No-regret learning in games with noisy feedback: Faster rates and adaptivity via learning rate separation

Yu-Guan Hsieh
Univ. of Grenoble Alpes
yu-guan.hsieh@univ-grenoble-alpes.fr

Kimon Antonakopoulos
EPFL
kimon.antonakopoulos@epfl.ch

Volkan Cevher
EPFL
volkan.cevher@epfl.ch

Panayotis Mertikopoulos
Univ. Grenoble Alpes, CNRS, LIG & Criteo AI Lab
panayotis.mertikopoulos@imag.fr

Abstract

We examine the problem of regret minimization when the learner is involved in a continuous game with other optimizing agents: in this case, if all players follow a no-regret algorithm, it is possible to achieve significantly lower regret relative to fully adversarial environments. We study this problem in the context of variationally stable games (a class of continuous games which includes all convex-concave and monotone games), and when the players only have access to noisy estimates of their individual payoff gradients. If the noise is additive, the game-theoretic and purely adversarial settings enjoy similar regret guarantees; however, if the noise is *multiplicative*, we show that the learners can, in fact, achieve *constant* regret. We achieve this faster rate via an optimistic gradient scheme with *learning rate separation* – that is, the method's extrapolation and update steps are tuned to different schedules, depending on the noise profile. Subsequently, to eliminate the need for delicate hyperparameter tuning, we propose a fully adaptive method that attains nearly the same guarantees as its non-adapted counterpart, while operating without knowledge of either the game or of the noise profile.

1 Introduction

Owing to its simplicity and versatility, the notion of regret has been the mainstay of online learning ever since the field's first steps [10, 27]. Stated abstractly, it concerns processes of the following form:

1. At each stage $t = 1, 2, \dots$, the learner selects an action x_t from some d-dimensional real space.
2. The environment determines a convex loss function ℓ_t and the learner incurs a loss of $\ell_t(x_t)$.
3. Based on this loss (and any other piece of information revealed), the learner updates their action $x_t \leftarrow x_{t+1}$ and the process repeats.

In this general setting, the agent's regret Reg_T is defined as the difference between the cumulative loss incurred by the sequence $(x_t)_t$ versus that of the best fixed action over the horizon of play T. Accordingly, the learner's objective is to minimize the growth rate of Reg_T, guaranteeing in this way that the chosen sequence of actions becomes asymptotically efficient over time.

Without further assumptions on the learner's environment or the type of loss functions encountered, it is not possible to go beyond the well-known minimax regret bound of $\Omega(\sqrt{T})$ [28, 57], which is achieved by the online gradient descent (OGD) policy of Zinkevich [60]. However, this lower bound concerns environments that are "adversarial" and loss functions that may vary arbitrarily from one

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

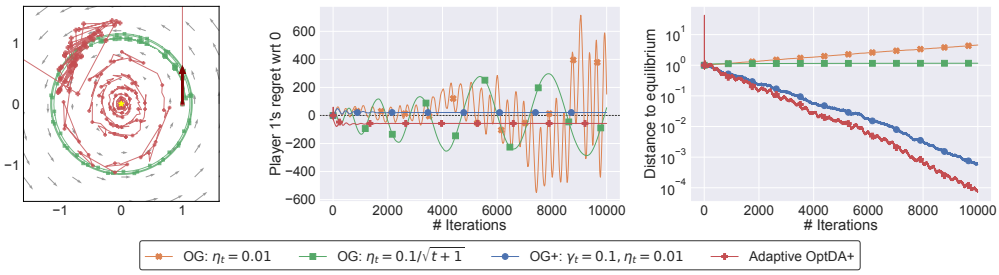


Figure 1: The behavior of different algorithms on the game $\min_{\theta \in \mathbb{R}} \max_{\phi \in \mathbb{R}} \theta\phi$ when the feedback is corrupted by noise. Left: trajectories of play. Center: regret of Player 1. Right: distance to equilibrium. Adaptive OptDA+ is run with $q = 1/4$. See Example 1 for the details of the model and Appendix C for additional figures.

stage to the next: if the environment is "smoother" – and not actively seeking to sabotage the learner's efforts – one could plausibly expect faster regret minimization rates.

This question is particularly relevant – and has received significant attention – in the backdrop of *multi-agent* learning in games. Here, the learners' environment is no longer arbitrary: instead, each player interacts with other regret minimizing players, and every player's individual loss function is determined by the actions chosen by all players via a fixed underlying mechanism – that of a *non-cooperative game*. Because of this mechanism – and the fact that players are changing their actions incrementally from one round to the next – the learners are facing a much more "predictable" sequence of events. As a result, there has been a number of research threads in the literature showing that it is possible to attain *near-constant regret* (i.e., at most polylogarithmic) in different classes of games, from the works of [16, 36] on finite two-player zero-sum games, to more recent works on general-sum finite games [1, 2, 18], extensive form games [23], and even continuous games [31].

Our contributions in the context of related work. The enabling technology for this range of near-constant regret guarantees is the *optimistic gradient* (OG) algorithmic template, itself a variant of the extra-gradient (EG) algorithm of Korpelevich [39]. The salient feature of this method – first examined by Popov [53] in a game-theoretic setting and subsequently popularized by Rakhlin and Sridharan [54] in the context of online learning – is that players use past gradient information to take a more informed "look-ahead" gradient step that stabilizes the method and leads to lower regret. This, however, comes with an important caveat: all of the above works crucially rely on the players' having access to exact payoff gradients, an assumption which is often violated in practice. When the players' feedback is corrupted by noise (or other uncertainty factors), the very same algorithms discussed above may incur *superlinear regret* (cf. Figure 1), leading to the following question:

Is it possible to achieve constant regret in the presence of noise and uncertainty?

