{t=p+1}^{q-1} \eta_t r_t \geq \frac{\varepsilon^2}{2K}.$$

Infinitely many disjoint upcrossings would therefore force $\sum_t \eta_t r_t = \infty$, contradicting (90). Thus $r_t \to 0$ almost surely.

Step 4: Identification of the limit set with Nash equilibrium. The proof follows the same lines as the proof of Lemma E.6, and hence omitted here.

□

E.4 FINITE-TIME FIRST-ORDER STATIONARITY

We conclude the unconstrained analysis with a finite-time first-order stationarity guarantee to complement the asymptotic convergence result of Theorem E.2 above.

Theorem E.3 (Finite-time stationarity)**.** *Suppose Φ is L-smooth and bounded below by $\Phi_{\inf}$. Then for any time horizon $T > 0$, with constant stepsize η_t and comparison radius δ_t of order $T^{-1/2}$, a uniformly sampled iterate x_τ from $\{x_1, \dots, x_T\}$ satisfies $\mathbb{E}\|\nabla\Phi(x_\tau)\|_{2,1} = \mathcal{O}(T^{-1/2})$.*

We provide a more detailed statement of this result in Theorem E.4 with the dependence on all constants of the problem (potential gap $\Phi(x_1) - \Phi_{\inf}$, dimension d_i, number N of players).

We first restate Theorem E.3 here in more details with explicit dependence on the problem constants.

Let $C_1 := \sum_{i=1}^N \frac{1}{\kappa_{d_i}} = \Theta\left(\sum_{i=1}^N \sqrt{d_i}\right), C_{2,\nu} := \sum_{i=1}^N \frac{1}{(2\nu_i\kappa_{d_i})^2} = \Theta\left(\sum_{i=1}^N d_i\right)$.

Theorem E.4 (Finite-time stationarity)**.** *Suppose Φ is L-smooth and bounded below by $\Phi_{\inf}$ along the iterates of the algorithm. Then for any time horizon $T > 0$,*

$$\mathbb{E}\|\nabla\Phi(x_\tau)\|_{2,1} \le \frac{\Delta_\Phi + LC_1\sum_{t=1}^T \eta_t\delta_t + \frac{LC_{2,\nu}}{2}\sum_{t=1}^T \eta_t^2}{\sum_{t=1}^T \eta_t}, \tag{95}$$

where $\Delta_\Phi = \Phi(x_1) - \Phi_{\inf}$. and τ is a random variable sampled according to $\mathbb{P}(\tau = t) = \frac{\eta_t}{\sum_{s=1}^T \eta_s}$. Setting $\eta_t = \frac{a}{\sqrt{T}}, \delta_t = \frac{b}{\sqrt{T}}$ for fixed constants $a, b > 0$ yields $\mathbb{E}\|\nabla\Phi(x_\tau)\|_{2,1} = \mathcal{O}(T^{-1/2})$ where τ is uniform on $\{1, \dots, T\}$.

Proof. Recall from (87) that

$$\mathbb{E}[\Phi(x_{t+1}) \mid \mathcal{F}_t] \le \Phi(x_t) - \eta_t\|\nabla\Phi(x_t)\|_{2,1} + LC_1\eta_t\delta_t + \frac{LC_{2,\nu}}{2}\eta_t^2. \tag{96}$$

Taking total expectation in the above inequality, summing over $t = 1, \dots, T$, using the lower bound $\Phi_{\inf}$, rearranging and dividing by $\sum_{t=1}^T \eta_t$ yields:

$$\frac{\sum_{t=1}^T \eta_t\mathbb{E}\left[\|\nabla\Phi(x_t)\|_{2,1}\right]}{\sum_{t=1}^T \eta_t} \le \frac{\Delta_\Phi + LC_1\sum_{t=1}^T \eta_t\delta_t + \frac{LC_{2,\nu}}{2}\sum_{t=1}^T \eta_t^2}{\sum_{t=1}^T \eta_t}, \tag{97}$$

where $\Delta_\Phi = \Phi(x_1) - \Phi_{\inf}$. The above can be rewritten equivalently using a random variable τ sampled according to $\mathbb{P}(\tau = t) = \frac{\eta_t}{\sum_{s=1}^T \eta_s}$, as follows:

$$\mathbb{E}\|\nabla\Phi(x_\tau)\|_{2,1} \le \frac{\Delta_\Phi + LC_1\sum_{t=1}^T \eta_t\delta_t + \frac{LC_{2,\nu}}{2}\sum_{t=1}^T \eta_t^2}{\sum_{t=1}^T \eta_t}. \tag{98}$$

□