# Multi-Agent Learning under Uncertainty: Recurrence vs. Concentration

**Anonymous Author(s)**
Affiliation
Address
email

## Abstract

In this paper, we examine the convergence landscape of multi-agent learning in continuous games under uncertainty. In more detail, we consider two stochastic models of regularized learning—in continuous and discrete time respectively—and we set out to characterize the long-run behavior of the induced sequence of play. In stark contrast to deterministic, full-information models of learning (or models with a vanishing learning rate), we show that, in general, the resulting dynamics *do not* converge. In lieu of this, we ask instead which actions are played more often in the long run, and by how much. To that end, we show that, despite wandering away from equilibrium infinitely often, the dynamics of regularized learning in strongly monotone games always return to its vicinity in *finite* time (which we estimate), and their long-run distribution is sharply concentrated around a neighborhood thereof. We quantify the degree of this concentration, and we show that these favorable properties may all break down if the underlying game is not strongly monotone—underscoring in this way the limits of regularized learning in the presence of persistent randomness and uncertainty.

## 1 Introduction

In its most abstract form, the basic model for multi-agent learning in games unfolds as follows: (*i*) At each stage of the process, every participating agent selects an action; (*ii*) the agents receive a reward determined by their chosen actions and their individual payoff functions; (*iii*) the agents update their actions and the process repeats. In this general context, the agents have to contend with various—and varying—degrees of uncertainty: (*a*) uncertainty about the game, the strategic interests of other players, and/or who else is involved in the game; (*b*) uncertainty about the outcomes of their actions, and which update directions may lead to better outcomes; and (*c*) uncertainty stemming from the environment, manifesting as random shocks to the players' payoffs and / or other disturbances.

Our goal in this paper is to quantify the impact of randomness and uncertainty on multi-agent learning; more precisely,we seek to understand the differences that arise in the players' long-run behavior when such uncertainty is present versus when it is not. A natural framework for exploring this question is within the greater setting of no-regret learning—and, in particular, the family of *"follow-the-regularized-leader"* (FTRL) algorithms and dynamics [34, 55, 56]. This class contains several mainstay learning methods—like online gradient descent (or, in our case, *ascent*) [65], the exponential / multiplicative weights (EW) algorithm and its variants (HEDGE, EXP3, and the like) [3, 4, 36, 60], the TSALLIS–INF forecaster [1, 2, 64], and many others—so it has become practically synonymous with the notion of online learning in games. In view of all this, it is only natural to ask:

*What is the long-run distribution of FTRL under uncertainty?*
*Which actions are played more often, and by how much?*
*Do the dynamics concentrate—and, if so, where?*

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

**Our contributions in the context of related work.** Of course, the interpretation of these questions is context-specific, and it depends on the specific learning setting at hand. In this paper, motivated by applications to machine learning and data science that typically involve continuous action spaces and rewards, we focus on continuous games, and we consider two models of regularized learning, in continuous and discrete time respectively.

In the continuous-time setting, we model the dynamics of FTRL in the presence of uncertainty as a stochastic differential equation (SDE) perturbed by a general Itô diffusion process, i.e., a continuous-time martingale with possibly colored and/or correlated components. In the context of *finite* games, models of this type have been studied by, among others, Bravo & Mertikopoulos [10], Foster & Young [17], Fudenberg & Harris [18] and Mertikopoulos & Moustakas [39], the first two in an evolutionary setting, the latter as a continuous-time model of the EW algorithm in the presence of random disturbances. Follow-up works in this direction include [10, 11, 15, 22, 24] on *finite* games, while [29–31, 42] considered a regularized learning model in convex minimization problems. The model which is closest to our own is that of [41, 43], who study the regret properties and guarantees of a stochastic version of the dual averaging dynamics of Nesterov [49].

At a high level, our findings reveal a crisp dichotomy between games that are *null-monotone* (like bilinear min-max games or zero-sum bimatrix games), and *strongly monotone* games (like Kelly auctions, Cournot competitions, joint signal covariance optimization problems, etc.). Specifically:

1. In null-monotone games, uncertainty induces a persistent drift *away from equilibrium*: the dynamics reach greater distances from equilibrium in finite time (which we estimate) and they require, on average, infinite time to return. In particular, if the game admits an interior equilibrium, the dynamics diffuse away—escaping toward infinity or to the boundary of the game's action space—and they exhibit *no concentration* in any region of interior actions.
2. In strongly monotone games, uncertainty still induces a persistent outward drift, but this is now partially countered by the dynamics' deterministic component. Thus, in stark contrast to the null-monotone case, the players' learning trajectories end up in a near-equilibrium region whose size scales with the level of uncertainty, and we estimate both the size of this region and the time required to reach it. Paradoxically, the dynamics return with probability 1 arbitrarily close to where they started, infinitely often, in a way reminiscent of Poincaré recurrence in bimatrix min-max games [45, 51]; however, these returns can be exceedingly far apart, so there is no antinomy.

In discrete time, we consider a standard implementation of FTRL with a constant learning rate and stochastic first-order oracle feedback. Variants with a vanishing learning rate have been studied extensively in the stochastic approximation literature, and they are known to exhibit favorable convergence guarantees in, among others, strongly monotone games, cf. [44, 46] and references therein. At the same time however, these properties typically come at the expense of the algorithm slowing down to a crawl; for this reason, owing to their simplicity, robustness, and superior empirical performance, constant / non-vanishing learning rate schedules are much more common in practice.

On the downside, the long-run behavior of FTRL is much less understood in this case. To the best of our knowledge, the most relevant results come from recent works by Loizou et al. [37] and Huang & Zhang [23], who established upper bounds on the mean distance to equilibrium for stochastic gradient descent / ascent in strongly monotone games, and Vlatakis et al. [59], who studied the ergodic properties of constant step-size variants of the stochastic extragradient and stochastic gradient descent–ascent algorithms for weakly quasi-strongly monotone variational inequalities.

One reason that results about the statistics of the long-run behavior of FTRL are particularly scarce in the literature is that, in discrete time, even the most basic tools of stochastic analysis are often inapplicable. Nevertheless, based in no small part on the insights gained by our continuous-time analysis, we manage to establish the following version of the strong-null dichotomy:

1. In null-monotone games with an unbounded action space, the sequence of play under FTRL drifts away to infinity on average (though not necessarily with probability 1).
2. In strongly monotone games, we show that the mean time required to reach a given distance from the game's equilibrium is finite, and we provide an explicit estimate thereof. If the game's equilibrium is interior, we also show that FTRL converges strongly to a unique invariant measure, which is concentrated in a certain region around the game's equilibrium, which we also estimate.

We find these results particularly appealing as they provide the first glimpse into the distributional properties of regularized learning in games under uncertainty.

## 2 Preliminaries

**2.1. Continuous games.** Throughout the sequel, we consider games with a finite number of players and a continuum of actions per player. Formally, players will be indexed by $i \in \mathcal{N} = \{1, \dots, N\}$ and, during play, each player will be selecting an action $x_i$ from a closed convex subset $\mathcal{X}_i$ of some $d_i$-dimensional normed space $\mathcal{V}_i$. Aggregating over all players, we will write $\mathcal{X} = \prod_i \mathcal{X}_i$ for the space of the players' joint action profiles $x = (x_1, \dots, x_N)$ and $d = \sum_i d_i$ for the dimension of the ambient space $\mathcal{V} = \prod_i \mathcal{V}_i$. Finally, we will use the shorthand $x = (x_i; x_{-i})$ when we want to highlight the action of player $i \in \mathcal{N}$ against the action profile $x_{-i} = (x_j)_{j \neq i}$ of all other players—and, in similar notation, $\mathcal{X}_{-i} = \prod_{j \neq i} \mathcal{X}_j$ for the space thereof.

The reward of each player $i \in \mathcal{N}$ in a given action profile will be determined by an associated payoff function $u_i \colon \mathcal{X} \to \mathbb{R}$, assumed here to be *individually concave*, i.e., $u_i(x_i; x_{-i})$ is concave in $x_i$ for all $x_{-i} \in \mathcal{X}_{-i}$. We will further assume that each $u_i$ is $\beta$-Lipschitz smooth, and we will write

$$v_i(x) = \nabla_{x_i} u_i(x_i; x_{-i}) \quad \text{and} \quad v(x) = (v_1(x), \dots, v_N(x)) \tag{1}$$

for the individual gradient field of each player and the ensemble thereof.[1]

The tuple $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ will be referred to as a *concave game* [54]. Mainstay examples of such games include (mixed extensions of) finite games, resource allocation problems, Kelly auctions, Cournot competitions, etc.; for completeness, we detail some of these applications in Appendix A.

**2.2. Nash equilibrium.** The leading solution concept in game theory is that of a *Nash equilibrium*, defined here as an action profile $x^* \in \mathcal{X}$ which discourages unilateral deviations, i.e.,

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

A concave game always admits a Nash equilibrium if $\mathcal{X}$ is compact, and it admits a *unique* equilibrium if the game is strongly monotone in the sense below:

**Definition 1.** A game $\mathcal{G} \equiv \mathcal{G}(\mathcal{N}, \mathcal{X}, u)$ is called *$\alpha$-monotone* if there exists some $\alpha \geq 0$ such that

$$\langle v(x') - v(x), x' - x \rangle \leq -\alpha \|x' - x\|^2 \quad \text{for all } x, x' \in \mathcal{X}. \tag{Mon}$$

If $\alpha > 0$, $\mathcal{G}$ will be called *strongly monotone*; otherwise, if (Mon) holds with $\alpha = 0$, $\mathcal{G}$ will be called (*merely*) *monotone*; finally, if (Mon) binds for $\alpha = 0$ and all $x, x' \in \mathcal{X}$, $\mathcal{G}$ will be called *null-monotone*.

**2.3. Regularized learning.** In the rest of our paper, we will consider a family of online learning schemes adhering to the following "regularized learning" model: players aggregate gradient feedback on their payoff functions over time and, at each instance of play, they choose the action which is most closely aligned to this aggregate. We provide a detailed description of this model in Sections 3 and 4—in continuous and discrete time respectively—and only describe here the core idea.

At a high level, the common denominator of these schemes is the way that players choose their actions based on the accumulation of payoff gradients over time. Formally, we will treat payoff gradients as dual vectors and we will write $\mathcal{Y}_i \coloneqq \mathcal{V}_i^*$ for the dual space of $\mathcal{V}_i$ and $\mathcal{Y} = \prod_i \mathcal{Y}_i = \mathcal{V}^*$ for the ensemble thereof. Then, given an aggregate of gradient steps $y_i \in \mathcal{Y}_i$, we will assume that the $i$-th player chooses an action via a "*generalized projection*"—or *mirror*—map $Q_i \colon \mathcal{Y}_i \to \mathcal{X}_i$ of the form

$$Q_i(y_i) = \arg\max_{x_i} \{\langle y_i, x_i \rangle - h_i(x_i)\} \quad \text{for all } y_i \in \mathcal{Y}_i. \tag{2}$$

In the above $h_i \colon \mathcal{X}_i \to \mathbb{R}$ is a continuous $K_i$-strongly convex function, that is,

$$h_i(\lambda x_i + (1-\lambda)x_i') \leq \lambda h_i(x_i) + (1-\lambda) h_i(x_i') - \tfrac{1}{2} K_i \lambda (1-\lambda) \|x_i' - x_i\|^2 \tag{3}$$

for all $x_i, x_i' \in \mathcal{X}_i$ and all $\lambda \in [0,1]$. This function is known as the *regularizer* of the method and it acts as a penalty term that smooths out the "hard" $\arg\max$ correspondence $y_i \mapsto \arg\max_i \langle y_i, x_i \rangle$. This regularization scheme has a very long and rich history in game theory and optimization, where $Q$ is often referred to as a "*quantal*" or "*regularized*" best response operator, cf. [34, 38, 44, 55, 58] and references therein. For concreteness, we describe below the two leading examples of this regularization setup (suppressing in both cases the player index $i \in \mathcal{N}$ for notational clarity):

[1] We are tacitly assuming here that the players' payoff functions are defined in an open neighborhood of $\mathcal{X}$ in $\mathcal{V}$; this assumption is done only for convenience, and it does not affect any of our results.

**Example 1** (Euclidean regularization)**.** Let $h(x) = \frac{1}{2}\|x\|_2^2$. Then (B.12) boils down to the Euclidean projection map

$$Q(y) = \Pi_{\mathcal{X}}(y) \equiv \arg\max_{x \in \mathcal{X}} \|y - x\|_2 \,. \tag{4}$$

Thus, in particular, if $\mathcal{X} = \mathcal{V}$, we readily recover the identity map $Q(y) = y$. ❦

**Example 2** (Entropic regularization)**.** Let $\mathcal{X} = \{x \in \mathbb{R}^d_+ : \sum_{k=1}^d x_k = 1\}$ be the unit simplex of $\mathbb{R}^d$, and let $h(x) = \sum_{k=1}^d x_k \log x_k$ denote the (negative) entropy on $\mathcal{X}$. Then (B.12) yields the *logit* map

$$Q(y) = \Lambda(y) \equiv \frac{(\exp(y_1), \dots, \exp(y_d))}{\exp(y_1) + \dots + \exp(y_d)} \,. \tag{5}$$

This map forms the basis of the seminal HEDGE and EXP3 algorithms in online learning, cf. [3, 4, 12, 34, 36, 55] and references therein. ❦

To ease notation in the sequel, we will write $h(x) \coloneqq \sum_i h_i(x_i)$ for the players' aggregate regularizer, $K \coloneqq \min_i K_i$ for the strong convexity modulus of $h$, and $Q \coloneqq \prod_i Q_i \colon \mathcal{Y} \to \mathcal{X}$ for the resulting ensemble mirror map. In the next sections, we describe in detail how this regularization setup is used in a learning context.

## 3 Learning under uncertainty in continuous time

To set the stage for the analysis to come, we begin with two simple games that will serve as "minimal working examples" for the more general model and results presented later in this section. We focus for the moment on continuous-time interactions; the discrete-time setting is presented in Section 4.

**3.1. A gentle start.** Consider the following 2-player, convex-concave min-max games:

$$(a) \text{ Bilinear min-max:} \qquad u_1(x_1, x_2) = -u_2(x_1, x_2) = -x_1 x_2 \qquad \text{for } x_1, x_2 \in \mathbb{R}. \tag{6a}$$

$$(b) \text{ Quadratic min-max:} \qquad u_1(x_1, x_2) = -u_2(x_1, x_2) = x_2^2/2 - x_1^2/2 \qquad \text{for } x_1, x_2 \in \mathbb{R}. \tag{6b}$$

Both games are monotone and they admit a unique Nash equilibrium at the origin. Their gradient fields are $v(x_1, x_2) = (-x_2, x_1)$ and $v(x_1, x_2) = -(x_1, x_2)$ respectively, so the first game is *null-monotone* and the second one is 1-*strongly monotone*. Accordingly, if each player follows their individual payoff gradient to increase their rewards, we obtain the gradient descent / ascent dynamics

$$(a)\ \dot{x}(t) = (-x_2(t), x_1(t)) \quad \text{and} \quad (b)\ \dot{x}(t) = -(x_1(t), x_2(t)) \tag{GDA}$$

for the bilinear and quadratic games (6a) and (6b) respectively. It is then trivial to see that, in the bilinear case, (GDA) cycles periodically at a constant distance from the game's equilibrium, whereas, in the quadratic case, the dynamics converge to the game's equilibrium at a geometric rate.

To model uncertainty in this setting, we will consider the stochastic gradient dynamics

$$dX(t) = v(X(t))\, dt + \sigma\, dW(t) \tag{S-GDA}$$

where $W(t) = (W_1(t), W_2(t))$ is a Brownian motion in $\mathbb{R}^2$ and $\sigma > 0$ is the magnitude of the noise entering the process. Intuitively, this SDE should be viewed as a rigorous formulation of the informal model $\dot{x}(t) = v(x(t)) +$ "noise", with the Brownian term $W(t)$ capturing all sources of noise, randomness and uncertainty in the players' environment.[2] Consequently, to understand the impact of uncertainty in each case of (GDA), we will examine the following quantities:

1. The distance $\|X(t)\|_2^2$ of $X(t)$ from the game's equilibrium (that is, the origin of $\mathbb{R}^2$).
2. The time $\tau_r = \inf\{t > 0 : \|X(t)\|_2 \leq r\}$ at which $X(t)$ gets within $r$ of the game's equilibrium.
3. The density $\mathcal{P}(x, t)$ of $X(t)$—and, if it exists, its long-run limit $\mathcal{P}_\infty(x) \coloneqq \lim_{t\to\infty} \mathcal{P}(x, t)$.

When it exists, $\mathcal{P}_\infty$ is known as the *stationary*—or *invariant*—*distribution* of $X$, and it is closely related to the *occupation measure* $\mu_t$ of the process, defined here as

$$\mu_t(\mathcal{B}) = \frac{1}{t} \int_0^t \mathbb{1}\{X(s) \in \mathcal{B}\}\, ds \quad \text{for every Borel } \mathcal{B} \subseteq \mathcal{X}. \tag{7}$$

Under mild ergodicity conditions [26, Cor. 25.9], we have $\lim_{t\to\infty} \mu_t(\mathcal{B}) = \int_{\mathcal{B}} \mathcal{P}_\infty$ so, concretely, $\mathcal{P}_\infty$ measures the fraction of time that $X(t)$ spends in a given subset of $\mathcal{X}$ in the long run.

Taken together, these metrics provide a fairly complete picture of the statistics of $X(t)$ so, in the rest of this section, we analyze them in the context of (S-GDA) applied to the games (6a) and (6b).

[2]For a primer on SDEs, see [32, 50]; for completeness, we also present some basic elements in Appendix C.

**Case 1: Bilinear saddles.** In this case, by a direct application of Itô's formula—the chain rule of stochastic calculus [50, Chap. 4]—we readily obtain

$$d\big(\|X(t)\|_2^2\big) = 2X(t)\cdot dX(t) + dX(t)\cdot dX(t) = 2\sigma^2\,dt + \sigma\,X(t)\cdot dW(t)\,. \tag{8}$$

This suggests that, on average, $\|X(t)\|_2^2$ increases as $\Theta(\sigma^2 t)$. Building on this observation, we show in Appendix D that the dynamics (S-GDA) for the bilinear game (6a) enjoy the following properties:

**Proposition 1.** *Suppose that (S-GDA) is run on the game (6a) with initial condition $x_0 \in \mathbb{R}^2$. Then:*

1. *$\lim_{t\to\infty}\mathbb{E}_{x_0}\big[\|X(t)\|_2^2\big] = \infty$, i.e., $X(t)$ escapes to infinity in mean square.*
2. *$\mathbb{E}_{x_0}[\tau_r] = \infty$ if $r < \|x_0\|$, i.e., $X(t)$ takes infinite time on average to get closer to equilibrium.*
3. *The limit $\mathcal{P}_\infty(x) = \lim_{t\to\infty}\mathcal{P}(x,t)$ does not exist, i.e., $X$ does not admit an invariant distribution.*

Proposition 1 shows that, in the presence of uncertainty, the periodicity of the deterministic dynamics (GDA) is completely destroyed. In fact, despite random fluctuations that occasionally bring $X(t)$ closer to equilibrium, (S-GDA) exhibits a consistent drift *away from equilibrium*, escaping any compact set in finite time and requiring infinite time to return. As a result, $X(t)$ becomes infinitely spread out in the long run, exhibiting no measurable concentration in *any* region of $\mathbb{R}^2$. For a partial illustration of this behavior—which we view as antithetical to convergence—cf. Fig. 1. ❦

**Case 2: Quadratic saddles.** We now proceed to examine the behavior of (S-GDA) in the quadratic min-max problem (6b), where (S-GDA) gives $dX(t) = -X(t)\,dt + \sigma\,dW(t)$. As is well known [32, Chap. 7.4], this SDE describes the 2-dimensional *Ornstein–Uhlenbeck* (OU) process

$$X(t) = X(0)e^{-t} + \sigma\int_0^t e^{-(t-s)}\,dW(s)\,. \tag{9}$$

Hence, by unfolding the stochastic integral in (9), we can draw the following conclusions for $X(t)$:

**Proposition 2.** *Suppose that (S-GDA) is run on the game (6b) with initial condition $x_0 \in \mathbb{R}^2$. Then:*

1. *$\lim_{t\to\infty}\mathbb{E}_{x_0}\big[\|X(t)\|_2^2\big] = \sigma^2$, i.e., the dynamics fluctuate at mean distance $\sigma$ from equilibrium.*
2. *The mean time required to get within distance $r$ of the game's equilibrium is bounded as*

$$\mathbb{E}_{x_0}[\tau_r] \leq \frac{1}{2}\,\frac{\|x_0\|_2^2 - r^2}{r^2 - \sigma^2} \qquad \textit{for all } \sigma < r < \|x_0\|_2. \tag{10}$$

3. *The density of $X(t)$ is $\mathcal{P}(x,t) = [\pi\sigma^2(1-e^{-2t})]^{-1}\exp\Big(-\frac{\|x-e^{-t}x_0\|_2^2}{(1-e^{-2t})\sigma^2}\Big)$. In particular, $X(t)$ converges in distribution to a Gaussian random variable centered at $0$, viz.*

$$\mathcal{P}_\infty(x) \equiv \lim_{t\to\infty}\mathcal{P}(x,t) = 1/(\pi\sigma^2)\cdot e^{-\|x\|_2^2/\sigma^2}\,. \tag{11}$$

Proposition 2 shows that the geometric convergence properties of the deterministic dynamics (GDA) are again destroyed in the presence of uncertainty. However, in stark contrast to Proposition 1 for the bilinear case, $X(t)$ now exhibits a consistent drift *toward equilibrium*, and it ends up being sharply concentrated at a distance of $\mathcal{O}(\sigma^2)$ from equilibrium. This interplay between recurrence and concentration will play a crucial role in the sequel, and our aim in the rest of this section will be to quantify the extent to which it holds in a more general setting.