Our paper seeks to address this question in a class of continuous games that satisfy a variational stability condition in the spirit of [31]. This class contains all bilinear min-max games (the unconstrained analogue of two-player, zero-sum finite games), cocoercive and monotone games, and it is one of the settings of choice when considering applications to generative models and robust reinforcement learning [13, 30, 35, 42]. As for the noise contaminating the players' gradient feedback, we consider two standard models that build on a classical distinction by Polyak [52] : *a*) *additive*; and *b*) *multiplicative* gradient noise. The first model is more common when dealing with problem-agnostic first-order oracles [47]; the latter arises naturally in the study of randomized coordinate descent [47], asynchronous player updating schemes [5], signal processing and control [56], etc.

In this general context, our contributions can be summarized as follows:

1. We introduce a learning rate separation mechanism that effectively disjoins the extrapolation and update steps of the OG algorithm. The resulting method, which we call OG+, guarantees $\mathcal{O}(\sqrt{T})$ regret in the presence of additive gradient noise; however, if the noise is multiplicative and the method is tuned appropriately, it achieves *constant* $\mathcal{O}(1)$ regret.
2. On the downside, OG+ may fail to achieve sublinear regret in an adversarial environment. To counter this, we propose a "primal-dual" variant of OG+, which we call OptDA+, and which retains the above properties of OG+, while achieving $\mathcal{O}(\sqrt{T})$ regret in the adversarial case.

	Adversarial	All players run the same algorithm			
	Bounded feedback	Additive noise		Multiplicative noise	
	Regret	Regret	Convergence	Regret	Convergence
OG	✗	✗	✗	✗	✗
OG+	✗	$\sqrt{t}\log t$	✓	constant	✓
OptDA+	$\sqrt{t}$	$\sqrt{t}$	✓	constant	-
AdaOptDA+ ($q = 1/4$)	$t^{3/4}$	$\sqrt{t}$	✓	constant	-

Table 1: Summary of the results obtained in the paper. The cross ✗ indicates a negative result (divergence of trajectory or potentially unbounded regret with decreasing stepsize) while - indicates that the behavior of the algorithm is unknown. Our methods improve upon vanilla OG through scale separation between the two steps.

3. Subsequently, to obviate the need for delicate hyperparameter tuning, we propose a fully adaptive method that enjoys nearly the the same regret guarantees as mentioned above, without any prior knowledge of the game or of the uncertainties involved. Interestingly, our method figures a trade-off between achieving small regret when facing adversarial opponents and achieving small regret when facing opponents that adopt the same prescribed strategy, which prevents us from obtaining the optimal $\mathcal{O}(\sqrt{T})$ regret bound in the former situation.
4. Finally, we complement our analysis with a series of equilibrium convergence results for the range of algorithms presented above under both additive and multiplicative noise.

To the best of our knowledge, our work is the first in the literature to point out that constant regret may still be achievable in the presence of stochasticity (even in the simplest case where the noise profile is known in advance). In this regard, it can be seen as a first estimation of the degree of uncertainty that can enter the process before the aspiration of constant (or polylogarithmic) regret becomes an impossible proposition.[1]

A summary of our results is presented in Table 1. In the paper's appendix, we discuss some further related works that are relevant but not directly related to our work. We also mention here that our paper focuses on the unconstrained setting, as this simplifies considerably the presentation and treatment of multiplicative noise models. We defer the constrained case (where players must project their actions to a convex subset of $\mathbb{R}^d$), to future work.

2 Problem Setup

Throughout this paper, we focus on deriving optimal regret minimization guarantees for multi-agent game-theoretic settings with noisy feedback. Starting with the single-agent case, given a sequence of actions $x_t \in \mathcal{X} = \mathbb{R}^d$ and a sequence of loss functions $f_t \colon \mathcal{X} \to \mathbb{R}$, we define the associated *regret* induced by x_t relative to a benchmark action $p \in \mathcal{X}$ as

$$\mathrm{Reg}_T(p) = \sum_{t=1}^{T} [f_t(x_t) - f_t(p)]. \tag{1}$$

We then say that learner has *no regret* if $\mathrm{Reg}_T(p) = o(T)$ for all $p \in \mathcal{X}$. In the sequel, we extend this basic framework to the multi-agent, game-theoretic case, and we discuss the various feedback model available to the optimizer(s).

No-regret learning in games. The game-theoretic analogue of the above framework is defined as follows. We consider a finite set of players indexed by $i \in \mathcal{N} = \{1, \dots, N\}$, each with their individual action space $\mathcal{X}^i = \mathbb{R}^{d^i}$ and their associated loss function $\ell^i \colon \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} = \Pi_{i \in \mathcal{N}} \mathcal{X}^i$ denotes the game's joint action space. For clarity, any ensemble of actions or functions whose definition involves multiple players will be typeset in bold. In particular, we will write

[1] We also note that under the additional assumption of *cocoercivity*, such constant regret bound can be readily derived from [42, Th. 4.4]. Extending this result to the broader family of variationally stable games that we address here however requires non-trivial modification to the algorithm and to the analysis thereof.

