

ROBUST EQUILIBRIA IN CONTINUOUS GAMES: FROM STRATEGIC TO DYNAMIC ROBUSTNESS

KYRIAKOS LOTIDIS^{*,c}, PANAYOTIS MERTIKOPOULOS[◊],
NICHOLAS BAMBOS^{*}, AND JOSE BLANCHET^{*}

ABSTRACT. In this paper, we examine the robustness of Nash equilibria in continuous games, under both strategic and dynamic uncertainty. Starting with the former, we introduce the notion of a *robust equilibrium* as those equilibria that remain invariant to small—but otherwise arbitrary—perturbations to the game’s payoff structure, and we provide a crisp geometric characterization thereof. Subsequently, we turn to the question of dynamic robustness, and we examine which equilibria may arise as stable limit points of the dynamics of “*follow the regularized leader*” (FTRL) in the presence of randomness and uncertainty. Despite their very distinct origins, we establish a structural correspondence between these two notions of robustness: strategic robustness implies dynamic robustness, and, conversely, the requirement of strategic robustness cannot be relaxed if dynamic robustness is to be maintained. Finally, we examine the rate of convergence to robust equilibria as a function of the underlying regularizer, and we show that entropically regularized learning converges at a geometric rate in games with affinely constrained action spaces.

1. INTRODUCTION

A fundamental requirement in game theory—which predates even the cornerstone notion of a Nash equilibrium—concerns the robustness that should be inherent in any axiomatization of rational behavior. To quote a famous passage by von Neumann & Morgenstern [53, p. 32]: “In whatever way we formulate the guiding principles and the objective justification of rational behavior, provisos will have to be made for every possible conduct of “the others.” If the superiority of rational behavior over any other kind is to be established, then its description must include rules of conduct for all conceivable situations—including those where “the others” behaved irrationally in the sense of the standards which the theory will set for them.”

As a byproduct of this tenet, there has been a flurry of activity since the 1970s in proposing refinements of the Nash equilibrium concept, all in an effort to dismiss equilibria that are highly fragile or otherwise implausible (e.g., because they involve threats that are not credible).¹ This pursuit of robustness has recently gained increased momentum owing to the applications of game theory to machine learning and data science, two fields where the notion of robustness has been likewise elusive. Here, even though many game-theoretic solutions perform extremely well on specific tasks—such as a well-trained generative adversarial network (GAN) at equilibrium—the resulting models tend to have a narrow performance

^{*} STANFORD UNIVERSITY.

[◊] UNIV. GRENOBLE ALPES, CNRS, INRIA, GRENOBLE INP, LIG, 38000 GRENOBLE, FRANCE.

^c CORRESPONDING AUTHOR.

E-mail addresses: klotidis@stanford.edu, panayotis.mertikopoulos@imag.fr, bambos@stanford.edu, jose.blanchet@stanford.edu.

2020 *Mathematics Subject Classification.* Primary 91A10, 91A26; secondary 68Q32, 68T05.

Key words and phrases. Robust equilibrium; regularized learning; stochastic stability.

¹For a masterful introduction to the topic, see the textbook of van Damme [52].

envelope, being brittle, and unable to adapt to situations that deviate from their initial configuration.

In game-theoretic terms, this highlights the fact that, even though a Nash equilibrium is resilient to unilateral deviations, it need not be *robust* to small perturbations in the *payoff data* of the game (which, in a machine learning context, could represent distributional shifts, incomplete observations, and/or other sources of uncertainty). In view of this, it is natural to ask

Which equilibria remain robust in the presence of strategic uncertainty?

This question has been the lodestar of the equilibrium refinement literature, and it has led to a wide array of proposals aiming to get rid of “unreasonable” equilibria that may disappear even under the most minute perturbation to the players’ payoffs—from Selten’s notion of *trembling hand perfection* [46], to Myerson’s concept of *properness* [38], and the various criteria of strategic stability introduced by Kohlberg & Mertens [28] (hyperstability, full stability, sequential stability, etc.).

Dually to the above theory of “strategic refinement”, an important alternative approach has been based on *dynamic* considerations: that is, the players of a game start off-equilibrium, and in one sense or another learn (or fail to learn) to play an equilibrium over time. Here, the focus is on the players’ learning protocol, the information available during play, and the presence (or absence) of players that may deviate from this protocol. By the so-called “folk theorem of evolutionary game theory” [23], it is well known that only *strict* equilibria are stable and attracting under the replicator dynamics, a result which was extended more recently to a broad class of “regularized learning” schemes, in both continuous [17] and discrete time [18, 19, 36].

These two viewpoints are not always compatible: for instance, in 2×2 games with two pure equilibria and one mixed (such as the Chicken / Hawk-Dove game), the mixed equilibrium is ruled out by almost all game-theoretic learning algorithms and dynamics, even though it survives a broad range of strategic refinement attacks. A point of hope here is the equivalence between (setwise) strategic and dynamic stability proved by Ritzberger & Weibull [40], who showed that a span of pure strategies in the mixed extension of a finite game is strategically stable in the sense of Kohlberg & Mertens [28] if and only if it is asymptotically stable under the replicator dynamics—see also [10, 13] for an extension to a wider class of discrete-time models for learning, with different information assumptions.

Notably, these considerations all concern finite games in normal (or extensive) form. By contrast, most applications of game theory to machine learning and data science involve *continuous games*, that is, games with a finite number of players and a continuum of actions per player—for example, GANs, multi-agent reinforcement learning, Kelly auctions, etc. In view of this, our paper seeks to answer the following questions in the context of continuous games:

Which equilibria arise as robust predictions of the players’ learning dynamics?

We refer to these two types of robustness as *strategic* and *dynamic robustness*, respectively. Our paper aims to quantify the interplay between the two, and the links that connect them.

Our contributions in the context of related work. Aiming for the strongest possible definition of robustness, we propose the following strategic refinement criterion:

*An equilibrium of a continuous game is strategically robust
if it remains an equilibrium in any slightly perturbed, nearby game.*

This requirement is similar in spirit to—but considerably stronger than—the classical notion of *essentiality* of Wu & Jiang [55], which posits that any nearby game has a nearby, possibly different equilibrium. Importantly, our results apply to *local* Nash equilibria, which are

especially relevant in machine learning applications where payoff landscapes are typically nonconcave. This distinction is crucial, as global Nash equilibria do not always exist in general continuous games, making local equilibrium guarantees both meaningful and necessary in practice.

An important point here is that, in contrast to finite games—where the notion of “nearby” is fairly unambiguous—perturbations to a continuous game involve functional variations and, as such, the metric that quantifies a “small” perturbation plays a crucial role. Importantly, albeit natural, our proposed robustness requirement becomes vacuous if distances are measured with respect to the players’ payoff functions: more precisely, it is always possible to find a payoff perturbation with arbitrarily small L^∞ -norm that ends up upsetting *any* equilibrium.

The underlying issue here is that a small payoff perturbation may exhibit very high local variability, which can disrupt the first-order stationarity conditions that characterize equilibria in continuous games, thereby eliminating them altogether. To circumvent this issue, we argue that deviations of continuous games should be measured by comparing their respective *gradient fields*, which encode all the strategic information in the game. This shift in perspective leads to a crisp geometric characterization of strategically robust equilibria: they are extreme points of the game’s action space, and they are sharp in the sense that the game’s individual payoff gradients form a strictly acute angle with any tangent direction (cf. Fig. 1 later in the paper).

From a dynamic standpoint, we focus throughout on the family of algorithms known as “*follow the regularized leader*” (FTRL) [31, 47–49]. This is arguably one of the most—if not the most—popular class of policies for online learning due to its strong regret minimization and convergence guarantees, and it contains as special cases gradient descent/ascent methods [3, 58], dual averaging [39, 56], the exponential / multiplicative weights algorithm [2, 5, 5, 33, 54], implicitly normalized forecasters [1, 4, 57], exponentiated gradient methods [7, 27, 51], and many stochastic approximation schemes, adaptive [24, 25] and non-adaptive alike [26, 34–36].

In this general context, we examine which equilibria admit robust convergence guarantees as stable limit points of the dynamics of FTRL in the presence of randomness and uncertainty. Our first main result is that *strategic robustness implies dynamic robustness*, i.e., any strategically robust equilibrium is stable and attracting with high probability under the dynamics of FTRL, for any choice of regularizer. Conversely, we also show that the strategic robustness requirement cannot be lifted, and we provide an example of a game with an extreme, non-robust equilibrium which attracts *all* FTRL orbits under a certain choice of regularizer, and *none* under another.

To the best of our knowledge, this is the first result of its kind for general continuous games. In the context of *finite* games, Flokas et al. [17] showed that a point is asymptotically stable under the continuous-time FTRL dynamics if and only if it is a strict Nash equilibrium, while [10] extended this equivalence to discrete-time models of regularized learning under uncertainty. Strict equilibria are prime examples of strategically robust equilibria, so this part of the analysis of [10] is subsumed in ours. In the context of concave games—that is, continuous games with individually concave payoff functions—Mertikopoulos & Zhou [34] showed that sharp global equilibria enjoy comparable convergence guarantees under FTRL with a vanishing step-size. While such step-size schedules are effective at suppressing noise in the long run, they do so at the cost of significantly slowing down the algorithm’s convergence. By contrast, we focus on fast, *constant* step-size schedules, which are widely used in practice due to their simplicity and often superior empirical performance. In this regime, we show that entropically regularized learning with a constant step-size converges to robust equilibria at a geometric rate, compared to distinctly subgeometric rates in the case

of vanishing step-size policies—subsuming in this way a range of previous results for finite [19] and stochastic games [20].

2. PRELIMINARIES

We start by briefly reviewing some basics of game theory and regularized learning, introducing the necessary context for our results.

2.1. The game-theoretic framework. Throughout our paper, we focus on a class of *continuous games* consisting of a finite set of players $i \in \mathcal{N} = \{1, \dots, N\}$, and defined by the following primitives:

- (1) Each player $i \in \mathcal{N}$ has access to a compact convex subset \mathcal{X}_i of some finite dimensional vector space \mathcal{V}_i , describing the set of *actions* available to said player. By $\mathcal{X} := \prod_i \mathcal{X}_i$ we denote the space of all ensembles $x = (x_1, \dots, x_N)$ of actions $x_i \in \mathcal{X}_i$ that are independently chosen by each player $i \in \mathcal{N}$. We will also write $x = (x_i; x_{-i})$ to emphasize the action of player $i \in \mathcal{N}$ against the joint action profile $x_{-i} \equiv (x_j)_{j \neq i}$ of all other players.
- (2) The players' rewards are determined by their individual *payoff functions* $u_i: \mathcal{X} \rightarrow \mathbb{R}$, assumed to be continuously differentiable for all $i \in \mathcal{N}$. Denoting by $\mathcal{Y}_i \equiv \mathcal{V}_i^*$ the dual space of \mathcal{V}_i , we define the individual gradient vector $v_i: \mathcal{X} \rightarrow \mathcal{Y}_i$ of player $i \in \mathcal{N}$ by

$$v_i(x) = \nabla_{x_i} u_i(x_i; x_{-i}) \quad (1)$$

and the ensemble $v(x) = (v_1(x), \dots, v_N(x)) \in \mathcal{Y} \equiv \prod_{i \in \mathcal{N}} \mathcal{Y}_i$ thereof.

A *continuous game* is then defined as a tuple $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ with players, actions and payoff functions as above.

Nash equilibrium. The best known solution concept in game theory is that of a *Nash equilibrium* (NE), which characterizes the actions $x^* \in \mathcal{X}$ from which no player has incentive to unilaterally deviate. Formally, $x^* \in \mathcal{X}$ is a Nash equilibrium if

$$u_i(x^*) \geq u_i(x_i; x_{-i}^*) \quad \text{for all } x_i \in \mathcal{X}_i, i \in \mathcal{N}. \quad (\text{NE})$$

A game $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ always admits a Nash equilibrium if \mathcal{X} is compact and each player's payoff function u_i is *individually concave* in the sense that $u_i(x_i; x_{-i})$ is concave in x_i for all $x_{-i} \in \mathcal{X}_{-i}$ [14, 44]. In this case, basic arguments from convex analysis [42, 43] show that x^* is an equilibrium of \mathcal{G} if and only if it satisfies the (Stampacchia) variational inequality

$$\langle v(x^*), x - x^* \rangle \leq 0 \quad \text{for all } x \in \mathcal{X}. \quad (\text{VI})$$

If the players' functions are not individually concave, a game may not admit a Nash equilibrium. In that case, it is more meaningful to consider *local* Nash equilibria, i.e., profiles $x^* \in \mathcal{X}$ such that

$$u_i(x^*) \geq u_i(x_i; x_{-i}^*) \quad \text{for all } x \text{ in a neighborhood } \mathcal{U} \text{ of } x^* \text{ in } \mathcal{X}. \quad (\text{LNE})$$

In stark contrast to games with individually concave payoff functions, (VI) no longer characterizes local Nash equilibria: specifically, by first-order stationarity, we have (LNE) \implies (VI) but the converse need not hold; in fact, a solution x^* of (VI) may be a global payoff *maximizer* for all $i \in \mathcal{N}$.

Note. In the sequel, we will work with general continuous games that may not admit a global equilibrium—but admit *local* Nash equilibria. To streamline our presentation, we will use the term “equilibrium” without any further qualification to refer to *local* equilibria, and we will say explicitly “global equilibria” for profiles satisfying (NE).

2.2. Regularized learning in games. The most widely used framework for learning in games, is the so called “*follow the regularized leader*” (FTRL) template, primarily because it leads to no regret in a wide variety of settings [48, 49]. The corresponding update rule hinges on the notion of a *regularized choice map*, and proceeds as

$$y_{t+1} = y_t + \gamma \hat{v}_t, \quad x_t = Q(y_t) \quad \text{for } t = 1, 2, \dots \quad (\text{FTRL})$$

where (i) $x_t \in \mathcal{X}$ denotes the players’ action profile at step t ; (ii) $y_t = (y_{i,t})_{i \in \mathcal{N}} \in \mathcal{Y}$ is an auxiliary process that aggregates historical feedback into a compact state representation, i.e., a proxy for the players’ empirical performance up to time t ; (iii) $\hat{v}_t = (\hat{v}_{i,t})_{i \in \mathcal{N}} \in \mathcal{Y}$ denotes the current gradient-like payoff signal; (iv) $\gamma > 0$ is the learning rate, or step-size parameter of the process; and (v) $Q: \mathcal{Y} \rightarrow \mathcal{X}$ is a mapping between the auxiliary process on the dual space \mathcal{Y} , and the players’ strategy space \mathcal{X} . In what follows, we analyze the key components of this framework.

The algorithm’s step-size. Throughout this work, we adopt a *constant* step-size routine. This stands in contrast to the stochastic approximation literature [8, 11, 29], where (FTRL) is typically implemented with a *vanishing* step-size satisfying the Robbins-Monro summability conditions $\sum_t \gamma_t = \infty$, $\sum_t \gamma_t^2 < \infty$ [41], which is known to promote convergence by gradually suppressing the effect of noise [34].

On the other hand, in practical applications, it is common to employ a *constant* (or non-diminishing) step-size for several reasons. First, constant step-sizes are easier to tune and maintain, making them more suitable for large-scale or production environments. Moreover, methods with vanishing step-sizes often experience long warm-up phases and converge slowly to a neighborhood of the equilibrium. In comparison, constant step-size methods in machine learning settings typically reach the vicinity of a solution much faster—often within 0.1% accuracy [16]. Indeed, many state-of-the-art architectures, including transformers and large language models, use step-size schedules that remain effectively constant over billions or even trillions of samples [15].

The mirror map. A central ingredient of regularized learning is the mirror map $Q \equiv (Q_i)_{i \in \mathcal{N}}$, with each $Q_i: \mathcal{Y}_i \rightarrow \mathcal{X}_i$ induced by a strongly convex *regularizer* $h_i: \mathcal{X}_i \rightarrow \mathbb{R}$ that promotes stability during the learning process. To streamline our presentation and letting $h(x) = \sum_{i \in \mathcal{N}} h_i(x_i)$, the players’ *mirror map* is defined as

$$Q(y) := \arg \max_{x \in \mathcal{X}} \{ \langle y, x \rangle - h(x) \} \quad (2)$$

In the rest of our paper, we will write $\mathcal{X}_h = \text{im } Q$ for the image of \mathcal{Y} under Q —and, likewise, $\mathcal{X}_{h_i} = \text{im } Q_i$ for each player $i \in \mathcal{N}$. In particular, if Q is interior-valued—that is, $\mathcal{X}_h = \text{ri } \mathcal{X}$ —we will say that h is *steep* because, in this case, the (sub)gradients of h explode to infinity as $x \rightarrow \text{bd } \mathcal{X}$ (i.e., h becomes “infinitely steep”); instead, if $\text{im } Q = \mathcal{X}$, we will say that h is *non-steep*. For a detailed discussion on this distinction and related concepts, see [Appendix A](#).

Different choices of the regularizer h induce different projection-like operations, adapted to the geometry of the underlying space. We describe two mainstay examples below.

Example 2.1 (Euclidean projection). The quadratic regularizer $h(x) = \|x\|_2^2/2$ gives rise to the Euclidean projection $Q(y) = \Pi_{\mathcal{X}}(y) = \arg \min_{x \in \mathcal{X}} \|y - x\|_2$. Here, h is non-steep. \blacklozenge

Example 2.2 (Exponential weights). For \mathcal{A}_i a *finite* set of actions per player $i \in \mathcal{N}$, and $\mathcal{X}_i \equiv \Delta(\mathcal{A}_i)$, the entropic regularizer $h_i(x_i) = \sum_{\alpha_i \in \mathcal{A}_i} x_{i\alpha_i} \log x_{i\alpha_i}$ gives rise to the logit map, defined via $Q_i(y_i) = \exp(y_i)/\|\exp(y_i)\|_1$, where $\exp(y_i)$ denotes the element-wise exponential of y_i . \blacklozenge

The feedback process. Throughout this work, we consider two distinct feedback models: (i) *stochastic gradients*; and (ii) *payoff-based* feedback. We describe both frameworks below.