**3.2. Learning in continuous time.** We now proceed to describe our general model for multi-agent learning under uncertainty, hinging on the stochastic *"follow-the-regularized-leader"* template

$$dY_i(t) = v_i(X(t))\,dt + dM_i(t) \qquad X_i(t) = Q_i(Y_i(t))\,. \tag{S-FTRL}$$

In the above, $(i)$ $Y_i(t) \in \mathcal{Y}_i$ is a "score" variable that tracks the aggregation of individual payoff gradients in $\mathcal{Y}_i$; $(ii)$ $M_i(t) \in \mathcal{Y}_i$ is a continuous square-integrable martingale acting as a catch-all, "colored noise" disturbance term; and $(iii)$ $Q_i\colon \mathcal{Y}_i \to \mathcal{X}_i$ is the regularized mirror map of player $i \in \mathcal{N}$, as per (B.12). In this regard, (S-FTRL) represents a noisy "stimulus-response" mechanism, where each player $i \in \mathcal{N}$ tracks the aggregation of payoff gradients under uncertainty—the "stimulus"—and "responds" to this aggregate via their individual regularized mirror map $Q_i$.

*Remark* 1. The terminology "follow-the-regularized-leader" is due to [55, 56], who first studied this scheme in the context of online convex optimization in discrete time. This family of algorithms and dynamics has been widely studied in the literature; we provide more details on this in Appendix B. ❦

Moving forward, we will assume that the noise term $M(t) = (M_i(t))_{i\in\mathcal{N}}$ in (S-FTRL) is of the form

$$dM(t) = \sigma(X(t)) \cdot dW(t) \quad \text{or, more explicitly} \quad dM_i(t) = \sigma_i(X(t)) \cdot dW(t) \tag{12}$$

where $W(t) = (W_1(t), \dots, W_m(t))$ is a standard Brownian motion in $\mathbb{R}^m$, and $\sigma(x) = (\sigma_i(x))_{i\in\mathcal{N}}$ is an ensemble of state-dependent *diffusion matrices* $\sigma_i\colon \mathcal{X}_i \to \mathbb{R}^{d_i\times m}$, $i \in \mathcal{N}$.[3] Importantly, the model (12) allows for *correlated uncertainty* between different components of the process—e.g., accounting for random disturbances on shared road segments in a congestion game—so it will be our base model for the sequel. Our only standing assumption will be that $\sigma\colon \mathcal{X} \to \mathbb{R}^{d\times m}$ is bounded and Lipschitz continuous, which ensures that (S-FTRL) is *well-posed*, i.e., it admits a unique strong solution that exists for all time for every initial condition $Y(0) \leftarrow y \in \mathcal{Y}$ (cf. Appendix C).

*Remark* 2. To connect the above with Section 3.1, note that (S-GDA) is recovered from (S-FTRL) by taking $\mathcal{X}_i = \mathbb{R}$, $M_i(t) = \sigma\, W_i(t)$, and $h_i(x_i) = x_i^2/2$ for $i = 1, 2$ (so $Q_i(y_i) = y_i$ by Example 1). ❦

**3.3. Analysis and results.** We now proceed to describe our main results for the stochastic dynamics (S-FTRL)—which, as we show shortly, reflect the dichotomy between bilinear and quadratic saddle-point problems that we noted in Section 3.1. To state them, it will be convenient to introduce a "primal-dual"generalization of the Euclidean distance that is more closely aligned with the regularization setup underlying the players' response scheme. Deferring the details to Appendix B, we define here the *Fenchel coupling* induced by the regularizer $h_i$ of player $i \in \mathcal{N}$ as

$$F_i(p_i, y_i) = h_i(p_i) + h_i^*(y_i) - \langle y_i, x_i\rangle \quad \text{for all } i \in \mathcal{N}, \text{and all } p_i \in \mathcal{X}_i,\, y_i \in \mathcal{Y}_i \tag{13}$$

where $h_i^*(y_i) \coloneqq \max_{x_i\in\mathcal{X}_i}\{\langle y_i, x_i\rangle - h_i(x_i)\}$ denotes the convex conjugate of $h_i$. For example, in the unconstrained Euclidean case (Example 1), we recover the Euclidean distance squared, viz. $F_i(p_i, y_i) = \frac{1}{2}\|Q_i(y_i) - p_i\|^2$; by comparison, under entropic regularization on the simplex (Example 2), we get a "dualized" version of the Kullback–Leibler divergence, cf. Appendix B. In all cases, $F_i$ is *positive-semidefinite* in the sense that $F_i(p_i, y_i) \geq 0$ for all $y_i \in \mathcal{Y}_i$, with equality if and ony if $Q_i(y_i) = p_i$. In view of this, the total coupling $F(x, y) \coloneqq \sum_i F_i(p_i, y_i)$ is a valid measure of "divergence" between $p \in \mathcal{X}$ and $y \in \mathcal{Y}$, and we will use it freely in the sequel as such.

The last ingredient that we will need is two measures of the amount of randomness in (S-FTRL), viz.

$$\sigma_{\min}^2 \coloneqq \min_{x\in\mathcal{X}} \lambda_{\min}(\Sigma(x)) \quad \text{and} \quad \sigma_{\max}^2 \coloneqq \max_{x\in\mathcal{X}} \lambda_{\max}(\Sigma(x)) \tag{14}$$

where $\Sigma \equiv \sigma\sigma^\top$ denotes the quadratic covariation matrix of the martingale $M(t)$, and $\lambda_{\min}$ (resp. $\lambda_{\max}$) denotes the minimum (resp. maximum) eigenvalue thereof.

With all this in hand, we will focus on two broad classes of games, *null-monotone* and *strongly monotone*, of which the bilinear and quadratic examples of Section 3.1 are archetypal examples. To state our results, we will assume that (S-FTRL) is initialized at $x_0 \leftarrow Q(y_0) \in \operatorname{ri}\mathcal{X}$ for some $y_0 \in \mathcal{Y}$, and we will write $F_t \equiv F(x^*, Y(t))$ where $x^*$ is an equilibrium of the game. We then have:

**Theorem 1** (Null-monotone games)**.** *Suppose that* (S-FTRL) *is run with a smooth mirror map $Q$ in a null-monotone game $\mathcal{G}$ that admits an interior equilibrium $x^*$, and consider the hitting times $\tau_\varepsilon^- \coloneqq \inf\{t > 0 : F_t \leq F_0 - \varepsilon\}$ and $\tau_\varepsilon^+ \coloneqq \inf\{t > 0 : F_t \geq F_0 + \varepsilon\}$. If $\sigma_{\min}^2 > 0$ and $\varepsilon > 0$ is small enough, then*

$$\mathbb{E}_{x_0}[\tau_\varepsilon^-] = \infty \quad \text{and} \quad \mathbb{E}_{x_0}[\tau_\varepsilon^+] \leq 2\varepsilon\big/\big(\kappa\,\sigma_{\min}^2\big) \tag{15}$$

*for some positive constant $\kappa \equiv \kappa_\varepsilon > 0$; in addition, $X(t)$ does not admit a limiting distribution.*

**Theorem 2** (Strongly monotone games)**.** *Suppose that* (S-FTRL) *is run in an $\alpha$-strongly monotone game $\mathcal{G}$, and consider the hitting time $\tau_r \coloneqq \inf\{t > 0 : X(t) \in \mathbb{B}_r(x^*)\}$ for a ball $\mathbb{B}_r(x^*) = \{x : \|x - x^*\| \leq r\}$ of radius $r$ centered on the (necessarily unique) equilibrium $x^*$ of $\mathcal{G}$. Then:*

$$\mathbb{E}_{x_0}[\tau_r] \leq (F_0/\alpha)\big/(r^2 - r_\sigma^2) \quad \textit{for all } r > r_\sigma, \tag{16}$$

*where $r_\sigma \coloneqq \sigma_{\max}/\sqrt{2K\alpha}$. If, in addition, $\sigma_{\min} > 0$ and $x^*$ is interior, $X(t)$ admits an invariant distribution concentrated in a ball of radius $\mathcal{O}(\sigma_{\max})$ around $x^*$, and we have*

$$\lim\nolimits_{t\to\infty} \mu_t(\mathbb{B}_r(x^*)) \geq 1 - r_\sigma^2/r^2 \quad \textit{for all } r > r_\sigma. \tag{17}$$

[3] By the martingale representation theorem [50, Thm. 4.3.4], the loss in generality is negligible in our case.

Conceptually, Theorems 1 and 2 reflect the dichotomy between the bilinear and quadratic examples studied in detail in Section 3.1. Indeed, we see that:

1. In *null-monotone games*, the stochastic dynamics (S-FTRL) exhibit a consistent drift *away from equilibrium*, moving to greater distances in finite time, and requiring infinite time to return. As a result, if the game has an interior equilibrium, $X(t)$ becomes infinitely spread out in the long run, exhibiting no concentration in *any* region of $\mathcal{X}$ other than, possibly, its boundary (if $\mathcal{X}$ is constrained).
2. In *strongly monotone games*, the dynamics drift *toward equilibrium*, and they end up being concentrated around the game's (necessarily unique) equilibrium. However, the players' learning trajectories continue to fluctuate at a distance which scales as $\mathcal{O}(\sigma_{\max})$ and, with probability 1, they return arbitrarily close to where they started, infinitely often.

These properties paint a sharp separation between null- and strongly monotone games, with uncertainty carrying drastically different consequences in each case; for an illustration, see Fig. 1.

The proof of Theorems 1 and 2 is detailed in Appendix D. From a technical standpoint, our analysis hinges on the use of the Fenchel coupling (13) as a "mean" energy function for the dynamics. In the null-monotone case, the hitting time estimates (15) rely on an application of Dynkin's formula [50, Chap. 7.4], coupled with an eigenvalue estimation for the growth of $F$. Then, by descending to a specific quotient of $\mathcal{Y}$ that compactifies the sublevel sets of $F$, we are able to leverage the fact that $\mathbb{E}_{x_0}[\tau_r] = \infty$ for $r < F_0$ to show that the dynamics are *not* positively recurrent—and hence, they *do not* admit an invariant distribution. The analysis for the strongly monotone case has the same starting point, but it then branches out almost immediately: the hitting time estimate (16) is again obtained via Dynkin's stopping time formula, but positive recurrence can no longer be established in $\mathcal{X}$, because the infinitesimal generator of $X(t)$ is not uniformly elliptic (that is, its eigenvalues are not bounded away from zero). Instead, we work directly with the infinetisimal generator of the score process $Y(t)$ whose generator *is* uniformly elliptic after taking a specific quotient in $\mathcal{Y}$. This allows us to deduce positive recurrence in $\mathcal{Y}$, which we then push forward to $\mathcal{X}$ via $Q$, and leverage the convergence of the occupation measures to the invariant distribution of the process to derive the concentration bound (17). We detail these steps in a series of technical lemmas in Appendix D.

## 4 Learning under uncertainty in discrete time

We now turn to the discrete-time setting—our ultimate goal from the outset owing to its high practical and algorithmic relevance. A key point to keep in mind here is that the discrete-time analysis is significantly more intricate and difficult because of the absence of closed-form solutions and the inapplicability of diffusion-based techniques. Nevertheless, as we shall see later in this section, the structural insights gained from the continuous-time analysis of Section 3 remain highly valuable in this context because they end up shaping the foundations of our tools and techniques.

**4.1. Learning in discrete time.** The most widely used algorithmic framework for multi-agent learning in games adheres again to the "follow-the-regularized-leader" template which unfolds over discrete time instances $t = 0, 1, \dots$ as

$$Y_{i,t+1} = Y_{i,t} + \gamma \hat{v}_{i,t} \qquad X_{i,t+1} = Q_i(Y_{i,t+1})\,. \tag{FTRL}$$

In addition to the notions already introduced and discussed in Section 3.2, $(i)$ $\hat{v}_{i,t}$ denotes here a stochastic estimate of the player's payoff gradient vector at $X_{i,t}$; and $(ii)$ $\gamma > 0$ is a *step-size* parameter, interchangeably referred to as the *learning rate* of the process. We discuss these new elements below.

**The feedback process.** In terms of feedback, we assume that, at each round $t = 0, 1, \dots,$ each player receives stochastic gradient feedback of the form

$$\hat{v}_{i,t} = \mathsf{V}_i(X_t;\omega_t) \quad \text{or, aggregating over all players} \quad \hat{v}_t = \mathsf{V}(X_t;\omega_t) \tag{18}$$

where $\hat{v}_t = (\hat{v}_{i,t})_{i\in\mathcal{N}}$ and $\mathsf{V}(x;\omega) = (\mathsf{V}_i(x;\omega))_{i\in\mathcal{N}}$ is a *stochastic first-order oracle* of the form

$$\mathsf{V}(x;\omega) = v(x) + \mathsf{U}(x;\omega)\,. \tag{SFO}$$

In the above, $\omega_t$, $t = 0, 1, \dots,$ is an i.i.d. sequence of random seeds drawn from some complete probability space $\Omega$, and $\mathsf{U}(x;\omega)$ is a random $\mathcal{Y}$-valued vector satisfying the standard assumptions

$$\mathbb{E}_\omega[\mathsf{U}(x;\omega)] = 0 \qquad \text{and} \qquad \mathbb{E}_\omega[\|\mathsf{U}(x;\omega)\|_*^2] \le \sigma^2 \tag{19}$$

for some $\sigma > 0$. In this way, letting $\mathcal{F}_t$, $t = 0, 1, \dots$, denote the history of the process up to time $t$, and writing $U_t \coloneqq \mathsf{U}(X_t; \omega_t)$ for the noise in the players' gradient feedback at time $t$, we get

$$\hat{v}_t = v(X_t) + U_t \quad \text{with} \quad \mathbb{E}[U_t \mid \mathcal{F}_t] = 0 \text{ and } \mathbb{E}[\|U_t\|_*^2 \mid \mathcal{F}_t] \leq \sigma^2. \tag{20}$$

Following standard practice in the field—see e.g., [59, 62] and references therein—we further assume that the probability distribution $\nu_x$ of $\mathsf{U}(x)$ decomposes as $\nu_x = \nu_x^c + \nu_x^\perp$ where: (*a*) $\nu_x^\perp$ is singular relative to the Lebesgue measure $\lambda_{\mathcal{Y}}$ on $\mathcal{Y}$; (*b*) $\nu_x^c$ is absolutely continuous relative to $\lambda_{\mathcal{Y}}$; and (*c*) the density $p_x(y)$ of $\nu_x^c$ is jointly continuous in $x$ and $y$, and it satisfies $\inf_{x\in\mathcal{K}} p_x(y) > 0$ for every compact set $\mathcal{K} \subseteq \mathcal{X}$ and all $y \in \mathcal{Y}$. This last assumption is relatively mild and ensures that the noise retains a non-degenerate, smooth component across $\mathcal{X}$, much like the assumption $\sigma_{\min} > 0$ for the diffusion matrix of (S-FTRL) in Section 3. This condition is trivially satisfied by most continuous error distributions in practice, and it can always be enforced by injecting a small uniform Gaussian noise component into the process, a technique which is widely used in both optimization and reinforcement learning to promote sufficient exploration and avoid degeneracy issues and saddle-points [6, 19, 35, 61]. In such cases, the density of the absolutely continuous component is strictly positive everywhere and independent of $x \in \mathcal{X}$, so the uniform lower bound condition holds trivially.

**The algorithm's learning rate.** The second feature which sets the discrete-time framework apart is the method's learning rate $\gamma$. Here and throughout, we consider a *constant* learning rate schedule; this should be contrasted to the stochastic approximation literature [6, 9, 33], where (FTRL) is run with a *vanishing* step-size $\gamma_t \to 0$, typically satisfying some form of the Robbins–Monro summability conditions $\sum_t \gamma_t = \infty$, $\sum_t \gamma_t^2 < \infty$. In many cases, the use of a vanishing step-size enables convergence of the algorithm because it dampens the impact of the noise over time [44]; at the same time however, this often comes at the price of slowing the algorithm down to a crawl over practical timescales. For this reason, we focus here exclusively on the constant step-size case, which is much more common in practice owing to its simplicity, robustness, and often superior empirical performance—but whose theoretical behavior remains far less understood.

**4.2. Analysis and results.** We now have the necessary groundwork in place to present our results for (FTRL). Before doing so, we should stress that the discrete-time analysis is, by necessity, more qualitative in nature than the more explicit, continuous-time results presented in Section 3. This gap is difficult to avoid: in continuous time, the rules of stochastic calculus comprise a very sharp set of tools with which to obtain closed-form estimates for the processes involved; by contrast, in discrete time, even the most basic tools of stochastic analysis—like Dynkin's formula—are dulled down because of measurability and subsampling issues.

As before, we split our focus between null- and strongly monotone games.

**The null-monotone regime.** A key take-away from the analysis of Section 3 is that, in null-monotone games, uncertainty causes the dynamics of regularized learning to spread out, diverging to infinity on average, without concentrating at any region of $\mathcal{X}$ other than its boundary. Our first result below shows that a version of this tenet continues to hold in discrete time:

**Theorem 3** (Null-monotone games)**.** *Suppose that* (FTRL) *is run in a null-monotone game $\mathcal{G}$ with a strongly convex $h^*$, let $x^*$ be an equilibrium of $\mathcal{G}$, and let $F_t = F(x^*, Y_t)$. Then* $\lim_{t\to\infty} \mathbb{E}[F_t] = \infty$.

This result shows that (FTRL) drifts away to infinity on average—though, of course, as in the continuous-time case, this does not mean that this occurs with probability 1. What is missing from Theorem 3 relative to Theorem 1 is a bound on the mean time required for $F_t$ to increase or decrease by $\varepsilon$. In the absence of a *consistent* drift component, our continuous-time estimates were only made possible through the use of stochastic calculus; in discrete time however, $X_t$ evolves in *discrete*, driftless jumps, introducing overshoots and upcrossings that render this question significantly harder. We conjecture that similar bounds do hold in discrete time, but we leave this open as a conjecture.

**The strongly monotone regime.** We now turn to the long-run behavior of (FTRL) in strongly monotone games. Based in no small part on the continuous-time analysis of the previous section, our goal will be to understand the distributional properties of the dynamics, with a particular focus on (*a*) the existence and uniqueness of an invariant measure; and (*b*) the extent to which this measure is concentrated around the game's equilibrium—which, in turn, quantifies the long-run proximity of the iterates of (FTRL) to equilibrium. With all this in mind, our results can be stated as follows:

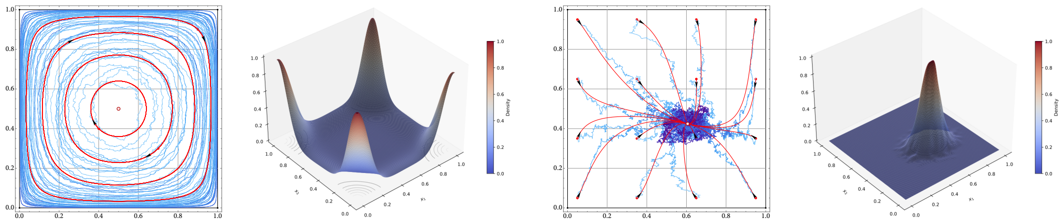


**Figure 1:** Trajectories and statistics of play under (FTRL) with entropic regularization in two min-max games over $\mathcal{X} = [0,1]^2$, a bilinear and a quadratic one (left vs. right half respectively). Deterministic orbits are plotted in red and stochastic trajectories in shades of blue, with darker hues indicating later points in time; the density plots depict the resulting visitation frequency in $\mathcal{X}$. In tune with Theorems 3 and 4, we see that learning in null-monotone games drifts toward the *extremes* of $\mathcal{X}$; by contrast, in strongly monotone games, learning orbits drift toward equilibrium, but continue to fluctuate around it. More details are provided in Appendix F.

**Theorem 4** (Strongly monotone games)**.** *Suppose that* (FTRL) *is run in an $\alpha$-strongly monotone game $\mathcal{G}$, let $r_\sigma \coloneqq \sqrt{\gamma(\sigma^2+\beta^2)/(\alpha K)}$, and consider the hitting time $\tau_r \coloneqq \inf\{t > 0 : X_t \in \mathbb{B}_r(x^*)\}$ for a ball $\mathbb{B}_r(x^*) = \{x : \|x - x^*\| \leq r\}$ of radius $r$ centered on the* (*necessarily unique*) *equilibrium $x^*$ of $\mathcal{G}$. Then, for all $r > r_\sigma$, we have*

$$\mathbb{E}[\tau_r] \leq \frac{1}{\alpha\gamma(r^2 - r_\sigma^2)} \times \begin{cases} F_0 & \text{if } X_0 \notin \mathbb{B}_r(x^*), \\ F_0 + \alpha\gamma r^2 & \text{if } X_0 \in \mathbb{B}_r(x^*), \end{cases} \tag{21}$$

*where $F_0 = F(x^*, Y_0)$. If, in addition, $x^*$ is interior, $X_t$ admits a unique invariant distribution to which it converges in total variation, and we have*

$$\lim_{t\to\infty} \frac{1}{t}\,\mathbb{E}\left[\sum_{s=0}^{t} \mathbb{1}\{X_t \in \mathbb{B}_r(x^*)\}\right] \geq 1 - r_\sigma^2/r^2 \tag{22}$$

*for all $r > r_\sigma$ such that $\mathbb{B}_r(x^*) \subseteq \operatorname{ri}\mathcal{X}$.*

*Remark.* Unlike the continuous-time setting of Section 3, we must treat the cases $X_0 \in \mathbb{B}_r(x^*)$ and $X_0 \notin \mathbb{B}_r(x^*)$ separately. This distinction arises only in discrete time, because the iterates may exhibit large jumps—so, returning to set $\mathbb{B}_r(x^*)$ is not guaranteed, even if the process is initialized within. ❦

We prove Theorem 4 in Appendix E following the main steps outlined below. First, shadowing our continuous-time analysis, we reduce the dynamics to a suitable quotient space of $\mathcal{Y}$, eliminating redundant directions and ensuring that the process evolves in a minimal, non-degenerate domain. Building on this, we show that the induced dynamics are Lebesgue-irreducible, i.e., every measurable set with positive Lebesgue measure is reachable with positive probability under the transition kernel of the process. At the same time, invoking (21), we also deduce that $\mathbb{P}(\tau_r < \infty) = 1$ for any initial condition, implying that $\mathbb{B}_r(x^*)$ is visited infinitely often. In a similar vein, we also show that $\mathbb{B}_r(x^*)$ satisfies a minorization condition, meaning that the transition kernel from any point in the ball dominates a fixed reference measure. This implies that, upon returning to $\mathbb{B}_r(x^*)$, the process has a nonzero chance of "forgetting" its past, allowing us to construct a regeneration structure via a coupling argument. Finally, leveraging the continuity of $F(x^*, y)$, we obtain a uniform bound on the expected return times $\mathbb{E}[\tau_r]$ over any initialization in $\mathbb{B}_r(x^*)$, which allows us to conclude that the process $Y_t$ is positive Harris recurrent. As a result, the iterates converge to a unique invariant measure, and we obtain quantitative control over their long-run concentration by means of our previous estimates.