$\mathbf{x} = (x^i, \mathbf{x}^{-i}) \in \mathcal{X}$ for the action profile of all players, where x^i and $\mathbf{x}^{-i}$ respectively denote the action of player i and the joint action of all players other than i. In this way, each player $i \in \mathcal{N}$ incurs at round t a loss $\ell^i(\mathbf{x}_t)$ which is determined not only by their individual action x^i_t, but also by the actions $\mathbf{x}^{-i}_t$ of all other players. Thus, by drawing a direct link with (1), given a sequence of play x^i_t, the *individual regret* of each player $i \in \mathcal{N}$ is defined as

$$\operatorname{Reg}^i_T(p^i) = \sum_{t=1}^{T} \ell(x^i_t, \mathbf{x}^{-i}_t) - \ell(p^i, \mathbf{x}^{-i}_t), \tag{2}$$

From a static viewpoint, the most widely spread solution concept in game theory is that of a *Nash equilibrium*, i.e., a state from which no player has incentive to deviate unilaterally. Formally, a point $\mathbf{x}_\star \in \mathcal{X}$ is a Nash equilibrium if for all $i \in \mathcal{N}$ and all $x^i \in \mathcal{X}^i$, we have $\ell^i(x^i_\star, \mathbf{x}^{-i}_\star) \leq \ell^i(x^i, \mathbf{x}^{-i}_\star)$. In particular, if the players' loss functions are assumed individually convex (see below), Nash equilibria coincide precisely with the zeros of the players' individual gradient field, denoted by $V^i = \nabla_{x^i} \ell^i$. That is, $\mathbf{x}_\star$ is a Nash equilibrium if and only if $\mathbf{V}(\mathbf{x}_\star) = 0$. We will make the following blanket assumptions for all this:

Assumption 1 (Convexity and Smoothness)**.** For all $i \in \mathcal{N}$, $\ell^i(\cdot, \mathbf{x}^{-i})$ is convex at all $\mathbf{x}^{-i}$ and the individual gradient of each player $\nabla_{x^i} \ell^i$ is L-Lipschitz continuous.

Assumption 2 (Variational Stability)**.** The solution set $\mathcal{X}_\star = \{\mathbf{x} \in \mathcal{X} : \mathbf{V}(\mathbf{x}) = 0\}$ of the game is nonempty, and for all $\mathbf{x} \in \mathcal{X}$, $\mathbf{x}_\star \in \mathcal{X}_\star$, we have $\langle \mathbf{V}(\mathbf{x}), \mathbf{x} - \mathbf{x}_\star \rangle = \sum_{i \in \mathcal{N}} \langle V^i(\mathbf{x}), x^i - x^i_\star \rangle \geq 0$.

Convexity as stated in Assumption 1 is a vital requirement in the literature of online learning; otherwise, it is not possible to transform iterative gradient bounds to bona fide regret guarantees. In a similar vein, variational stability can be seen as a variant of the convexity assumption for multi-agent environments, where unilateral convexity assumptions do not suffice to give rise to a learnable game– for example, finite games are unilaterally linear, but finding a Nash equilibrium of a finite game is a PPAD-complete problem [15]. Our work thus focuses on games that satisfy the variational stability condition. Some important families of games that are covered by this criterion are monotone games (i.e., $\mathbf{V}$ is monotone), which in their turn include convex-concave zero-sum games, zero-sum polymatrix games, Cournot oligopolies, etc.

It is also worth noticing that several recent works [1, 18, 22] have managed to bypass Assumption 2 in achieving nearly constant regret when the players have access to perfect feedback. Whether these techniques are applicable in the stochastic setup are to be explored. In any case, Assumption 2 seems crucial for the last-iterate convergence presented in Section 6.

Oracle feedback and noise models. In terms of feedback, we will assume that players have access to noisy estimates of their individual payoff gradients, and we will consider two noise models, *additive noise* and *multiplicative noise*. To illustrate the difference between these two models, suppose we wish to estimate the value of some quantity $v \in \mathbb{R}$. Then, an estimate of v with additive noise is a random variable $\hat{v}_{\text{add}}$ of the form $\hat{v}_{\text{add}} = v + \xi_{\text{add}}$ for some zero-mean noise variable ξ_{add}; analogously, a multiplicative noise model for v is a random variable of the form $\hat{v}_{\text{mult}} = v(1 + \xi_{\text{mult}})$ for some zero-mean noise variable ξ_{mult}. The two models can be compared directly via the additive representation of the multiplicative noise model as $\hat{v}_{\text{mult}} = v + \xi_{\text{mult}} v$, which gives $\operatorname{Var}[\xi_{\text{add}}] = v^2 \operatorname{Var}[\xi_{\text{mult}}]$.

With all this in mind, we will consider the following oracle feedback model: let $g^i_t = V^i(\mathbf{x}_t) + \xi^i_t$ denote the gradient feedback to player i at round t, where ξ^i_t represents the aggregate measurement error relative to $V^i(\mathbf{x}_t)$. Then, with $(\mathcal{F}_t)_{t \in \mathbb{N}}$ denoting the natural filtration associated to $(\mathbf{x}_t)_{t \in \mathbb{N}}$ and $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_t]$ representing the corresponding conditional expectation, we make the following standard assumption for the measurement error vector $\boldsymbol{\xi}_t = (\xi^i_t)_{i \in \mathcal{N}}$.

Assumption 3. The noise vector $(\boldsymbol{\xi}_t)_{t \in \mathbb{N}}$ satisfies the following requirements for some $\sigma_A, \sigma_M \geq 0$.

(a) *Zero-mean:* For all $i \in \mathcal{N}$ and $t \in \mathbb{N}$, $\mathbb{E}_t[\xi^i_t] = 0$.

(b) *Finite variance:* For all $i \in \mathcal{N}$ and $t \in \mathbb{N}$, $\mathbb{E}_t[\|\xi^i_t\|^2] \leq \sigma_A^2 + \sigma_M^2 \|V^i(\mathbf{x}_t)\|^2$.