Stochastic gradient feedback. At every time step t , each player $i \in \mathcal{N}$ has access to a *stochastic first-order oracle* (SFO)—that is, a noisy version of their individual gradient vector of the form:

$$\hat{v}_t = v(x_t) + U_t \quad \text{with} \quad \mathbb{E}[U_t | \mathcal{F}_t] = 0 \quad (\text{SFO})$$

where U_t is zero-mean and conditionally sub-Gaussian given the information \mathcal{F}_t generated up to time $t \in \mathbb{N}$. In other words, players observe unbiased estimates of their individual gradient vectors.

Payoff-based feedback. Unlike (SFO) where players have access to a black-box oracle that provides noisy gradient information, it is often more realistic to consider a payoff-based paradigm where players observe *only* their realized payoffs—that is, a single scalar value—and have to reconstruct an estimate of their individual gradient vectors.

The most widely used method in this setting is the *single-point stochastic approximation* (SPSA) framework of [12, 50], which is based on finite differences along randomly sampled directions. Specifically, denoting the set of unit directions $\mathcal{E}_i := \{\pm e_1, \dots, \pm e_{d_i}\}$ that span the affine hull of \mathcal{X}_i of dimension d_i , each player $i \in \mathcal{N}$ draws a direction $w_{i,t} \in \mathcal{E}_i$ uniformly at random in every round $t \in \mathbb{N}$. Since the perturbed action $x_{i,t} + \delta_t w_{i,t}$ may lie outside \mathcal{X}_i for a perturbation radius $\delta_t > 0$, we introduce a pivot element $p_i \in \text{ri}(\mathcal{X}_i)$ and a radius $r_i > \delta_t$ such that $p_i + r_i w_i \in \mathcal{X}_i$ for all $w_i \in \mathcal{E}_i$. Based on these, we define the feasibility-adjusted action $x_{i,t}^\delta := x_{i,t} + (\delta_t/r_i)(p_i - x_{i,t}) \in \mathcal{X}_i$. Finally, each player queries the perturbed action $\hat{x}_{i,t} \equiv x_{i,t}^\delta + \delta_t w_{i,t}$ which is an element of \mathcal{X}_i , and observes the realized payoff value $u_i(\hat{x}_t)$.² The gradient vector is, then, estimated via the single-point stochastic approximation scheme:

$$\hat{v}_{i,t} := (d_i/\delta_t) u_i(\hat{x}_t) w_{i,t} \quad (\text{SPSA})$$

Importantly, the feasibility adjustment ensures that the perturbed action \hat{x}_t remains within the players' action set \mathcal{X} , while preserving the direction of the original perturbation w_t . As we show in [Appendix A](#), (SPSA) enjoys the bounds

$$\|\mathbb{E}[\hat{v}_t | \mathcal{F}_t] - v(x_t)\|_* = \mathcal{O}(\delta_t) \quad \text{and} \quad \|\hat{v}_t\|_* = \mathcal{O}(1/\delta_t). \quad (3)$$

These statistical properties of (SPSA) will play a crucial role in establishing its convergence guarantees; we will revisit them in [Section 4](#).

3. STRATEGIC ROBUSTNESS: GEOMETRIC AND VARIATIONAL CHARACTERIZATION

We begin in this section by addressing the strategic aspects of the equilibrium robustness question, namely:

*Which equilibria remain invariant under
small perturbations of the underlying game?*

We take this desideratum as the starting point for our definition of *strategic robustness*, that is, action profiles that remain (local) equilibria under small disturbances in the underlying game. This leads to a delicate interplay between the variational and geometric aspects of the underlying game, which we detail below.

²Since $r_i > \delta_t$, we write $x_{i,t}^\delta = x_{i,t}(1 - \delta_t/r_i) + (\delta_t/r_i)p_i$ which is a convex combination of points in \mathcal{X}_i . Regarding $\hat{x}_{i,t}$, note it can be written as $\hat{x}_{i,t} = x_{i,t}(1 - \delta_t/r_i) + (\delta_t/r_i)(p_i + r_i w_{i,t})$, which is also a convex combination of points in \mathcal{X}_i . Thus, both belong to \mathcal{X}_i .

3.1. A first approach and insights. The first step in our analysis is to quantify the meaning of “small”. In this regard, a natural way to measure the distance between two concave games, $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ and $\tilde{\mathcal{G}} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, \tilde{u})$, would be via the uniform distance

$$\rho(\mathcal{G}, \tilde{\mathcal{G}}) := \max_{i \in \mathcal{N}} \sup_{x \in \mathcal{X}} |u_i(x) - \tilde{u}_i(x)|. \quad (4)$$

Intuitively, if this quantity is small enough, the two games are nearly indistinguishable from a strategic perspective, since for every strategy profile $x \in \mathcal{X}$, the payoffs in \mathcal{G} and $\tilde{\mathcal{G}}$ are almost the same. Thus, one might expect that at least *some* equilibria of \mathcal{G} should persist under sufficiently small perturbations, especially given that a Nash equilibrium is defined in terms of the game’s payoff functions themselves.

Perhaps surprisingly, as we show below, this definition of distance *cannot* provide a meaningful concept of equilibrium robustness.

Proposition 1. *For any game \mathcal{G} and any equilibrium $x^* \in \mathcal{X}$ of \mathcal{G} , there exists a perturbed game $\tilde{\mathcal{G}}$, arbitrarily close to \mathcal{G} in the uniform metric (4) such that $x^* \in \mathcal{X}$ is not an equilibrium of $\tilde{\mathcal{G}}$.*

To show this, we provide [Examples 3.1](#) and [3.2](#) which, taken together, cover all possible types of equilibria in continuous games in the sense of [\(VI\)](#).

Example 3.1. Let \mathcal{G} be a continuous game, and let $x^* \in \mathcal{X}$ be an equilibrium of \mathcal{G} such that $\langle v_i(x^*), p_i - x_i^* \rangle < 0$ for some player $i \in \mathcal{N}$ and $p_i \in \mathcal{X}_i$. For arbitrary $\varepsilon > 0$, define $\tilde{u}_i : \mathcal{X} \rightarrow \mathbb{R}$ as

$$\tilde{u}_i(x) := u_i(x) - \varepsilon \exp(2\varepsilon^{-1} \langle v_i(x^*), x_i - x_i^* \rangle) \quad (5)$$

which is a continuously differentiable concave function in x_i , and let $\tilde{u}_j \equiv u_j$ for all $j \neq i$, $j \in \mathcal{N}$. Since $x^* \in \mathcal{X}$ is an equilibrium of \mathcal{G} , it holds $\langle v_i(x^*), x_i - x_i^* \rangle \leq 0$ for all $x_i \in \mathcal{X}_i$, which implies that

$$\rho(\mathcal{G}, \tilde{\mathcal{G}}) = \sup_{x \in \mathcal{X}} |u_i(x) - \tilde{u}_i(x)| = \varepsilon \sup_{x_i \in \mathcal{X}_i} \exp(2\varepsilon^{-1} \langle v_i(x^*), x_i - x_i^* \rangle) = \varepsilon. \quad (6)$$

Computing the individual gradient vector of player $i \in \mathcal{N}$, we obtain

$$\tilde{v}_i(x) = v_i(x) - 2v_i(x^*) \exp(2\varepsilon^{-1} \langle v_i(x^*), x_i - x_i^* \rangle) \quad (7)$$

and, evaluating it at $x^* \in \mathcal{X}$, we get, $\tilde{v}_i(x^*) = -v_i(x^*)$. Therefore, for $x = (p_i; x_{-i}^*) \in \mathcal{X}$, we have

$$\langle \tilde{v}(x^*), x - x^* \rangle = -\langle v_i(x^*), p_i - x_i^* \rangle > 0 \quad (8)$$

i.e., $x^* \in \mathcal{X}$ is not an equilibrium point of the perturbed game $\tilde{\mathcal{G}}$. \blacklozenge

Example 3.2. Let \mathcal{G} be a continuous game, and let $x^* \in \mathcal{X}$ be an equilibrium of \mathcal{G} such that $\langle v(x^*), x - x^* \rangle = 0$ for all $x \in \mathcal{X}$. Fix a player $i \in \mathcal{N}$ and $p_i \in \mathcal{X}_i$, and let $y_i \in \mathcal{V}_i^*$ with $\langle y_i, p_i - x_i^* \rangle > 0$. For arbitrary $\varepsilon > 0$, let $\tilde{u}_i : \mathcal{X} \rightarrow \mathbb{R}$ be defined as

$$\tilde{u}_i(x) := u_i(x) + \varepsilon \operatorname{diam}(\mathcal{X}_i)^{-1} \|y_i\|_*^{-1} \langle y_i, x_i - x_i^* \rangle \quad (9)$$

which is a concave function in x_i , and $\tilde{u}_j \equiv u_j$ for all $j \neq i$, $j \in \mathcal{N}$. Then, we readily get that

$$\rho(\mathcal{G}, \tilde{\mathcal{G}}) = \sup_{x \in \mathcal{X}} |u_i(x) - \tilde{u}_i(x)| = \varepsilon \operatorname{diam}(\mathcal{X}_i)^{-1} \|y_i\|_*^{-1} \sup_{x_i \in \mathcal{X}_i} |\langle y_i, x_i - x_i^* \rangle| \leq \varepsilon. \quad (10)$$

Computing the individual gradient vector of player $i \in \mathcal{N}$, we obtain

$$\tilde{v}_i(x) = v_i(x) + \varepsilon \operatorname{diam}(\mathcal{X}_i)^{-1} \|y_i\|_*^{-1} y_i \quad (11)$$

Therefore, for $x = (p_i; x_{-i}^*) \in \mathcal{X}$, it holds by the example’s assumptions that

$$\langle \tilde{v}(x^*), x - x^* \rangle = \langle v_i(x^*), p_i - x_i^* \rangle + \varepsilon \operatorname{diam}(\mathcal{X}_i)^{-1} \|y_i\|_*^{-1} \langle y_i, p_i - x_i^* \rangle > 0 \quad (12)$$

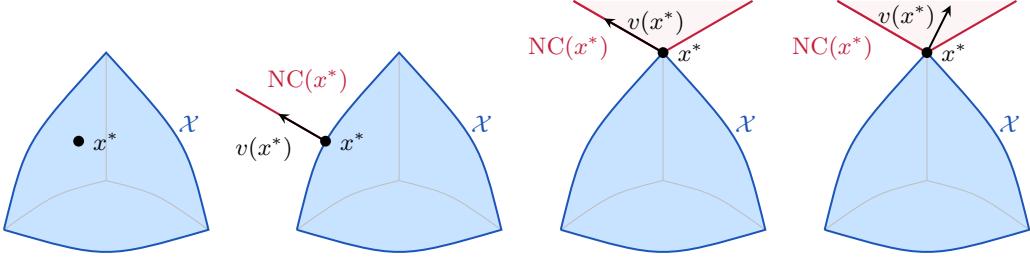


Figure 1. Different equilibrium configurations: an interior equilibrium ($v(x^*) = 0$); a boundary, non-extreme equilibrium (normal cone with empty topological interior); an extreme, non-robust equilibrium ($v(x^*)$ on the boundary of the normal cone); a robust equilibrium ($v(x^*)$ in the interior of the normal cone). Only the robust equilibrium remains invariant under strategic perturbations of the underlying game.

where we used that $\langle v_i(x^*), p_i - x_i^* \rangle = 0$ and $\langle y_i, p_i - x_i^* \rangle > 0$, as per our original assumptions. Thus, $x^* \in \mathcal{X}$ is not an equilibrium of the perturbed game $\tilde{\mathcal{G}}$. \blacklozenge

Remark 1. In Examples 3.1 and 3.2, if \mathcal{G} is concave, so is $\tilde{\mathcal{G}}$, indicating that this notion of distance is not proper even within the class of concave games.

The preceding examples demonstrate that under the distance (4), even an arbitrarily small perturbation to the payoff function of a *single* player can destroy *any* equilibrium.³ This phenomenon arises because, although an equilibrium is defined in terms of payoff functions, the first-order stationarity condition in (VI) shows that it fundamentally depends on the individual gradient vectors. Therefore, any meaningful notion of distance between two games must likewise be aware of the behavior of the individual gradient vectors.

3.2. Defining strategic robustness. As illustrated in Examples 3.1 and 3.2, small changes in the payoffs, though negligible in the uniform norm, can alter the equilibrium landscape quite significantly. To address this, we refine the notion of distance between games \mathcal{G} and $\tilde{\mathcal{G}}$ as follows:

$$\text{dist}(\mathcal{G}, \tilde{\mathcal{G}}) := \sup_{x \in \mathcal{X}} \|v(x) - \tilde{v}(x)\|_* \quad (13)$$

With this definition in hand, we are now ready to state the concept of strategic robustness in the class of continuous games.

Definition 1. An equilibrium $x^* \in \mathcal{X}$ of a game \mathcal{G} is called *strategically robust* if there exists $\varepsilon > 0$ such that for *any* game $\tilde{\mathcal{G}}$ with $\text{dist}(\mathcal{G}, \tilde{\mathcal{G}}) < \varepsilon$, x^* is also an equilibrium of $\tilde{\mathcal{G}}$.

As we explore next, this definition offers a meaningful notion of “closeness” for equilibrium stability, one that is grounded not in the payoff values themselves, but in the geometry they induce.

Geometric characterization. To provide a geometric characterization, we zoom in on the variational structure that governs Nash equilibria. Specifically, we show that strategically robust equilibria $x^* \in \mathcal{X}$ are precisely those solutions of (VI) for which the inequality is strict for all feasible deviations. Formally, we have the following characterization:

Theorem 1. Let $x^* \in \mathcal{X}$ be a joint action profile in $\mathcal{G}(\mathcal{N}, \mathcal{X}, u)$. Then the following are equivalent:

- (i) x^* is a strategically robust equilibrium.

³Such variations are not possible in the class of finite games, so, in this much more restrictive class, the sup-norm of the payoff differences is a valid metric to measure robustness.

- (ii) $\langle v(x^*), z \rangle \leq -m\|z\|$ for some $m > 0$ and all $z \in \text{TC}(x^*)$, where $\text{TC}(x^*)$ is the closure of all rays emanating from x^* and intersecting \mathcal{X} in at least one other point.
- (iii) $v(x^*) \in \text{int}(\text{PC}(x^*))$, where $\text{PC}(x^*) := \{y \in \mathcal{Y} : \langle y, z \rangle \leq 0, \text{ for all } z \in \text{TC}(x^*)\}$.

Intuitively, [Theorem 1](#) suggests that strategically robust equilibria are precisely those points $x^* \in \mathcal{X}$ where the associated gradient vector $v(x^*)$ lies in the topological interior of the polar cone $\text{PC}(x^*)$, i.e.,

$$\langle v(x^*), z \rangle < 0 \quad \text{for all } z \in \text{TC}(x^*). \quad (14)$$

We thus conclude that strategic robustness can only occur at boundary points where the tangent cone is pointed; if the feasible set is locally flat at $x^* \in \mathcal{X}$, the corresponding polar cone has empty interior, and robustness is not possible. This phenomenon is illustrated in [Fig. 1](#), and the full proof of [Theorem 1](#) is provided in [Appendix B](#).

Remark 2. Both [Examples 3.1](#) and [3.2](#) violate the condition in [Definition 1](#), but for different reasons. In the former case, although the perturbed payoffs can be made arbitrarily close to the original, the perturbed gradient vector at equilibrium can become arbitrarily large, making the distance $\text{dist}(\mathcal{G}, \tilde{\mathcal{G}})$ exceed any $\varepsilon > 0$. In the latter case, the polar cone $\text{PC}(x^*)$ at the equilibrium has empty interior, so strategic robustness cannot hold at x^* . \blacklozenge

Remark 3. Several important classes of games admit robust equilibria for all but a measure-zero set of instances. Examples include nonatomic, non-splittable routing games with arbitrary increasing cost functions [45], Markov potential games arising in multi-agent reinforcement learning [32], Cournot oligopolies in which firms have no or limited price-setting power [37], etc.

In the next section, we examine the dynamic implications of this result by studying the robustness of such equilibria under [\(FTRL\)](#).

4. FROM STRATEGIC TO DYNAMIC ROBUSTNESS: CONVERGENCE RESULTS

So far, we focused on strategic robustness, a static notion determined solely by the underlying game and the local geometry around the equilibrium in question. In this section, we shift to the dynamic perspective of our central question and explore which equilibria admit robust convergence guarantees, namely, equilibria that can emerge as stable outcomes of regularized learning under feedback and initialization uncertainty, regardless of the specific choice of regularizer.

To this end, we first establish that non-equilibrium points cannot arise as limits of [\(FTRL\)](#), even with *perfect* gradient feedback. Formally, we have the following proposition, whose proof is provided in [Appendix C](#).

Proposition 2. *Suppose that [\(FTRL\)](#) is run with perfect gradient feedback of the form $\hat{v}_t = v(x_t)$ for all $t = 1, 2, \dots$, and assume that x_t converges to some $\hat{x} \in \mathcal{X}$. Then \hat{x} is an equilibrium of \mathcal{G} .*

Having excluded non-equilibrium points as positive probability outcomes of a learning process, we now turn to identifying equilibria that are robust from a dynamic standpoint, and more precisely, under that of [\(FTRL\)](#). In this regard, strategically robust equilibria serve as natural candidates, as their stability with respect to game perturbations suggests they may also admit robust convergence guarantees. This is further supported by the finding that equilibrium points in the interior of the strategy space \mathcal{X} cannot be limit points: in particular, we show below that, even under i.i.d. stochastic noise, the iterates of [\(FTRL\)](#) diverge from such equilibria almost surely.