## 5 Concluding remarks

Our aim in this paper was to quantify the impact of noise and uncertainty on the dynamics of multi-agent regularized learning. Our findings reveal a sharp separation between games that are *null-monotone* (like bilinear min-max games), and *strongly monotone* games (like Kelly auctions or Cournot competitions). In the former case, the quasi-periodic profile of the deterministic dynamics is destroyed, and learning under uncertainty drifts away on average toward extreme points (or escapes to infinity); in the latter, the sharp convergence guarantees of the deterministic dynamics are diluted by noise, and the resulting dynamics end up concentrated in a region around the game's equilibrium (which we estimate). This paves the way for further explorations of the long-run statistics of regularized learning in games—especially pertaining to the invariant measure of the process—a topic which we find particularly promising for advancing our understanding of the field.

# 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] 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.

[3] 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.

[4] Auer, P., Cesa-Bianchi, N., Freund, Y., and Schapire, R. E. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

[5] Belmega, E. V., Lasaulce, S., and Debbah, M. Power allocation games for MIMO multiple access channels with coordination. *IEEE Trans. Wireless Commun.*, 8(5):3182–3192, June 2009.

[6] 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.

[7] Bertsekas, D. P. *Convex optimization algorithms*. Athena Scientific, 2015.

[8] Bhattacharya, R. N. Criteria for recurrence and existence of invariant measures for multidimensional diffusions. *The Annals of Probability*, 6:541–553, 1978.

[9] Borkar, V. S. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press and Hindustan Book Agency, 2008.

[10] Bravo, M. and Mertikopoulos, P. On the robustness of learning in games with stochastically perturbed payoff observations. *Games and Economic Behavior*, 103(John Nash Memorial issue):41–66, May 2017.

[11] Cabrales, A. Stochastic replicator dynamics. *International Economic Review*, 41(2):451–81, May 2000.

[12] Cesa-Bianchi, N. and Lugosi, G. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[13] Chen, G. and Teboulle, M. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, August 1993.

[14] Douc, R., Moulines, E., Priouret, P., and Soulier, P. *Markov chains*. Operation research and financial engineering. Springer, 2018. doi: 10.1007/978-3-319-97704-1. URL https://hal.science/hal-02022651.

[15] Engel, M. and Piliouras, G. A stochastic variant of replicator dynamics in zero-sum games and its invariant measures. *Physica D: Nonlinear Phenomena*, 456:133940, December 2023.

[16] Folland, G. B. *Real Analysis*. Wiley-Interscience, 2 edition, 1999.

[17] Foster, D. and Young, H. P. Stochastic evolutionary game dynamics. *Theoretical Population Biology*, 38: 219–232, 1990.

[18] Fudenberg, D. and Harris, C. Evolutionary dynamics with aggregate shocks. *Journal of Economic Theory*, 57(2):420–441, August 1992.

[19] Ge, R., Huang, F., Jin, C., and Yuan, Y. Escaping from saddle points – Online stochastic gradient for tensor decomposition. In *COLT '15: Proceedings of the 28th Annual Conference on Learning Theory*, 2015.

[20] Hernández-Lerma, O. and Lasserre, J.-B. *Markov Chains and Invariant Probabilities*. Birkhäuser Basel, 2003. doi: 10.1007/978-3-0348-8024-4.

[21] Hiriart-Urruty, J.-B. and Lemaréchal, C. *Fundamentals of Convex Analysis*. Springer, Berlin, 2001.

[22] Hofbauer, J. and Imhof, L. A. Time averages, recurrence and transience in the stochastic replicator dynamics. *The Annals of Applied Probability*, 19(4):1347–1368, 2009.

[23] Huang, K. and Zhang, S. New first-order algorithms for stochastic variational inequalities. *SIAM Journal on Optimization*, 32(4):2745–2772, 2022.

[24] Imhof, L. A. The long-run behavior of the stochastic replicator dynamics. *The Annals of Applied Probability*, 15(1B):1019–1045, 2005.

[25] Itô, K. Stochastic integral. *Proceedings of the Imperial Academy of Tokyo*, 20:519–524, 1944.

[26] Kallenberg, O. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, 2002. ISBN 0-387-95313-2.

[27] Karatzas, I. and Shreve, S. E. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, Berlin, 1998.

[28] Khasminskii, R. Z. *Stochastic Stability of Differential Equations*. Number 66 in Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2 edition, 2012.

[29] Krichene, W. *Continuous and discrete dynamics for online learning and convex optimization*. PhD thesis, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, 2016.

[30] Krichene, W. and Bartlett, P. Acceleration and averaging in stochastic descent dynamics. In *NIPS '17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.

[31] Krichene, W., Drighès, B., and Bayen, A. M. Online learning of Nash equilibria in congestion games. *SIAM Journal on Control and Optimization*, 53(2):1056–1081, 2015.

[32] Kuo, H.-H. *Introduction to Stochastic Integration*. Springer, Berlin, 2006.

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

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

[35] Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. Continuous control with deep reinforcement learning. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.

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

[37] Loizou, N., Berard, H., Gidel, G., Mitliagkas, I., and Lacoste-Julien, S. Stochastic gradient descent-ascent and consensus optimization for smooth games: Convergence analysis under expected co-coercivity. In *NeurIPS '21: Proceedings of the 35th International Conference on Neural Information Processing Systems*, 2021.

[38] McKelvey, R. D. and Palfrey, T. R. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(6):6–38, 1995.

[39] Mertikopoulos, P. and Moustakas, A. L. The emergence of rational behavior in the presence of stochastic perturbations. *The Annals of Applied Probability*, 20(4):1359–1388, July 2010.

[40] Mertikopoulos, P. and Sandholm, W. H. Learning in games via reinforcement and regularization. *Mathematics of Operations Research*, 41(4):1297–1324, November 2016.

[41] Mertikopoulos, P. and Staudigl, M. Convergence to Nash equilibrium in continuous games with noisy first-order feedback. In *CDC '17: Proceedings of the 56th IEEE Annual Conference on Decision and Control*, 2017.

[42] Mertikopoulos, P. and Staudigl, M. On the convergence of gradient-like flows with noisy gradient input. *SIAM Journal on Optimization*, 28(1):163–197, January 2018.

[43] Mertikopoulos, P. and Staudigl, M. Stochastic mirror descent dynamics and their convergence in monotone variational inequalities. *Journal of Optimization Theory and Applications*, 179(3):838–867, December 2018.

[44] 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.

[45] Mertikopoulos, P., Papadimitriou, C. H., and Piliouras, G. Cycles in adversarial regularized learning. In *SODA '18: Proceedings of the 29th annual ACM-SIAM Symposium on Discrete Algorithms*, 2018.

[46] Mertikopoulos, P., Hsieh, Y.-P., and Cevher, V. A unified stochastic approximation framework for learning in games. *Mathematical Programming*, 203:559–609, January 2024.

[47] Meyn, S., Tweedie, R. L., and Glynn, P. W. *Markov Chains and Stochastic Stability*. Cambridge Mathematical Library. Cambridge University Press, 2 edition, 2009.

[48] Michel, D. and Pardoux, É. An introduction to Malliavin calculus and some of its applications. In *Recent advances in stochastic calculus*, pp. 65–104. Springer, 1990.

[49] Nesterov, Y. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120(1): 221–259, 2009.

[50] Øksendal, B. *Stochastic Differential Equations*. Springer-Verlag, Berlin, 6 edition, 2013.

[51] Piliouras, G. and Shamma, J. S. Optimization despite chaos: Convex relaxations to complex limit sets via Poincaré recurrence. In *SODA '14: Proceedings of the 25th annual ACM-SIAM Symposium on Discrete Algorithms*, 2014.

[52] Rockafellar, R. T. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[53] Rockafellar, R. T. and Wets, R. J. B. *Variational Analysis*, volume 317 of *A Series of Comprehensive Studies in Mathematics*. Springer-Verlag, Berlin, 1998.

[54] Rosen, J. B. Existence and uniqueness of equilibrium points for concave $N$-person games. *Econometrica*, 33(3):520–534, 1965.

[55] Shalev-Shwartz, S. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011.

[56] 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*, pp. 1265–1272. MIT Press, 2006.

[57] Telatar, I. E. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications and Related Technologies*, 10(6):585–596, 1999.

[58] van Damme, E. *Stability and perfection of Nash equilibria*. Springer-Verlag, Berlin, 1987.

[59] Vlatakis, E. V., Giannou, A., Chen, Y., and Xie, Q. Stochastic methods in variational inequalities: Ergodicity, bias and refinements. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 4123–4131. PMLR, 02–04 May 2024.

[60] Vovk, V. G. Aggregating strategies. In *COLT '90: Proceedings of the 3rd Workshop on Computational Learning Theory*, pp. 371–383, 1990.

[61] Welling, M. and Teh, Y. W. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, pp. 681–688, 2011.

[62] Yu, L., Balasubramanian, K., Volgushev, S., and Erdogdu, M. A. An analysis of constant step size sgd in the non-convex regime: asymptotic normality and bias. NIPS '21, Red Hook, NY, USA, 2021. Curran Associates Inc. ISBN 9781713845393.

[63] Yu, W., Rhee, W., Boyd, S. P., and Cioffi, J. M. Iterative water-filling for Gaussian vector multiple-access channels. *IEEE Trans. Inf. Theory*, 50(1):145–152, 2004.

[64] 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.

[65] 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.

# A Examples

In this appendix, we provide a series of examples of games adhering to our basic assumptions regarding concavity and monotonicity.

**Example A.1** (Zero-sum bimatrix games)**.** A bimatrix game consists of two players, each with a finite set of actions $\mathcal{A}_i$, $i = 1, 2$, and a min-max objective function $L\colon \mathcal{A}_1 \times \mathcal{A}_2 \to \mathbb{R}$, typically encoded in a matrix $M \in \mathbb{R}^{\mathcal{A}_1 \times \mathcal{A}_2}$ with $M_{\alpha\beta} = L(\alpha, \beta)$ for all $\alpha \in \mathcal{A}_1$, $\beta \in \mathcal{A}_2$. The first player is cast in the role of the minimizer and the second player in that of the maximizer, so their corresponding payoff functions are defined as $u_1 = -L = -u_2$.

In the mixed extension of the game, each player can mix their actions by selecting a probability distribution—a *mixed strategy*—over $\mathcal{A}_i$, that is, an element $x_i$ of the probability simplex $\mathcal{X}_i \equiv \Delta(\mathcal{A}_i) = \{x_i \in \mathbb{R}_+^{\mathcal{A}_i} : \|x_i\|_1 = 1\}$. Accordingly, in matrix notation, the players' corresponding mixed payoffs are given by

$$u_1(x_1, x_2) = -x_1^\top M x_2 = -u_2(x_1, x_2) \tag{A.1}$$

so their individual gradient fields can be expressed as

$$v_1(x_1, x_2) = -M x_2 \qquad \text{and} \qquad v_2(x_1, x_2) = M^\top x_1 \tag{A.2}$$

for all $x_1 \in \mathcal{X}_1$ and all $x_2 \in \mathcal{X}_2$.

By definition, a mixed-strategy Nash equilibrium of a bimatrix zero-sum game satisfies

$$L(x_1^*, x_2) \leq L(x_1^*, x_2^*) \leq L(x_1, x_2^*) \quad \text{for all } x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2. \tag{A.3}$$

If, in addition, $x_1^*, x_2^*$ both have full support—that is, $x_1^* \in \operatorname{ri} \mathcal{X}_1$ and $x_2^* \in \operatorname{ri} \mathcal{X}_2$—we also have the "equalizing payoffs" condition

$$L(x_1^*, x_2) = L(x_1, x_2^*) \quad \text{for all } x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2 \tag{A.4}$$

which means that (A.3) binds identically. In this case, we readily get

$$\begin{aligned}\langle v_1(x_1, x_2), x_1 - x_1^* \rangle &+ \langle v_2(x_1, x_2), x_2 - x_2^* \rangle \\ &= u_1(x_1, x_2) - u_1(x_1^*, x_2) + u_2(x_1, x_2) - u_2(x_1, x_2^*) = 0\end{aligned} \tag{A.5}$$

for all $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2$, i.e., the game is null-monotone in the sense of Definition 1.

**Example A.2** (Cournot competition)**.** In the standard Cournot competition model, there is a finite set of *firms*, indexed by $i \in \mathcal{N} = \{1, \dots, N\}$, each providing the market with a quantity $x_i \in [0, B_i]$ of some good (or service) up to the firm's production budget $B_i$. Following the law of supply and demand, this good is priced following the simple linear model $P(x) = a - b \sum_i x_i$, i.e., as a linearly decreasing function of the total supply. Accordingly, in this model, the utility of firm $i$ is given by

$$u_i(x) = x_i P(x) - c_i x_i = \left[a - b \sum\nolimits_{j \in \mathcal{N}} x_j - c_i\right] x_i, \tag{A.6}$$

where $c_i$ represents the marginal production cost of firm $i$.

By a straightforward derivation, the players' individual payoff gradients are given by

$$v_i(x) = \frac{\partial u_i}{\partial x_i} = \left[a - b \sum\nolimits_{j \in \mathcal{N}} x_j - c_i\right] - b x_i \tag{A.7}$$

and hence, the Hessian matrix of the game will be

$$H_{ij}(x) \coloneqq \frac{1}{2} \frac{\partial^2 u_i}{\partial x_j \partial x_i} + \frac{1}{2} \frac{\partial^2 u_j}{\partial x_i \partial x_j} = -b - b\delta_{ij} \tag{A.8}$$

where $\delta_{ij}$ is the standard Kronecker delta. Since $H$ is circulant, standard linear algebra considerations show that its eigenvalues are $-b$ and $-(N+1)b$ (with multiplicity $N-1$ and 1 respectively), so it follows by a well-known second-order criterion that the Cournot competition game is $b$-strongly monotone [44, 54].

**Example A.3** (Signal covariance optimization)**.** Consider a vector Gaussian channel of the form

$$\mathbf{y} = \sum_{i\in\mathcal{N}} \mathbf{H}_i \mathbf{x}_i + \mathbf{z} \tag{A.9}$$

where $\mathbf{x}_i \in \mathbb{C}^{m_i}$ is the (complex-valued) signal transmitted by the $i$-th user of the channel, $\mathbf{H} \in \mathbb{C}^{n\times m_i}$ is the transfer matrix of the channel, $\mathbf{z} \in \mathbb{C}^n$ is the noise in the channel (assumed zero-mean Gaussian and, without loss of generality, with unit covariance), and $\mathbf{y} \in \mathbb{C}^n$ is the aggregate signal output of the channel [63]. In this context, each user $i \in \mathcal{N}$ controls the covariance matrix $\mathbf{X}_i = \mathbb{E}[\mathbf{x}_i\mathbf{x}_i^\dagger]$ subject to the power constraint $\operatorname{tr}(\mathbf{X}_i) = \mathbb{E}[\|\mathbf{x}_i\|^2] \leq P_i$, where $P_i$ denotes the user's maximum transmit power. In this case, by the celebrated Shannon–Telatar formula [57], and assuming a single-user decoding scheme at the receiver, the achievable rate of the $i$-th user is

$$u_i(\mathbf{X}_i;\mathbf{X}_{-i}) = \log\det\Big(\mathbf{I} + \textstyle\sum_j \mathbf{H}_j\mathbf{X}_j\mathbf{H}_j^\dagger\Big) - \log\det\Big(\mathbf{I} + \textstyle\sum_{j\neq i} \mathbf{H}_j\mathbf{X}_j\mathbf{H}_j^\dagger\Big). \tag{A.10}$$

Putting everything together, this defines a continuous game with players $i \in \mathcal{N} = \{1,\dots,N\}$, spectrahedral action sets of the form

$$\mathcal{Q}_i = \{\mathbf{X}_i \in \mathbb{C}^{m_i\times m_i} : \mathbf{X}_i \succcurlyeq 0 \text{ and } \operatorname{tr}\mathbf{X}_i \leq P_i\} \tag{A.11}$$

for all $i \in \mathcal{N}$, and payoff functions given by (A.10). By a calculation of Belmega et al. [5], it is known that this game is concave and monotone—and, in fact, *strongly monotone* if the linear mapping $(\mathbf{X}_1,\dots,\mathbf{X}_N) \mapsto \sum_i \mathbf{H}_i\mathbf{X}_i\mathbf{H}_i^\dagger$ is not rank-deficient. ❦

# B Mirror maps and regularization

In this appendix, we collect some background material, properties and examples regarding the regularization machinery underlying (FTRL) and (S-FTRL). To lighten notation—especially with respect to the player index $i \in \mathcal{N}$—we base everything in this appendix on an abstract closed convex subset of some $d$-dimensional vector space, which could either be $\mathcal{X}_i$ or $\mathcal{X}$, depending on the context.

The results presented below (or a version thereof) are known in the literature; nevertheless, we provide detailed proofs for completeness and to resolve any conflicts or ambiguities with different conventions in the literature.

**B.1. Preliminaries.** Let $\mathcal{V}$ be a $d$-dimensional normed space, let $\mathcal{Y} \coloneqq \mathcal{V}^*$ denote the (algebraic) dual of $\mathcal{V}$, and let $\langle y, x\rangle$ denote the canonical bilinear pairing between $x \in \mathcal{V}$ and $y \in \mathcal{V}^*$. If $\|\cdot\|$ is a norm on $\mathcal{V}$ will also write

$$\|y\|_* = \max\{\langle y, x\rangle : \|x\| \leq 1\} \tag{B.1}$$

for the induced dual norm on $\mathcal{Y}$, so $|\langle y, x\rangle| \leq \|x\|\|y\|_*$ for all $x \in \mathcal{V}$ and all $y \in \mathcal{Y}$ by construction.

Given a closed convex subset $\mathcal{C}$ of $\mathcal{V}$, we also define:

1. The *tangent cone* to $\mathcal{C}$ at $p \in \mathcal{C}$ as

$$\operatorname{TC}(p) = \operatorname{cl}\{z \in \mathcal{V} : p + tz \in \mathcal{C} \text{ for some } t > 0\} \tag{B.2}$$

   i.e., as the closure of the set of rays emanating from $p$ and meeting $\mathcal{C}$ in at least one other point.

2. The *dual cone* to $\mathcal{C}$ at $p \in \mathcal{C}$ as

$$\operatorname{TC}^*(p) = \{w \in \mathcal{Y} : \langle w, z\rangle \geq 0 \text{ for all } z \in \operatorname{TC}(p)\} \tag{B.3}$$

3. The *polar cone* to $\mathcal{C}$ at $p \in \mathcal{C}$ as

$$\operatorname{PC}(p) = \{w \in \mathcal{Y} : \langle w, z\rangle \leq 0 \text{ for all } z \in \operatorname{TC}(p)\} \tag{B.4}$$

Following standard conventions in the field [52], convex functions will be allowed to take values in the extended real line $\mathbb{R} \cup \{\infty\}$, and we will denote the *effective domain* of a convex function $f\colon \mathcal{V} \to \mathbb{R} \cup \{\infty\}$ as

$$\operatorname{dom} f \coloneqq \{x \in \mathcal{V} : f(x) < \infty\}. \tag{B.5}$$

When there is no danger of confusion, we will identify a convex function $f\colon \mathcal{V} \to \mathbb{R}$ with its restriction on $\operatorname{dom} f$; in other words, we will treat $f$ interchangeably as a function on $\operatorname{dom} f$ with values in $\mathbb{R}$, or as a function on $\mathcal{V}$ with values in $\mathbb{R} \cup \{\infty\}$ (and finite on $\operatorname{dom} f$).