As an example of the above, the case $\sigma_A, \sigma_M = 0$ corresponds to "perfect information", i.e., when players have full access to their payoff gradients. The case $\sigma_A > 0$, $\sigma_M = 0$, is often referred to as "absolute noise", and it is a popular context-agnostic model for stochastic first-order methods, cf. [34, 46] and references therein. Conversely, the case $\sigma_A = 0$, $\sigma_M > 0$, is sometimes called "relative noise" [52], and it is widely used as a model for randomized coordinate descent methods [47], randomized player updates in game theory [5], physical measurements in signal processing and

control [56], etc. In the sequel, we will treat both models concurrently, and we will use the term "noise" to tacitly refer to the presence of both additive and multiplicative components.

3 Optimistic gradient methods: Definitions, difficulties, and a test case

To illustrate some of the difficulties faced by first-order methods in a game-theoretic setting, consider the standard bilinear problem $\min_{\theta\in\mathbb{R}} \max_{\phi\in\mathbb{R}} \theta\phi$, i.e., $\ell^1(\theta,\phi) = \theta\phi = -\ell^2(\theta,\phi)$. This simple game has a unique Nash equilibrium at $(0,0)$ but, despite this uniqueness, it is well known that standard gradient descent/ascent methods diverge on this simple problem [17, 44]. To remedy this failure, one popular solution consists of incorporating an additional *extrapolation step* at each iteration of the algorithm, leading to the *optimistic gradient* (OG) method

$$x^i_{t+1} = x^i_t - 2\eta^i_{t+1} g^i_t + \eta^i_t g^i_{t-1},$$

where η^i_t is player i's learning rate at round t. For posterity, it will be convenient to introduce the auxiliary iterate X^i_t and write $X^i_{t+\frac{1}{2}} = x^i_t$ for the actual sequence of actions. The above update rule then becomes

$$X^i_{t+\frac{1}{2}} = X^i_t - \eta^i_t g^i_{t-1}, \qquad X^i_{t+1} = X^i_t - \eta^i_{t+1} g^i_t. \tag{OG}$$

This form of the algorithm effectively decouples the learner's *extrapolation step* (performed with g^i_{t-1}, which acts here as an *optimistic* guess for the upcoming feedback), and the bona fide *update step*, which exploits the received feedback g^i_t to update the player's action state from X^i_t to X^i_{t+1}. This mechanism helps the players attain *a*) *lower regret* when their utilities vary slowly (from an online learning viewpoint) [14, 54]; and *b*) *near-constant regret* when all players employ the said algorithm in certain classes of games [1, 2, 18, 23, 31].

However, the above guarantees concern only the case of *perfect gradient feedback*, and may fail completely when the feedback is contaminated by noise, as illustrated in the following example.

Example 1. Suppose that the game's objective is an expectation over $\mathcal{L}_1(\theta,\phi) = 3\theta\phi$ and $\mathcal{L}_2(\theta,\phi) = -\theta\phi$ so that $\ell^1 = -\ell^2 = (\mathcal{L}_1 + \mathcal{L}_2)/2$. At each round, we randomly draw $\mathcal{L}_1$ or $\mathcal{L}_2$ with probability $1/2$ and return the gradient of the sampled function as feedback. Assumption 3 is clearly satisfied here with $\sigma_A = 0$ and $\sigma_M = 2$, i.e., the noise is multiplicative; however, as shown in Figure 1, running (OG) with either constant or decreasing learning rate leads to *i*) divergent trajectories of play; and *ii*) regret oscillations that grow linearly or even superlinearly in magnitude over time.[2]

In view of the above negative results, we propose in the next section a simple fix of the algorithm that allows us to retain its constant regret guarantees in the presence of multiplicative noise.

4 Regret minimization with noisy feedback

In this section, we introduce OG+ and OptDA+, our backbone algorithms for learning under uncertainty, and we present their guarantees in different settings. All proofs are deferred to the appendix.

Learning rate separation and the role of averaging. Viewed abstractly, the failure of OG in the face of uncertainty should be attributed to its inability of separating noise from the expected variation of utilities. In fact, in a noisy environment, the two consecutive pieces of feedback are only close *in expectation*, so a player can only exploit this similarity when the noise is mitigated appropriately.

To overcome this difficulty, we adopt a learning rate separation strategy originally proposed for the EG algorithm by Hsieh et al. [30]. The key observation here is that by taking a larger extrapolation step, the noise effectively becomes an order of magnitude smaller relative to the expected variation of utilities. We refer to this generalization of OG as OG+, and we define it formally as

$$X^i_{t+\frac{1}{2}} = X^i_t - \gamma^i_t g^i_{t-1}, \qquad X^i_{t+1} = X^i_t - \eta^i_{t+1} g^i_t, \tag{OG+}$$

where $\gamma^i_t \geq \eta^i_t > 0$ are the player's learning rates (assumed $\mathcal{F}_{t-1}$-measurable throughout the sequel). Nonetheless, the design of OG+ is somehow counter-intuitive because the players' feedback enters

[2] By superlinear we mean that the regret grows faster than $\Theta(T)$, and this is possible here because neither the action set nor the feedback magnitude is bounded.

the algorithm with decreasing weights. This feature opens up the algorithm to adversarial attacks that can drive it to a suboptimal regime in early iterations, as formally shown in [50, Thm. 3].