Proposition 3. *Let $x^* \in \text{ri}(\mathcal{X})$ be a Nash equilibrium of $\mathcal{G}(\mathcal{N}, \mathcal{X}, u)$, and $(x_t)_{t \in \mathbb{N}}$ be the sequence of play induced by **(FTRL)** with $\hat{v}_t = v(x_t) + U_t$, where U_t i.i.d. with $\mathbb{E}[U_t] = 0$ and $\text{cov}(U_t) \succ 0$ for all $t \in \mathbb{N}$. Then:*

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} x_t = x^*\right) = 0 \quad \text{for any } x_1 \in \mathcal{X}_h. \quad (15)$$

Remark 4. The condition $\text{cov}(U_t) \succ 0$ is not necessary. In fact, it suffices to have $\text{cov}(U_t)$ non-degenerate in a direction $p - x^*$ for $p \in \mathcal{X}$, but we state our result under stronger assumptions for simplicity. \blacklozenge

The key idea of the proof, which is deferred to [Appendix C](#), is that, since $x^* \in \text{ri}(\mathcal{X})$, we have $\langle v(x^*), x - x^* \rangle = 0$ for all $x \in \mathcal{X}$. At the same time, as $\text{cov}(U_t) \succ 0$, the quantity $\langle U_t, x - x^* \rangle$ fluctuates and remains bounded away from zero infinitely often, thereby preventing convergence.

4.1. Learning with gradient-based feedback. In view of the impossibility result of [Proposition 3](#), we shift our focus on the convergence of **(FTRL)** toward strategically robust equilibria. We first consider the gradient feedback model, where each player receives an unbiased estimate of their individual gradient vector via **(SFO)**. Specifically, we analyze the behavior of **(FTRL)** and we establish local convergence guarantees toward strategically robust equilibria with high probability. This is encoded in the following theorem:

Theorem 2. *Let $x^* \in \mathcal{X}$ be a strategically robust equilibrium of $\mathcal{G}(\mathcal{N}, \mathcal{X}, u)$. Fix a confidence level $\delta > 0$, and let $(x_t)_{t \in \mathbb{N}}$ be the iterates of **(FTRL)** with feedback provided by **(SFO)**, and step-size $\gamma > 0$ sufficiently small. Then, there exists a neighborhood \mathcal{U} of x^* in \mathcal{X}_h such that:*

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} x_t = x^*\right) \geq 1 - \delta \quad \text{if } x_1 \in \mathcal{U}. \quad (16)$$

Before proceeding, a few remarks are in order. Since continuous games may admit multiple Nash equilibria, global convergence guarantees are in general unattainable. As such, our analysis focuses on the local convergence landscape of **(FTRL)**. As a sidenote, it is important to emphasize that the convergence result is robust to the choice of regularizer, relying solely on the general conditions outlined in [Section 2](#) rather than any particular functional form. We outline below the main steps of the proof, with full details provided in [Appendix C](#).

Proof Sketch. The key idea is that the auxiliary process y_t , which aggregates the players' gradient updates, diverges to infinity in a direction that steers the induced sequence $x_t = Q(y_t)$ toward the equilibrium in question. More formally, the proof relies on the following intuition. From [Theorem 1](#) and the continuity of the players' payoffs, strategic robustness implies that, in a neighborhood of a robust equilibrium x^* , the players' individual gradient fields point toward x^* . Consequently, the process y_t accumulates gradient steps that, on average, are aligned with the interior of the normal cone $\text{NC}(x^*)$ to the action space \mathcal{X} at x^* . As a result, y_t exhibits a consistent drift that carries it deeper into a "copy" of the normal cone $\text{NC}(x^*)$ embedded in the gradient space \mathcal{Y} . Moreover, we show that, with high probability, once y_t enters this region, it remains there, provided that the algorithm is not initialized too far from x^* . Combining the above, [Proposition C.1](#) establishes that the sequence of actions $x_t = Q(y_t)$ generated by **(FTRL)** converges to x^* . \blacklozenge

While our main focus lies on the qualitative convergence behavior of **(FTRL)**, stronger guarantees can be obtained under additional structural assumptions on the strategy space and the regularizer. In particular, suppose that \mathcal{X} is a polyhedral domain of the form $\mathcal{X} := \{x \in \mathbb{R}_+^d \mid Ax = b\}$ for some $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$, and h is decomposable with kernel function θ , i.e., h can be written as $h(x) = \sum_{j=1}^d \theta(x_j)$ for some continuous function $\theta : \mathbb{R}_+ \rightarrow \mathbb{R}$ with locally Lipschitz θ'' and $\theta'' > 0$. Under these conditions, we obtain the explicit convergence rates for the **(FTRL)** dynamics, as follows.

Theorem 3. *If, in addition, \mathcal{X} is a polyhedral domain and h is decomposable with kernel θ , on the event $E := \{\lim_{t \rightarrow \infty} x_t = x^*\}$ it holds:*

$$\|x_t - x^*\| = \phi(-\Theta(t)) \quad (17)$$

where ϕ is the rate function defined via

$$\phi(z) := \begin{cases} (\theta')^{-1}(z) & \text{if } z > \theta'(0^+) \\ 0 & \text{if } z \leq \theta'(0^+) \end{cases} \quad (18)$$

Remark 5. For the setting of [Example 2.2](#), with $\mathcal{X} = \Delta(\mathcal{A})$, $\theta(z) = z \log z$ and x^* a strict Nash equilibrium, the convergence rate of [\(FTRL\)](#) as per [Theorem 3](#), becomes $\|x_t - x^*\| = \exp(-\Theta(t))$. \blacklozenge

Remark 6. For finite games, [\[19\]](#) showed that under a step-size schedule of the form $\gamma_t \propto 1/t^p$, the Robins-Monro summability conditions require $p \in (1/2, 1]$, leading to convergence rates from $\phi(-\Theta(t^{1-p}))$ to $\phi(-\Theta(\log t))$. Our convergence guarantees remain valid under these step-size schedules, though they yield the slower aforementioned rates. \blacklozenge

Remark 7. It is important to highlight the different behavior of [\(FTRL\)](#), often referred to as a “lazy” variant of mirror descent [\[22\]](#), with that of the mirror descent algorithm, defined via the update

$$x_{t+1} = Q(\nabla h(x_t) + \gamma \hat{v}_t) \quad \text{for } t = 1, 2, \dots \quad (\text{MD})$$

where $\nabla h(x)$ denotes a continuous selection of $\partial h(x)$ [\[12\]](#). To illustrate the difference, consider the single-agent problem of maximizing $u(x) = x$ over $\mathcal{X} = [0, 1]$, where $x^* = 1$ is a robust equilibrium. Using the Euclidean regularizer (see [Example 2.1](#)), [\(MD\)](#) reduces to the projected gradient algorithm

$$x_{t+1} = \Pi(x_t + \gamma \hat{v}_t) \quad (\text{SGA})$$

where \hat{v}_t is a stochastic gradient of u at x_t , i.e., $\hat{v}_t = 1 + U_t$ where U_t is a Bernoulli process with $U_t = \pm 1$ with probability $1/2$. However, even if $x_t = x^*$ for some t , we then have that $x_{t+1} = 1 - \gamma$ with probability $1/2$. Thus, by a straightforward application of the Borel-Cantelli lemma, we conclude that, with probability 1, [\(SGA\)](#) does not converge to x^* . Through this toy example, note that although [\(MD\)](#) and [\(FTRL\)](#) use the same mirror map Q to select actions, they differ fundamentally in how feedback is processed. The “eager” nature of [\(MD\)](#) makes it more sensitive to noise, whereas [\(FTRL\)](#) maintains a cumulative dual variable y_t that aggregates all past feedback, effectively smoothing out fluctuations over time. Also, note that for steep regularizers, the iterations [\(MD\)](#) and [\(FTRL\)](#) coincide, as the mirror map Q is essentially injective (see [Appendix A](#) for more details). Therefore, differences in their behavior arise in the case where steepness does not hold. \blacklozenge

4.2. Learning with payoff-based feedback. We now turn to the payoff-based feedback model, where players observe only their realized payoffs and use them to estimate gradients indirectly via [\(SPSA\)](#). This feedback model introduces higher variance and structural bias due to the diminishing sampling radius and the feasibility corrections. Nevertheless, we show that strategically robust equilibria retain their dynamic robustness: they still locally attract the [\(FTRL\)](#) dynamics with high probability, as the following theorem suggests.

Theorem 4. *Let $x^* \in \mathcal{X}$ be a strategically robust equilibrium of \mathcal{G} . Fix a confidence level $\delta > 0$, and let $(x_t)_{t \in \mathbb{N}}$ be the iterates of [\(FTRL\)](#) run with [\(SPSA\)](#) with $\delta_t \propto 1/t^p$ for some $p \in (0, 1/2)$ and step-size $\gamma > 0$ sufficiently small. Then, there exists a neighborhood \mathcal{U} of x^* such that:*

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} x_t = x^*\right) \geq 1 - \delta \quad \text{if } x_1 \in \mathcal{U}. \quad (19)$$

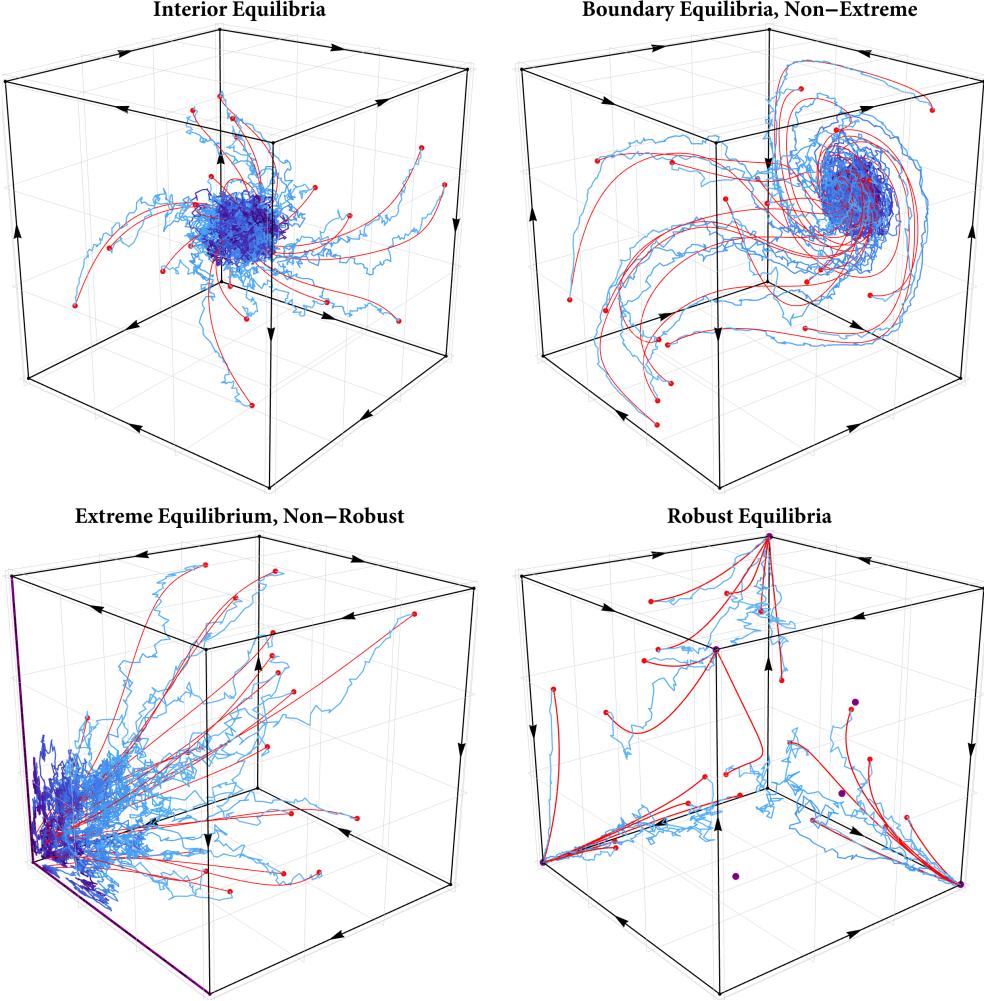


Figure 2. Convergence and non-convergence to different type of equilibria. Only robust equilibria are stochastically asymptotically stable under (FTRL).

If, in addition, \mathcal{X} is affinely constrained and h is decomposable with kernel θ , then, whenever x_t converges to x^* , we have:

$$\|x_t - x^*\| = \phi(-\Theta(t)). \quad (20)$$

Despite the scarcity of information inherent in the payoff-based feedback model, strategically robust equilibria retain not only their convergence properties but also their convergence speed, under the additional structural assumptions on the regularizer and domain, matching that of the (SFO) feedback setting. This is further discussed along with the proof of the theorem in [Appendix C](#).

4.3. Convergence landscape beyond strategic robustness. Having established the robust convergence properties of strategically robust equilibria, a natural question arises: *Can we expect robust convergence guarantees toward equilibria that lack this structural property?* As we show below, the answer is not encouraging: strategic robustness is essentially necessary for robust convergence.

To make this limitation precise, we move beyond the interior of the strategy space, where [Proposition 3](#) rules out equilibria as potential limit points, and shift our focus to non-robust equilibria on the boundary. To illustrate the behavior of [\(FTRL\)](#) in this setting, we construct a game with a unique equilibrium that exhibits fundamentally different long-run behavior depending on the regularizer.

Proposition 4. *Consider the 1-player game \mathcal{G} with $\mathcal{X} = [0, 1]$, $u(x) = -\frac{3}{4}x^{4/3}$ and $x^* = 0$. Let $(x_t)_{t \in \mathbb{N}}$ be the iterates of [\(FTRL\)](#) with $\gamma < 1$, and $\hat{v}_t = v(x_t) + U_t$, where U_t are i.i.d. standard normal random variables for all $t \in \mathbb{N}$. Then, for any initial condition $y_1 \in \mathbb{R}$, we have:*

- (i) *For $h(x) = x \log x$, it holds $\mathbb{P}(\lim_{t \rightarrow \infty} x_t = x^*) = 0$.*
- (ii) *For $h(x) = -2\sqrt{x}$, it holds $\mathbb{P}(\lim_{t \rightarrow \infty} x_t = x^*) = 1$.*

The core idea of the proof of [Proposition 4](#) (which we present in detail in [Appendix C](#)) is to construct a process z_t that dominates y_t . Importantly, the process z_t can be then viewed as a random walk with a diminishing drift whose rate of decay depends on the choice of regularizer. Depending on the magnitude of this drift, the process exhibits two sharply contrasting long-term behaviors: if the drift decays sufficiently fast, the process behaves like a zero-mean random walk and returns infinitely often with probability 1 (recurrence); conversely, if the drift diminishes at a slower rate, the process behaves like a random walk with constant drift and escapes to infinity with probability 1 (transience).

In view of the above, we conclude that strategic robustness cannot be relaxed without compromising convergence guarantees, even when the equilibrium lies on the boundary of the game's action space. For a graphical illustration, cf. [Fig. 2](#).

5. CONCLUDING REMARKS

Our aim in this paper was to examine the robustness of Nash equilibria in continuous games, under both strategic and dynamic uncertainty. From a strategic standpoint, the notion of *strategic robustness* characterizes those (local) equilibria which remain invariant under small perturbations of the underlying game, and we derived a tight geometric characterization thereof in terms of the variational geometry of the game. From a dynamic standpoint, we focused on the stability of regularized learning under uncertainty, and we established a deep structural connection between the two notions. Strategic robustness guarantees dynamic robustness under [\(FTRL\)](#), and this implication is essentially tight: without strategic robustness, dynamic robustness cannot be ensured. To the best of our knowledge, this is the first study of its kind for continuous games, and we believe that this two-way implication elucidates the delicate interplay between static and dynamic considerations.

APPENDIX A. AUXILIARY RESULTS

As a preamble to our analysis, we provide some basic properties of the regularizers and the mirror maps, and present some auxiliary results from martingale theory and Markov processes that we will use throughout the sequel.

A.1. Mirror maps and results from convex analysis. In this section, we provide a more detailed discussion of key notions from convex analysis, including mirror maps and regularizers.

To begin, let $(\mathcal{V}, \|\cdot\|)$ be a finite-dimensional normed vector space. Its dual space is denoted by $(\mathcal{V}^*, \|\cdot\|_*)$, where the dual norm is defined as

$$\|y\|_* \equiv \max\{\langle y, x \rangle : \|x\| \leq 1\}, \quad (\text{A.1})$$

and $\langle y, x \rangle$ denotes the canonical pairing between $y \in \mathcal{V}^*$ and $x \in \mathcal{V}$. To maintain consistency with the notation used throughout the paper, we will refer to \mathcal{V}^* as \mathcal{Y} from this point onward.

Given a closed convex set $\mathcal{X} \subseteq \mathcal{V}$ and a point $p \in \mathcal{X}$, we define the tangent cone $\text{TC}(p)$ and the polar cone $\text{PC}(p)$ as follows:

$$\text{TC}(p) = \text{cl}\{z \in \mathcal{V} : p + tz \in \mathcal{X} \text{ for some } t > 0\} \quad (\text{A.2})$$

and

$$\text{PC}(p) = \{y \in \mathcal{Y} : \langle y, z \rangle \leq 0, \text{ for all } z \in \text{TC}(p)\} \quad (\text{A.3})$$

For a strongly convex regularizer $h : \mathcal{X} \rightarrow \mathbb{R}$, the *subdifferential* of h at $x \in \mathcal{X}$ is defined as

$$\partial h(x) := \{y \in \mathcal{Y} : h(x') \geq h(x) + \langle y, x' - x \rangle \text{ for all } x' \in \mathcal{X}\} \quad (\text{A.4})$$

and we denote the *domain of subdifferentiability* of h as

$$\mathcal{X}_h = \{x \in \mathcal{X} : \partial h(x) \neq \emptyset\}. \quad (\text{A.5})$$

In addition, the mirror map Q , defined via

$$Q(y) = \arg \max_{x \in \mathcal{X}} \{\langle y, x \rangle - h(x)\} \quad (\text{A.6})$$

is single-valued on \mathcal{Y} , since the maximization problem admits a unique solution, as h is strongly convex. Finally, by the optimality conditions of (A.6), we get that

$$x = Q(y) \quad \text{if and only if} \quad y \in \partial h(x). \quad (\text{A.7})$$

since $0 \in y - \partial h(x)$. This readily implies that $\mathcal{X}_h = \text{im } Q$. In general, we have

$$\text{ri}(\mathcal{X}) \subseteq \mathcal{X}_h \subseteq \mathcal{X}, \quad (\text{A.8})$$

where the first inclusion follows from standard results on the subdifferentiability of convex functions [42, Chap. 26], whereas the second is immediate from the definition of \mathcal{X}_h . This leads to two contrasting regimes: (i) $\mathcal{X}_h = \text{ri}(\mathcal{X})$, in which case h is called *steep*; and (ii) $\mathcal{X}_h = \mathcal{X}$, in which case h is called *non-steep*.