Throughout the sequel, we will assume that all functions under study are *proper*, that is, $\operatorname{dom} f \neq \varnothing$. Then, given a proper function $f : \mathcal{V} \to \mathbb{R} \cup \{\infty\}$, the *subdifferential* of $f$ at $x \in \operatorname{dom} f$ is defined as

$$\partial f(x) := \{y \in \mathcal{Y} : f(x') \geq f(x) + \langle y, x' - x\rangle \text{ for all } x' \in \mathcal{V}\} \tag{B.6}$$

and we denote the *domain of subdifferentiability* of $f$ as

$$\operatorname{dom} \partial f = \{x \in \mathcal{V} : \partial f(x) \neq \varnothing\}\,. \tag{B.7}$$

With all this in hand, a *regularizer* on a closed convex subset $\mathcal{C}$ of $\mathcal{V}$ is a continuous function $h : \mathcal{C} \to \mathbb{R}$ which is *strongly convex*, i.e., there exists some $K > 0$ such that

$$h(\lambda x + (1-\lambda)x') \leq t h(x) + (1-\lambda)h(x') - \frac{K}{2}\lambda(1-\lambda)\|x' - x\|^2 \tag{B.8}$$

for all $x, x' \in \mathcal{C}$ and for all $\lambda \in [0,1]$. By standard arguments [7, 53], this immediately implies that

$$h(x') \geq h(x) + \partial h(x; x' - x) + \frac{K}{2}\|x' - x\|^2 \quad \text{for all } x, x' \in \mathcal{X}, \tag{B.9}$$

where

$$\partial h(x; x' - x) = \lim_{\theta \to 0^+} [h(x + \theta(x' - x)) - h(x)]/\theta \tag{B.10}$$

denotes the one-sided directional derivative of $h$ at $x$ along the direction of $x' - x$. In addition, we also define the following objects associated to $h$:

1. The *prox-domain* of $h$:

$$\mathcal{C}_h := \operatorname{dom} \partial h \tag{B.11}$$

2. The *mirror map* $Q : \mathcal{Y} \to \mathcal{X}$ induced by $h$:

$$Q(y) := \arg\max_{x \in \mathcal{X}} \{\langle y, x\rangle - h(x)\} \qquad \text{for all } y \in \mathcal{Y}. \tag{B.12}$$

3. The *convex conjugate* $h^* : \mathcal{Y} \to \mathbb{R}$ of $h$:

$$h^*(y) := \max_{x \in \mathcal{X}} \{\langle y, x\rangle - h(x)\} \qquad \text{for all } y \in \mathcal{Y}. \tag{B.13}$$

The proposition below provides some basic properties linking all the above:

**Proposition B.1.** *Let $h$ be a $K$-strongly convex regularizer on $\mathcal{C}$. Then:*

(*a*) *$Q$ is single-valued on $\mathcal{Y}$.*

(*b*) *For all $x \in \mathcal{C}_h$ and all $y \in \mathcal{Y}$, we have*

$$x = Q(y) \quad \textit{if and only if} \quad y \in \partial h(x)\,. \tag{B.14}$$

(*c*) *The image* $\operatorname{im} Q$ *of $Q$ is equal to the prox-domain of $h$, and we have*

$$\operatorname{ri} \mathcal{C} \subseteq \operatorname{im} Q = \mathcal{C}_h \subseteq \mathcal{C}\,. \tag{B.15}$$

(*d*) *The convex conjugate $h^* : \mathcal{Y} \to \mathbb{R}$ of $h$ is differentiable and satisfies*

$$Q(y) = \nabla h^*(y) \quad \textit{for all } y \in \mathcal{Y}. \tag{B.16}$$

(*e*) *$Q$ is $(1/K)$-Lipschitz continuous, that is,*

$$\|Q(y') - Q(y)\| \leq (1/K)\|y' - y\|_* \quad \textit{for all } y, y' \in \mathcal{Y}. \tag{B.17}$$

(*f*) *Fix some $y \in \mathcal{Y}$ and let $x = Q(y)$. Then, for all $x' \in \mathcal{X}$ we have:*

$$\partial h(x; x' - x) \geq \langle y, x' - x\rangle\,. \tag{B.18}$$

(*g*) *Fix some $y \in \mathcal{Y}$, and let $x = Q(y)$. Then $Q(y + w) = x$ for all $w \in \operatorname{PC}(x)$.*

*Proof.* For the most part, these properties are well known in the literature (except possibly the last one), so we only provide a pointer or a short sketch for most of them.

($a$) This readily follows from the fact that $h$ is strongly convex, so the $\arg\max$ in (B.12) is attained and is unique for all $y \in \mathcal{Y}$.

($b$) By Fermat's rule [52, Chap. 26], we readily see that $x$ solves (B.12) if and only if $y - \partial h(x) \ni 0$, that is, if and only if $y \in \partial h(x)$. Since this implies that $\partial h$, our claim follows.

($c$) By (B.14), we readily get $\operatorname{im} Q = \mathcal{C}_h$. As for the second part of our claim, it follows from basic properties of the subdifferential, cf. Rockafellar [52, Chap. 26].

($d$) This is simply Danskin's theorem, see e.g., Bertsekas [7, Proposition 5.4.8, Appendix B].

($e$) This is a consequence of the fact that $h^*$ is $(1/K)$-Lipschitz smooth, cf. Rockafellar & Wets [53, Theorem 12.60(b)].

($f$) Since $y \in \partial h(x)$ by (B.14), we readily get that

$$h(x + \theta(x' - x)) \geq h(x) + \theta\langle y, x' - x\rangle \quad \text{for all } \theta \in [0,1]\,. \tag{B.19}$$

Hence, by rearranging and taking the limit $\theta \to 0^+$, we conclude that

$$\partial h(x; x' - x) = \lim_{\theta \to 0^+} \frac{h(x + \theta(x' - x)) - h(x)}{\theta} \geq \langle y, x' - x\rangle \tag{B.20}$$

as claimed.[4]

($g$) By (B.14) it suffices to show that $y + w \in \partial h(x)$ for all $w \in \mathrm{PC}(x)$. However, if $w \in \mathrm{PC}(x)$, we also have $\langle w, x' - x\rangle \leq 0$ for all $x' \in \mathcal{X}$, and hence, with $y \in \partial h(x)$, we readily get

$$\begin{aligned} h(x') &\geq h(x) + \langle y, x' - x\rangle \\ &\geq h(x) + \langle y + w, x' - x\rangle \quad \text{for all } x' \in \mathcal{X}. \end{aligned} \tag{B.21}$$

This shows that $y + w \in \partial h(x)$ and completes our proof. ■

Following [40, 44], we also define the *Fenchel coupling* associated to $h$ as

$$F(p, y) = h(p) + h^*(y) - \langle y, p\rangle \quad \text{for all } p \in \mathcal{X}, y \in \mathcal{Y}. \tag{B.22}$$

The next proposition shows that the Fenchel coupling can be seen as a "primal-dual" measure of divergence between $p \in \mathcal{C}$ and $y \in \mathcal{Y}$:

**Proposition B.2.** *Let $h$ be a $K$-strongly convex regularizer on $\mathcal{C}$. Then, for all $p \in \mathcal{X}$ and all $y \in \mathcal{Y}$, we have:*

($a$) $F(p, y) \geq 0$ *with equality if and only if* $p = Q(y)$. (B.23a)

($b$) $F(p, y) \geq \frac{1}{2}K\,\|Q(y) - p\|^2$. (B.23b)

*Proof.* These properties are known in the literature, but we provide a quick proof for completeness.

($a$) By the Fenchel–Young inequality, we have $h(p) + h^*(y) \geq \langle y, p\rangle$ for all $p \in \mathcal{X}$, $y \in \mathcal{Y}$, with equality if and only if $y \in \partial h(p)$. Our claim then follows from (B.14).

($b$) Let $x = Q(y)$ so $y \in \partial h(x)$ by (B.14). Then, by the definition of $F$, we have

$$\begin{aligned} F(p, y) &= h(p) + h^*(y) - \langle y, p\rangle \\ &= h(p) + \langle y, x\rangle - h(x) - \langle y, p\rangle && \text{\# since } y \in \partial h(x) \\ &\geq h(p) - h(x) - \partial h(x; p - x) && \text{\# by Proposition B.1} \\ &\geq \tfrac{1}{2}K\|x - p\|^2 && \text{\# by (B.8)} \end{aligned}$$

so our proof is complete. ■

Our last result at this point is a useful differentiation formula for the Fenchel coupling:

**Lemma B.1.** *For all $p \in \mathcal{X}$ and all $y \in \mathcal{Y}$, we have:*

$$\nabla_y F(p, y) = Q(y) - p\,. \tag{B.24}$$

*Proof.* The proof follows immediately from Danskin's theorem, cf. Eq. (B.16) of Proposition B.1. ■

[4] The existence of the limit is guaranteed by elementary convex analysis arguments, cf. Bertsekas [7, App. B].

**B.2. Update lemmas.** Moving forward, we note that the basic update step of (FTRL) can be written as

$$y^+ = y + w \quad \text{and} \quad x^+ = Q(y^+) \tag{B.25}$$

for some $y, w \in \mathcal{Y}$. With this in mind, we state below a series of identities and estimates for the Fenchel coupling before and after an update of the form (B.25).

The first is a primal-dual version of the so-called "three-point identity" for Bregman functions [13]:

**Lemma B.2.** *Fix some $p \in \mathcal{X}$, $y \in \mathcal{Y}$, and let $x = Q(y)$. Then, for all $y^+ \in \mathcal{Y}$, we have:*

$$F(p, y^+) = F(p, y) + F(x, y^+) + \langle y^+ - y, x - p \rangle. \tag{B.26}$$

*Proof.* By definition, we have:

$$F(p, y^+) = h(p) + h^*(y^+) - \langle y^+, p \rangle \tag{B.27a}$$
$$F(p, y) = h(p) + h^*(y) - \langle y, p \rangle \tag{B.27b}$$
$$F(x, y^+) = h(x) + h^*(y^+) - \langle y^+, x \rangle \tag{B.27c}$$

Thus, subtracting (B.27b) and (B.27c) from (B.27a), and rearranging, we get

$$F(p, y^+) = F(p, y) + F(x, y^+) - h(x) - h^*(y) + \langle y^+, x \rangle - \langle y^+ - y, p \rangle \,. \tag{B.28}$$

Our assertion then follows by recalling that $x = Q(y)$, so $h(x) + h^*(y) = \langle y, x \rangle$. ■

The next result we present concerns the Fenchel coupling before and after a direct update step; similar results exist in the literature, but we again provide a proof for completeness.

**Lemma B.3.** *Fix some $p \in \mathcal{X}$ and $y, w \in \mathcal{Y}$. Then, letting $x = Q(y)$, $y^+ = y + w$, and $x^+ = Q(y^+)$ as per (B.25), we have:*

$$F(p, y^+) = F(p, y) + \langle w, x^+ - p \rangle - F(x^+, y) \tag{B.29a}$$
$$\leq F(p, x) + \langle w, x - p \rangle + \frac{1}{2K}\|w\|_*^2 \,. \tag{B.29b}$$

*Proof.* By the three-point identity (B.26), we have

$$F(x, y) = F(x, y^+) + F(x^+, x) + \langle y - y^+, x^+ - p \rangle \tag{B.30}$$

so our first claim is immediate. For our second claim, rearranging terms and employing the Fenchel–Young inequality gives

$$\begin{aligned} &F(p, y) + \langle w, x^+ - p \rangle - F(x^+, y) \\ &\quad = F(p, y) + \langle w, x - p \rangle + \langle w, x^+ - x \rangle - F(p, y) \\ &\quad \leq F(p, y) + \langle w, x - p \rangle + \frac{1}{2K}\|w\|_*^2 + \frac{K}{2}\|x - p\|^2 - F(p, y) \end{aligned} \tag{B.31}$$

so our claim follows from Proposition B.2. ■

# C A short primer on stochastic analysis

In this appendix, we collect some standard results from stochastic analysis in order to provide a degree of self-completeness to the main text. For an introduction to stochastic analysis and the theory of SDEs, we refer the reader to the masterful accounts of Øksendal [50] and Kuo [32].

The main focus of the theory is the study of ordinary differential equations (ODEs) perturbed by noise, modeled informally after the *Langevin equation*

$$\frac{dZ}{dt} = b(Z(t)) + \eta(t) \tag{LE}$$

where $Z(t)$ is a stochastic process in $\mathbb{R}$, $b\colon \mathbb{R} \to \mathbb{R}$ is the *drift* of the process, and $\eta(t)$ is the "noise" perturbing the deterministic ODE $\dot{z} = b(z)$. Unfortunately, albeit natural, the problem with (LE) is that any reasonable continuous-time model of noise would lead to trajectories that are almost nowhere differentiable, so the meaning of "$dZ/dt$" in (LE) is rather precarious.[5]

In lieu of this, to give formal meaning to (LE), we consider instead the *stochastic differential equation*

$$dZ(t) = b(Z(t))\, dt + \sigma(Z(t))\, dW(t) \tag{SDE}$$

which is shorthand for the integral equation

$$Z(t) = \int_0^t b(Z(s))\, ds + \int_0^t \sigma(Z(s))\, dW(s)\,. \tag{C.1}$$

for some state-dependent *diffusion coefficient* $\sigma\colon \mathbb{R} \to \mathbb{R}$. The key element in the above formulation is the so-called *Itô integral* that appears in the right-hand side of (SDE), and which is defined relative to what is known as a standard *Brownian motion* on $\mathbb{R}$. Intuitively, what this means is that the integral $\int_0^t \sigma(Z(s))\, dW(s)$ is obtained in the limit $\delta t = t_{k+1} - t_k \to 0$ of the discrete-time approximation

$$\int_0^t \sigma(Z(s))\, dW(s) \approx \sum_{k=1}^{\lceil t/\delta t \rceil} \sigma(Z(t_k))\, [W(t_{k+1}) - W(t_k)] \tag{C.2}$$

where $W(t)$ is some stochastic process that satisfies what one would expect from a "white noise" process (zero-mean, with independent increments), but is still "regular enough" to possess a reasonable behavior in the limit $\delta t \to 0$. These considerations lead to the formal definition of a Brownian motion—or, more precisely, the *Wiener process*—which is characterized by the following properties:

1. The increments of $W$ are *independent*, that is, for all $t, \tau > 0$, the future increments $W(t+\tau) - W(t)$ of $W$ are independent of its past values $W(s)$, $s < t$.
2. The increments of $W$ are *Gaussian*, that is, for all $t, \tau > 0$, the future increments $W(t+\tau) - W(t)$ of $W$ are normally distributed with mean 0 and variance $\tau$, i.e., $W(t+\tau) - W(t) \sim \mathcal{N}(0, \tau)$.
3. The sample paths of $W$ are *continuous* (a.s.), i.e., $W(t)$ is a continuous function of $t$ for almost every realization of $W$.

The existence of a process with the above properties is by no means a trivial affair, but it can constructed e.g., as the scaling limit of a random walk, or some other discrete-time stochastic processes with stationary independent increments.

Providing a more detailed account of the definition of $W(t)$ and the associated stochastic integral which appears in (SDE) is well beyond the scope of our paper; for an accessible introduction, we refer the reader to Øksendal [50, Chap. 2]. What is more important for our purposes is that, albeit non-differentiable, the solution $Z(t)$ still satisfies a certain version of the chain rule, known as *Itô's formula* [25]. Specifically, for any $C^2$ function $f\colon \mathbb{R} \to \mathbb{R}$, we have

$$df(Z(t)) = f'(Z(t)) b(Z(t))\, dt + \tfrac{1}{2} f''(Z(t)) \sigma^2(t)\, dt + f'(Z(t))\, \sigma(Z(t))\, dW(t) \tag{C.3}$$

or, more compactly:

$$df(Z(t)) = f'(Z(t))\, dZ(t) + \tfrac{1}{2} f''(Z(t))\, dZ(t) \cdot dZ(t) \tag{C.4}$$

where the product $dZ \cdot dZ$ is computed according to the rules of stochastic calculus [50]:

$$dt \cdot dt = 0 \quad dt \cdot dW(t) = 0 \quad \text{and} \quad dW(t) \cdot dW(t) = dt\,. \tag{C.5}$$

Thanks to Itô's formula, we can still do calculus with stochastic processes satisfying (SDE); the resulting set of differentiation rules is known as *Itô*—or *stochastic*—*calculus*.

For our purposes, we will consider multi-dimensional analogues of (SDE) where, mutatis mutandis, $(i)$ $Z(t)$ evolves in $\mathbb{R}^n$; $(ii)$ the drift of the process is given by a vector field $b\colon \mathbb{R}^n \to \mathbb{R}^n$; $(iii)$ $W(t)$ is an $m$-dimensional Brownian motion evolving in $\mathbb{R}^m$; and $(iv)$ $\sigma\colon \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is the *diffusion matrix* of the SDE. In this case, Itô's formula for a $C^2$ function $f\colon \mathbb{R}^n \to \mathbb{R}$ becomes

$$\begin{aligned} df(Z(t)) = \sum_{i=1}^n b_i(Z(t)) \frac{\partial f}{\partial z_i}\, dt + \frac{1}{2} \sum_{i,j=1}^n \sum_{k=1}^m \sigma_{ik}(Z(t)) \sigma_{jk}(Z(t)) \frac{\partial^2 f}{\partial z_i \partial z_j}\, dt \\ + \sum_{i=1}^n \sum_{k=1}^m \sigma_{ik}(Z(t)) \frac{\partial f}{\partial z_i}\, dW_k(t) \end{aligned} \tag{C.6}$$

[5] In particular, consider a noise process $\eta(t)$ which is *a*) *zero-mean:* $\mathbb{E}[\eta(t)] = 0$; *b*) *uncorrelated:* $\mathbb{E}[\eta(t_1)\eta(t_2)] = 0$ if $t_2 \neq t_1$; and *c*) *stationary*, in the sense that $\eta(t+s)$ and $\eta(t)$ are identically distributed for all $s > 0$. Then, *any* such process does not have continuous paths [50, p. 21].

In our analysis, we will also require a weaker version of Itô's formula for convex functions $f\colon \mathbb{R}^n \to \mathbb{R}$ that are not $C^2$ but are only $L$-Lipschitz smooth, i.e., $C^1$-smooth with $L$-Lipschitz continuous derivatives. We borrow the precise statement from [42, Proposition C.2] which, in our notation, gives

$$f(Z(t)) \leq f(Z(0)) + \sum_{i=1}^{n} \int_0^t \partial_{z_i} f(Z(s))\, dZ_i(s) + \frac{L}{2} \int_0^s \operatorname{tr}[\sigma(Z(s))\, \sigma(Z(s))^\top]\, ds \tag{C.7}$$

or, more explicitly,

$$\begin{aligned} f(Z(t)) \leq f(Z(0)) &+ \sum_{i=1}^{n} \int_0^t b_k(Z(s))\, \partial_{z_i} f(Z(s))\, ds + \frac{L}{2} \int_0^s \operatorname{tr}[\sigma(Z(s))\, \sigma(Z(s))^\top]\, ds \\ &+ \sum_{i=1}^{n} \sum_{k=1}^{m} \int_0^t \sigma_{ik}(Z(s))\, \partial_{z_i} f(Z(s))\, dW_k(s)\,. \end{aligned} \tag{C.8}$$

The deterministic part of (the strong version of) Itô's formula for $C^2$-smooth functions is captured by the so-called *infinitesimal generator* of (SDE), defined here as the differential operator $\mathcal{L}$ whose action on $f$ is given by

$$\mathcal{L} f(x) = \sum_{i=1}^{n} b_i(z) \frac{\partial f}{\partial z_i} + \frac{1}{2} \sum_{i,j=1}^{n} \sum_{k=1}^{m} \sigma_{ik}(z) \sigma_{jk}(z) \frac{\partial^2 f}{\partial z_i \partial z_j} \quad \text{for all } z \in \mathbb{R}^n. \tag{C.9}$$

Accordingly, Itô's formula can be written more compactly as

$$df(Z(t)) = \mathcal{L} f(Z(t))\, dt + \nabla_z f(Z(t))^\top\, \sigma(Z(t))\, dW(t)\,. \tag{C.10}$$

Thus, letting $\mathbb{P}_z(\cdot)$ denote the law of $Z$ initialized at $Z(0) \leftarrow z \in \mathbb{R}^n$, and writing $\mathbb{E}_z[\cdot]$ for the corresponding expectation, we readily get

$$\mathbb{E}_z[f(Z(t))] = f(z) + \mathbb{E}_z\left[\int_0^t \mathcal{L} f(Z(s))\, ds\right] \quad \text{for all } t \geq 0. \tag{C.11}$$

This shows that the infinitesimal generator of $Z$ captures precisely the mean part of the evolution of $f(Z(t))$ under (SDE). In fact, this simple expression admits a far-reaching generalization known as *Dynkin's formula* [50, Chap. 7.4]:

**Proposition C.1** (Dynkin's formula)**.** *Suppose that $Z(t)$ is initialized at $Z(0) \leftarrow z \in \mathbb{R}^n$. Then, for every bounded stopping time $\tau$ and every $C^2$-smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$, we have*

$$\mathbb{E}_z[f(Z(\tau))] = f(z) + \mathbb{E}_z\left[\int_0^\tau \mathcal{L} f(Z(s))\, ds\right]. \tag{C.12}$$

Moving forward, the matrix

$$A(z) = \sigma(z)\sigma(z)^\top \tag{C.13}$$

or, in components,

$$A_{ij}(z) = \sum_{k=1}^{m} \sigma_{ik}(z)\sigma_{jk}(z) \quad i, j = 1, \dots, n, \tag{C.14}$$

is known as the *principal symbol* of $\mathcal{L}$, and we say that $\mathcal{L}$ is *uniformly elliptic* if there exists some $c > 0$ such that $u^\top A(z) u \geq c\|u\|^2$ for all $z, u \in \mathbb{R}^n$ (that is, if the eigenvalues of $A(z)$ are positive and uniformly bounded away from 0). If this is the case, the noise in (SDE) is "uniformly exciting" in the sense that it does not vanish along any direction at any point of the state space of the process. Concretely, by standard results—see e.g., [48, Sec. 3.3.6.1] and references therein—this implies that every region of $\mathbb{R}^n$ is visited by $Z(t)$ with positive probability, viz.

$$\mathbb{P}_z(Z(t) = z' \text{ for some } t > 0) > 0 \quad \text{for all } z, z' \in \mathbb{R}^n. \tag{C.15}$$

If (SDE) is uniformly elliptic—i.e., if the infinitesimal generator thereof is uniformly elliptic—the behavior of $Z(t)$ can be further classified as *transient* or *recurrent*. Formally, these two fundamental notions are defined as follows:

**Definition C.1.** Suppose that (SDE) is initialized at some $z \in \mathbb{R}^n$. Then:

1. $Z(t)$ is *transient* from $z \in \mathbb{R}^n$ if it escapes every compact subset $\mathcal{K}$ of $\mathbb{R}^n$ in finite time, i.e., there exists some (possibly random) $T_{\mathcal{K}} < \infty$ such that

$$\mathbb{P}_z(Z(t) \notin \mathcal{K} \text{ for all } t \geq T_{\mathcal{K}}) = 1\,. \tag{C.16}$$

2. $Z(t)$ is *recurrent* relative to a compact subset $\mathcal{K}$ of $\mathbb{R}^n$ if the hitting time

$$\tau_{\mathcal{K}} = \inf\{t > 0 : Z(t) \in \mathcal{K}\} \tag{C.17}$$

is finite (a.s.). If, in addition, $\mathbb{E}[\tau_{\mathcal{K}}] < \infty$, we will say that $Z(t)$ is *positive recurrent*; otherwise, $Z(t)$ will be called *null recurrent*.