To circumvent this issue, we also consider a *dual averaging* variant of OG+ that we refer to as OptDA+, and which treats the gradient feedback used to update the players' chosen actions with the same weight. Specifically, OptDA+ combines the mechanisms of optimism [17, 54], dual averaging [31, 49, 58], and learning rate separation [30] as follows

$$X^i_{t+\frac{1}{2}} = X^i_t - \gamma^i_t g^i_{t-1}, \qquad X^i_{t+1} = X^i_1 - \eta^i_{t+1} \sum_{s=1}^{t} g^i_s. \tag{OptDA+}$$

As we shall see below, these mechanisms dovetail in an efficient manner and allow the algorithm to achieve sublinear regret even in the adversarial regime. [Of course, OG+ and OptDA+ coincide when the update learning rate η^i_t is taken constant.]

Quasi-descent inequality. Before stating our main results on the regret incurred by OG+ and OptDA+, we present the key quasi-descent inequality that underlies our analysis, as it provides theoretical evidence on how the separation of learning rates can lead to concrete performance benefits.

Lemma 1. *Let Assumptions 1 and 3 hold and all players run either (OG+) or (OptDA+) with non-increasing learning rate sequences γ^i_t, η^i_t. Then, for all $i \in \mathcal{N}$, $t \geq 2$, and $p^i \in \mathcal{X}^i$, we have*

$$\mathbb{E}_{t-1}\left[\frac{\|X^i_{t+1} - p^i\|^2}{\eta^i_{t+1}}\right] \leq \mathbb{E}_{t-1}\Bigg[\frac{\|X^i_t - p^i\|^2}{\eta^i_t} + \left(\frac{1}{\eta^i_{t+1}} - \frac{1}{\eta^i_t}\right)\|u^i_t - p^i\|^2 \tag{3a}$$

$$- 2\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle \tag{3b}$$

$$- \gamma^i_t(\|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \|V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2) \tag{3c}$$

$$- \|X^i_t - X^i_{t+1}\|^2/2\eta^i_t + \gamma^i_t\|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 \tag{3d}$$

$$+ (\gamma^i_t)^2 L\|\xi^i_{t-\frac{1}{2}}\|^2 + L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{(\boldsymbol{\eta}_t+\boldsymbol{\gamma}_t)^2} + 2\eta^i_t\|g^i_t\|^2\Bigg], \tag{3e}$$

where i) $\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{(\boldsymbol{\eta}_t+\boldsymbol{\gamma}_t)^2} := \sum_{j=1}^N (\eta^j_t + \gamma^j_t)^2\|\xi^j_{t-\frac{1}{2}}\|^2$, and ii) $u^i_t = X^i_t$ if player i runs (OG+) and $u^i_t = X^i_1$ if player i runs (OptDA+).

Lemma 1 indicates how the (weighted) distance between the player's chosen actions and a fixed benchmark action evolves over time. In order to provide some intuition on how this inequality will be used to derive our results, we sketch below the role that each term plays in our analysis.

1. Thanks to the convexity of the players' loss functions, the regret of each player can be bounded by the sum of the pairing terms in (3b). On the other hand, taking $\mathbf{x}_\star \in \mathcal{X}_\star$, $p^i = x^i_\star$, and summing from $i = 1$ to N, we obtain $-2\langle \mathbf{V}(\mathbf{X}_{t+\frac{1}{2}}), \mathbf{X}_{t+\frac{1}{2}} - \mathbf{x}_\star\rangle$, which is non-positive by Assumption 2, and can thus be dropped from the inequality.
2. The weighted squared distance to p^i, i.e., $\|X^i_t - p^i\|^2/\eta^i_t$, telescopes when controlling the regret (Section 4) and serves as a Lyapunov function for equilibrium convergence (Section 6).
3. The negative term in (3c) provides a consistent negative drift that partially cancels out the noise.
4. The difference in (3d) can be bounded using the smoothness assumption and leaves out terms that are in the order of $\gamma^i_t(\gamma^j_t)^2$.
5. Line (3e) contains a range of positive terms of the order $(\gamma^j_t)^2 + \eta^i_t$. To ensure that they are sufficiently small with respect to the decrease of (3c), both $(\gamma^j_t)_{j\in\mathcal{N}}$ and η^i_t/γ^i_t should be small. Applying Assumption 3 gives $\mathbb{E}[\|g^i_t\|^2] \leq \mathbb{E}[(1+\sigma^2_M)\|V^i(\mathbf{X}_{t+\frac{1}{2}})\| + \sigma^2_A]$, revealing that γ^i_t/η^i_t needs to be at least in the order of $(1+\sigma^2_M)$.
6. Last but not least, $(1/\eta^i_{t+1} - 1/\eta^i_t)\|u^i_t - p^i\|^2$ simply telescopes for OptDA+ (in which case $u^i_t = X^i_1$) but is otherwise difficult to control for OG+ when η^i_{t+1} differs from η^i_t. This additional difficulty forces us to use a global learning rate common across all players when analysing OG+.

To summarize, OG+ and OptDA+ are more suitable for learning in games with noisy feedback because the scale separation between the extrapolation and the update steps delivers a consistent negative drift (3c) that is an order of magnitude greater relative to the deleterious effects of the noise. We will exploit this property to derive our main results for OG+ and OptDA+ below.