Finally, we include here for future reference an elementary result concerning solid (convex) cones.

Lemma A.1. *Let \mathcal{K} be a convex cone in \mathbb{R}^d with nonempty topological interior, and let $z \in \text{int}(\mathcal{K})$. Then there exists a finitely generated cone \mathcal{K}' such that $z \in \text{int } \mathcal{K}' \subseteq \text{int } \mathcal{K}$.*

Remark. We stress here that, by $\text{int } \mathcal{K}$ we mean the topological interior of \mathcal{K} (which is nonempty by assumption), not the relative interior $\text{ri } \mathcal{K}$ thereof (whis is always nonempty).

Proof. Since \mathcal{K} is closed and $z \in \text{int } \mathcal{K}$, there exists a closed ball \mathcal{B} centered at z , which is entirely contained in $\text{int } \mathcal{K}$ (an immediate consequence of the fact that z is well-separated from the boundary $\text{bd } \mathcal{K}$ of \mathcal{K}). Since \mathcal{B} is not contained in any lower-dimensional subspace of \mathbb{R}^d , it is possible to find inductively d linearly independent vectors $z_1, \dots, z_d \in \mathcal{B}$ on the boundary $\text{bd } \mathcal{B}$ of \mathcal{B} such that z is contained in the convex hull $\Delta(z_1, \dots, z_d)$ (and, in particular, in the relative interior thereof). Thus, letting $\mathcal{K}' \equiv \mathcal{K}(z_1, \dots, z_d)$ be the polyhedral cone generated by z_1, \dots, z_d , we have $\mathcal{K}' \subseteq \mathcal{K}$ and $z \in \text{int } \mathcal{K}'$ by construction, and our proof is complete. \blacksquare

A.2. Statistical bounds and results from probability theory. In this section, we provide some basic statistical bounds for (SPSA), and we present some results that we will use freely in the sequel. We start with our bounds for (SPSA), specifically:

Proposition A.1. *The estimator (SPSA) enjoys the following bounds:*

$$\|\mathbb{E}[\hat{v}_t | \mathcal{F}_t] - v(x_t)\|_* = \mathcal{O}(\delta_t) \quad \text{and} \quad \|\hat{v}_t\|_* = \mathcal{O}(1/\delta_t). \quad (\text{A.9})$$

Proof. Letting $\zeta_t := \delta_t(w_t + (p - x_t)/r)$, we write $\hat{x}_t = x_t + \zeta_t$, and we have for player $i \in \mathcal{N}$:

$$u_i(\hat{x}_t)w_{i,t} = u_i(x_t)w_{i,t} + \langle \nabla u_i(x_t), \zeta_t \rangle w_{i,t} + \int_0^1 \langle \nabla u_i(x_t + \tau\zeta_t) - \nabla u_i(x_t), \zeta_t \rangle d\tau w_{i,t} \quad (\text{A.10})$$

Now, the middle term can be unfolded as

$$\langle \nabla u_i(x_t), \zeta_t \rangle w_{i,t} = \langle \nabla_i u_i(x_t), \zeta_{i,t} \rangle w_{i,t} + \sum_{j \neq i} \langle \nabla_j u_i(x_t), \zeta_{j,t} \rangle w_{i,t} \quad (\text{A.11})$$

and, noting that $\mathbb{E}[w_{i,t} | \mathcal{F}_t] = 0$, we take conditional expectation, and we get:

$$\mathbb{E}[\langle \nabla u_i(x_t), \zeta_t \rangle w_{i,t} | \mathcal{F}_t] = \delta_t \mathbb{E}[\langle \nabla_i u_i(x_t), w_{i,t} \rangle w_{i,t} | \mathcal{F}_t] = (\delta_t/d_i) v_i(x_t) \quad (\text{A.12})$$

and

$$\mathbb{E}[u_i(x_t)w_{i,t} | \mathcal{F}_t] = 0 \quad (\text{A.13})$$

Therefore, we have:

$$\|\mathbb{E}[\hat{v}_{i,t} | \mathcal{F}_t] - v_i(x_t)\| = \left\| \mathbb{E} \left[\int_0^1 \langle \nabla u_i(x_t + \tau\zeta_t) - \nabla u_i(x_t), \zeta_t \rangle d\tau w_{i,t} \mid \mathcal{F}_t \right] \right\| = \mathcal{O}(\delta_t) \quad (\text{A.14})$$

Now, for the second bound, since u_i is continuous on a compact domain, it is bounded, and we readily get that:

$$\|\hat{v}_{i,t}\|_* = \mathcal{O}(1/\delta_t) \quad (\text{A.15})$$

as was to be shown. \blacksquare

Moving forward, we provide some useful results from probability theory. The first two statements below are adapted from the classical textbook of Hall & Heyde [21], while the third one is a simplified version of [30, Theorem 3.2] on the recurrence of a nonnegative Markov process with diminishing drift. Namely, we have:

Theorem A.1. (Doob's maximal inequality, [21, Corollary 2.1]) *If S_t is a martingale, we have:*

$$\mathbb{P} \left(\sup_{\tau \leq t} |S_\tau| > t \right) \leq \frac{\mathbb{E}[|S_t|]}{t} \quad \text{for all } t > 0. \quad (\text{A.16})$$

Theorem A.2. (Burkholder's inequality, [21, Theorem 2.10]) *Let $S_t := \sum_{\tau=1}^t D_\tau$, where $(D_\tau)_{\tau \in \mathbb{N}}$ is a martingale difference sequence, and let $q \in (1, \infty)$. Then, there exists a constant C that depends only on q such that:*

$$\mathbb{E}[|S_t|^q] \leq C \mathbb{E} \left[\left| \sum_{\tau=1}^t D_\tau^2 \right|^{q/2} \right] \quad (\text{A.17})$$

Theorem A.3. (Lamperti [30, Theorem 3.2]) *Let the non-negative stochastic process $(x_t)_{t \in \mathbb{N}}$ be defined as*

$$x_{t+1} = (x_t + f(x_t) + \xi_t)^+ \quad (\text{A.18})$$

for some $x \mapsto f(x)$ bounded measurable function, and ξ_t i.i.d. with $\mathbb{E}[\xi_t] = 0$, $\mathbb{V}(\xi_t) = \sigma^2 \neq 0$ and finite $2 + \varepsilon$ moment for some $\varepsilon > 0$. Then:

(i) if $f(x) \leq \sigma^2/2x$ for all x large enough, the process is recurrent in the sense there exists $c < \infty$ such that

$$\mathbb{P} \left(\liminf_{t \rightarrow \infty} x_t \leq c \right) = 1 \quad (\text{A.19})$$

(ii) if $f(x) \geq \theta\sigma^2/2x$ for some $\theta > 1$ and all x large enough, the process is transient in the sense that

$$\mathbb{P} \left(\lim_{t \rightarrow \infty} x_t = \infty \right) = 1 \quad (\text{A.20})$$

APPENDIX B. ANALYSIS AND RESULTS FOR STRATEGIC ROBUSTNESS

Our aim in this appendix is to provide a detailed proof for [Theorem 1](#), which we restate below for convenience.

Theorem 1. *Let $x^* \in \mathcal{X}$ be a joint action profile in $\mathcal{G}(\mathcal{N}, \mathcal{X}, u)$. Then the following are equivalent:*

- (i) x^* is a strategically robust equilibrium.
- (ii) $\langle v(x^*), z \rangle \leq -m\|z\|$ for some $m > 0$ and all $z \in \text{TC}(x^*)$, where $\text{TC}(x^*)$ is the closure of all rays emanating from x^* and intersecting \mathcal{X} in at least one other point.
- (iii) $v(x^*) \in \text{int}(\text{PC}(x^*))$, where $\text{PC}(x^*) := \{y \in \mathcal{Y} : \langle y, z \rangle \leq 0, \text{ for all } z \in \text{TC}(x^*)\}$.

Proof. We will go full-circle by showing (i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i).

(i) \Rightarrow (ii). Suppose that $x^* \in \mathcal{X}$ is a strategically robust equilibrium, and let $\varepsilon > 0$ be such that x^* is an equilibrium of any $\tilde{\mathcal{G}}$ with $\text{dist}(\mathcal{G}, \tilde{\mathcal{G}}) \leq \varepsilon$.

For the sake of contradiction, suppose that there exists $z \neq 0, z \in \text{TC}(x^*)$ such that

$$\langle v(x^*), z \rangle = 0. \quad (\text{B.1})$$

which readily implies that there exists player $i \in \mathcal{N}$ and $z_i \neq 0, z_i \in \text{TC}_i(x^*)$, such that

$$\langle v_i(x^*), z_i \rangle = 0. \quad (\text{B.2})$$

Fix some $y_i \in \mathcal{Y}_i$ such that $\langle y_i, z_i \rangle > 0$, and let $y \equiv (y_1, \dots, y_N) \in \mathcal{Y}$ with $y_j \equiv 0$ for $j \neq i$, $j \in \mathcal{N}$. Using (B.1) and the definition of y , we get that $\langle v(x^*) + \varepsilon\|y\|_*^{-1}y, z \rangle > 0$, and therefore, there exists $p \in \mathcal{X}$ such that:

$$\langle v(x^*) + \varepsilon\|y\|_*^{-1}y, p - x^* \rangle > 0 \quad (\text{B.3})$$

Now, define the game $\tilde{\mathcal{G}}$ with payoff functions

$$\tilde{u}_i(x) := u_i(x) + \varepsilon\|y\|_*^{-1}\langle y_i, x_i - x^* \rangle \quad (\text{B.4})$$

and $\tilde{u}_j \equiv u_j$ for all $j \in \mathcal{N}, j \neq i$. Then, the individual gradient vector of player $i \in \mathcal{N}$ is given by

$$\tilde{v}_i(x) = v_i(x) + \varepsilon\|y\|_*^{-1}y_i \quad (\text{B.5})$$

and the distance between \mathcal{G} and $\tilde{\mathcal{G}}$ is equal to

$$\text{dist}(\mathcal{G}, \tilde{\mathcal{G}}) = \sup_{x \in \mathcal{X}} \|v(x) - \tilde{v}(x)\|_* = \varepsilon\|y\|_*^{-1}\|y\|_* = \varepsilon. \quad (\text{B.6})$$

Finally, we conclude that $x^* \in \mathcal{X}$ is not an equilibrium of $\tilde{\mathcal{G}}$, since for $p \in \mathcal{X}$ as above, we have

$$\langle \tilde{v}(x^*), p - x^* \rangle = \langle v(x^*) + \varepsilon\|y\|_*^{-1}y, p - x^* \rangle > 0 \quad (\text{B.7})$$

where the last inequality holds by (B.3). Thus, we arrive at a contradiction, i.e., $\langle v(x^*), z \rangle < 0$ for all $z \neq 0, z \in \text{TC}(x^*)$. Finally, since $\{z \in \mathcal{Y} : z \in \text{TC}(x^*), \|z\| = 1\}$ is compact, we readily obtain

$$\sup\{\langle v(x^*), z \rangle : z \in \text{TC}(x^*), \|z\| = 1\} \leq -m \quad (\text{B.8})$$

for some $m > 0$. Therefore, for all $z \in \text{TC}(x^*)$, we have:

$$\langle v(x^*), z \rangle \leq -m\|z\| \quad (\text{B.9})$$

as was to be shown.

(ii) \implies (iii). First, note that the $\|\cdot\|_*$ -ball of radius $\varepsilon > 0$ centered at $v(x^*)$ can be written as:

$$\mathbb{B}_\varepsilon(v(x^*)) = v(x^*) + \varepsilon \mathbb{B}_1(0) \quad (\text{B.10})$$

where $\mathbb{B}_\varepsilon(y) := \{y' \in \mathcal{Y} : \|y' - y\|_* \leq r\}$ for $y \in \mathcal{Y}$. Now, take any $y \in \mathbb{B}_1(0)$ and $z \in \text{TC}(x^*)$. Then, for $\varepsilon > 0$ we have

$$\begin{aligned} \langle v(x^*) + \varepsilon y, z \rangle &= \langle v(x^*), z \rangle + \varepsilon \langle y, z \rangle \\ &\leq -m\|z\| + \varepsilon \|y\|_* \|z\| \\ &\leq -(m - \varepsilon)\|z\| \end{aligned} \quad (\text{B.11})$$

Setting $\varepsilon = m/2$, we have for all $z \in \text{TC}(x^*)$

$$\langle v(x^*) + (m/2)y, z \rangle < -(m/2)\|z\| \quad (\text{B.12})$$

which implies that $v(x^*) + (m/2)y \in \text{PC}(x^*)$. Thus, we readily get that $\mathbb{B}_{m/2}(v(x^*)) \subseteq \text{PC}(x^*)$, i.e., $v(x^*) \in \text{int}(\text{PC}(x^*))$.

(iii) \implies (i). Suppose that $v(x^*) \in \text{int}(\text{PC}(x^*))$. First, it directly implies that $\langle v(x^*), z \rangle \leq 0$, for all $z \in \text{TC}(x^*)$, i.e., x^* is an equilibrium of \mathcal{G} . In addition, there exists $\varepsilon > 0$ such that $\mathbb{B}_\varepsilon(v(x^*)) \subseteq \text{PC}(x^*)$. Therefore, for any game $\tilde{\mathcal{G}}$ with $\text{dist}(\mathcal{G}, \tilde{\mathcal{G}}) < \varepsilon$, we immediately get that $\tilde{v}(x^*) \in \text{PC}(x^*)$, which implies that $x^* \in \mathcal{X}$ is an equilibrium of $\tilde{\mathcal{G}}$. Thus, x^* is strategically robust, and our proof is complete. \blacksquare

APPENDIX C. ANALYSIS AND RESULTS FOR DYNAMIC ROBUSTNESS

In this appendix, we provide detailed proofs of the statements presented in Section 4, along with several intermediate results that will serve as key building blocks.

C.1. Intermediate results. We begin this section with two results establishing sufficient conditions for convergence, followed by a high-probability deviation bound for martingales. We conclude with a variant of Farkas' Lemma, which will be instrumental in deriving convergence rates.

Proposition C.1. *Let $x^* \in \mathcal{X}$ and $\mathcal{Z} := \{z_1, \dots, z_m\} \subseteq \mathcal{V}$ be a set of unit vectors, such that any $z \in \text{TC}(x^*)$ can be written as $z = \sum_{j=1}^m \lambda_j z_j$ for some $\lambda_j \geq 0$. If $\lim_{t \rightarrow \infty} \langle y_t, z_j \rangle = -\infty$ for all $z_j \in \mathcal{Z}$, then $\lim_{t \rightarrow \infty} Q(y_t) = x^*$.*

Proof. Denote $Q(y_t)$ by x_t , and suppose that $\limsup_{t \rightarrow \infty} \|x_t - x^*\| > 0$. Then, there exists a subsequence $(x_{t_\tau})_{\tau \in \mathbb{N}}$ such that $\|x_{t_\tau} - x^*\|$ stays bounded away from zero, i.e., $\|x_{t_\tau} - x^*\| \geq c$ for some $c > 0$ and all $\tau \in \mathbb{N}$. Since $y_{t_\tau} \in \partial h(x_{t_\tau})$, we readily get for $z_{t_\tau} = (x_{t_\tau} - x^*)/\|x_{t_\tau} - x^*\|$:

$$\begin{aligned} h(x^*) &\geq h(x_{t_\tau}) + \langle y_{t_\tau}, x^* - x_{t_\tau} \rangle \\ &= h(x_{t_\tau}) - \langle y_{t_\tau}, z_{t_\tau} \rangle \|x_{t_\tau} - x^*\| \\ &\geq \min h - \langle y_{t_\tau}, z_{t_\tau} \rangle \|x_{t_\tau} - x^*\|. \end{aligned} \quad (\text{C.1})$$

Now, we have $z_{t_k} \in \text{TC}(x^*)$, and by assumption, $z_{t_\tau} = \sum_{j=1}^m \lambda_{j,\tau} z_j$ for some coefficients $\lambda_{j,\tau} \geq 0$. Therefore, the above inequality can be written as:

$$\begin{aligned} h(x^*) &\geq \min h - \|x_{t_\tau} - x^*\| \sum_{j=1}^m \lambda_{j,\tau} \langle y_{t_\tau}, z_j \rangle \\ &\geq \min h - \left(\max_{j'} \langle y_{t_\tau}, z_{j'} \rangle \right) \|x_{t_\tau} - x^*\| \sum_{j=1}^m \lambda_{j,\tau} \end{aligned} \quad (\text{C.2})$$

Now, note that by the definition of z_{t_τ} , we have $\|z_{t_\tau}\| = 1$, and, thus:

$$1 = \|z_{t_\tau}\| = \left\| \sum_{j=1}^m \lambda_{j,\tau} z_j \right\| \leq \sum_{j=1}^m \lambda_{j,\tau} \|z_j\| = \sum_{j=1}^m \lambda_{j,\tau} \quad (\text{C.3})$$

where we used that $\|z_j\| = 1$ for all j . Now, since $\lim_{t \rightarrow \infty} \langle y_t, z_j \rangle = -\infty$ for all $z_j \in \mathcal{Z}$, it readily implies that

$$\lim_{t \rightarrow \infty} \max_{j'} \langle y_t, z_{j'} \rangle = -\infty \quad (\text{C.4})$$

Therefore, for all τ large enough, we have $-\max_{j'} \langle y_{t_\tau}, z_{j'} \rangle > 0$, and, using that $\sum_{j=1}^m \lambda_{j,\tau} \geq 1$ and $\|x_{t_\tau} - x^*\| \geq c$ for all $\tau \in \mathbb{N}$, we obtain:

$$\begin{aligned} h(x^*) &\geq \min h + \left(-\max_{j'} \langle y_{t_\tau}, z_{j'} \rangle \right) \|x_{t_\tau} - x^*\| \sum_{j=1}^m \lambda_{j,\tau} \\ &\geq \min h - c \max_{j'} \langle y_{t_\tau}, z_{j'} \rangle \end{aligned} \quad (\text{C.5})$$

Finally, letting $\tau \rightarrow \infty$, we get that $h(x^*) \geq \infty$, which is a contradiction. Thus, the result follows. \blacksquare

Finally, using the above proposition, we establish the following corollary.