If (SDE) is uniformly elliptic, we have the following fundamental dichotomy:

**Theorem C.1** (Transience / recurrence dichotomy)**.** *Suppose that* (SDE) *is uniformly elliptic. Then:*

1. *If* (SDE) *is positive recurrent* (*resp. null recurrent*) *for some initial condition $z \in \mathbb{R}^n$ and some compact subset $\mathcal{K}$ of $\mathbb{R}^n$, then it is positive recurrent* (*resp. null recurrent*) *for every initial condition and every compact subset of $\mathbb{R}^n$.*
2. *If* (SDE) *is transient from some initial condition $z \in \mathbb{R}^n$, it is transient from every initial condition.*

For a more detailed version of Theorem C.1, we refer the reader to Bhattacharya [8, Proposition 3.1] who, to the best of our knowledge, was the first to state and prove this criterion. In words, Theorem C.1 simply states that, as long as (SDE) is uniformly elliptic, then it is *either transient or recurrent*; and if it is recurrent, it is *either positive or null recurrent*; no other outcome is possible. The choice of initialization or compact set in Definition C.1 does not matter (so, in particular, $Z$ cannot be transient from some region of $\mathbb{R}^n$ and recurrent from another). This crisp separation of regimes will play a major role in our analysis, and we will refer to it as the *transience / recurrence dichotomy*.

An important consequence of positive recurrence is that, under uniform ellipticity, $Z(t)$ admits a unique *invariant measure*, that is, a probability measure $\nu$ on $\mathbb{R}^n$ such that $Z(t) \sim \nu$ for all $t \geq 0$ whenever $Z(0) \sim \nu$. Importantly, the proviso that $\nu$ is a probability measure implies that $\nu(\mathbb{R}^n) < \infty$; if the process is null-recurrent, the semigroup of flows of (SDE) still admits an invariant meassure in the sense of Khasminskii [28], but this measure is no longer finite, i.e., $\nu(\mathbb{R}^n) = \infty$. Finally, if the process is transient, (SDE) does not admit such a measure.

# D Analysis and results in continuous time

We now proceed to prove the continuous-time results for (S-FTRL) that we presented in Section 3.

**D.1. Proofs omitted from Section 3.1.** We begin with the "gentle start" results of Section 3.1, which we restate below for convenience.

**Proposition 1.** *Suppose that* (S-GDA) *is run on the game* (6a) *with initial condition $x_0 \in \mathbb{R}^2$. Then:*

1. *$\lim_{t\to\infty} \mathbb{E}_{x_0}\left[\|X(t)\|_2^2\right] = \infty$, i.e., $X(t)$ escapes to infinity in mean square.*
2. *$\mathbb{E}_{x_0}[\tau_r] = \infty$ if $r < \|x_0\|$, i.e., $X(t)$ takes infinite time on average to get closer to equilibrium.*
3. *The limit $\mathcal{P}_\infty(x) = \lim_{t\to\infty} \mathcal{P}(x,t)$ does not exist, i.e., $X$ does not admit an invariant distribution.*

**Proposition 2.** *Suppose that* (S-GDA) *is run on the game* (6b) *with initial condition $x_0 \in \mathbb{R}^2$. Then:*

1. *$\lim_{t\to\infty} \mathbb{E}_{x_0}\left[\|X(t)\|_2^2\right] = \sigma^2$, i.e., the dynamics fluctuate at mean distance $\sigma$ from equilibrium.*
2. *The mean time required to get within distance $r$ of the game's equilibrium is bounded as*

$$\mathbb{E}_{x_0}[\tau_r] \leq \frac{1}{2}\frac{\|x_0\|_2^2 - r^2}{r^2 - \sigma^2} \qquad \text{for all } \sigma < r < \|x_0\|_2. \tag{10}$$

3. *The density of $X(t)$ is $\mathcal{P}(x,t) = [\pi\sigma^2(1 - e^{-2t})]^{-1} \exp\!\left(-\frac{\|x - e^{-t}x_0\|_2^2}{(1-e^{-2t})\sigma^2}\right)$. In particular, $X(t)$ converges in distribution to a Gaussian random variable centered at $0$, viz.*

$$\mathcal{P}_\infty(x) \equiv \lim_{t\to\infty} \mathcal{P}(x,t) = 1/(\pi\sigma^2) \cdot e^{-\|x\|_2^2/\sigma^2}\,. \tag{11}$$

*Proof of Proposition 1.* For our first claim, note that Itô's formula (C.6) applied to the function $f(x) = \|x\|_2^2$ under the dynamics (S-GDA) for the game (6a) readily yields the expression

$$d\big(\|X(t)\|_2^2\big) = 2X(t)\cdot dX(t) + \, dX(t)\cdot dX(t) = 2\sigma^2\, dt + \sigma\, X(t)\cdot dW(t)\,. \tag{D.1}$$

Hence, by (C.11), we get

$$\mathbb{E}_{x_0}[\|X(t)\|_2^2] = 2\sigma^2 t\,, \tag{D.2}$$

which proves our claim.

For our second claim, consider the hitting time $\tau = \inf\{t > 0 : \|X(t)\|_2 \le r\}$ with $r < \|x_0\|_2$, and assume that $\mathbb{E}[\tau] < \infty$. Then, by Dynkin's formula (Proposition C.1) applied to $f(x) = \|x\|_2^2$ and $\tau$, we readily get

$$\mathbb{E}_{x_0}[f(X(\tau))] = f(x_0) + \mathbb{E}_{x_0}\left[\int_0^\tau 2\sigma^2\, ds\right] = f(x_0) + 2\sigma^2\,\mathbb{E}_{x_0}[\tau] \ge \|x_0\|_2^2\,. \tag{D.3}$$

However, since $f(X(\tau)) = r^2$ by construction, we readily get $r^2 \ge \|x_0\|_2^2$, a contradiction. This shows that $\mathbb{E}_{x_0}[\tau] = \infty$, as asserted.

Finally, for our third claim, it is easy to check that (S-GDA) is uniformly elliptic under the stated assumptions. Thus, by Theorem C.1 and the fact that $\mathbb{E}_{x_0}[\tau] = \infty$, it follows that $X(t)$ cannot be positive recurrent. By the discussion following Theorem C.1, this implies that $X(t)$ does not admit an inveriant measure, so the density $\mathcal{P}(x,t)$ of $X(t)$ does not converge to a limit either. ■

*Proof of Proposition 2.* Under the dynamics (S-GDA) for the game (6b), each coordinate of $X(t)$ evolves as an Ornstein–Uhlenbeck process, viz.

$$dX_i(t) = -X_i(t)\, dt + \sigma\, dW_i(t) \quad \text{for } i = 1,2. \tag{D.4}$$

Since the processes are decoupled, we conclude by standard stochastic analysis arguments [32, Example 7.4.5] that

$$X_i(t) = X_i(0)e^{-t} + \sigma \int_0^t e^{-(t-s)}\, dW_i(s)\,. \tag{D.5}$$

In turn, by [32, Theorem 7.4.7], this implies that the transition probability kernel of $X_i(t)$ is given by

$$\mathcal{P}_i(x_i,t) = \frac{1}{\sigma\sqrt{\pi(1-e^{-2t})}} \exp\!\left(-\frac{(x_i - e^{-t}x_{i,0})^2}{(1-e^{-2t})\sigma^2}\right) \quad \text{for } i = 1,2, \tag{D.6}$$

that is, $X_i(t)$ follows a Gaussian distribution with mean $\mathbb{E}_{x_{i,0}}[X_i(t)] = x_{i,0}e^{-t}$ and variance

$$\mathbb{E}[X_i^2(t)] = \frac{\sigma^2}{2}\left(1 - e^{-2t}\right). \tag{D.7}$$

Our first and third claims then follow immediately.

For our second claim, note that the infinitesimal generator of $X(t)$ is now given by

$$\mathcal{L}f(x) = -\langle \nabla f(x), x\rangle + \frac{1}{2}\sigma^2 \Delta f(x)\,, \tag{D.8}$$

where $\Delta f \equiv \operatorname{tr} \nabla^2 f$ denotes the Laplacian of $f$. Then, Dynkin's formula (Proposition C.1) applied to $f(x) = \|x\|_2^2$ at the truncated hitting time $\tau_r \wedge t \equiv \min\{\tau_r, t\}$, $t > 0$, readily yields

$$\begin{aligned}
\mathbb{E}_{x_0}[\|X(\tau_r \wedge t)\|_2^2] &= \|x_0\|_2^2 + \mathbb{E}_{x_0}\left[\int_0^{\tau_r\wedge t} 2[\sigma^2 - \|X(s)\|_2^2]\, ds\right] \\
&\le \|x_0\|_2^2 + \mathbb{E}_{x_0}\left[\int_0^{\tau_r\wedge t} 2(\sigma^2 - r^2)\, ds\right] \\
&= \|x_0\|_2^2 + 2(\sigma^2 - r^2)\,\mathbb{E}_{x_0}[\tau_r \wedge t]\,.
\end{aligned} \tag{D.9}$$

Since $\|X(\tau_r \wedge r)\|_2^2 \ge r^2$ by construction (recall that $\|x_0\|_2 > r$), we get

$$\mathbb{E}_{x_0}[\tau_r \wedge t] \le \frac{\|x_0\|_2^2 - r^2}{2(r^2 - \sigma^2)} \quad \text{for all } t > 0. \tag{D.10}$$

Since $\mathbb{E}_{x_0}[\tau_r \wedge t]$ is uniformly bounded, our claim follows by taking the limit $t \to \infty$ (so $\tau_r \wedge t \to \tau_r$ pointwise), and invoking the dominated convergence theorem. ■

**D.2. General properties of the dynamics (S-FTRL).** We now proceed to establish the properties of the stochastic dynamics (S-FTRL) in the general case, for null- and strongly monotone games respectively. Before doing so, we begin with a result of a book-keeping nature (which is, however, necessary to ensure that the ensuing questions are meaningful).

**Proposition D.1.** *Suppose that $\sigma$ is Lipschitz continuous. Then, for every initial condition $X(0) \leftarrow x_0 = Q(y_0) \in \mathcal{X}$, the dynamics (S-FTRL) admit a unique strong solution that exists for all time.*

*Proof.* Note that the dynamics (S-FTRL) can be recast in fully autonomous form as

$$dY(t) = v(Q(Y(t)))\, dt + \sigma(Q(Y(t))) \cdot dW(t)\,. \tag{D.11}$$

Note further that $v$, $\sigma$ and $Q$ are all Lipschitz, by our standing assumptions for the game, our assumptions here, and Proposition B.1 respectively. In turn, this implies that the compositions $\tilde{v} = v \circ Q$ and $\tilde{\sigma} = \sigma \circ Q$ are likewise Lipschitz continuous, so our claim follows from the existence and uniqueness theorem for SDEs with Lipschitz data, see e.g., [50, Theorem 5.2.1]. ■

Our next result is an ancillary calculation responsible for much of the heavy lifting in the upcoming analysis.

**Proposition D.2.** *Fix a base point $p \in \mathcal{X}$ and consider the energy function*

$$E(y) \coloneqq F(p, y) = h(p) + h^*(y) - \langle y, p\rangle \quad \text{for } y \in \mathcal{Y}. \tag{D.12}$$

*Then, for every stopping time $\tau \geq 0$, we have*

$$\begin{aligned} E(Y(\tau)) - E(Y(0)) \leq \int_0^\tau \langle v(X(s)), X(s) - p\rangle\, ds + \frac{\sigma_{\max}^2}{2K}\tau \\ + \int_0^\tau (X(s) - p)^\top \sigma(X(s))\, dW(s)\,. \end{aligned} \tag{D.13}$$

*If, in particular, $Q$ is smooth, we have*

$$\begin{aligned} E(Y(\tau)) - E(Y(0)) = \int_0^\tau &\langle v(X(s)), X(s) - p\rangle\, ds \\ &+ \frac{1}{2}\int_0^\tau \operatorname{tr}[\Sigma(X(s))\ \operatorname{Jac} Q(Y(s))]\, ds \\ &+ \int_0^\tau (X(s) - p)^\top \sigma(X(s))\, dW(s)\,. \end{aligned} \tag{D.14}$$

*Proof.* Assume first that $Q$ is $C^1$-smooth; In this case, by Lemma B.1, we have $\nabla F(p, y) = Q(y) - p$, and hence,

$$\nabla^2 E(y) = \nabla^2 h^*(y) = \operatorname{Jac} Q(y)\,. \tag{D.15}$$

Thus, by Itô's formula (C.6), we readily get

$$\begin{aligned} dE(Y(t)) &= (X(t) - p) \cdot dY(t) + \frac{1}{2}\operatorname{tr}[\sigma^\top(X(t))\, \nabla^2 E(Y(t))\, \sigma(X(t))]\, dt \\ &= \langle v(X(t)), X(t) - p\rangle\, dt \\ &\quad + \frac{1}{2}\operatorname{tr}[\Sigma(X(t))\ \operatorname{Jac} Q(Y(t))]\, dt \\ &\quad + (X(t) - p)^\top \sigma(X(t))\, dW(t) \end{aligned} \tag{D.16}$$

so (D.14) follows.

Now, if $Q$ is not smooth, Proposition B.1 shows that it is still $(1/K)$-Lipschitz continuous, which, equivalently, means that $h^*$ is $(1/K)$-Lipschitz smooth. Thus, (D.13) follows by the weak Itô formula for Lipschitz smooth functions (C.7) applied to $h^*$, and noting that $\operatorname{tr}[\sigma(x)\sigma(x)^\top] \leq d\sigma_{\max}^2$. ■

**D.3. The null-monotone case.** We begin our analysis proper with our result for null-monotone games, which we restate below for convenience.

**Theorem 1** (Null-monotone games)**.** *Suppose that* (S-FTRL) *is run with a smooth mirror map $Q$ in a null-monotone game $\mathcal{G}$ that admits an interior equilibrium $x^*$, and consider the hitting times $\tau_\varepsilon^- := \inf\{t > 0 : F_t \leq F_0 - \varepsilon\}$ and $\tau_\varepsilon^+ := \inf\{t > 0 : F_t \geq F_0 + \varepsilon\}$. If $\sigma_{\min}^2 > 0$ and $\varepsilon > 0$ is small enough, then*

$$\mathbb{E}_{x_0}[\tau_\varepsilon^-] = \infty \quad \text{and} \quad \mathbb{E}_{x_0}[\tau_\varepsilon^+] \leq 2\varepsilon \big/ \big(\kappa\, \sigma_{\min}^2\big) \tag{15}$$

*for some positive constant $\kappa \equiv \kappa_\varepsilon > 0$; in addition, $X(t)$ does not admit a limiting distribution.*

*Proof.* We start with the decreasing case, where we argue by contradiction. Specifically, let $x^*$ be an equilibrium of $\mathcal{G}$, and assume that $\mathbb{E}_{x_0}[\tau_\varepsilon^-] < \infty$. Then, by applying Dynkin's formula to the energy function $E(y)$ at $\tau_\varepsilon^-$ for $p \leftarrow x^*$ (cf. Propositions C.1 and D.2), we readily get

$$\begin{aligned}
\mathbb{E}_{x_0}[E(Y(\tau_\varepsilon^-))] &= E(y_0) + \mathbb{E}_{x_0}\left[\int_0^{\tau_\varepsilon^-} \Big(\langle v(X(s)), X(s) - x^*\rangle + \tfrac{1}{2}\operatorname{tr}[\Sigma(X(s))\ \operatorname{Jac} Q(Y(s))]\Big)\, ds\right] \\
&= F_0 + \frac{1}{2}\,\mathbb{E}_{x_0}\left[\int_0^{\tau_\varepsilon^-} \operatorname{tr}[\Sigma(X(s))\ \operatorname{Jac} Q(Y(s))]\, ds\right] && \text{\# by null monotonicity} \\
&\geq F_0
\end{aligned} \tag{D.17}$$

where the last line follows from the fact that $\Sigma$ and $\operatorname{Jac} Q$ are both positive-semidefinite. However, since $\mathbb{E}_{x_0}[E(Y(\tau_\varepsilon^-))] = F_0 - \varepsilon$ by the definition of $\tau_\varepsilon^-$, we get $F_0 - \varepsilon \geq F_0$, a contradiction which establishes our claim.

Since $\sigma_{\min} > 0$, we further conclude that $Y(t)$ is uniformly elliptic. Thus, for any compact set $\mathcal{K} \subseteq \{y \in \mathcal{Y} : F(x^*, y) \leq F_0 - \varepsilon\}$, the hitting time $\tau_{\mathcal{K}} = \inf\{t > 0 : Y(t) \in \mathcal{K}\}$ will be infinite on average (because $\mathbb{E}_{x_0}[\tau_{\mathcal{K}}] \geq \mathbb{E}_{x_0}[\tau_\varepsilon^-] = \infty$), so, by Theorem C.1, $Y(t)$ cannot be positive recurrent. In turn, this implies that $Y(t)$ does not admit an invariant measure on $\mathcal{Y}$, which proves our claim.

Finally, for the second part of (15), applying Dynkin's formula to the energy function $E(y)$ for $p \leftarrow x^*$ at the truncated hitting times $\tau_\varepsilon^+ \wedge t$, $t > 0$, we get:

$$\begin{aligned}
\mathbb{E}_{x_0}[E(Y(\tau_\varepsilon^+) \wedge t)] &= E(y_0) + \mathbb{E}_{x_0}\left[\int_0^{\tau_\varepsilon^+ \wedge t} \Big(\langle v(X(s)), X(s) - x^*\rangle + \tfrac{1}{2}\operatorname{tr}[\Sigma(X(s))\ \operatorname{Jac} Q(Y(s))]\Big)\, ds\right] \\
&= F_0 + \frac{1}{2}\,\mathbb{E}_{x_0}\left[\int_0^{\tau_\varepsilon^+ \wedge t} \operatorname{tr}[\Sigma(X(s))\ \operatorname{Jac} Q(Y(s))]\, ds\right] && \text{\# by null monotonicity} \\
&\geq F_0 + \frac{\sigma_{\min}^2}{2}\,\mathbb{E}_{x_0}\left[\int_0^{\tau_\varepsilon^+ \wedge t} \operatorname{tr}[\operatorname{Jac} Q(Y(s))]\, ds\right]
\end{aligned} \tag{D.18}$$

where the last line follows from the estimate

$$\begin{aligned}
\operatorname{tr}[\Sigma \operatorname{Jac} Q] &= \operatorname{tr}[(\operatorname{Jac} Q)^{1/2}\Sigma(\operatorname{Jac} Q)^{1/2}] \\
&= (1, \dots, 1) \cdot (\operatorname{Jac} Q)^{1/2}\Sigma(\operatorname{Jac} Q)^{1/2} \cdot (1, \dots, 1)^\top \\
&\geq \sigma_{\min}^2 (1, \dots, 1) \cdot (\operatorname{Jac} Q)^{1/2} \cdot (\operatorname{Jac} Q)^{1/2} \cdot (1, \dots, 1)^\top \\
&= \sigma_{\min}^2 \operatorname{tr}[\operatorname{Jac} Q]\,.
\end{aligned} \tag{D.19}$$

By the assumptions of the theorem (smooth $Q$ and interior initialization), it follows that the (necessarily compact) set $\mathcal{D}_\varepsilon := \{x = Q(y) : F(x^*, y) \leq F_0 + \varepsilon\}$ is contained in the relative interior $\operatorname{ri}\mathcal{X}$ of $\mathcal{X}$. In turn, this implies that $\kappa \equiv \kappa_\varepsilon := \min\{\operatorname{tr}[\operatorname{Jac} Q(y)] : F(x^*, y) \leq F_0 + \varepsilon\} > 0$, so (D.18) becomes

$$F_0 + \varepsilon \geq F_0 + \frac{1}{2}\kappa\sigma_{\min}^2\, \mathbb{E}_{x_0}[\tau_\varepsilon^+ \wedge t] \tag{D.20}$$

and hence

$$\mathbb{E}_{x_0}[\tau_\varepsilon^+ \wedge t] \leq \frac{2\varepsilon}{\kappa\sigma_{\min}^2} \quad \text{for all } t \geq 0. \tag{D.21}$$

This shows that $\mathbb{E}_{x_0}[\tau_\varepsilon^+ \wedge t]$ is uniformly bounded, so the upper bound in (15) follows by letting $t \to \infty$ (which implies $\tau_\varepsilon^+ \wedge t \to \tau_\varepsilon^+$ pointwise), and invoking the dominated convergence theorem. ∎

**D.4. The strongly monotone case.** We now turn to our main result for strongly monotone games. Our proof strategy draws on methods related to the analysis of (S-FTRL) in the context of convex minimization, as explored by [42], and incorporating ideas that can be traced back to [24].