Constant regret under uncertainty. We are now in a position to state our regret guarantees:

Theorem 1. *Suppose that Assumptions 1–3 hold and all players run (OG+) with non-increasing learning rate sequences γ_t and η_t such that*

$$\gamma_t \leq \min\left(\frac{1}{3L\sqrt{2N(1+\sigma_M^2)}}, \frac{1}{2(4N+1)L\sigma_M^2}\right) \quad \text{and} \quad \eta_t \leq \frac{\gamma_t}{2(1+\sigma_M^2)} \quad \text{for all } t \in \mathbb{N}. \tag{4}$$

Then, for all $i \in \mathcal{N}$ and all $p^i \in \mathcal{X}^i$, we have

(a) If $\gamma_t = \mathcal{O}(1/(t^{\frac{1}{4}}\sqrt{\log t}))$ and $\eta_t = \Theta(1/(\sqrt{t}\log t))$, then $\mathbb{E}\left[\mathrm{Reg}_T^i(p^i)\right] = \tilde{\mathcal{O}}(\sqrt{T})$.

(b) If the noise is multiplicative and the learning rates are constant, then $\mathbb{E}\left[\mathrm{Reg}_T^i(p^i)\right] = \mathcal{O}(1)$.

The first part of Theorem 1 guarantees the standard $\tilde{\mathcal{O}}(\sqrt{T})$ regret in the presence of additive noise, in accordance with existing results in the literature. What is far more surprising is the second part of Theorem 1 which shows that when the noise is multiplicative (i.e., when $\sigma_A = 0$), it is *still* possible to achieve constant regret. This represents a dramatic improvement in performance, which we illustrate in Figure 1: by simply taking the extrapolation step to be 10 times larger, the player's regret becomes completely stabilized. In this regard, Theorem 1 provides fairly conclusive evidence that having access to exact gradient payoffs *is not* an absolute requisite for achieving constant regret in a game-theoretic context.

On the downside, the above result requires all players to use the same learning rate sequences, a technical difficulty that we overcome below by means of the dual averaging mechanism of OptDA+.

Theorem 2. *Suppose that Assumptions 1–3 hold and all players run (OptDA+) with non-increasing learning rate sequences γ_t^i and η_t^i such that*

$$\gamma_t^i \leq \frac{1}{2L}\min\left(\frac{1}{\sqrt{3N(1+\sigma_M^2)}}, \frac{1}{(4N+1)\sigma_M^2}\right) \quad \text{and} \quad \eta_t^i \leq \frac{\gamma_t^i}{4(1+\sigma_M^2)} \quad \text{for all } t \in \mathbb{N},\ i \in \mathcal{N}. \tag{5}$$

Then, for any $i \in \mathcal{N}$ and $p^i \in \mathcal{X}^i$, we have:

(a) If $\gamma_t^j = \mathcal{O}(1/t^{\frac{1}{4}})$ and $\eta_t^j = \Theta(1/\sqrt{t})$ for all $j \in \mathcal{N}$, then $\mathbb{E}\left[\mathrm{Reg}_T^i(p^i)\right] = \mathcal{O}(\sqrt{T})$.

(b) If the noise is multiplicative and the learning rates are constant, then $\mathbb{E}\left[\mathrm{Reg}_T^i(p^i)\right] = \mathcal{O}(1)$.

The similarity between Theorems 1 and 2 suggests that OptDA+ enjoys nearly the same regret guarantee as OG+ while allowing for the use of *player-specific* learning rates. As OptDA+ and OG+ coincide when run with constant learning rates, Theorem 1(b) is in fact a special case of Theorem 2(b). However, when the algorithms are run with decreasing learning rates, they actually lead to different trajectories. In particular, when the feedback is corrupted by additive noise, this difference translates into the removal of logarithmic factors in the regret bound. More importantly, as we show below, it also helps to achieve sublinear regret when the opponents do not follow the same learning strategy, i.e., in the fully arbitrary, adversarial case.

Proposition 1. *Suppose that Assumption 3 holds and player i runs (OptDA+) with non-increasing learning rates $\gamma_t^i = \Theta(1/t^{\frac{1}{2}-q})$ and $\eta_t^i = \Theta(1/\sqrt{t})$ for some $q \in [0, 1/4]$. If $\sup_{x^i \in \mathcal{X}^i}\|V^i(x^i)\| < +\infty$, we have $\mathbb{E}[\mathrm{Reg}_T^i(p^i)] = \mathcal{O}(T^{\frac{1}{2}+q})$ for every benchmark action $p^i \in \mathcal{X}^i$.*

We introduce the exponent q in Proposition 1 because, as suggested by Theorem 2(a), the whole range of $q \in [0, 1/4]$ leads to the optimal $\mathcal{O}(\sqrt{T})$ regret bound for additive noise when all the players adhere to the use of OptDA+. However, it turns out that taking smaller q (i.e., smaller extrapolation step) is more favorable in the adversarial regime. This is because arbitrarily different successive feedback may make the extrapolation step harmful rather than helpful. On the other hand, our previous discussion also suggests that taking larger q (i.e., larger extrapolation steps), should be more beneficial when all the players use OptDA+. We will quantify this effect in Section 6; however, before doing so, we proceed in the next section to show how the learning rates of Proposition 1 can lead to the design of a fully adaptive, parameter-agnostic algorithm.

5 Adaptive learning rates

So far, we have focused exclusively on algorithms run with *predetermined* learning rates, whose tuning requires knowledge of the various parameters of the model. Nonetheless, even though a player

might be aware of their own loss function, there is little hope that the noise-related parameters are also known by the player. Our goal in this section will be to address precisely this issue through the design of *adaptive* methods enjoying the following desirable properties:

- The method should be implementable by every individual player using only local information and without any prior knowledge of the setting's parameters (for the noise profile and the game alike).
- The method should guarantee sublinear individual regret against any bounded feedback sequence.
- When employed by all players, the method should guarantee $\mathcal{O}(\sqrt{T})$ regret under additive noise and $\mathcal{O}(1)$ regret under multiplicative noise.