Corollary C.1. *Let $\mathcal{W}(M) := \{y \in \mathcal{Y} : \max_{z \in \mathcal{Z}} \langle y, z \rangle < -M\}$ for $M > 0$. Then, for any $\varepsilon > 0$, there exists $M_\varepsilon > 0$ such that for all $y \in \mathcal{W}(M_\varepsilon)$ it holds $\|x^* - Q(y)\| < \varepsilon$.*

Proof. Suppose it does not hold. Then, there exists $\varepsilon > 0$ such that for any $t \in \mathbb{N}$, one can find $y_t \in \mathcal{Y}$ such that $\max_{z \in \mathcal{Z}} \langle y_t, z \rangle < -t$ and $\|x^* - Q(y_t)\| \geq \varepsilon$. Taking $t \rightarrow \infty$ leads to a contradiction with [Proposition C.1](#). \blacksquare

Lemma C.1. *Let $S_t := \gamma \sum_{\tau=1}^t \xi_\tau$ be a martingale with respect to a filtration $(\mathcal{F}_t)_{t \in \mathbb{N}}$ such that $\mathbb{E}[|\xi_t|^q] \leq \sigma_t^q$ for all $t \in \mathbb{N}$ and some $q \geq 2$. Then, for any $\mu \in (0, 1)$ and $c > 0$, it holds:*

$$\mathbb{P}\left(\sup_{s \leq t} |S_s| > c(\gamma t)^\mu\right) \leq C_q \frac{\gamma^{q(1-\mu)} \sum_{\tau=1}^t \sigma_\tau^q}{t^{1+q(\mu-1/2)}} \quad (\text{C.6})$$

where C_q is a constant that depends only on c and q .

Proof. To bound the maximum absolute deviation of S_t , we apply Doob's maximal inequality (see [Theorem A.1](#)), and obtain:

$$\mathbb{P}\left(\sup_{s \leq t} |S_s| > c(\gamma t)^\mu\right) \leq \frac{\mathbb{E}[|S_t|^q]}{c^q (\gamma t)^{q\mu}} \quad (\text{C.7})$$

Now, we invoke Burkholder's inequality (see [Theorem A.2](#)), from which we get:

$$\mathbb{E}[|S_t|^q] \leq C'_q \mathbb{E}\left[\left(\sum_{\tau=1}^t \gamma^2 |\xi_\tau|^2\right)^{q/2}\right] \leq C'_q \gamma^q \mathbb{E}\left[\left(\sum_{\tau=1}^t |\xi_\tau|^2\right)^{q/2}\right] \quad (\text{C.8})$$

where C_q is a constant that depends only on q . Since $q \geq 2$, applying Jensen's inequality, we obtain:

$$\left(\frac{1}{t} \sum_{\tau=1}^t |\xi_\tau|^2\right)^{q/2} \leq \frac{1}{t} \left(\sum_{\tau=1}^t |\xi_\tau|^q\right) \quad (\text{C.9})$$

and, therefore,

$$\mathbb{E} \left[\left(\sum_{\tau=1}^t |\xi_{\tau}|^2 \right)^{q/2} \right] \leq t^{q/2-1} \mathbb{E} \left[\sum_{\tau=1}^t |\xi_{\tau}|^q \right] \leq t^{q/2-1} \sum_{\tau=1}^t \sigma_{\tau}^q \quad (\text{C.10})$$

Thus, combining the above with (C.7) and (C.8), we obtain:

$$\begin{aligned} \mathbb{P} \left(\sup_{s \leq t} |S_s| > c(\gamma t)^{\mu} \right) &\leq \frac{C'_q \gamma^q t^{q/2-1} \sum_{\tau=1}^t \sigma_{\tau}^q}{c^q (\gamma t)^{q\mu}} \\ &\leq C_q \frac{\gamma^{q(1-\mu)} \sum_{\tau=1}^t \sigma_{\tau}^q}{t^{1+q(\mu-1/2)}} \end{aligned} \quad (\text{C.11})$$

for $C_q \equiv C'_q/c^q$, and the proof is complete. \blacksquare

We finally provide a separation result in the spirit of Farkas' lemma, that we will need for establishing the convergence rates.

Lemma C.2. *Let $\mathcal{X} = \{x \in \mathcal{V} : Ax = b, x \geq 0\}$ for $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^d$. Then, for all $x^* \in \mathcal{X}$ with $\text{act}(x^*) := \{\beta \in \{1, \dots, d\} : x_{\beta}^* = 0\}$, there exists $P \equiv P(x^*) \geq 1$ such that for all $\mathcal{I} \subseteq \text{act}(x^*)$ at least one of the following is true:*

- (i) $\mathcal{I} \neq \emptyset$ and there exists $\beta \in \text{act}(x^*) \setminus \mathcal{I}$ such that $x_{\beta} \leq P \max\{x_{\alpha} : \alpha \in \mathcal{I}\}$ for all $x \in \mathcal{X}$.
- (ii) There exists $z \in \ker(A)$ such that $\|z\| \leq P$, $z_{\beta} = 0$ for $\beta \in \mathcal{I}$ and $1 \leq z_{\beta} \leq P$ for $\beta \in \text{act}(x^*) \setminus \mathcal{I}$. Then, there exists

Proof. For the proof, see Azizian et al. [6, Lemma 6]. \blacksquare

C.2. Main results of Section 4. With the necessary tools in place, we proceed to prove the main results stated in Section 4. We start with the first result, establishing that a non-equilibrium point cannot arise as a limit point of the sequence of play induced by (FTRL).

Proposition 2. *Suppose that (FTRL) is run with perfect gradient feedback of the form $\hat{v}_t = v(x_t)$ for all $t = 1, 2, \dots$, and assume that x_t converges to some $\hat{x} \in \mathcal{X}$. Then \hat{x} is an equilibrium of \mathcal{G} .*

Proof. Since \hat{x} is not an equilibrium, there exists $p \in \mathcal{X}$ with $\langle v(\hat{x}), p - \hat{x} \rangle > 0$. Therefore, by continuity of the function $x \mapsto \langle v(x), p - x \rangle$, there exists a neighborhood \mathcal{U} of \hat{x} and $c > 0$ such that $\langle v(x), p - x \rangle \geq c$ for all $x \in \mathcal{U}$.

Moreover, since $\text{cl}(\mathcal{U})$ compact, we have $\sup_{x \in \text{cl}(\mathcal{U})} \|v(x)\|_* = B < \infty$. For the sake of contradiction, suppose that $x_t \rightarrow \hat{x}$. Then, $x_t \in \mathcal{U} \cap \mathbb{B}_{c/4B}(\hat{x})$ eventually, i.e., there exists n_0 such that $x_t \in \mathcal{U}$ and $\|x_t - \hat{x}\| < c/4B$ for all $t \geq n_0$.

Finally, since $y_t \in \partial h(x_t)$, we have for $t > t_0$:

$$\begin{aligned} h(p) &\geq h(x_t) + \langle y_t, p - x_t \rangle \\ &\geq h(x_t) + \langle y_{t_0}, p - x_t \rangle + \gamma \sum_{\tau=t_0}^{t-1} \langle v(x_{\tau}), p - x_t \rangle \\ &\geq h(x_t) + \langle y_{t_0}, p - x_t \rangle + \gamma \sum_{\tau=t_0}^{t-1} \langle v(x_{\tau}), p - x_{\tau} + x_{\tau} - x_t \rangle \\ &\geq h(x_t) + \langle y_{t_0}, p - x_t \rangle + \gamma \sum_{\tau=t_0}^{t-1} (\langle v(x_{\tau}), p - x_{\tau} \rangle + \langle v(x_{\tau}), x_{\tau} - x_t \rangle) \end{aligned}$$

$$\begin{aligned}
&\geq h(x_t) + \langle y_{t_0}, p - x_t \rangle + \gamma \sum_{\tau=t_0}^{t-1} (\langle v(x_\tau), p - x_\tau \rangle - \|v(x_\tau)\|_* \|x_\tau - x_t\|) \\
&\geq h(x_t) + \langle y_{t_0}, p - x_t \rangle + \gamma \sum_{\tau=t_0}^{t-1} (c - B \|x_\tau - x_t\|) \\
&\geq h(x_t) - \|y_{t_0}\|_* \|p - x_t\| + \gamma \sum_{\tau=t_0}^{t-1} (c - c/2) \\
&\geq \min h - \|y_{t_0}\|_* \text{diam}(\mathcal{X}) + \gamma c(t - t_0)/2
\end{aligned} \tag{C.12}$$

Taking $t \rightarrow \infty$, we obtain $h(p) \geq \infty$, which is a contradiction. Therefore, the result follows. \blacksquare

Moving forward, we show that equilibrium points in the relative interior cannot be limit points of (FTRL), either. Formally, we have:

Proposition 3. *Let $x^* \in \text{ri}(\mathcal{X})$ be a Nash equilibrium of $\mathcal{G}(\mathcal{N}, \mathcal{X}, u)$, and $(x_t)_{t \in \mathbb{N}}$ be the sequence of play induced by (FTRL) with $\hat{v}_t = v(x_t) + U_t$, where U_t i.i.d. with $\mathbb{E}[U_t] = 0$ and $\text{cov}(U_t) \succ 0$ for all $t \in \mathbb{N}$. Then:*

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} x_t = x^*\right) = 0 \quad \text{for any } x_1 \in \mathcal{X}_h. \tag{15}$$

Proof. Since x^* an equilibrium point in $\text{ri}(\mathcal{X})$, we readily get that $\langle v(x^*), x - x^* \rangle = 0$ for all $x \in \mathcal{X}$, and $x^* \in \mathcal{X}_h$. In view of this, there exists $y^* \in \mathcal{Y}$ such that $y^* \in \partial h(x^*)$, i.e., $x^* = Q(y^*)$. Our goal is to show that the auxiliary process y_t does not converge to $\partial h(x^*)$. However, there are infinitely many points in \mathcal{Y} that belong to $\partial h(x^*)$, so this attempt is insufficient, in the sense that, showing that y^* is not a limit point of the y_t dynamics, does not preclude that some other $y \in \partial h(x^*)$ is not.

To tackle this issue, we will show that the space \mathcal{Y} can be decomposed as $\mathcal{Y} = \hat{\mathcal{Y}} \oplus \bar{\mathcal{Y}}$ where all the “essential” deviations of the problem is in $\bar{\mathcal{Y}}$. For this, we define the set

$$\hat{\mathcal{Y}} = \{y \in \mathcal{Y} : \langle y, p - x \rangle = 0, \text{ for all } x, p \in \mathcal{X}\}. \tag{C.13}$$

which is a subspace of \mathcal{Y} , and as the following lemma suggests, is equal to the polar cone at any point in the relative interior.

Lemma C.3. *Let $x_0 \in \text{ri}(\mathcal{X})$ and $\hat{\mathcal{Y}} = \{y \in \mathcal{Y} : \langle y, p - x \rangle = 0, \text{ for all } x, p \in \mathcal{X}\}$. Then $\hat{\mathcal{Y}} = \text{PC}(x_0)$.*

To preserve the clarity of the argument, we defer the proof of Lemma C.3 until the end of this proposition. Letting $\bar{\mathcal{Y}}$ be the orthocomplement of $\hat{\mathcal{Y}}$, we readily get that $\mathcal{Y} = \hat{\mathcal{Y}} \oplus \bar{\mathcal{Y}}$, and any point y in \mathcal{Y} can be uniquely written as $y = \hat{y} + \bar{y}$ with $\hat{y} \in \hat{\mathcal{Y}}$ and $\bar{y} \in \bar{\mathcal{Y}}$. Defining the linear map $\Pi : \mathcal{Y} \rightarrow \mathcal{Y}$ as $\Pi y = \bar{y}$, and more importantly, under all points in $\partial h(x^*)$ under Π are essentially unique.

This is formalized in the following lemma, whose proof is relegated after this proposition.

Lemma C.4. *Let $x_0 \in \text{ri}(\mathcal{X})$ and $y, y' \in \partial h(x_0)$. Then $\Pi y \in \partial h(x_0)$, and $\Pi y = \Pi y'$.*

In view of the above, we are now ready to prove the result. Namely, fix some $p \in \mathcal{X}$, $p \neq x^*$ and let $\xi_t := \langle \Pi U_t, p - x^* \rangle$.

Then, setting $\sigma^2 \equiv (p - x^*)^\top \Sigma (p - x^*) > 0$, we have $\xi_t \sim (0, \sigma^2)$ i.i.d., and, so, there exists $\varepsilon, \delta > 0$ such that $\mathbb{P}(\xi_t > \varepsilon) = \delta$ for all $t \in \mathbb{N}$. Therefore, by the second Borel-Cantelli lemma [9], we get $\mathbb{P}(A) = 1$ for $A \equiv \{\xi_t > \varepsilon \text{ infinitely often}\}$. For the sake of contradiction, suppose

that $\mathbb{P}(B) > 0$ for $B \equiv \{\lim_{t \rightarrow \infty} x_t = x^*\}$. Fix some $\omega \in B$. Then, for all t large enough, we readily get that $x_t(\omega) \in \text{ri}(\mathcal{X})$, and, denoting $z_t := \Pi y_t$ and $z^* := \Pi y^*$, we readily get that

$$\lim_{t \rightarrow \infty} z_t(\omega) = z^* \quad (\text{C.14})$$

Thus, setting $\alpha_t \equiv \langle z_t - z^*, p - x^* \rangle$ we conclude by the above equality that $\lim_{t \rightarrow \infty} \alpha_t = 0$, and therefore it holds

$$\begin{aligned} 0 &= \lim_{t \rightarrow \infty} (\alpha_t - \alpha_{t-1}) \\ &= \lim_{t \rightarrow \infty} \langle \Pi \hat{v}_t, p - x^* \rangle \\ &= \lim_{t \rightarrow \infty} \langle \Pi v(x_t), p - x^* \rangle + \langle \Pi U_t, p - x^* \rangle \\ &= \langle \Pi v(x^*), p - x^* \rangle + \lim_{t \rightarrow \infty} \xi_t \\ &= \lim_{t \rightarrow \infty} \xi_t \end{aligned} \quad (\text{C.15})$$

Therefore, $\omega \notin A$, which implies that $B \subseteq A^c$, with $\mathbb{P}(A^c) = 0$. Thus, $\mathbb{P}(B) = 0$, which is a contradiction, and the result follows. \blacksquare

We now prove the two auxiliary lemmas presented in the proof of [Proposition 3](#).

Lemma C.3. *Let $x_0 \in \text{ri}(\mathcal{X})$ and $\hat{\mathcal{Y}} = \{y \in \mathcal{Y} : \langle y, p - x \rangle = 0, \text{ for all } x, p \in \mathcal{X}\}$. Then $\hat{\mathcal{Y}} = \text{PC}(x_0)$.*

Proof. First, we will show that

$$\text{PC}(x_0) = \{y \in \mathcal{Y} : \langle y, p - x_0 \rangle = 0 \text{ for all } p \in \mathcal{X}\} \quad (\text{C.16})$$

For this, suppose that there exist $y \in \text{PC}(x_0)$ and $p' \in \mathcal{X}$ such $\langle y, p' - x_0 \rangle < 0$. Then, since $x_0 \in \text{ri}(\mathcal{X})$, there exists $\alpha > 0$ such that $x_0 - \alpha(p' - x_0) \in \mathcal{X}$. By the definition of the polar cone, $\langle y, x_0 - \alpha(p' - x_0) - x_0 \rangle \leq 0$, or equivalently, $\langle y, p' - x_0 \rangle \geq 0$, which is a contradiction. Therefore, [\(C.16\)](#) holds, which implies that $\hat{\mathcal{Y}} \subseteq \text{PC}(x_0)$.