For convenience, we begin by restating Theorem 2.

**Theorem 2** (Strongly monotone games)**.** *Suppose that* (S-FTRL) *is run in an $\alpha$-strongly monotone game $\mathcal{G}$, and consider the hitting time $\tau_r := \inf\{t > 0 : X(t) \in \mathbb{B}_r(x^*)\}$ for a ball $\mathbb{B}_r(x^*) = \{x : \|x - x^*\| \leq r\}$ of radius $r$ centered on the (necessarily unique) equilibrium $x^*$ of $\mathcal{G}$. Then:*

$$\mathbb{E}_{x_0}[\tau_r] \leq (F_0/\alpha)\big/(r^2 - r_\sigma^2) \quad \textit{for all } r > r_\sigma, \tag{16}$$

*where $r_\sigma := \sigma_{\max}/\sqrt{2K\alpha}$. If, in addition, $\sigma_{\min} > 0$ and $x^*$ is interior, $X(t)$ admits an invariant distribution concentrated in a ball of radius $\mathcal{O}(\sigma_{\max})$ around $x^*$, and we have*

$$\lim_{t\to\infty} \mu_t(\mathbb{B}_r(x^*)) \geq 1 - r_\sigma^2/r^2 \quad \textit{for all } r > r_\sigma. \tag{17}$$

*Proof.* Our proof proceeds along the following basic steps:

*Step 1.* Deriving an estimate for the mean hitting time $\mathbb{E}_{x_0}[\tau_r]$.

*Step 2.* Descending to a restricted process $\tilde{Y}(t)$ where any redundant degrees of freedom in $Y(t)$ have been "modded out".

*Step 3.* Showing that the restricted process is positive recurrent.

*Step 4.* Estimating the resulting invariant distribution and pushing the result forward to $X(t)$.

In what follows, we go through the steps outlined above, one at a time.

**Step 1: Estimating the hitting time.** We begin with the hitting time estimate (16). To that end, setting $p \leftarrow x^*$ in Proposition D.2, we get

$$\begin{aligned} E(Y(\tau)) - E(y_0) &= \int_0^\tau \langle v(X(s)), X(s) - x^*\rangle\, ds + \frac{1}{2K}\int_0^\tau \operatorname{tr}[\Sigma(X(s))]\, ds \\ &\quad + \int_0^\tau (X(s) - x^*)^\top \sigma(X(s))\, dW(s) \\ &\leq -\alpha \int_0^\tau \|X(s) - x^*\|^2\, ds + \frac{\sigma_{\max}^2}{2K}\tau + M(\tau)\,. \end{aligned} \tag{D.22}$$

where we set

$$M(t) = \int_0^t (X(s) - x^*)^\top \sigma(X(s))\, dW(s)\,. \tag{D.23}$$

Thus, by a quick rearrangement, we obtain

$$\alpha \int_0^\tau \|X(s) - x^*\|^2\, ds \leq E(y_0) - E(Y(\tau)) + \frac{\sigma_{\max}^2 \tau}{2K} + M(\tau) \tag{D.24}$$

and hence, with $E \geq 0$:

$$\int_0^\tau \|X(s) - x^*\|^2\, ds \leq \frac{F_0}{\alpha} + \frac{\sigma_{\max}^2 \tau}{2\alpha K} + \frac{M(\tau)}{\alpha}\,. \tag{D.25}$$

Thus, applying the above to the truncated hitting time $\tau \leftarrow \tau_r \wedge t \equiv \min\{\tau_r, t\}$, $t > 0$, we get

$$\begin{aligned} r^2\, \mathbb{E}_{x_0}[\tau_r \wedge t] &\leq \mathbb{E}_{x_0}\left[\int_0^{\tau_r \wedge t} \|X(s) - x^*\|^2\, ds\right] && \text{\# b/c } \|X(s) - x^*\| \geq r \text{ for } s \leq \tau_r \wedge t \\ &\leq \frac{F_0}{\alpha} + r_\sigma^2\, \mathbb{E}_{x_0}[\tau_r \wedge t] + \frac{1}{\alpha}\mathbb{E}_{x_0}[M(\tau_r \wedge t)]\,. \end{aligned} \tag{D.26}$$

Since $\tau_r \wedge t \leq t$ is uniformly bounded, we will have $\mathbb{E}_{x_0}[M(\tau_r \wedge t)] = \mathbb{E}[M(0)] = 0$ by the optional sampling theorem for continuous-time martingales [27, Theorem 3.22]. Thus, a simple rearrangement gives

$$\mathbb{E}_{x_0}[\tau_r \wedge t] \leq \frac{F_0/\alpha}{r^2 - r_\sigma^2} \quad \text{for all } t \geq 0. \tag{D.27}$$

This shows that $\mathbb{E}_{x_0}[\tau_r \wedge t]$ is uniformly bounded, so the bound (16) follows by letting $t \to \infty$ (which implies $\tau_r \wedge t \to \tau_r$ pointwise), and invoking the dominated convergence theorem.

**Step 2: Descending to the restricted process.** We now proceed to examine the recurrence properties of $X(t)$. To that end, note first that the assumption $\sigma_{\min} > 0$ directly implies that (S-FTRL) is uniformly elliptic in the sense discussed in Appendix C. As such, consider the set

$$\mathcal{D}_r \coloneqq Q^{-1}(\mathbb{B}_r(x^*)) = \{y \in \mathcal{Y} : \|Q(y) - x^*\| \leq r\}\,. \tag{D.28}$$

and note that

$$\tau_r = \inf\{t > 0 : X(t) \in \mathbb{B}_r(x^*)\} = \inf\{t > 0 : Y(t) \in \mathcal{D}_r\} \tag{D.29}$$

so $Y(t)$ is positive recurrent relative to $\mathcal{D}_r$. Thus, if $\mathcal{D}_r$ is compact, Theorem C.1 immediately shows that $Y(t)$ is positive recurrent, and hence admits an invariant measure $\nu$ on $\mathcal{Y}$. In general however, $\mathcal{D}_r$ *need not be* compact, so we cannot conclude that $Y(t)$ is recurrent from the fact that it hits $\mathcal{D}_r$ in finite time on average.

To circumvent this difficulty, we will consider a "restricted" process which *is* positive recurrent, while retaining all information present in $Y(t)$. The main idea here will be to "collapse" the fibers of $Q$, that is, those directions in $\mathcal{Y}$ which map to the same point in $\mathcal{X}$ under $Q$: since the dynamics (S-FTRL) factor through $X(t) = Q(Y(t))$, these directions carry no relevant information, so they can be effectively discarded.

To carry all this out, let $\tilde{\mathcal{V}}$ denote the *"tangent hull"* of $\mathcal{X}$ in $\mathcal{V}$, viz.

$$\tilde{\mathcal{V}} \coloneqq \operatorname{aff}(\mathcal{X} - \mathcal{X}) = \{z \in \mathcal{V} : x + tz \in \mathcal{X} \text{ for all sufficiently small } t > 0 \text{ and all } x \in \operatorname{ri}\mathcal{X}\}\,. \tag{D.30}$$

In words, $\tilde{\mathcal{V}}$ is the smallest subspace of $\mathcal{V}$ which contains $\mathcal{X}$ when the latter is translated to the origin so, by construction, $\mathcal{X}$ is full-dimensional when viewed as a subset of $\tilde{\mathcal{V}}$.[6] In this sense, $\tilde{\mathcal{V}}$ contains all the "essential" directions of motion of the problem.

Dually to the above, we also consider the corresponding dual space $\tilde{\mathcal{Y}} \equiv \tilde{\mathcal{V}}^*$ of $\tilde{\mathcal{V}}$; this is not a subspace of $\mathcal{Y}$, but there exists a canonical surjection $\Pi\colon \mathcal{Y} \twoheadrightarrow \tilde{\mathcal{Y}}$ defined by restricting the action of $y \in \mathcal{Y}$ to $\tilde{\mathcal{V}}$, that is,

$$\langle \Pi(y), z\rangle = \langle y, z\rangle \quad \text{for all } z \in \tilde{\mathcal{V}}. \tag{D.31}$$

The kernel of $\Pi$ is precisely the *annihilator* $\operatorname{Ann}(\tilde{\mathcal{V}})$ of $\tilde{\mathcal{V}}$, i.e.,

$$\ker\Pi = \operatorname{Ann}(\tilde{\mathcal{V}}) \equiv \{w \in \mathcal{Y} : \langle w, z\rangle = 0 \text{ for all } z \in \tilde{\mathcal{V}}\} \tag{D.32}$$

so, by the first isomorphism theorem, we get a canonical identification $(\mathcal{V}/\tilde{\mathcal{V}})^* \cong \ker\Pi$.

The main reason for descending from $\mathcal{Y}$ to $\tilde{\mathcal{Y}}$ is the following: in the original space $\mathcal{Y}$, we have $Q(y + w) = Q(y)$ whenever $w$ annihilates $\tilde{\mathcal{V}}$, cf. Proposition B.1. As a result, the inverse image of *any* compact subset of $\mathcal{X}$ under $Q$ will always contain a copy of $\operatorname{Ann}(\tilde{\mathcal{V}})$, so it can never be compact itself. By contrast, by "modding out" $\operatorname{Ann}(\tilde{\mathcal{V}})$ and descending to the restricted space $\tilde{\mathcal{Y}}$, this is no longer the case.

To move forward, consider the *restricted mirror map* $\tilde{Q}\colon \tilde{\mathcal{Y}} \to \mathcal{X}$ given by

$$\tilde{Q}(\tilde{y}) = Q(y) \quad \text{whenever } \Pi(y) = \tilde{y}. \tag{D.33}$$

By the last item of Proposition B.1 we have $Q(y) = Q(y + w)$ whenever $w \in \operatorname{Ann}(\tilde{\mathcal{V}})$; this means that the choice of representative in (D.33) does not matter, so $\tilde{Q}$ is well-defined. Accordingly, letting

$$\tilde{Y}(t) = \Pi(Y(t)) \tag{D.34}$$

and applying $\Pi$ to (S-FTRL) yields the "restricted" dynamics

$$d\tilde{Y}(t) = d(\Pi \cdot Y(t)) = \Pi \cdot v(X(t))\,dt + \Pi \cdot \sigma(X(t)) \cdot dW(t) \tag{D.35}$$

where $X(t) = Q(Y(t)) = \tilde{Q}(\tilde{Y}(t))$ and, in a slight abuse of notation, we are overloading the symbol $\Pi$ to denote both the linear map $\Pi\colon \mathcal{Y} \to \tilde{\mathcal{Y}}$ and its representation as a matrix. These dynamics represent a time-homogeneous SDE in terms of $\tilde{Y}$, and they will be our main object of study in the rest of our proof.

[6] Specifically, unless $\mathcal{X}$ is a singleton, it has nonempty topological interior when viewed as a subset of $\tilde{\mathcal{V}}$.

**Step 3: Positive recurrence of the restricted process.** With all this in hand, positive recurrence for the restricted process $\tilde{Y}(t)$ boils down to the following: *a*) verifying that the infinitesimal generator of $\tilde{Y}$ is uniformly elliptic; and *b*) showing that the mean time required for $\tilde{Y}(t)$ to reach some compact set of $\tilde{\mathcal{Y}}$ is finite.

We begin by establishing uniform ellipticity. In view of (D.35), the principal symbol (C.13) of the infinitesimal generator of $\tilde{Y}(t)$ is

$$A = (\Pi \cdot \sigma) \cdot (\Pi \cdot \sigma)^\top = \Pi\sigma\sigma^\top\Pi^\top = \Pi\Sigma\Pi^\top\,. \tag{D.36}$$

Since $\Sigma \succcurlyeq \sigma_{\min}^2 I$, we readily get

$$A \succcurlyeq \sigma_{\min}^2 \Pi\Pi^\top \succcurlyeq \sigma_{\min}^2 \pi_{\min}^2 I \tag{D.37}$$

with $\sigma_{\min} > 0$ (by assumption) and $\pi_{\min} \coloneqq \lambda_{\min}(\Pi\Pi^\top) > 0$ (because $\Pi$ is surjective, so it has full rank). This shows that the principal symbol $\Pi\Sigma\Pi^\top$ of the generator of $\tilde{Y}$ is uniformly positive-definite, so $\tilde{Y}$ is itself uniformly elliptic.

For the second component of our proof of positive recurrence, recall that $x^* \in \operatorname{ri}\mathcal{X}$, so there exists some sufficiently small $r > 0$ such that the (compact convex) set

$$\mathcal{K}_r \coloneqq \mathbb{B}_r \cap \mathcal{X} = \{x \in \mathcal{X} : \|x - x^*\| \leq r\} \tag{D.38}$$

lies in its entirety within $\operatorname{ri}\mathcal{X}$. We then claim that the inverse image

$$\tilde{\mathcal{D}}_r \coloneqq \tilde{Q}^{-1}(\mathcal{K}_r) = \{\tilde{y} \in \tilde{\mathcal{Y}} : \|\tilde{Q}(\tilde{y}) - x^*\| \leq r\} \tag{D.39}$$

of $\mathcal{K}_r$ under the restricted mirror map $\tilde{Q}$ is compact. To see this, note first that $\tilde{\mathcal{D}}_r = \partial h(\mathcal{K}_r)$ by Proposition B.1.[7] Thus, given that $\mathcal{K}_r$ is a convex body in $\tilde{\mathcal{V}}$ that is entirely contained in the (relative) interior of the prox-domain $\mathcal{X}_h$ of $h$ (because $\operatorname{ri}\mathcal{X} \subseteq \operatorname{dom}\partial h \equiv \mathcal{X}_h$), it follows that $\partial h(\mathcal{K}_r)$ is itself compact by the upper hemicontinuity of $\partial h$ [21, Remark 6.2.3].

To conclude, note that

$$\begin{aligned}
\inf\{t > 0 : \tilde{Y}(t) \in \tilde{\mathcal{D}}_r\} &= \inf\{t > 0 : \|\tilde{Q}(\tilde{Y}(t)) - x^*\| \leq r\} \\
&= \inf\{t > 0 : \|Q(Y(t)) - x^*\| \leq r\} \\
&= \inf\{t > 0 : X(t) \in \mathbb{B}_r(x^*)\} \\
&= \tau_r
\end{aligned} \tag{D.40}$$

so it follows from (16) that $\tilde{Y}(t)$ hits $\tilde{\mathcal{D}}_r$ in finite time on average. Since $\tilde{Y}(t)$ is uniformly elliptic, Theorem C.1 shows that it is positive recurrent, as claimed.

**Step 4: Estimating the long-run occupation measure.** Since the restricted process $\tilde{Y}(t)$ is a positive recurrent Itô diffusion, standard results show that it admits an invariant distribution $\tilde{\nu}$ on $\tilde{\mathcal{Y}}$ which satisfies the law of large numbers

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(\tilde{Y}(s))\,ds = \int_{\tilde{\mathcal{Y}}} f\,d\tilde{\nu} \tag{D.41}$$

for every $\tilde{\nu}$-integrable test function $f$ on $\tilde{\mathcal{Y}}$. Thus, letting $\nu = \tilde{Q}_*\tilde{\nu} \equiv \tilde{\nu} \circ \tilde{Q}^{-1}$ denote the corresponding push-forward measure on $\mathcal{X}$, we get

$$\begin{aligned}
\lim_{t\to\infty} \mu_t(\mathbb{B}_r) &= \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}\{X(s) \in \mathbb{B}_r\}\,ds \\
&= \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}\{\tilde{Q}(\tilde{Y}(s)) \in \mathbb{B}_r\}\,ds \\
&= \lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}\{\tilde{Y}(s) \in \tilde{\mathcal{D}}_r\}\,ds \\
&= \int_{\tilde{\mathcal{Y}}} \mathbb{1}\{\tilde{y} \in \tilde{\mathcal{D}}_r\}\,d\tilde{\nu}(\tilde{y})
\end{aligned}$$

[7] Strictly speaking, we are viewing here $\partial h$ as taking values in $\tilde{\mathcal{Y}}$ instead of $\mathcal{Y}$; this is a simple matter of identifying $h\colon \mathcal{V} \to \mathbb{R}$ with its canonical restriction to $\tilde{\mathcal{V}} \subseteq \mathcal{V}$.

$$= \tilde{\nu}(\tilde{\mathcal{D}}_r) \,. \tag{D.42}$$

In a similar manner, we also get

$$\begin{aligned}
1 - \tilde{\nu}(\tilde{\mathcal{D}}_r) &= \lim_{t\to\infty} \frac{1}{t}\,\mathbb{E}\left[\int_0^t \mathbb{1}\{X(s) \notin \mathbb{B}_r\}\,ds\right] && \text{# b/c } \lim_{t\to\infty}\mu_t \text{ is deterministic} \\
&\leq \lim_{t\to\infty} \frac{1}{t}\,\mathbb{E}\left[\int_0^t \frac{\|X(s)-x^*\|^2}{r^2}\,ds\right] && \text{# b/c } \tfrac{\|X(s)-x^*\|^2}{r^2} \geq 1 \text{ outside } \mathbb{B}_r \\
&\leq \lim_{t\to\infty} \frac{1}{r^2}\left[\frac{F_0}{\alpha t} + \frac{\sigma_{\max}^2}{2\alpha K}\right] && \text{# by (D.25)} \\
&= \frac{r_\sigma^2}{r^2} \,. && \text{(D.43)}
\end{aligned}$$

Our claim then follows by combining Eqs. (D.42) and (D.43). ∎

# E Analysis and results in discrete time

In this appendix, we proceed to prove the discrete-time results presented in Section 4.

**E.1. The null-monotone case.** We begin with our analysis for for null-monotone games. For convenience, we restate Theorem 3 below:

**Theorem 3** (Null-monotone games)**.** *Suppose that* (FTRL) *is run in a null-monotone game $\mathcal{G}$ with a strongly convex $h^*$, let $x^*$ be an equilibrium of $\mathcal{G}$, and let $F_t = F(x^*, Y_t)$. Then $\lim_{t\to\infty} \mathbb{E}[F_t] = \infty$.*

*Proof.* By a second-order Taylor expansion with Lagrange remainder, there exists $w_t \in [Y_t, Y_{t+1}]$ such that:

$$F_{t+1} = F_t + \gamma\langle \hat{v}_t, X_t - x^*\rangle + \frac{\gamma^2}{2}\hat{v}_t \nabla^2 h^*(w_t)\hat{v}_t \,. \tag{E.1}$$

Since $\mathcal{G}$ is null-monotone, we have $\langle v(X_t), X_t - x^*\rangle = 0$ by assumption, and thus

$$\begin{aligned}
F_{t+1} &= F_t + \gamma\langle U_t, X_t - x^*\rangle + \frac{\gamma^2}{2}\hat{v}_t \nabla^2 h^*(w_t)\hat{v}_t && \text{(E.2)}\\
&\geq F_t + \gamma\langle U_t, X_t - x^*\rangle + \frac{m}{2}\gamma^2\|\hat{v}_t\|_*^2 && \text{(E.3)}
\end{aligned}$$

where $m$ denotes here the strong convexity modulus of $h^*$. Moving forward, note that $(i)$ $\mathbb{E}[\langle U_t, X_t - x^*\rangle] = \mathbb{E}[\langle \mathbb{E}[U_t \mid \mathcal{F}_t], X_t - x^*\rangle] = 0$; and $(ii)$ $\inf_t \mathbb{E}\big[\|\hat{v}_t\|_*^2\big] \geq \inf \mathbb{E}[\|\mathsf{V}(x;\omega)\|_*^2] > 0$, so there exists some $V_* > 0$ such that $\mathbb{E}[\|\hat{v}_t\|_*^2] \geq V_*^2$ for all $t$. We thus get

$$\begin{aligned}
\mathbb{E}[F_{t+1}] &\geq \mathbb{E}[F_t] + \frac{m}{2}\gamma^2\,\mathbb{E}\big[\|\hat{v}_t\|_*^2\big] \\
&\geq \mathbb{E}[F_t] + m\gamma^2 V_*^2/2 \\
&\geq F_0 + m\gamma^2 V_*^2 t/2 && \text{(E.4)}
\end{aligned}$$

Our result then follows by taking the limit $t \to \infty$. ∎

**E.2. The strongly monotone case.** We now turn to our main result for strongly monotone games, which we restate below for convenience.

**Theorem 4** (Strongly monotone games)**.** *Suppose that* (FTRL) *is run in an $\alpha$-strongly monotone game $\mathcal{G}$, let $r_\sigma \coloneqq \sqrt{\gamma(\sigma^2+\beta^2)/(\alpha K)}$, and consider the hitting time $\tau_r \coloneqq \inf\{t > 0 : X_t \in \mathbb{B}_r(x^*)\}$ for a ball $\mathbb{B}_r(x^*) = \{x : \|x - x^*\| \leq r\}$ of radius $r$ centered on the* (*necessarily unique*) *equilibrium $x^*$ of $\mathcal{G}$. Then, for all $r > r_\sigma$, we have*

$$\mathbb{E}[\tau_r] \leq \frac{1}{\alpha\gamma(r^2 - r_\sigma^2)} \times \begin{cases} F_0 & \text{if } X_0 \notin \mathbb{B}_r(x^*), \\ F_0 + \alpha\gamma r^2 & \text{if } X_0 \in \mathbb{B}_r(x^*), \end{cases} \tag{21}$$

*where $F_0 = F(x^*, Y_0)$. If, in addition, $x^*$ is interior, $X_t$ admits a unique invariant distribution to which it converges in total variation, and we have*

$$\lim_{t\to\infty} \frac{1}{t}\, \mathbb{E}\left[\sum_{s=0}^{t} \mathbb{1}\{X_t \in \mathbb{B}_r(x^*)\}\right] \geq 1 - r_\sigma^2/r^2 \tag{22}$$

*for all $r > r_\sigma$ such that $\mathbb{B}_r(x^*) \subseteq \operatorname{ri}\mathcal{X}$.*

*Proof.* The main theme of our proof shadows the continuous-time analysis, but it requires distinct tools and techniques to address the specific challenges that arise in the discrete-time Markov chain setting (where, among others, the main tools of stochastic calculus cannot be applied). To streamline our presentation, we follow a step-by-step approach, as outlined below.