In order to achieve the above, inspired by the learning rate requirements of Theorem 2 and Proposition 1, we fix $q \in (0, 1/4]$ and consider the following Adagrad-style [20] learning rate schedule.

$$\gamma_t^i = \frac{1}{\left(1 + \sum_{s=1}^{t-2} \|g_s^i\|^2\right)^{\frac{1}{2}-q}}, \quad \eta_t^i = \frac{1}{\sqrt{1 + \sum_{s=1}^{t-2} \left(\|g_s^i\|^2 + \|X_s^i - X_{s+1}^i\|^2\right)}}. \tag{Adapt}$$

As in Adagrad, the sum of the squared norm of the feedback appears in the denominator. This helps controlling the various positive terms appearing in Lemma 1, such as $L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{(\boldsymbol{\eta}_t+\boldsymbol{\gamma}_t)^2}$ and $2\eta_t^i\|g_t^i\|^2$. Nonetheless, this sum is not taken to the same exponent in the definition of the two learning rates. This scale separation ensures that the contribution of the term $-\gamma_t^i\|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2$ appearing in (3c) remains negative, and it is the key for deriving constant regret under multiplicative noise. As a technical detail, the term $\|X_s^i - X_{s+1}^i\|^2$ is involved in the definition of η_t^i for controlling the difference of (3d). Finally, we do not include the previous received feedback g_{t-1}^i in the definition of γ_t^i and η_t^i. This makes these learning rates $\mathcal{F}_{t-1}$-measurable, which in turn implies $\mathbb{E}[\gamma_t^i\eta_t^i\xi_{t-\frac{1}{2}}^i] = 0$.

From a high-level perspective, the goal with (Adapt) is to recover automatically the learning rate schedules of Theorem 2. This in particular means that γ_t^i and η_t^i should at least be in the order of $\Omega(1/t^{\frac{1}{2}-q})$ and $\Omega(1/\sqrt{t})$, suggesting the following boundedness assumptions on the feedback.

Assumption 4. There exists $G, \bar{\sigma} \geq 0$ such that *i)* $\|V^i(x^i)\| \leq G$ for all $i \in \mathcal{N}$, $x^i \in \mathcal{X}^i$; and *ii)* $\|\xi_t^i\| \leq \bar{\sigma}$ for all $i \in \mathcal{N}$, $t \in \mathbb{N}$ with probability 1.

These assumptions are standard in the literature on adaptive methods, cf. [3, 9, 21, 38].

Regret. We begin with the method's fallback guarantees, deferring all proofs to the appendix.

Proposition 2. *Suppose that Assumption 4 holds and a player $i \in \mathcal{N}$ follows* (OptDA+) *with learning rates given by* (Adapt)*. Then, for any benchmark action $p^i \in \mathcal{X}^i$, we have $\mathbb{E}[\mathrm{Reg}_T^i(p^i)] = \mathcal{O}(T^{\frac{1}{2}+q})$.*

Proposition 2 provides exactly the same rate as Proposition 1, illustrating in this way the benefit of taking a smaller q for achieving smaller regret against adversarial opponents. Nonetheless, as we see below, taking smaller q may incur higher regret when adaptive OptDA+ is employed by all players. In particular, we require $q > 0$ in order to obtain constant regret under multiplicative noise, and this prevents us from obtaining the optimal $\mathcal{O}(\sqrt{T})$ regret in fully adversarial environments.

Theorem 3. *Suppose that Assumptions 1–4 hold and all players run* (OptDA+) *with learning rates given by* (Adapt)*. Then, for any $i \in \mathcal{N}$ and point $p^i \in \mathcal{X}^i$, we have $\mathbb{E}[\mathrm{Reg}_T^i(p^i)] = \mathcal{O}(\sqrt{T})$. Moreover, if the noise is multiplicative ($\sigma_A = 0$), we have $\mathbb{E}[\mathrm{Reg}_T^i(p^i)] = \mathcal{O}(\exp(1/(2q)))$.*

The proof of Theorem 3 is based on Lemma 1; we also note that the $\mathcal{O}(\sqrt{T})$ regret guarantee can in fact be derived for any $q \leq 1/4$ (even negative ones). The main difficulty here consists in bounding (3d), which does not directly cancel out since $\gamma_t^i\eta_t^i$ might not be small enough. To overcome this challenge, we have involved the squared difference $\|X_s^i - X_{s+1}^i\|^2$ in the definition of η_t^i so that the sum of these terms cannot be too large when η_t^i is not small enough. More details on this aspect can be found in the proof of Lemma 18 in the appendix.

Importantly, the $\mathcal{O}(\sqrt{T})$ guarantee above does not depend on the choice of q. This comes in sharp contrast to the constant regret bounds (in T) that we obtain for multiplicative noise. In fact, a key step for proving this is to show that for some (environment-dependent) constant C, we have

$$\sum_{i=1}^N \mathbb{E}\left[\left(1 + \sum_{s=1}^t \|g_s^i\|^2\right)^{\frac{1}{2}+q}\right] \leq C \sum_{i=1}^N \mathbb{E}\left[\sqrt{1 + \sum_{s=1}^t \|g_s^i\|^2}\right] \qquad \text{for all } t \in \mathbb{N} \tag{6}$$

This inequality is derived from Lemma 1 by carefully bounding (3a), (3d), (3e) from above and bounding (3b), (3c) from below. Applying Jensen's inequality, we then further deduce that the right-hand side of inequality (6) is bounded by some constant. This constant, however, is exponential in $1/q$. This leads to an inherent trade-off in the choice of q: larger values of q favor the situation where all players adopt adaptive OptDA+ under multiplicative noise, while smaller values of q provide better fallback guarantees in adversarial environments.