Now, for the inverse inclusion, let $y \in \text{PC}(x_0)$ and $p, x \in \mathcal{X}$. Then, we have:

$$\begin{aligned} \langle y, p - x \rangle &= \langle y, p - x_0 + x_0 - x \rangle \\ &= \langle y, p - x_0 \rangle + \langle y, x_0 - x \rangle \\ &= 0 \end{aligned} \quad (\text{C.17})$$

where the last equality follows by [\(C.16\)](#). Thus, $y \in \hat{\mathcal{Y}}$, and we conclude the result. \blacksquare

Lemma C.4. *Let $x_0 \in \text{ri}(\mathcal{X})$ and $y, y' \in \partial h(x_0)$. Then $\Pi y \in \partial h(x_0)$, and $\Pi y = \Pi y'$.*

Proof. For the first part, note that

$$\begin{aligned} \langle y, p - x_0 \rangle &= \langle \hat{y} + \bar{y}, p - x_0 \rangle = \langle \hat{y}, p - x_0 \rangle + \langle \bar{y}, p - x_0 \rangle \\ &= \langle \bar{y}, p - x_0 \rangle \\ &= \langle \Pi y, p - x_0 \rangle \end{aligned} \quad (\text{C.18})$$

which directly implies that $\Pi y \in \partial h(x_0)$. For the second part, since $x_0 \in \text{ri}(\mathcal{X})$, and $y, y' \in \partial h(x_0)$, we have that

$$\langle y - y', p - x_0 \rangle = 0 \quad \text{for all } p \in \mathcal{X} \quad (\text{C.19})$$

Thus $y - y' \in \text{PC}(x_0)$, and invoking [Lemma C.3](#) we obtain that $y - y' \in \hat{\mathcal{Y}}$. Therefore, applying the linear projection operator Π , we readily get that $\Pi(y - y') = 0$, and, using linearity, the result follows. \blacksquare

We now turn to our main convergence theorems, showing that the iterates of **(FTRL)** converge with high probability under both gradient-based and payoff-based feedback

Theorem 2. *Let $x^* \in \mathcal{X}$ be a strategically robust equilibrium of $\mathcal{G}(\mathcal{N}, \mathcal{X}, u)$. Fix a confidence level $\delta > 0$, and let $(x_t)_{t \in \mathbb{N}}$ be the iterates of **(FTRL)** with feedback provided by **(SFO)**, and step-size $\gamma > 0$ sufficiently small. Then, there exists a neighborhood \mathcal{U} of x^* in \mathcal{X}_h such that:*

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} x_t = x^*\right) \geq 1 - \delta \quad \text{if } x_1 \in \mathcal{U}. \quad (16)$$

Proof. Since x^* is strategically robust, $v(x^*)$ lies in the interior of the $\text{PC}(x^*)$. By [Lemma A.1](#), this implies in turn that there exists a polyhedral cone \mathcal{K} generated by $\mathcal{Z} \equiv \{z_1, \dots, z_r\}$ for $r \in \mathbb{N}$, such that $\text{TC}(x^*) \subseteq \mathcal{K}$ and $\langle v(x^*), z \rangle < 0$ for all $z \in \mathcal{Z}$.⁴ Therefore, for all $z \in \mathcal{Z}$, we have $\langle v(x^*), z \rangle \leq -m$, and by continuity of the vector field v , there exists a neighborhood \mathcal{U} of x^* and $c > 0$ such that $\langle v(x), z \rangle \leq -c$ for all $z \in \mathcal{Z}$ and $x \in \mathcal{U}$.

Fixing some $z \in \mathcal{Z}$, we obtain:

$$\begin{aligned} \langle y_{t+1}, z \rangle &= \langle y_t, z \rangle + \gamma \langle \hat{v}_t, z \rangle \\ &= \langle y_t, z \rangle + \gamma \langle v(x_t), z \rangle + \gamma \langle U_t, z \rangle \\ &= \langle y_1, z \rangle + \gamma \sum_{\tau=1}^t \langle v(x_\tau), z \rangle + \gamma \sum_{\tau=1}^t \langle U_\tau, z \rangle \end{aligned} \quad (\text{C.20})$$

Now, we define the stochastic process $(S_t)_{t \in \mathbb{N}}$ via $S_t := \gamma \sum_{\tau=1}^t \langle U_\tau, z \rangle$, which is a martingale, since $\mathbb{E}[\langle U_\tau, z \rangle | \mathcal{F}_\tau] = 0$.

Therefore, by [Lemma C.1](#) for $\sigma_t \equiv \sigma$, $q > 2$ and $\mu \in (0, 1)$, whose value is determined later, we get:

$$\delta_t := \mathbb{P}\left(\sup_{s \leq t} |S_s| > c(\gamma t)^\mu\right) \leq C_q \frac{\gamma^{q(1-\mu)} \sigma^q}{t^{q(\mu-1/2)}} \quad (\text{C.21})$$

where C_q is a constant that depends only on c and q . Thus, we readily have that:

$$\begin{aligned} \mathbb{P}\left(\bigcap_{t \geq 1} \left\{ \sup_{s \leq t} |S_s| \leq c(\gamma t)^\mu \right\}\right) &= 1 - \mathbb{P}\left(\bigcup_{t \geq 1} \left\{ \sup_{s \leq t} |S_s| > c(\gamma t)^\mu \right\}\right) \\ &\geq 1 - \sum_{t=1}^{\infty} \mathbb{P}\left(\sup_{s \leq t} |S_s| > c(\gamma t)^\mu\right) \\ &\geq 1 - \sum_{t=1}^{\infty} \delta_t \end{aligned} \quad (\text{C.22})$$

where the second inequality comes from the union bound. Now, we need to ensure that $\sum_{t=1}^{\infty} \delta_t \leq \delta/r$. For this, we need the sequence to be summable, which, using [\(C.21\)](#), is guaranteed for $q(\mu - 1/2) > 1$, or equivalently, $\mu \in (1/2 + 1/q, 1)$. Therefore, for $\gamma > 0$ small enough, we obtain that $\sum_{t=1}^{\infty} \delta_t \leq \delta/r$.

Therefore, with probability at least $1 - \delta/r$, the template inequality becomes:

$$\langle y_{t+1}, z \rangle \leq \langle y_1, z \rangle + \gamma \sum_{\tau=1}^t \langle v(x_\tau), z \rangle + c(\gamma t)^\mu \quad (\text{C.23})$$

If we initialize y_1 such that $\langle y_1, z' \rangle < -M - c$ for all $z' \in \mathcal{Z}$, we get that $\langle y_t, z \rangle < -M$ for all $t \in \mathbb{N}$ with probability at least $1 - \delta/r$. To see this, suppose that $\langle y_\tau, z \rangle < -M$ for all

⁴To resolve any ambiguities, the cone in question here is the polar of the cone provided by [Lemma A.1](#).

$\tau = 1, \dots, t$. Then

$$\begin{aligned} \langle y_{t+1}, z \rangle &= \langle y_1, z \rangle + \gamma \sum_{\tau=1}^t \langle v(x_\tau), z \rangle + \gamma \sum_{\tau=1}^t \langle U_\tau, z \rangle \\ &\leq -M - c - c\gamma t + c(\gamma t)^\mu \end{aligned} \quad (\text{C.24})$$

For $t \in \mathbb{N}$ with $\gamma t < 1$, we have $-c + c(\gamma t)^\mu < 0$, while for $\gamma t \geq 1$, it holds $-c\gamma t + c(\gamma t)^\mu < 0$. In both cases, we conclude that $\langle y_{t+1}, z \rangle < -M$, and by induction, we get the inequality.

Therefore, with probability at least $1 - \delta/r$, we have:

$$\langle y_{t+1}, z \rangle \leq -M - c - c\gamma t + c(\gamma t)^\mu \quad (\text{C.25})$$

and sending $t \rightarrow \infty$, we get $\langle y_t, z \rangle \rightarrow -\infty$.

Finally, repeating the same argument for all $z \in \mathcal{Z}$ and applying a union bound, we readily get that $\langle y_t, z \rangle \rightarrow -\infty$ with probability at least $1 - \delta$, and invoking [Proposition C.1](#), the result follows. \blacksquare

Having established the local convergence to x^* with high probability, we proceed to the convergence rate in the case of affinely constrained \mathcal{X} and decomposable regularizer h .

Theorem 3. *If, in addition, \mathcal{X} is a polyhedral domain and h is decomposable with kernel θ , on the event $E := \{\lim_{t \rightarrow \infty} x_t = x^*\}$ it holds:*

$$\|x_t - x^*\| = \phi(-\Theta(t)) \quad (17)$$

where ϕ is the rate function defined via

$$\phi(z) := \begin{cases} (\theta')^{-1}(z) & \text{if } z > \theta'(0^+) \\ 0 & \text{if } z \leq \theta'(0^+) \end{cases} \quad (18)$$

Proof. By the definition of the iterates of [\(FTRL\)](#), we have:

$$Q(y_t) = \arg \min_{x \in \mathcal{X}} \{h(x) - \langle y_t, x \rangle : Ax = b, x \geq 0\} \quad (\text{C.26})$$

Introducing the Lagrangian

$$\mathcal{L}(x, \lambda, \mu) = h(x) - \langle y_t, x \rangle + \sum_{i=1}^m \lambda_i (a_i^\top x - b_i) - \sum_{j=1}^d \mu_j x_j \quad (\text{C.27})$$

with $\lambda_i \in \mathbb{R}$ and $\mu_j \geq 0$, by the KKT conditions, we readily obtain:

$$y_t = \nabla h(x_t) + \sum_{s=1}^m \lambda_s a_s - \mu \quad (\text{C.28})$$

where $\nabla h(x) = \sum_{\beta=1}^d \theta'(x_{\beta,t}) e_\beta$, since θ is continuously differentiable.

For the sequel, we define the set of active constraints at x^* as $\text{act}(x^*) := \{\beta \in \{1, \dots, d\} : x_\beta^* = 0\}$. Note that on the event of $\{\lim_{t \rightarrow \infty} x_t = x^*\}$, the iterates x_t lie in a neighborhood of x^* , as shown in [Theorem 2](#). Thus, all non-active indices $\alpha \notin \text{act}(x^*)$ stay bounded away from zero, and so $|\theta(x_{\alpha,t})|$ remains bounded for all t .

We treat the two cases separately: (i) the steep case, where h is steep – equivalently $\theta'(0^+) = -\infty$, and (ii) the non-steep case, where h is not steep, i.e., $\theta'(0^+) > -\infty$.

The steep case. We define the set of “good” indices \mathcal{I} at step t as: $\beta \in \mathcal{I}$ if $\theta'(x_{\beta,t}) \leq -\Theta(t)$. Our goal is to show that all indices $\text{act}(x^*)$ of x_t are “good”. Fix some $t \in \mathbb{N}$.

Suppose that $\text{act}(x^*) \setminus \mathcal{I} \neq \emptyset$, and let $P \geq 1$, as per [Lemma C.2](#). Then,

- If condition (i) of [Lemma C.2](#) holds, there exists β' such that $x_{\beta',t} \leq P \max\{x_{\alpha,t} : \alpha \in \mathcal{I}\}$, and thus, $\mathcal{I} \leftarrow \mathcal{I} \cup \{\beta'\}$.
- If condition (ii) of [Lemma C.2](#) holds, there exists $z' \in \ker(A)$ such that $\|z'\| \leq P$, $z'_\beta = 0$ if $\beta \in \mathcal{I}$ and $1 \leq z'_\beta \leq P$ if $\beta \in \text{act}(x^*) \setminus \mathcal{I}$. By [\(C.28\)](#), and noting that $z' \in \ker(A)$ and $\mu = 0$, since all constraints are non-active due to steepness of h , we have:

$$\langle \nabla h(x_t), z' \rangle = \langle y_t, z' \rangle \quad (\text{C.29})$$

Moreover, it holds:

$$\begin{aligned} \langle \nabla h(x_t), z' \rangle &= \sum_{\beta=1}^d \theta'(x_{\beta,t}) z'_\beta = \sum_{\beta \in \mathcal{I}} \theta'(x_{\beta,t}) z'_\beta + \sum_{\beta \in \text{act}(x^*) \setminus \mathcal{I}} \theta'(x_{\beta,t}) z'_\beta + \sum_{\beta \notin \text{act}(x^*)} \theta'(x_{\beta,t}) z'_\beta \\ &= \sum_{\beta \in \text{act}(x^*) \setminus \mathcal{I}} \theta'(x_{\beta,t}) z'_\beta + C \end{aligned} \quad (\text{C.30})$$

for a constant C , since all non-active indices remain bounded away from zero, as explained in the beginning. Now, note that $z' \in \text{TC}(x^*)$, and thus, by [Lemma A.1](#), we can write z' as $z' = \sum_{i=1}^r \ell_i z_i$ with $\ell_i \geq 0$, such that $\langle y_t, z_i \rangle \leq -\Theta(t)$ for all $i = 1, \dots, r$ as in the proof of [Theorem 2](#). So, combining it with [\(C.35\)](#), [\(C.30\)](#), we obtain:

$$\sum_{\beta \in \text{act}(x^*) \setminus \mathcal{I}} \theta'(x_{\beta,t}) z'_\beta \leq -\Theta(t) \quad (\text{C.31})$$

and therefore, there exists at least one $\beta' \in \text{act}(x^*) \setminus \mathcal{I}$ such that

$$\theta'(x_{\beta',t}) z'_{\beta'} \leq -\Theta(t) \quad (\text{C.32})$$

Thus, $\mathcal{I} \leftarrow \mathcal{I} \cup \{\beta'\}$.

Therefore, as $\text{act}(x^*)$ is finite, we conclude inductively that $\theta'(x_{\beta,t}) \leq -\Theta(t)$ for all $\beta \in \text{act}(x^*)$. Finally, we have that $\mathbb{R}^d = \text{row}(A) + \text{span}\{e_\beta : \beta \in \text{act}(x^*)\}$, and thus, for all i , we can write the standard basis vector e_i as $e_i = \sum_{\beta \in \text{act}(x^*)} \lambda_{i,\beta} e_\beta + a_i$ for some $a_i \in \text{row}(A)$

$$\begin{aligned} x_{i,t} - x_i^* &= \langle x_t - x^*, e_i \rangle = \left\langle x_t - x^*, \sum_{\beta \in \text{act}(x^*)} \lambda_{i,\beta} e_\beta + a_i \right\rangle \\ &= \left\langle x_t - x^*, \sum_{\beta \in \text{act}(x^*)} \lambda_{i,\beta} e_\beta \right\rangle \\ &= \sum_{\beta \in \text{act}(x^*)} \lambda_{i,\beta} x_{\beta,t} \end{aligned} \quad (\text{C.33})$$

where we used that $\langle x_t - x^*, a_i \rangle = 0$. Thus, since $\theta'(x_{\beta,t}) \leq -\Theta(t)$ for all $\beta \in \text{act}(x^*)$, by the equivalence of norms and the above, we conclude that

$$\|x_t - x^*\| = (\theta')^{-1}(-\Theta(t)) \quad (\text{C.34})$$

The non-steep case. For the non-steep case, we follow a similar approach, but with some modifications since the iterates of **(FTRL)** are not always in the interior of \mathcal{X} .

Specifically, let the set of “good” indices \mathcal{I} be defined as: $\beta \in \mathcal{I}$ if $x_{\beta,t} = 0$ or $\theta'(x_{\beta,t}) \leq -\Theta(t)$. Our goal is to show that all indices $\text{act}(x^*)$ of x_t are “good”. We construct \mathcal{I} sequentially, as before.

Suppose that $\text{act}(x^*) \setminus \mathcal{I} \neq \emptyset$, and let $P \geq 1$, as per [Lemma C.2](#). Then,

- If condition (i) of [Lemma C.2](#) holds, there exists β' such that $x_{\beta',t} \leq P \max\{x_{\alpha,t} : \alpha \in \mathcal{I}\}$, and thus, $\mathcal{I} \leftarrow \mathcal{I} \cup \{\beta'\}$.
- If condition (ii) of [Lemma C.2](#) holds, there exists $z' \in \ker(A)$ such that $\|z'\| \leq P$, $z'_\beta = 0$ if $\beta \in \mathcal{I}$ and $1 \leq z'_\beta \leq P$ if $\beta \in \text{act}(x^*) \setminus \mathcal{I}$. Therefore, we have

$$\begin{aligned} \langle \nabla h(x_t), z' \rangle &= \langle y_t, z' \rangle + \langle \mu, z' \rangle = \langle y_t, z' \rangle + \sum_{\beta \in \mathcal{I}} \mu_\beta z'_\beta + \sum_{\beta \in \text{act}(x^*) \setminus \mathcal{I}} \mu_\beta z'_\beta + \sum_{\beta \notin \text{act}(x^*)} \mu_\beta z'_\beta \\ &= \langle y_t, z' \rangle \end{aligned} \quad (\text{C.35})$$

where, in this case, we used that (i) $z'_\beta = 0$ for $\beta \in \mathcal{I}$, (ii) $\mu_\beta = 0$ by complementary slackness for $\beta \notin \text{act}(x^*)$ since these constraints remain non-active for the whole process, and (iii) $\mu_\beta = 0$, again by complementary slackness for $\beta \in \text{act}(x^*) \setminus \mathcal{I}$ since if they were active, we would have $\beta \in \mathcal{I}$. This, with the same argument as before, we conclude that

$$\sum_{\beta \in \text{act}(x^*) \setminus \mathcal{I}} \theta'(x_{\beta,t}) z'_\beta \leq -\Theta(t) \quad (\text{C.36})$$

and therefore, there exists at least one $\beta' \in \text{act}(x^*) \setminus \mathcal{I}$ such that

$$\theta'(x_{\beta',t}) z'_{\beta'} \leq -\Theta(t) \quad (\text{C.37})$$

This holds until β' vanishes, which can lead to $\mu_{\beta'} > 0$. In either case, we have $\mathcal{I} \leftarrow \mathcal{I} \cup \{\beta'\}$.

Finally, since $\text{act}(x^*)$ is finite, we conclude inductively that all for all $\beta \in \text{act}(x^*)$, we have either $\theta'(x_{\beta,t}) \leq -\Theta(t)$ or $x_{\beta,t} = 0$. As in the steep case, we conclude

$$\|x_t - x^*\| = \phi(-\Theta(t)) \quad (\text{C.38})$$

and our proof is complete. \blacksquare

We now shift to the payoff-based setting. The relevant result is restated below.