**Step 1: Deriving a hitting estimate.** Due to measurability issues, we cannot apply Dynkin's lemma directly in the discrete-time setting, which makes the proof more involved. Moreover, unlike in the continuous-time regime, we need to distinguish between different initializations. Specifically, we consider two cases depending on whether the initial state $X_0$ lies within the ball $\mathbb{B}_r(x^*)$ or not.

- *Case 1: $X_0 \notin \mathbb{B}_r(x^*)$.*

  Letting $F_t \coloneqq F(x^*, Y_t)$ and unfolding (B.29b), we readily obtain:

$$F_t \leq F_0 + \gamma \sum_{s=0}^{t-1} \langle \hat{v}_s, X_s - x^* \rangle + \frac{\gamma^2}{2K} \sum_{s=0}^{t-1} \|\hat{v}_s\|_*^2 \tag{E.5}$$

  and, setting $t \leftarrow \tau_r \wedge t$, we get:

$$F_{\tau_r\wedge t} \leq F_0 + \gamma \sum_{s=0}^{(\tau_r\wedge t)-1} \langle \hat{v}_s, X_s - x^* \rangle + \frac{\gamma^2}{2K} \sum_{s=0}^{(\tau_r\wedge t)-1} \|\hat{v}_s\|_*^2 \tag{E.6}$$

  Thus, taking expectation conditional on the initial state $Y_0 = y$, we have

$$\begin{aligned}
\mathbb{E}\big[F_{\tau_r\wedge t}\big] &\leq F_0 + \gamma\, \mathbb{E}\left[\sum_{s=0}^{(\tau_r\wedge t)-1} \langle \hat{v}_s, X_s - x^* \rangle\right] + \frac{\gamma^2}{2K}\, \mathbb{E}\left[\sum_{s=0}^{(\tau_r\wedge t)-1} \|\hat{v}_s\|_*^2\right] \\
&\leq F_0 + \sum_{s=0}^{t-1} \mathbb{E}\left[\left(\gamma\langle \hat{v}_s, X_s - x^* \rangle + \frac{\gamma^2}{2K}\|\hat{v}_s\|_*^2\right) \mathbb{1}(\tau_r \geq s+1)\right]
\end{aligned} \tag{E.7}$$

  For notational convenience, we denote each summand above per

$$D_s \coloneqq \mathbb{E}\left[\left(\gamma\langle \hat{v}_s, X_s - x^* \rangle + \frac{\gamma^2}{2K}\|\hat{v}_s\|_*^2\right) \mathbb{1}(\tau_r \geq s+1)\right]$$

  and noting that the random variable $\mathbb{1}(\tau_r \geq s+1)$ is $\mathcal{F}_s$-measurable, we get

$$\begin{aligned}
D_s &= \mathbb{E}\left[\mathbb{E}\left[\left(\gamma\langle \hat{v}_s, X_s - x^* \rangle + \frac{\gamma^2}{2K}\|\hat{v}_s\|_*^2\right) \mathbb{1}(\tau_r \geq s+1) \,\middle|\, \mathcal{F}_s\right]\right] \\
&= \mathbb{E}\left[\mathbb{1}(\tau_r \geq s+1)\, \mathbb{E}\left[\gamma\langle \hat{v}_s, X_s - x^* \rangle + \frac{\gamma^2}{2K}\|\hat{v}_s\|_*^2 \,\middle|\, \mathcal{F}_s\right]\right] \\
&= \mathbb{E}\left[\mathbb{1}(\tau_r \geq s+1)\left(\gamma\langle v(X_s), X_s - x^* \rangle + \frac{\gamma^2}{2K}\, \mathbb{E}\big[\|\hat{v}_s\|_*^2 \,\big|\, \mathcal{F}_s\big]\right)\right] \\
&= \mathbb{E}\left[\mathbb{1}(\tau_r \geq s+1)\left(-\gamma\alpha\|X_s - x^*\|^2 + \frac{\gamma^2}{2K}\, \mathbb{E}\big[\|\hat{v}_s\|_*^2 \,\big|\, \mathcal{F}_s\big]\right)\right]
\end{aligned} \tag{E.8}$$

  where we used that $\mathbb{E}[U(X_s, \omega_s) \,|\, \mathcal{F}_s] = 0$. At this point, we note that $\mathbb{E}\big[\|\hat{v}_s\|_*^2 \,|\, \mathcal{F}_s\big] \leq 2\,\mathbb{E}\big[\|v(X_s)\|_*^2 + \|U(X_s, \omega_s)\|_*^2 \,|\, \mathcal{F}_s\big] \leq 2(\beta^2 + \sigma^2)$, and since $r_\sigma^2 \equiv \gamma(\beta^2 + \sigma^2)/(\alpha K)$, we get

$$\begin{aligned}
D_s &\leq \mathbb{E}\left[\left(-\gamma\alpha\|X_s - x^*\|^2 + \alpha\gamma r_\sigma^2\right) \mathbb{1}(\tau_r \geq s+1)\right] \\
&\leq \mathbb{E}\left[\left(-\gamma\alpha r^2 + \alpha\gamma r_\sigma^2\right) \mathbb{1}(\tau_r \geq s+1)\right]
\end{aligned} \tag{E.9}$$

where in the last step we used that $\|X_s - x^*\| \geq r^2$ on $\{\tau_r \geq s+1\}$. Thus, plugging the above bound into (E.7), we obtain

$$\begin{aligned}
\mathbb{E}\big[F_{\tau_r \wedge t}\big] \leq F_0 + \sum_{s=0}^{t-1} D_s &\leq F_0 + \sum_{s=0}^{t-1} \mathbb{E}\Big[\Big(-\gamma\alpha r^2 + \alpha\gamma r_\sigma^2\Big)\mathbb{1}(\tau_r \geq s+1)\Big] \\
&= F_0 - \alpha\gamma(r^2 - r_\sigma^2)\,\mathbb{E}\left[\sum_{s=0}^{(\tau_r \wedge t)-1} 1\right] \\
&= F_0 - \alpha\gamma(r^2 - r_\sigma^2)\,\mathbb{E}[(\tau_r \wedge t)]
\end{aligned} \tag{E.10}$$

As $F$ is nonnegative, we readily obtain that

$$\alpha\gamma(r^2 - r_\sigma^2)\,\mathbb{E}[(\tau_r \wedge t)] \leq F_0 \tag{E.11}$$

and, since $r_\sigma < r$, we get:

$$\mathbb{E}[(\tau_r \wedge t)] \leq \frac{1}{\alpha\gamma(r^2 - r_\sigma^2)} F_0 \tag{E.12}$$

Finally, taking $t \to \infty$, and invoking the monotone convergence theorem [16], we get

$$\mathbb{E}[\tau_r] \leq \frac{1}{\alpha\gamma(r^2 - r_\sigma^2)} F_0 \tag{E.13}$$

- *Case 2:* $X_0 \in \mathbb{B}_r(x^*)$. In this case, we have:

$$\begin{aligned}
\mathbb{E}[\tau_r] &= \mathbb{E}[\mathbb{1}(Q(Y_1) \in \mathbb{B}_r(x^*)) + \mathbb{1}(Q(Y_1) \notin \mathbb{B}_r(x^*))\tau_r] \\
&= \mathbb{P}(Q(Y_1) \in \mathbb{B}_r(x^*)) + \mathbb{E}\big[\mathbb{1}(Q(Y_1) \notin \mathbb{B}_r(x^*))(1 + \mathbb{E}_{Y_1}[\tau_r])\big] \\
&= \mathbb{P}(Q(Y_1) \in \mathbb{B}_r(x^*)) + \mathbb{P}(Q(Y_1) \notin \mathbb{B}_r(x^*)) + \mathbb{E}\big[\mathbb{1}(Q(Y_1) \notin \mathbb{B}_r(x^*))\,\mathbb{E}_{Y_1}[\tau_r]\big] \\
&= 1 + \mathbb{E}\big[\mathbb{1}(Q(Y_1) \notin \mathbb{B}_r(x^*))\,\mathbb{E}_{Y_1}[\tau_r]\big] \\
&\leq 1 + \mathbb{E}\left[\mathbb{1}(Q(Y_1) \notin \mathbb{B}_r(x^*))\Big(\alpha\gamma(r^2 - r_\sigma^2)\Big)^{-1} F(x^*, Y_1)\right] \\
&\leq 1 + \Big(\alpha\gamma(r^2 - r_\sigma^2)\Big)^{-1}\,\mathbb{E}[F(x^*, Y_1)] \\
&\leq 1 + \Big(\alpha\gamma(r^2 - r_\sigma^2)\Big)^{-1}\,\mathbb{E}\left[F_0 + \gamma\langle \hat{v}_0, X_0 - x^*\rangle + \frac{\gamma^2}{2K}\|\hat{v}_0\|_*^2\right] \\
&\leq 1 + \Big(\alpha\gamma(r^2 - r_\sigma^2)\Big)^{-1}\Big(F_0 + \gamma\langle v(X_0), X_0 - x^*\rangle + \alpha\gamma r_\sigma^2\Big) \\
&\leq 1 + \Big(\alpha\gamma(r^2 - r_\sigma^2)\Big)^{-1}\Big(F_0 - \gamma\alpha\|X_0 - x^*\|^2 + \alpha\gamma r_\sigma^2\Big) \\
&\leq 1 + \Big(\alpha\gamma(r^2 - r_\sigma^2)\Big)^{-1}\Big(F_0 + \alpha\gamma r_\sigma^2\Big) \\
&= \frac{F_0 + \alpha\gamma r^2}{\alpha\gamma(r^2 - r_\sigma^2)}
\end{aligned} \tag{E.14}$$

Thus, collectively, we get:

$$\mathbb{E}[\tau_r] \leq \frac{1}{\alpha\gamma(r^2 - r_\sigma^2)} \times \begin{cases} F_0 & \text{if } X_0 \notin \mathbb{B}_r(x^*), \\ F_0 + \alpha\gamma r^2 & \text{if } X_0 \in \mathbb{B}_r(x^*), \end{cases} \tag{E.15}$$

**Step 2: Descending to the restricted process.** As in the continuous-time case, establishing positive recurrence requires analyzing a "restricted" version of the process. To that end, we follow the same construction as in *Step 2 of Theorem 2*, and we define the canonical surjection $\Pi\colon \mathcal{Y} \twoheadrightarrow \tilde{\mathcal{Y}}$ by restricting the action of $y \in \mathcal{Y}$ to $\tilde{\mathcal{V}}$, that is,

$$\langle \Pi(y), z\rangle = \langle y, z\rangle \quad \text{for all } z \in \tilde{\mathcal{V}}. \tag{E.16}$$

whose kernel of $\Pi$ is precisely the *annihilator* $\operatorname{Ann}(\tilde{\mathcal{V}})$ of $\tilde{\mathcal{V}}$, i.e.,

$$\ker\Pi = \operatorname{Ann}(\tilde{\mathcal{V}}) = \{y \in \mathcal{Y} : \langle y, z\rangle = 0 \text{ for all } z \in \tilde{\mathcal{V}}\} \tag{E.17}$$

In addition, we consider the *restricted mirror map* $\tilde{Q}\colon \tilde{\mathcal{Y}} \to \mathcal{X}$ given by

$$\tilde{Q}(\tilde{y}) = Q(y) \quad \text{whenever } \Pi(y) = \tilde{y}. \tag{E.18}$$

Accordingly, letting

$$\tilde{Y}_t = \Pi(Y_t) \tag{E.19}$$

and applying $\Pi$ to (FTRL) yields the "restricted" process

$$\tilde{Y}_{t+1} = \Pi \cdot Y_{t+1} = \Pi \cdot Y_t + \gamma(\Pi \cdot v(X_t) + \Pi \cdot U_t) = \tilde{Y}_t + \gamma(\tilde{v}(X_t) + \tilde{U}_t) \tag{E.20}$$

where $X_t = Q(Y_t) = \tilde{Q}(\tilde{Y}_t)$ and, in a slight abuse of notation, we are overloading the symbol $\Pi$ to denote both the linear map $\Pi\colon \mathcal{Y} \to \tilde{\mathcal{Y}}$ and its representation as a matrix. Finally, writing $\tilde{Y}$ as

$$\tilde{Y}_{t+1} = \tilde{Y}_t + \gamma\tilde{v}\big(\tilde{Q}(\tilde{Y}_t)\big) + \gamma\mathsf{U}\big(\tilde{Q}(\tilde{Y}_t), \omega_t\big) \tag{E.21}$$

we conclude that it is a time-homogeneous Markov process, and we denote its kernel by $\tilde{q}$, where for any $\tilde{y} \in \tilde{\mathcal{Y}}$ and Borel set $\mathcal{A} \subseteq \tilde{\mathcal{Y}}$, we have $\tilde{q}(\tilde{y}, \mathcal{A}) = \mathbb{P}\big(\tilde{Y}_{t+1} \in \mathcal{A} \,\big|\, \tilde{Y}_t = \tilde{y}\big)$.

**Step 3: Recurrence of the restricted process.** To establish the recurrence of the restricted process, we first need to understand the effect of $\Pi$ on the distribution of $\mathsf{U}(x)$. As stated in the assumptions in Section 4, the probability distribution $\nu_x$ of $\mathsf{U}(x)$ decomposes as $\nu_x = \nu_x^c + \nu_x^\perp$. Noting the push-forward measure is linear, we readily obtain that $\Pi_*\nu_x = \Pi_*\nu_x^c + \Pi_*\nu_x^\perp$, where $\Pi_*\nu_x$ denotes the push-forward measure $\mathcal{A} \mapsto (\nu_x \circ \Pi^{-1})(\mathcal{A})$. For notational convenience, we denote $\widehat{\mathcal{Y}} \equiv \operatorname{Ann}(\tilde{\mathcal{V}})$ and $p(x, y) \equiv p_x(y)$. Then, each $y \in \mathcal{Y}$ can be decomposed as $y = \tilde{y} + \hat{y}$, and since $\Pi$ has full column-rank, the measure $\Pi_*\nu_x^c$ has density with respect to the Lebesgue measure $\lambda_{\tilde{\mathcal{Y}}}$ on $\tilde{\mathcal{Y}}$, given by

$$\tilde{p}(x, \tilde{y}) = \int_{\widehat{\mathcal{Y}}} p(x, \tilde{y}, \hat{y})\, d\lambda_{\widehat{\mathcal{Y}}}(\hat{y}) \tag{E.22}$$

where $\lambda_{\widehat{\mathcal{Y}}}$ is the Lebesgue measure on $\widehat{\mathcal{Y}}$. Importantly, the density $\tilde{p}$ satisfies the following properties, which will be crucial for establishing the recurrence of the process. We formalize these in the proposition below, whose proof is deferred until after the theorem.

**Proposition E.1.** *Let the function $\tilde{p}$ as defined in* (E.22)*. Then:*

*(i) For any compact set $\mathcal{K} \subseteq \mathcal{X}$ and every $\tilde{y} \in \tilde{\mathcal{Y}}$, it holds $\inf_{x \in \mathcal{K}} \tilde{p}(x, \tilde{y}) > 0$.*

*(ii) The function $\tilde{p}$ is (jointly) lower semi-continuous.*

**Lebesgue irreducibility.** We now show that the restricted process $\tilde{Y}_t$ is Lebesgue irreducible; that is, starting from any point in its domain, the process has a positive probability of reaching any open set with nonzero Lebesgue measure. This property is crucial for establishing recurrence, as it ensures that the process does not avoid regions of the space indefinitely.

For this, let a Borel measurable set $\mathcal{A} \subseteq \tilde{\mathcal{Y}}$ with $\lambda_{\tilde{\mathcal{Y}}}(\mathcal{A}) > 0$. We will show that $\tilde{q}(\tilde{y}, \mathcal{A}) > 0$ for all $\tilde{y} \in \tilde{\mathcal{Y}}$, which implies that $\mathcal{A}$ can be reached from any state $\tilde{y}$ in one step with positive probability.

$$\begin{aligned}
\tilde{q}(\tilde{y}, \mathcal{A}) &= \mathbb{P}\big(\tilde{Y}_{t+1} \in \mathcal{A} \,\big|\, \tilde{Y}_t = \tilde{y}\big) \\
&= \mathbb{P}\Big(\tilde{y} + \gamma\tilde{v}(\tilde{Q}(\tilde{y})) + \gamma\tilde{\mathsf{U}}(\tilde{Q}(\tilde{y}), \omega) \in \mathcal{A}\Big) \\
&= \mathbb{P}\Big(\tilde{\mathsf{U}}(\tilde{Q}(\tilde{y}), \omega) \in \big\{\gamma^{-1}\mathcal{A} - \gamma^{-1}\tilde{y} - \tilde{v}(\tilde{Q}(\tilde{y}))\big\}\Big) \\
&= \mathbb{P}\Big(\tilde{\mathsf{U}}(\tilde{Q}(\tilde{y}), \omega) \in \mathcal{A}_{\tilde{y}}\Big) \\
&\geq \int_{\mathcal{A}_{\tilde{y}}} \tilde{p}(\tilde{Q}(\tilde{y}), z)\, d\lambda_{\tilde{\mathcal{Y}}}(z)
\end{aligned} \tag{E.23}$$

where $\mathcal{A}_{\tilde{y}} \equiv \gamma^{-1}\mathcal{A} - \gamma^{-1}\tilde{y} - \tilde{v}(\tilde{Q}(\tilde{y}))$ with $\lambda_{\tilde{\mathcal{Y}}}(\mathcal{A}_{\tilde{y}}) = \gamma^{-d}\lambda_{\tilde{\mathcal{Y}}}(\mathcal{A}) > 0$. Finally, since $\tilde{p}(\tilde{Q}(\tilde{y}), \cdot)$ strictly positive, we conclude that $\tilde{q}(\tilde{y}, \mathcal{A}) > 0$, which shows that $\tilde{Y}_t$ induced by (E.20) is Lebesgue-irreducible.

**Harris recurrence.** Our next step is to show that $\tilde{Y}_t$ is Harris recurrent. This means that the process returns to every set of positive Lebesgue measure infinitely often with probability one. Establishing Harris recurrence is a key step toward proving ergodicity, as it ensures that the process does not drift away or get trapped. For this, we will show that $\tilde{\mathcal{D}}_r = \{\tilde{y} \in \tilde{\mathcal{Y}} : \|Q(\tilde{y}) - x^*\| \leq r\}$ is a recurrent set from which we can go "everywhere" with positive probability. Importantly, the set $\tilde{\mathcal{D}}_r$ is compact as shown in *Step 3 of Theorem 2*.

The first part to prove Harris recurrence is immediate from *Step 1* of our proof; namely, since $\mathbb{E}_{\tilde{y}}[\tau_r] < \infty$ for any initial condition $\tilde{y} \in \tilde{\mathcal{Y}}$, we readily get that $\mathbb{P}_{\tilde{y}}(\tau_r < \infty) = 1$.

For the second part, we will prove the so-called minorization property; that is, there exists a nontrivial measure $\mu$ and a constant $\alpha > 0$ such that

$$\tilde{q}(\tilde{y}, \mathcal{A}) \geq \alpha \mu(\mathcal{A}) \qquad \text{for all } \tilde{y} \in \tilde{\mathcal{D}}_r \text{ and Borel sets } \mathcal{A} \subseteq \tilde{\mathcal{Y}}. \tag{E.24}$$

This condition implies that, from any point in $\tilde{\mathcal{D}}_r$ the process has a uniformly lower-bounded probability of reaching any set $\mathcal{A}$ in one step according to the reference measure $\mu$.

To establish the minorization condition (E.24), we define for notational convenience the function $f : \tilde{\mathcal{Y}} \times \tilde{\mathcal{Y}} \to \mathcal{X}_h \times \tilde{\mathcal{Y}}$ as

$$f(\tilde{y}, z) = \left(\tilde{Q}(\tilde{y}), \gamma^{-1}(z - \tilde{y}) - \tilde{v}(\tilde{Q}(\tilde{y}))\right) \tag{E.25}$$

which is continuous as a composition of continuous functions. With this definition in hand, we perform the change of variables in (E.23), and we have:

$$\begin{aligned} \tilde{q}(\tilde{y}, \mathcal{A}) &\geq \gamma^{-d} \int_{\mathcal{A}} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}}(z) \\ &\geq \gamma^{-d} \int_{\mathcal{A}} \inf_{\tilde{y} \in \tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}}(z) \end{aligned} \tag{E.26}$$

To finally construct the measure $\mu$, we need to ensure that

$$0 < \int_{\tilde{\mathcal{Y}}} \inf_{\tilde{y} \in \tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}}(z) < \infty \tag{E.27}$$

To this end, we state the following proposition, whose proof is deferred until after the theorem to maintain the flow.