Theorem 4. *Let $x^* \in \mathcal{X}$ be a strategically robust equilibrium of \mathcal{G} . Fix a confidence level $\delta > 0$, and let $(x_t)_{t \in \mathbb{N}}$ be the iterates of **(FTRL)** run with **(SPSA)** with $\delta_t \propto 1/t^p$ for some $p \in (0, 1/2)$ and step-size $\gamma > 0$ sufficiently small. Then, there exists a neighborhood \mathcal{U} of x^* such that:*

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} x_t = x^*\right) \geq 1 - \delta \quad \text{if } x_1 \in \mathcal{U}. \quad (\text{19})$$

If, in addition, \mathcal{X} is affinely constrained and h is decomposable with kernel θ , then, whenever x_t converges to x^ , we have:*

$$\|x_t - x^*\| = \phi(-\Theta(t)). \quad (\text{20})$$

Proof. First of all, we write \hat{v}_t in the following convenient form:

$$\hat{v}_t = v(x_t) + U_t + b_t \quad (\text{C.39})$$

with

$$U_t = \hat{v}_t - \mathbb{E}[\hat{v}_t | \mathcal{F}_t] \quad \text{and} \quad b_t = \mathbb{E}[\hat{v}_t | \mathcal{F}_t] - v(x_t) \quad (\text{C.40})$$

which, by [Proposition A.1](#), satisfy the bounds $\|U_t\|_* = \mathcal{O}(1/\delta_t)$ and $\|b_t\|_* = \mathcal{O}(\delta_t)$. Now, as in the proof of [Theorem 2](#), $v(x^*)$ lies in the interior of the $\text{PC}(x^*)$. By [Lemma A.1](#) this in

turn implies that there exists a polyhedral cone \mathcal{K} generated by $\mathcal{Z} \equiv \{z_1, \dots, z_r\}$ for $r \in \mathbb{N}$, such that $\text{TC}(x^*) \subseteq \mathcal{K}$ and $\langle v(x^*), z \rangle < 0$ for all $z \in \mathcal{Z}$.⁵ Therefore, for all $z \in \mathcal{Z}$, we have $\langle v(x^*), z \rangle \leq -m$, and by continuity of the vector field v , there exists a neighborhood \mathcal{U} of x^* and $c > 0$ such that $\langle v(x), z \rangle \leq -c$ for all $z \in \mathcal{Z}$ and $x \in \mathcal{U}$. Fix some $z \in \mathcal{Z}$. Then, unfolding the evolution of y_t , we have:

$$\begin{aligned} \langle y_{t+1}, z \rangle &= \langle y_t, z \rangle + \gamma \langle \hat{v}_t, z \rangle \\ &= \langle y_t, z \rangle + \gamma \langle v(x_t), z \rangle + \gamma \langle U_t, z \rangle + \gamma \langle b_t, z \rangle \\ &= \langle y_1, z \rangle + \gamma \sum_{\tau=1}^t \langle v(x_\tau), z \rangle + \gamma \sum_{\tau=1}^t \langle U_\tau, z \rangle + \gamma \sum_{\tau=1}^t \langle b_\tau, z \rangle \end{aligned} \quad (\text{C.41})$$

Now, we define the stochastic process $(S_t)_{t \in \mathbb{N}}$ via $S_t := \gamma \sum_{\tau=1}^t \langle U_\tau, z \rangle$, which is a martingale, since $\mathbb{E}[\langle U_\tau, z \rangle | \mathcal{F}_\tau] = 0$, and $\mathbb{E}[|\langle U_\tau, z \rangle|^q | \mathcal{F}_\tau] \leq \mathbb{E}[\|U_\tau\|_*^q | \mathcal{F}_\tau] = \mathcal{O}((1/\delta_t)^q)$.

Therefore, by Lemma C.1 for $\sigma_t = \Theta(1/\delta_t)$, $q > 2$ and $\mu \in (0, 1)$, whose value is determined later, we get:

$$\delta_t := \mathbb{P}\left(\sup_{s \leq t} |S_s| > (c/2)(\gamma t)^\mu\right) \leq C_q \frac{\gamma^{q(1-\mu)} \sum_{\tau=1}^t \sigma_\tau^q}{t^{1+q(\mu-1/2)}} \quad (\text{C.42})$$

where C_q is a constant that depends only on c and q . Thus, for $\delta_t = \delta/t^p$, there exist $B > 0$ such that:

$$\sum_{\tau=1}^t \sigma_\tau^q \leq B \delta^{-q} \sum_{\tau=1}^t \tau^{pq} \leq B' \delta^{-q} t^{1+pq} \quad (\text{C.43})$$

where we used that $\sum_{\tau=1}^t \tau^{pq} = \Theta(t^{1+pq})$. So, using the above bound, (C.44) becomes:

$$\delta_t \leq C'_q \frac{\gamma^{q(1-\mu)} \delta^{-q} t^{1+pq}}{t^{1+q(\mu-1/2)}} \leq C'_q \frac{\gamma^{q(1-\mu)} \delta^{-q}}{t^{q(\mu-1/2-p)}} \quad (\text{C.44})$$

Thus, we readily have that:

$$\begin{aligned} \mathbb{P}\left(\bigcap_{t \geq 1} \left\{ \sup_{s \leq t} |S_s| \leq c(\gamma t)^\mu \right\}\right) &= 1 - \mathbb{P}\left(\bigcup_{t \geq 1} \left\{ \sup_{s \leq t} |S_s| > c(\gamma t)^\mu \right\}\right) \\ &\geq 1 - \sum_{t=1}^{\infty} \mathbb{P}\left(\sup_{s \leq t} |S_s| > c(\gamma t)^\mu\right) \\ &\geq 1 - \sum_{t=1}^{\infty} \delta_t \end{aligned} \quad (\text{C.45})$$

Now, we need to show that there exists $\mu \in (0, 1)$ and $q > 2$ such that $\sum_{t=1}^{\infty} \delta_t \leq \delta/r$. In order for the sum to be finite, we need $q(\mu - 1/2 - p) > 1$ which readily implies that $p < \mu - 1/2 - 1/q$.

For the bias term, since $\|b_\tau\|_* = \Theta(\delta_\tau)$, we have:

$$\sum_{\tau=1}^t \langle b_\tau, z \rangle \leq \sum_{\tau=1}^t \|b_\tau\|_* \|z\| \leq \sum_{\tau=1}^t \|b_\tau\|_* \leq D \sum_{\tau=1}^t \delta_\tau \leq D \sum_{\tau=1}^t \delta / \tau^p \leq D' \delta t^{1-p} \quad (\text{C.46})$$

for some $D' > 0$, where in the last inequality we used that $\sum_{\tau=1}^t 1/\tau^p = \Theta(t^{1-p})$.

⁵As before, to resolve any ambiguities, the cone in question here is the polar of the cone provided by Lemma A.1.

Therefore, for $1 - \mu < p$, and $\gamma < 1$, we readily get that

$$\gamma \sum_{\tau=1}^t \langle b_\tau, z \rangle \leq D' \gamma \delta t^\mu \leq D' \delta (\gamma t)^\mu \quad (\text{C.47})$$

Therefore, we need to satisfy

$$1 - \mu < p < \mu - 1/2 - 1/q \quad (\text{C.48})$$

from which we obtain $\mu \in (3/4, 1)$. Thus, for $p \in (0, 1/2)$, there exist $\mu \in (3/4, 1)$ and $q > 2$ that satisfy (C.47). So, for δ, γ sufficiently small we can guarantee that

$$\gamma \sum_{\tau=1}^t \langle b_\tau, z \rangle \leq (c/2)(\gamma t)^\mu \quad \text{and} \quad \sum_{t=1}^{\infty} \delta_t \leq \delta/r \quad (\text{C.49})$$

Therefore, with probability at least $1 - \delta/r$, the template inequality becomes:

$$\begin{aligned} \langle y_{t+1}, z \rangle &\leq \langle y_1, z \rangle + \gamma \sum_{\tau=1}^t \langle v(x_\tau), z \rangle + (c/2)(\gamma t)^\mu + (c/2)(\gamma t)^\mu \\ &\leq \langle y_1, z \rangle + \gamma \sum_{\tau=1}^t \langle v(x_\tau), z \rangle + c(\gamma t)^\mu \end{aligned} \quad (\text{C.50})$$

Initializing y_1 such that $\langle y_1, z' \rangle < -M - c$ for all $z' \in \mathcal{Z}$, we have $\langle y_t, z \rangle < -M$ for all $t \in \mathbb{N}$ with probability at least $1 - \delta/r$. To see this, we proceed by induction, and suppose that $\langle y_\tau, z \rangle < -M$ for all $\tau = 1, \dots, t$. Then

$$\begin{aligned} \langle y_{t+1}, z \rangle &= \langle y_1, z \rangle + \gamma \sum_{\tau=1}^t \langle v(x_\tau), z \rangle + \gamma \sum_{\tau=1}^t \langle U_\tau, z \rangle + \gamma \sum_{\tau=1}^t \langle b_\tau, z \rangle \\ &\leq -M - c - c\gamma t + c(\gamma t)^\mu \end{aligned} \quad (\text{C.51})$$

For $t \in \mathbb{N}$ with $\gamma t < 1$, we have $-c + c(\gamma t)^\mu < 0$, while for $\gamma t \geq 1$, it holds $-c\gamma t + c(\gamma t)^\mu < 0$. In both cases, we conclude that $\langle y_{t+1}, z \rangle < -M$, and by induction, we get the inequality.

Therefore, with probability at least $1 - \delta/r$, we have:

$$\langle y_{t+1}, z \rangle \leq -M - c - c\gamma t + c(\gamma t)^\mu \quad (\text{C.52})$$

and sending $t \rightarrow \infty$, we get $\langle y_t, z \rangle \rightarrow -\infty$.

As a final step, using the same argument for all $z \in \mathcal{Z}$ and applying a union bound, we have that $\langle y_t, z \rangle \rightarrow -\infty$ with probability at least $1 - \delta$, and by [Proposition C.1](#), we get that

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} x_t = x^*\right) \geq 1 - \delta. \quad (\text{C.53})$$

If, in addition, \mathcal{X} is affinely constrained and h is decomposable with kernel θ , then the argument in the proof of [Theorem 3](#) applies verbatim, yielding:

$$\|x_t - x^*\| = \phi(-\Theta(t)). \quad (\text{C.54})$$

whenever x_t converges to x^* . ■

We conclude this appendix with [Proposition 4](#), which illustrates that an extreme, non-strategically robust equilibrium may exhibit fundamentally different behavior depending on the choice of regularizer.

Proposition 4. *Consider the 1-player game \mathcal{G} with $\mathcal{X} = [0, 1]$, $u(x) = -\frac{3}{4}x^{4/3}$ and $x^* = 0$. Let $(x_t)_{t \in \mathbb{N}}$ be the iterates of [\(FTRL\)](#) with $\gamma < 1$, and $\hat{v}_t = v(x_t) + U_t$, where U_t are i.i.d. standard normal random variables for all $t \in \mathbb{N}$. Then, for any initial condition $y_1 \in \mathbb{R}$, we have:*

- (i) For $h(x) = x \log x$, it holds $\mathbb{P}(\lim_{t \rightarrow \infty} x_t = x^*) = 0$.
- (ii) For $h(x) = -2\sqrt{x}$, it holds $\mathbb{P}(\lim_{t \rightarrow \infty} x_t = x^*) = 1$.

Proof. We show each case separately.

(i) Writing down the (FTRL) dynamics, we have

$$\begin{aligned} y_{t+1} &= y_t + \gamma(-x_t^{1/3} + U_t) \\ x_t &= \sup_{x \in [0,1]} (y_t x - x \log x) \end{aligned} \tag{C.55}$$

Solving the maximization problem in the definition of x_t , we obtain:

$$x_t = \begin{cases} \exp(y_t - 1), & \text{if } y_t \leq 1 \\ 1, & \text{if } y_t > 1 \end{cases}$$

or, equivalently, $x_t = \mathbb{1}(y_t > 1) + \mathbb{1}(y_t \leq 1) \exp(y_t - 1)$ and the dual process can be written as:

$$y_{t+1} = y_t - \gamma \mathbb{1}(y_t > 1) - \gamma \mathbb{1}(y_t \leq 1) \exp((y_t - 1)/3) + \gamma U_t \tag{C.56}$$

It is clear that $x_t \rightarrow 0$ if and only if $y_t \rightarrow -\infty$ as t goes to infinity. For notational convenience, set $z_n \equiv -y_t$. Then, the evolution of the dual process becomes:

$$z_{t+1} = z_t + \gamma \mathbb{1}(z_t < -1) + \gamma \mathbb{1}(z_t \geq -1) \exp((-z_t - 1)/3) - \gamma U_t \tag{C.57}$$

Now, define the process

$$z'_{t+1} \equiv (z'_t + \gamma \mathbb{1}(z'_t < -1) + \gamma \mathbb{1}(z'_t \geq -1) \exp((-z'_t - 1)/3) - \gamma U_t)^+, \quad z'_1 = z_1 \tag{C.58}$$

where U_t is the same random variable as in (C.57).

The rest of our proof relies on a series of claims, which we state and prove one-by-one.

Claim 1. *The process $(z'_t)_{t \in \mathbb{N}}$ dominates $(z_t)_{t \in \mathbb{N}}$, i.e., $z'_t \geq z_t$ for all $t \in \mathbb{N}$.*

The proof of **Claim 1** lies at the end. Now, invoking [Theorem A.3](#) with

$$f(z) \equiv \gamma \mathbb{1}(z < -1) + \gamma \mathbb{1}(z \geq -1) \exp((-z - 1)/3) \tag{C.59}$$

bounded and $\sigma^2 = \gamma^2$, it holds that

$$f(z) \leq \frac{\sigma^2}{2z} \quad \text{for all } z \text{ large enough} \tag{C.60}$$

Thus, $(z'_t)_{t \in \mathbb{N}}$ is recurrent, which implies that

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} z'_t = \infty\right) = 0 \tag{C.61}$$

Finally, since $(z'_t)_{t \in \mathbb{N}}$ dominates $(z_t)_{t \in \mathbb{N}}$ by **Claim 1**, we obtain

$$\left\{\lim_{t \rightarrow \infty} z_t = \infty\right\} \subseteq \left\{\lim_{t \rightarrow \infty} z'_t = \infty\right\} \tag{C.62}$$

which implies that

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} x_t = x^*\right) = \mathbb{P}\left(\lim_{t \rightarrow \infty} z_t = \infty\right) \leq \mathbb{P}\left(\lim_{t \rightarrow \infty} z'_t = \infty\right) = 0 \tag{C.63}$$

and the result follows.

Proof of [Claim 1](#). Consider the function

$$g(z) := z + \gamma \mathbb{1}(z < -1) + \gamma \mathbb{1}(z \geq -1) \exp((-z - 1)/3) \tag{C.64}$$

Then:

- For $z < -1$: $g'(z) = 1$
- For $z > -1$: $g'(z) = 1 - \gamma/3 \exp((-z - 1)/3) > 1 - \gamma/3 > 0$

Thus, g is strictly increasing for all $z \in \mathbb{R}$. Now, for the sake of contradiction, suppose that there exists $\omega \in \Omega$ and a first time $k+1 \in \mathbb{N}$ where the dominance does not hold, i.e.,

$$z_k(\omega) \leq z'_k(\omega) \quad \text{and} \quad z'_{k+1}(\omega) < z_{k+1}(\omega) \quad (\text{C.65})$$

By the monotonicity property of g , we get that

$$g(z_k(\omega)) \leq g(z'_k(\omega)) \quad (\text{C.66})$$

and, therefore, adding $-\gamma U_\tau(\omega)$ in both sides

$$\begin{aligned} z_{\tau+1}(\omega) &\leq z'_\tau + \gamma \mathbb{1}(z'_\tau < -1) + \gamma \mathbb{1}(z'_\tau \geq -1) \exp((-z'_\tau - 1)/3) - \gamma U_\tau \\ &\leq (z'_\tau + \gamma \mathbb{1}(z'_\tau < -1) + \gamma \mathbb{1}(z'_\tau \geq -1) \exp((-z'_\tau - 1)/3) - \gamma U_\tau)^+ \\ &\leq z'_{\tau+1}(\omega) \end{aligned} \quad (\text{C.67})$$

which is a contradiction. Thus, the proof of [Claim 1](#) is complete.

(ii) In this setup, the [\(FTRL\)](#) dynamics are described by the system

$$\begin{aligned} y_{t+1} &= y_t + \gamma(-x_t^{1/3} + U_t) \\ x_t &= \sup_{x \in [0,1]} (y_t x + 2\sqrt{x}) \end{aligned} \quad (\text{C.68})$$

Solving the maximization problem in the definition of x_t , we obtain:

$$x_t = \begin{cases} (-y_t)^{-2}, & \text{if } y_t \leq -1 \\ 1, & \text{if } y_t > -1 \end{cases} \quad (\text{C.69})$$

or, equivalently, $x_t = \mathbb{1}(y_t > -1) + \mathbb{1}(y_t \leq -1)(-y_t)^{-2}$ and the dual process can be written as:

$$y_{t+1} = y_t - \gamma \mathbb{1}(y_t > -1) - \gamma \mathbb{1}(y_t \leq -1)(-y_t)^{-2/3} + \gamma U_t \quad (\text{C.70})$$

For notational convenience, set $z_n \equiv -y_t$. Then, the evolution of the dual process becomes:

$$z_{t+1} = z_t + \gamma \mathbb{1}(z_t < 1) + \gamma \mathbb{1}(z_t \geq 1) z_t^{-2/3} - \gamma U_t \quad (\text{C.71})$$

It is clear that $x_t \rightarrow 0$ if and only if $z_t \rightarrow \infty$ as t goes to infinity. Now, define the process

$$z'_{t+1} = \left(z'_t + \gamma \mathbb{1}(z'_t < 1) + \gamma \mathbb{1}(z'_t \geq 1) z_t^{-2/3} - \gamma U_t \right)^+, \quad z'_1 = z_1 \quad (\text{C.72})$$

where U_t is the same randomness as in [\(C.71\)](#).

Claim 2. *The process $(z'_t)_{t \in \mathbb{N}}$ dominates $(z_t)_{t \in \mathbb{N}}$, i.e., $z'_t \geq z_t$ for all $t \in \mathbb{N}$.*

The proof of [Claim 2](#) lies at the end. Now, invoking [Theorem A.3](#) with

$$f(z) \equiv \gamma \mathbb{1}(z < 1) + \gamma \mathbb{1}(z \geq 1) z^{-2/3} \quad (\text{C.73})$$

bounded, $\sigma^2 = \gamma^2$, and $\theta > 1$, we have

$$f(z) \geq \frac{\sigma^2 \theta}{2z} \quad \text{for all } z \text{ large enough} \quad (\text{C.74})$$

Thus, $(z'_t)_{t \in \mathbb{N}}$ is transient, which implies that $\mathbb{P}(A) = 1$ for $A = \{\omega \in \Omega : \lim_{t \rightarrow \infty} z'_t(\omega) = \infty\}$.

Now, fix some $\omega \in A$. Since $\lim_{t \rightarrow \infty} z'_t(\omega) = \infty$, there exists $n_\omega \in \mathbb{N}$ such that $z'_t > 1$ for all $n \geq n_\omega$, and therefore

$$\begin{aligned} z'_{t+1} &= z'_t + \gamma z_t'^{-2/3} - \gamma U_t \\ &= z'_{n_\omega} + \gamma \sum_{\tau=n_\omega+1}^t \left(z_\tau'^{-2/3} - U_\tau \right) \end{aligned} \quad (\text{C.75})$$

from which we conclude that

$$\sum_{\tau=n_\omega+1}^t \left(z_\tau'^{-2/3} - U_\tau \right) \rightarrow \infty \quad \text{as } t \rightarrow \infty \quad (\text{C.76})$$

Finally, we have that

$$z_{t+1} = z_{n_\omega} + \gamma \sum_{\tau=n_\omega+1}^t \left(\mathbb{1}(z_\tau < 1) + \mathbb{1}(z_\tau \geq 1) z_\tau'^{-2/3} - U_\tau \right) \quad (\text{C.77})$$

and, since $(z'_t)_{t \in \mathbb{N}}$ dominates $(z_t)_{t \in \mathbb{N}}$, and $z'_t > 1$ for all $t \geq t_\omega$, we readily get that

$$\sum_{\tau=n_\omega+1}^t \left(\mathbb{1}(z_\tau < 1) + \mathbb{1}(z_\tau \geq 1) z_\tau'^{-2/3} - U_\tau \right) \geq \sum_{\tau=n_\omega+1}^t \left(z_\tau'^{-2/3} - U_\tau \right) \quad (\text{C.78})$$

Thus, by (C.76), we conclude that

$$\sum_{\tau=n_\omega+1}^t \left(\mathbb{1}(z_\tau < 1) + \mathbb{1}(z_\tau \geq 1) z_\tau'^{-2/3} - U_\tau \right) \rightarrow \infty \quad \text{as } t \rightarrow \infty \quad (\text{C.79})$$

which implies that $\lim_{t \rightarrow \infty} z_t(\omega) = \infty$. Therefore, we obtain that $\lim_{t \rightarrow \infty} z_t(\omega) = \infty$ for all $\omega \in A$, and since $\mathbb{P}(A) = 1$, it follows that

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} x_t = x^*\right) = \mathbb{P}\left(\lim_{t \rightarrow \infty} z_t = \infty\right) = 1 \quad (\text{C.80})$$

and the proof is complete.

Proof of Claim 2. Consider the function

$$g(z) := z + \gamma \mathbb{1}(z < 1) + \gamma \mathbb{1}(z \geq 1) z^{-2/3} \quad (\text{C.81})$$

Then:

- For $z < 1$: $g'(z) = 1 + \gamma > 0$
- For $z > 1$: $g'(z) = 1 - 2\gamma z^{-5/3}/3 > 1 - 2\gamma/3 > 0$

Thus, g is strictly increasing for all $z \in \mathbb{R}$. Now, for the sake of contradiction, suppose that there exists $\omega \in \Omega$ and a first time $k+1 \in \mathbb{N}$ where the dominance does not hold, i.e.,

$$z_k(\omega) \leq z'_k(\omega) \quad \text{and} \quad z'_{k+1}(\omega) < z_{k+1}(\omega) \quad (\text{C.82})$$

By the monotonicity property of g , we get that

$$g(z_k(\omega)) \leq g(z'_k(\omega)) \quad (\text{C.83})$$

and therefore, adding $-\gamma U_\tau(\omega)$ in both sides

$$\begin{aligned} z_{\tau+1}(\omega) &\leq z'_\tau + \gamma \mathbb{1}(z'_\tau < 1) + \gamma \mathbb{1}(z'_\tau \geq 1) z'_\tau^{-2/3} - \gamma U_\tau \\ &\leq \left(z'_\tau + \gamma \mathbb{1}(z'_\tau < 1) + \gamma \mathbb{1}(z'_\tau \geq 1) z'_\tau^{-2/3} - \gamma U_\tau \right)^+ \\ &\leq z'_{\tau+1}(\omega) \end{aligned} \quad (\text{C.84})$$

which is a contradiction. Thus, the proof of *Claim 2* is complete. ■

ACKNOWLEDGMENTS

Jose Blanchet gratefully acknowledges support from the Department of Defense through the Air Force Office of Scientific Research (Award FA9550-20-1-0397) and the Office of Naval Research (Grant 1398311), as well as from the National Science Foundation (Grants 2229012, 2312204, and 2403007). Panayotis Mertikopoulos is also a member of Archimedes/Athena RC and acknowledges financial support by the French National Research Agency (ANR) in the framework of the PEPR IA FOUNDRY project (ANR-23-PEIA-0003), the project IRGA-SPICE (G7H-IRG24E90), and project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0, funded by the European Union under the NextGenerationEU Program.

REFERENCES

- [1] Abernethy, J., Lee, C., and Tewari, A. Fighting bandits with a new kind of smoothness. In *NIPS '15: Proceedings of the 29th International Conference on Neural Information Processing Systems*, 2015.
- [2] Arora, S., Hazan, E., and Kale, S. The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.
- [3] Arrow, K. J., Hurwicz, L., and Uzawa, H. *Studies in linear and non-linear programming*. Stanford University Press, 1958.
- [4] Audibert, J.-Y., Bubeck, S., and Lugosi, G. Minimax policies for combinatorial prediction games. In *COLT '11: Proceedings of the 24th Annual Conference on Learning Theory*, 2011.
- [5] Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, 1995.
- [6] Azizian, W., Iutzeler, F., Malick, J., and Mertikopoulos, P. On the rate of convergence of Bregman proximal methods in constrained variational inequalities. <http://arxiv.org/abs/2211.08043>, 2022.
- [7] Beck, A. and Teboulle, M. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167–175, 2003.
- [8] Benaïm, M. Dynamics of stochastic approximation algorithms. In Azéma, J., Émery, M., Ledoux, M., and Yor, M. (eds.), *Séminaire de Probabilités XXXIII*, volume 1709 of *Lecture Notes in Mathematics*, pp. 1–68. Springer Berlin Heidelberg, 1999.
- [9] Billingsley, P. *Probability and measure*. A Wiley-Interscience publication. Wiley, 3. ed edition, 1995. ISBN 0471007102.
- [10] Boone, V. and Mertikopoulos, P. The equivalence of dynamic and strategic stability under regularized learning in games. In *NeurIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems*, 2023.
- [11] Borkar, V. S. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press and Hindustan Book Agency, 2008.
- [12] Bravo, M., Leslie, D. S., and Mertikopoulos, P. Bandit learning in concave N -person games. In *NeurIPS '18: Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018.
- [13] Cauvin, P.-L., Legacci, D., and Mertikopoulos, P. The impact of uncertainty on regularized learning in games. In *ICML '25: Proceedings of the 42nd International Conference on Machine Learning*, 2025.
- [14] Debreu, G. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences of the USA*, October 1952.
- [15] DeepSeek-AI, Liu, A., Feng, B., Xue, B., Wang, B., Wu, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., Dai, D., Guo, D., Yang, D., Chen, D., Ji, D., Li, E., Lin, F., Dai, F., Luo, F., Hao, G., Chen, G., Li, G., Zhang, H., Bao, H., Xu, H., Wang, H., Zhang, H., Ding, H., Xin, H., Gao, H., Li, H., Qu, H., Cai, J. L., Liang, J., Guo, J., Ni, J., Li, J., Wang, J., Chen, J., Chen, J., Yuan, J., Qiu, J., Li, J., Song, J., Dong, K., Hu, K., Gao, K., Guan, K., Huang, K., Yu, K., Wang, L., Zhang, L., Xu, L., Xia, L., Zhao, L., Wang, L., Zhang, L., Li, M., Wang, M., Zhang, M., Tang, M., Li, M., Tian, N., Huang, P., Wang, P., Zhang, P., Wang, Q., Zhu, Q., Chen, Q., Du, Q., Chen, R. J., Jin, R. L., Ge, R., Zhang, R., Pan, R., Wang, R., Xu, R., Zhang, R., Chen, R., Li, S. S., Lu, S., Zhou, S., Chen, S., Wu, S., Ye, S., Ye, S., Ma, S., Wang, S., Zhou, S., Yu, S., Zhou, S., Pan, S., Wang, T., Yun, T., Pei, T., Sun, T., Xiao, W. L., Zeng, W., Zhao, W., An, W., Liu, W., Liang, W., Gao, W., Yu, W., Zhang, W., Li, X. Q., Jin, X., Wang, X., Bi, X., Liu, X., Wang, X., Shen, X., Chen, X., Zhang, X., Chen, X., Nie, X., Sun, X., Wang, X., Cheng, X., Liu, X., Xie, X., Liu, X., Yu, X., Song, X., Shan, X., Zhou, X.,

Yang, X., Li, X., Su, X., Lin, X., Li, Y. K., Wang, Y. Q., Wei, Y. X., Zhu, Y. X., Zhang, Y., Xu, Y., Xu, Y., Huang, Y., Li, Y., Zhao, Y., Sun, Y., Li, Y., Wang, Y., Yu, Y., Zheng, Y., Zhang, Y., Shi, Y., Xiong, Y., He, Y., Tang, Y., Piao, Y., Wang, Y., Tan, Y., Ma, Y., Liu, Y., Guo, Y., Wu, Y., Ou, Y., Zhu, Y., Wang, Y., Gong, Y., Zou, Y., He, Y., Zha, Y., Xiong, Y., Ma, Y., Yan, Y., Luo, Y., You, Y., Liu, Y., Zhou, Y., Wu, Z. F., Ren, Z. Z., Ren, Z., Sha, Z., Fu, Z., Xu, Z., Huang, Z., Zhang, Z., Xie, Z., Zhang, Z., Hao, Z., Gou, Z., Ma, Z., Yan, Z., Shao, Z., Xu, Z., Wu, Z., Zhang, Z., Li, Z., Gu, Z., Zhu, Z., Liu, Z., Li, Z., Xie, Z., Song, Z., Gao, Z., and Pan, Z. Deepseek-v3 technical report, 2025. URL <https://arxiv.org/abs/2412.19437>.

[16] Dieuleveut, A., Durmus, A., and Bach, F. R. Bridging the gap between constant step size stochastic gradient descent and markov chains. *The Annals of Statistics*, 2017. URL <https://api.semanticscholar.org/CorpusID:4605591>.

[17] Flokas, L., Vlatakis-Gkaragkounis, E. V., Lianeas, T., Mertikopoulos, P., and Piliouras, G. No-regret learning and mixed Nash equilibria: They do not mix. In *NeurIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

[18] Giannou, A., Vlatakis-Gkaragkounis, E. V., and Mertikopoulos, P. Survival of the strictest: Stable and unstable equilibria under regularized learning with partial information. In *COLT '21: Proceedings of the 34th Annual Conference on Learning Theory*, 2021.

[19] Giannou, A., Vlatakis-Gkaragkounis, E. V., and Mertikopoulos, P. The convergence rate of regularized learning in games: From bandits and uncertainty to optimism and beyond. In *NeurIPS '21: Proceedings of the 35th International Conference on Neural Information Processing Systems*, 2021.

[20] Giannou, A., Lotidis, K., Mertikopoulos, P., and Vlatakis-Gkaragkounis, E. V. On the convergence of policy gradient methods to Nash equilibria in general stochastic games. In *NeurIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems*, 2022.

[21] Hall, P. and Heyde, C. C. *Martingale Limit Theory and Its Application*. Probability and Mathematical Statistics. Academic Press, New York, 1980.

[22] Hazan, E. Introduction to online convex optimization. *CoRR*, abs/1909.05207, 2019. URL <https://arxiv.org/abs/1909.05207>.

[23] Hofbauer, J. and Sigmund, K. Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, 40(4), July 2003.

[24] Hsieh, Y.-G., Antonakopoulos, K., and Mertikopoulos, P. Adaptive learning in continuous games: Optimal regret bounds and convergence to Nash equilibrium. In *COLT '21: Proceedings of the 34th Annual Conference on Learning Theory*, 2021.

[25] Hsieh, Y.-G., Antonakopoulos, K., Cevher, V., and Mertikopoulos, P. No-regret learning in games with noisy feedback: Faster rates and adaptivity via learning rate separation. In *NeurIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems*, 2022.

[26] Hsieh, Y.-P., Mertikopoulos, P., and Cevher, V. The limits of min-max optimization algorithms: Convergence to spurious non-critical sets. In *ICML '21: Proceedings of the 38th International Conference on Machine Learning*, 2021.

[27] Kivinen, J. and Warmuth, M. K. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, 1997.

[28] Kohlberg, E. and Mertens, J.-F. On the strategic stability of equilibria. *Econometrica*, 54(5):1003–1037, September 1986.

[29] Kushner, H. J. and Clark, D. S. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer, 1978.

[30] Lamperti, J. Criteria for the recurrence or transience of stochastic process. i. *Journal of Mathematical Analysis and Applications*, 1(3):314–330, 1960. ISSN 0022-247X. doi: [https://doi.org/10.1016/0022-247X\(60\)90005-6](https://doi.org/10.1016/0022-247X(60)90005-6). URL <https://www.sciencedirect.com/science/article/pii/0022247X60900056>.

[31] Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, Cambridge, UK, 2020.

[32] Leonardos, S., Overman, W., Panageas, I., and Piliouras, G. Global convergence of multi-agent policy gradient in Markov potential games. In *ICLR '22: Proceedings of the 2022 International Conference on Learning Representations*, 2022.

[33] Littlestone, N. and Warmuth, M. K. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

- [34] Mertikopoulos, P. and Zhou, Z. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1-2):465–507, January 2019.
- [35] Mertikopoulos, P., Lecouat, B., Zenati, H., Foo, C.-S., Chandrasekhar, V., and Piliouras, G. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *ICLR '19: Proceedings of the 2019 International Conference on Learning Representations*, 2019.
- [36] Mertikopoulos, P., Hsieh, Y.-P., and Cevher, V. A unified stochastic approximation framework for learning in games. *Math. Program.*, 203(1–2):559–609, August 2023. ISSN 0025-5610. doi: 10.1007/s10107-023-02001-y. URL <https://doi.org/10.1007/s10107-023-02001-y>.
- [37] Monderer, D. and Shapley, L. S. Potential games. *Games and Economic Behavior*, 14(1):124 – 143, 1996.
- [38] Myerson, R. B. Refinements of the Nash equilibrium concept. *International Journal of Game Theory*, 7 (2):73–80, 1978.
- [39] Nesterov, Y. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120 (1):221–259, 2009.
- [40] Ritzberger, K. and Weibull, J. W. Evolutionary selection in normal-form games. *Econometrica*, 63(6): 1371–99, November 1995.
- [41] Robbins, H. and Monro, S. A stochastic approximation method. *Annals of Mathematical Statistics*, 22: 400–407, 1951.
- [42] Rockafellar, R. T. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.
- [43] Rockafellar, R. T. and Wets, R. J. B. *Variational Analysis*, volume 317 of *A Series of Comprehensive Studies in Mathematics*. Springer-Verlag, Berlin, 1998.
- [44] Rosen, J. B. Existence and uniqueness of equilibrium points for concave N -person games. *Econometrica*, 33(3):520–534, 1965.
- [45] Rosenthal, R. W. A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory*, 2:65–67, 1973.
- [46] Selten, R. Re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4:25–55, 1975.
- [47] Shalev-Shwartz, S. *Online learning: Theory, algorithms, and applications*. PhD thesis, Hebrew University of Jerusalem, 2007.
- [48] Shalev-Shwartz, S. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2), 2011.
- [49] Shalev-Shwartz, S. and Singer, Y. Convex repeated games and Fenchel duality. In *NIPS' 06: Proceedings of the 19th Annual Conference on Neural Information Processing Systems*. MIT Press, 2006.
- [50] Spall, J. C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Control*, 37(3):332–341, March 1992.
- [51] Tsuda, K., Rätsch, G., and Warmuth, M. K. Matrix exponentiated gradient updates for on-line Bregman projection. *Journal of Machine Learning Research*, 6:995–1018, 2005.
- [52] van Damme, E. *Stability and perfection of Nash equilibria*. Springer-Verlag, Berlin, 1987.
- [53] von Neumann, J. and Morgenstern, O. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
- [54] Vovk, V. G. Aggregating strategies. In *COLT '90: Proceedings of the 3rd Workshop on Computational Learning Theory*, pp. 371–383, 1990.
- [55] Wu, W.-T. and Jiang, J.-H. Essential equilibrium points of N -person non-cooperative games. *Scientia Sinica*, 11(10):1307–1322, 1962.
- [56] Xiao, L. Dual averaging methods for regularized stochastic learning and online optimization. *Journal of Machine Learning Research*, 11:2543–2596, October 2010.
- [57] Zimmert, J. and Seldin, Y. Tsallis-INF: An optimal algorithm for stochastic and adversarial bandits. *Journal of Machine Learning Research*, 22(28):1–49, 2021.
- [58] Zinkevich, M. Online convex programming and generalized infinitesimal gradient ascent. In *ICML '03: Proceedings of the 20th International Conference on Machine Learning*, pp. 928–936, 2003.