**Proposition E.2.** *The density $\tilde{p}$ satisfies:*

$$0 < \int_{\tilde{\mathcal{Y}}} \inf_{\tilde{y} \in \tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}}(z) < \infty \tag{E.28}$$

With Proposition E.1 in hand, we define the measure $\mu$ as

$$\mu(\mathcal{A}) \coloneqq \frac{\int_{\mathcal{A}} \inf_{\tilde{y} \in \tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}}(z)}{\int_{\tilde{\mathcal{Y}}} \inf_{\tilde{y} \in \tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}}(z)} \qquad \text{for all Borel } \mathcal{A} \subseteq \tilde{\mathcal{Y}}. \tag{E.29}$$

Therefore, (E.26) becomes:

$$\tilde{q}(\tilde{y}, \mathcal{A}) \geq \gamma^{-d} \int_{\mathcal{A}} \inf_{\tilde{y} \in \tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}} = \gamma^{-d} \int_{\tilde{\mathcal{Y}}} \inf_{\tilde{y} \in \tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}}(z) \cdot \mu(\mathcal{A}) \tag{E.30}$$

Thus, setting $\alpha \equiv \gamma^{-d} \int_{\tilde{\mathcal{Y}}} \inf_{\tilde{y} \in \tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}}(z)$, we conclude the minorization condition (E.24).

Therefore, the set $\tilde{\mathcal{D}}_r$ is recurrent and $\mu(\tilde{\mathcal{D}}_r) > 0$ (since $\lambda_{\tilde{\mathcal{Y}}}(\tilde{\mathcal{D}}_r) > 0$), and thus by [14, Proposition 11.2.1] the Markov process $\tilde{Y}_t$ admits an invariant measure. In addition, based on the equivalence (D.40) the expected return time $\mathbb{E}[\tau_r]$ to $\tilde{\mathcal{D}}_r$ is uniformly bounded for all initial conditions $\tilde{y}$ on $\tilde{\mathcal{D}}_r$, due to the continuity of the Fenchel coupling $F$. Therefore, invoking [47, Theorem 13.0.1], we conclude that the process $\tilde{Y}_t$ admits a unique invariant probability measure $\tilde{\nu}$, and the law of $\tilde{Y}_t$ converges to $\tilde{\nu}$ *in total variation* for every initial condition $\tilde{y} \in \tilde{\mathcal{Y}}$.

**Step 4: Estimating the long-run occupation measure.** Finally, for the last part, letting $F_t \coloneqq F(x^*, Y_t)$ and unfolding (B.29b), we obtain:

$$F_t \leq F_0 + \gamma \sum_{s=0}^{t-1} \langle \hat{v}_s, X_s - x^* \rangle + \frac{\gamma^2}{2K} \sum_{s=0}^{t-1} \|\hat{v}_s\|_*^2 \tag{E.31}$$

Taking expectations in both sides, we readily get

$$\begin{aligned} 0 \leq \mathbb{E}[F_t] &\leq \mathbb{E}\left[F_0 + \gamma \sum_{s=0}^{t-1} \langle \hat{v}_s, X_s - x^* \rangle + \frac{\gamma^2}{2K} \sum_{s=0}^{t-1} \|\hat{v}_s\|_*^2\right] \\ &\leq F_0 - \alpha\gamma\, \mathbb{E}\left[\sum_{s=0}^{t-1} \|X_s - x^*\|^2\right] + t\alpha\gamma r_\sigma^2 \end{aligned} \tag{E.32}$$

Therefore, by rearranging terms and dividing both sides by $t$, we have:

$$\frac{1}{t}\, \mathbb{E}\left[\sum_{s=0}^{t-1} \|X_s - x^*\|^2\right] \leq \frac{1}{\alpha\gamma t} F_0 + r_\sigma^2 \tag{E.33}$$

Moreover, we have:

$$\frac{1}{t}\, \mathbb{E}\left[\sum_{s=0}^{t-1} \mathbb{1}\{X_s \notin \mathbb{B}_r(x^*)\}\right] \leq \frac{1}{r^2 t}\, \mathbb{E}\left[\sum_{s=0}^{t-1} \|X_s - x^*\|^2\right] \leq \frac{1}{\alpha\gamma t r^2} F_0 + \frac{r_\sigma^2}{r^2} \tag{E.34}$$

Now, note that $\{X_s \notin \mathbb{B}_r(x^*)\} \equiv \{\tilde{Y}_t \notin \tilde{\mathcal{D}}_r\}$ by construction, and thus

$$\frac{1}{t}\, \mathbb{E}\left[\sum_{s=0}^{t-1} \mathbb{1}\{X_s \notin \mathbb{B}_r(x^*)\}\right] = \frac{1}{t}\, \mathbb{E}\left[\sum_{s=0}^{t-1} \mathbb{1}\{\tilde{Y}_t \notin \tilde{\mathcal{D}}_r\}\right] \tag{E.35}$$

Taking $t \to \infty$, and invoking Birkhoff's individual ergodic theorem [20, Theorem 2.3.4], we readily get that the mean occupation measure $\mathcal{A} \mapsto t^{-1}\, \mathbb{E}\big[\sum_{s=0}^{t-1} \mathbb{1}\{\tilde{Y}_t \in \mathcal{A}\}\big]$ converges strongly to the invariant measure $\tilde{\nu}$, and therefore

$$\lim_{t\to\infty} \frac{1}{t}\, \mathbb{E}\left[\sum_{s=0}^{t-1} \mathbb{1}\{X_s \notin \mathbb{B}_r(x^*)\}\right] = \lim_{t\to\infty} \frac{1}{t}\, \mathbb{E}\left[\sum_{s=0}^{t-1} \mathbb{1}\{\tilde{Y}_t \notin \tilde{\mathcal{D}}_r\}\right] = 1 - \tilde{\nu}(\tilde{\mathcal{D}}_r) \tag{E.36}$$

and, using (E.34), we have:

$$\tilde{\nu}(\tilde{\mathcal{D}}_r) \geq 1 - \frac{r_\sigma^2}{r^2} \tag{E.37}$$

and our proof is complete. ■

To keep the presentation self-contained, we restate and prove Proposition E.1 and Proposition E.2 below.

**Proposition E.1.** *Let the function $\tilde{p}$ as defined in (E.22). Then:*

*(i) For any compact set $\mathcal{K} \subseteq \mathcal{X}$ and every $\tilde{y} \in \tilde{\mathcal{Y}}$, it holds $\inf_{x\in\mathcal{K}} \tilde{p}(x, \tilde{y}) > 0$.*

*(ii) The function $\tilde{p}$ is (jointly) lower semi-continuous.*

*Proof.* (i) For the first part, let $\tilde{y} \in \tilde{\mathcal{Y}}$. Then

$$\inf_{x\in\mathcal{K}} \tilde{p}(x, \tilde{y}) = \inf_{x\in\mathcal{K}} \int_{\hat{\mathcal{Y}}} p(x, \tilde{y}, \hat{y})\, d\lambda_{\hat{\mathcal{Y}}}(\hat{y}) \geq \int_{\hat{\mathcal{Y}}} \inf_{x\in\mathcal{K}} p(x, \tilde{y}, \hat{y})\, d\lambda_{\hat{\mathcal{Y}}}(\hat{y}) > 0 \tag{E.38}$$

(ii) For the second part, let $(x, \tilde{y}) \in \mathcal{X} \times \tilde{\mathcal{Y}}$, and let a sequence $\{(x_t, \tilde{y}_t)\}_{t\in\mathbb{N}}$ with $\lim_{t\to\infty}(x_t, \tilde{y}_t) = (x, \tilde{y})$. Since $p$ is jointly continuous, applying Fatou's lemma [16], we get

$$\tilde{p}(x, \tilde{y}) = \int_{\hat{\mathcal{Y}}} p(x, \tilde{y}, \hat{y})\, d\lambda_{\hat{\mathcal{Y}}}(\hat{y}) = \int_{\hat{\mathcal{Y}}} \liminf_{t\to\infty} p(x_t, \tilde{y}_t, \hat{y})\, d\lambda_{\hat{\mathcal{Y}}}(\hat{y})$$

$$\leq \liminf_{t\to\infty} \int_{\widehat{\mathcal{Y}}} p(x_t, \tilde{y}_t, \hat{y}) d\lambda_{\widehat{\mathcal{Y}}}(\hat{y})$$
$$= \liminf_{t\to\infty} \tilde{p}(x_t, \tilde{y}_t) \tag{E.39}$$

i.e.,

$$\tilde{p}(x, \tilde{y}) \leq \liminf_{t\to\infty} \tilde{p}(x_t, \tilde{y}_t) \tag{E.40}$$

and the result follows. ∎

**Proposition E.2.** *The density $\tilde{p}$ satisfies:*

$$0 < \int_{\tilde{\mathcal{Y}}} \inf_{\tilde{y}\in\tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}}(z) < \infty \tag{E.28}$$

*Proof.* The upper bound is trivial since $\tilde{p}$ is a probability density and

$$\int_{\tilde{\mathcal{Y}}} \inf_{\tilde{y}\in\tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}}(z) \leq \int_{\tilde{\mathcal{Y}}} \tilde{p}(f(\tilde{y}, z)) d\lambda_{\tilde{\mathcal{Y}}}(z) \leq 1 \tag{E.41}$$

For the lower bound, we will show that

$$\inf_{\tilde{y}\in\tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z)) > 0 \qquad \text{for all } z \in \tilde{\mathcal{Y}}. \tag{E.42}$$

Suppose not, i.e., there exists $z_0 \in \tilde{\mathcal{Y}}$ such that $\inf_{\tilde{y}\in\tilde{\mathcal{D}}_r} \tilde{p}(f(\tilde{y}, z_0)) = 0$. Since $\tilde{\mathcal{D}}_r$ is compact and $\tilde{p} \circ f$ is lower semi-continuous, the infimum over $\tilde{\mathcal{D}}_r$ is realized, meaning that there exists $\tilde{y}_0 \in \tilde{\mathcal{D}}_r$ such that $\tilde{p}(f(\tilde{y}_0, z_0)) = 0$, or, equivalently,

$$\tilde{p}\Big(\tilde{Q}(\tilde{y}_0), \gamma^{-1}(z_0 - \tilde{y}_0) - \tilde{v}(\tilde{Q}(\tilde{y}_0))\Big) = 0 \tag{E.43}$$

This contradicts Proposition E.1 for $\mathcal{K} \leftarrow \mathcal{K}_r$. Finally, since we integrating over a set with positive measure, our result follows. ∎

# F Further numerical results and details

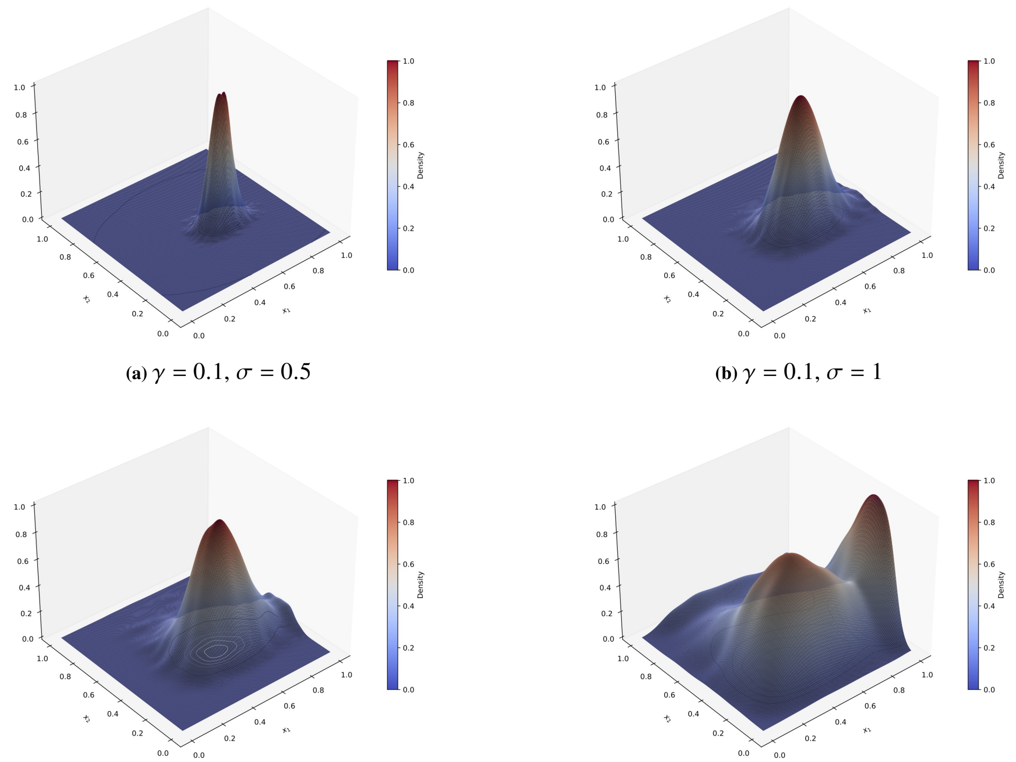
In this section, we present some additional numerical simulations to illustrate and validate our theoretical findings. To this end, we consider two simple yet representative examples: $(i)$ a strongly monotone two-player min-max game on the unit square; and $(ii)$ a finite zero-sum game (as an example of a null-monotone game).

**Strongly monotone.** We consider the strongly monotone two-player min-max game defined by $f : [0, 1] \times [0, 1] \to \mathbb{R}$ with

$$f(x_1, x_2) = -(x_1 - 0.5)^2 + 0.5x_1x_2 + 2(x_2 - 0.5)^2 \tag{F.1}$$

and entropic regularization. To be more precise, the payoff functions of the two players are given by $u_1(x_1, x_2) = f(x_1, x_2) = -u_2(x_1, x_2)$, and $x^* = (20/33, 14/33)$ is the unique Nash equilibrium point.

Fig. 2 demonstrates the behavior of (FTRL) under varying step sizes and noise levels for the min-max game defined by the function $f(x_1, x_2)$. Specifically, we consider step-sizes $\gamma \in \{0.1, 0.5\}$, and stochastic feedback of the form $\hat{v}_t = v(X_t) + \sigma\omega_t$, where $\omega \sim N(0, I_2)$ for $\sigma \in \{0.5, 1\}$. For each $(\gamma, \sigma)$ configuration, we perform $10^5$ independent trials, each running for $10^2$ steps. The initial state $Y_0$ for each trial was drawn uniformly at random from $[0, 1]^2$. Each surface represents the empirical density of the final (FTRL) iterates, while the color overlay visualizes their distribution across the $10^5$ independent trials. Warmer (red) regions indicate higher concentration of final iterates, whereas cooler (blue) regions correspond to lower probability of ending in those regions, as indicated by the colorbar on the side. We observe that smaller step sizes and lower noise levels lead to a tighter concentration of the final iterates around the Nash equilibrium. In contrast, increasing either the step size or the noise variance results in a more dispersed distribution. This behavior aligns with both intuition and our theoretical findings: higher noise introduces greater stochastic variability, while larger step sizes amplify this effect by inducing more aggressive updates that are prone to overshooting, ultimately increasing the spread of the iterates.



(a) $\gamma = 0.1, \sigma = 0.5$

(b) $\gamma = 0.1, \sigma = 1$

(c) $\gamma = 0.5, \sigma = 0.5$

(d) $\gamma = 0.5, \sigma = 1$

**Figure 2:** Visualization of the long-run occupancy measure for the min-max game with loss-gain function $f(x_1, x_2)$. Each plot shows the empirical density of the final iterates of $10^5$ runs of (FTRL) for $10^2$ steps, starting from uniformly random initial conditions. The surface plot encodes density via both height and color. Each row corresponds to a different step-size $\gamma \in \{0.1, 0.5\}$, while the columns vary the noise level $\sigma \in \{0.5, 1\}$.

**Null-monotone.** Fig. 3 shows the empirical distribution of the final iterates under the (FTRL) dynamics in the classic matching pennies game with entropic regularization, played over the probability simplex with payoff matrix

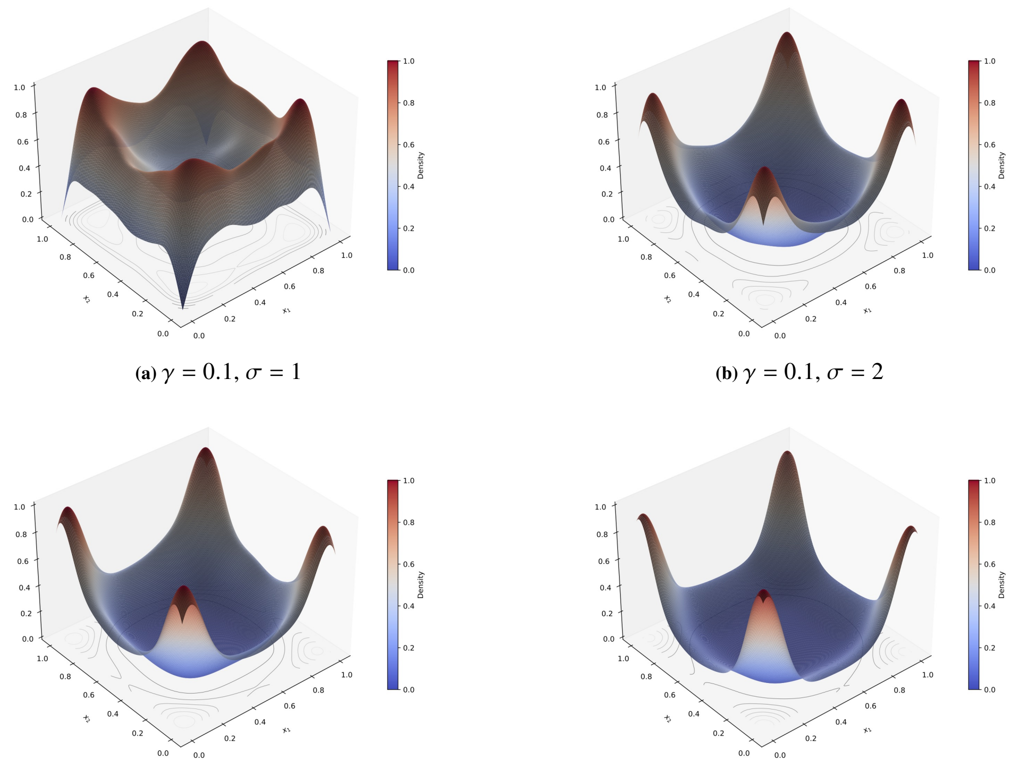
$$P = \begin{bmatrix} (+1, -1) & (-1, +1) \\ (-1, +1) & (+1, -1) \end{bmatrix}.$$

The unique Nash equilibrium of the game is the mixed strategy $(0.5, 0.5)$ for both players. As before, we consider stochastic feedback of the form $\hat{v}_t = v(X_t) + \sigma\omega_t$, where $\omega \sim N(0, I_2)$ for $\gamma \in \{0.1,\ 0.2\}$ and $\sigma \in \{1,\ 2\}$. For each $(\gamma, \sigma)$ configuration, we perform $10^5$ independent trials, each running for $10^2$ steps. Each surface plot corresponds to a different combination of step-size and noise variance, with the empirical density of the final iterates represented through both height and color over the simplex domain. We see that across all configurations, the iterates tend to concentrate near the corners of the simplex, reflecting the instability of the interior equilibrium in the presence of noise. This consistent shift toward extreme points highlights the system's inherent tendency to escape the central equilibrium under stochastic perturbations.

## G Errata

During the preparation of the supplementary material, we noticed a number of typographic errors and omissions in the main paper that could possibly cause confusion. We clarify those below:

- **L174:** The phrase should read "(S-GDA) is run on the game (6a) with initial condition $x_0$" (the existing syntax is correct, but confusing).
- **L141:** The "ensemble map" should read "ensemble mirror map".
- **L186:** "Uhlebeck" should read "Uhlenbeck"




**(a)** $\gamma = 0.1$, $\sigma = 1$

**(b)** $\gamma = 0.1$, $\sigma = 2$

**(c)** $\gamma = 0.2$, $\sigma = 1$

**(d)** $\gamma = 0.2$, $\sigma = 2$

**Figure 3:** Visualization of the long-run occupancy measure for the bilinear game with entropic regularization. Each plot shows the empirical density of the final iterates of $10^5$ runs of (FTRL) for $10^2$ steps, starting from uniformly random initial conditions. The surface plot encodes density via both height and color. Each row corresponds to a different step-size $\gamma \in \{0.1, 0.2\}$, while the columns vary the noise level $\sigma \in \{1, 2\}$.

- **L188:** The phrase should read "(S-GDA) is run on the game (6b) with initial condition $x_0$" (the existing syntax is correct, but confusing).
- **L192:** The phrase should read "centered at 0" (instead of just "centered")
- **L240–241:** The phrase should "is run with a smooth mirror map $Q$" (this was omitted by mistake) and the phrase "admits an interior equilibrium" should appear in L241.
- **L255:** The phrase should read "except, possibly, its boundary" ($\mathcal{X}$ may not have a boundary).
- **L282:** The phrase should read "Nevertheless, as we shall see later in this section" (added for clarity)
- **L302:** The phrase should read "the density. . . is jointly continuous in $x$ and $y$" (this was omitted by mistake).
- **L306:** The phrase should read "small uniform Gaussian" (added for clarity).
- **References:** Because the submitted file did not contain references cited in the appendix, the number—and numbering—of references has changed.

The errata identified above have all been corrected in the file at hand, and highlighted in red (except for the new references cited in the appendix).

# NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: It can be found in Section 3, Section 4 and the appendix.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: It can be found in Section 1.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: It can be found in Section 3, Section 4 and the appendix.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.

- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: It can be found in Appendix F, and the code is included in the supplemental material.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code is included in the supplemental material.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.

- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: It can be found in Appendix F.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: No statistical significance applicable.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: It can be found in Appendix F.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

Answer: [Yes]

Justification: The paper conforms with the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: There is no societal impact.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: There are no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: The paper does not use existing assets.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (https://neurips.cc/Conferences/2025/LLM) for what should or should not be described.