6 Trajectory analysis

In this section, we shift our focus to the analysis of the joint trajectory of play when all players follow the same learning strategy. We derive the convergence of the trajectory of play induced by the algorithms (cf. Figure 1) and provide bounds on the sum of the players' payoff gradient norms $\sum_{t=1}^{T}\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2$. This may be regarded as a relaxed convergence criterion, and by the design of the algorithms, a feedback sequence of smaller magnitude also suggests a more stable trajectory.

Convergence of trajectories under multiplicative noise. When the noise is multiplicative, its effect is in expectation absorbed by the progress brought by the extrapolation step. We thus expect convergence results that are similar to the noiseless case. This is confirmed by the following theorem.

Theorem 4. *Suppose that Assumptions 1–3 hold with $\sigma_A = 0$ and all players run* (OG+) / (OptDA+) *with learning rates given in Theorem 2(b).*[3] *Then, $\mathbf{X}_{t+\frac{1}{2}}$ converges almost surely to a Nash equilibrium and enjoys the stabilization guarantee $\sum_{t=1}^{+\infty}\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] < +\infty$.*

Idea of proof. The proof of Theorem 4 follows the following steps.

1. We first show $\sum_{t=1}^{+\infty}\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] < +\infty$ using Lemma 1. This implies $\sum_{t=1}^{\infty}\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2$ is finite almost surely, and thus with probability 1, $\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|$ converges to 0 and all cluster point of $(\mathbf{X}_{t+\frac{1}{2}})_{t\in\mathbb{N}}$ is a solution.
2. Applying the Robbins–Siegmund theorem to a suitable quasi-descent inequality then gives the almost sure convergence of $\mathbb{E}_{t-1}[\sum_{i\in\mathcal{N}}\|X_t^i - x_\star^i\|^2/\eta^i]$ to finite value for any $\mathbf{x}_\star \in \mathcal{X}_\star$.
3. The conditioning on $\mathcal{F}_{t-1}$ makes the above quantity not directly amenable to analysis. This difficulty is specific to the optimistic algorithms that we consider here as they make use of past feedback in each iteration. We overcome this issue by introducing a virtual iterate $\tilde{\mathbf{X}}_t = (\tilde{X}_t^i)_{i\in\mathcal{N}}$ with $\tilde{X}_t^i = X_t^i + \eta^i\xi_{t-\frac{1}{2}}^i$ that serves as a $\mathcal{F}_{t-1}$-measurable surrogate for X_t^i. We then derive the almost sure convergence of $\sum_{i\in\mathcal{N}}\|\tilde{X}_t^i - x_\star^i\|^2/\eta^i$.
4. To conclude, along with the almost sure convergence of $\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\|$ and $\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|$ to 0 we derive the almost sure convergence of $\mathbf{X}_{t+\frac{1}{2}}$ to a Nash equilibrium. □

In case where the players run the adaptive variant of OptDA+, we expect the learning rates to behave as constants asymptotically and thus similar reasoning can still apply. Formally, we show in the appendix that under multiplicative noise the learning rates of the players converge almost surely to positive constants, and prove the following results concerning the induced trajectory.

Theorem 5. *Suppose that Assumptions 1–4 hold with $\sigma_A = 0$ and all players run* (OptDA+) *with learning rates* (Adapt)*. Then, i) $\sum_{t=1}^{+\infty}\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 < +\infty$ with probability 1, and ii) $\mathbf{X}_{t+\frac{1}{2}}$ converges almost surely to a Nash equilibrium.*

Compared to Theorem 4, we can now only bound $\sum_{t=1}^{\infty}\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2$ in an almost sure sense. This is because in the case of adaptive learning rates, our proof relies on inequality (6), and deriving a bound on $\sum_{t=1}^{\infty}\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2]$ from this inequality does not seem possible. Nonetheless, with the almost sure convergence of the learning rates to positive constants, we still manage to prove almost sure last-iterate convergence of the trajectory of play towards a Nash equilibrium.

Such last-iterate convergence results for adaptive methods are relatively rare in the literature, and most of them assume perfect oracle feedback. To the best of our knowledge, the closest antecedents

[3] Recall that OG+ and OptDA+ are equivalent when run with constant learning rates.

to our result are [2, 42], but both works make the more stringent cocoercive assumptions and consider adaptive learning rate that is the same for all the players. In particular, their learning rates are computed with global feedback and are thus less suitable for the learning-in-game setup.

Convergence of trajectories under additive noise. To ensure small regret under additive noise, we take vanishing learning rates. This makes the analysis much more difficult as the term $(1/\eta_{t+1}^i - 1/\eta_t^i)\|u_t^i - p^i\|^2$ appearing on the right-hand side of inequality (3a) is no longer summable. Nonetheless, it is still possible to provide bound on the sum of the squared operator norms.

Theorem 6. *Suppose that Assumptions 1–3 hold and either i) all players run (OG+) with learning rates described in Theorem 1(a) and $\gamma_t = \Omega(1/t^{\frac{1}{2}-q})$ for some $q \in [0, 1/4]$; ii) all players run (OptDA+) with learning rates described in Theorem 2(a) and $\gamma_t^i = \Omega(1/t^{\frac{1}{2}-q})$ for all $i \in \mathcal{N}$ for some $q \in [0, 1/4]$; or iii) all players run (OptDA+) with learning rates (Adapt) and Assumption 4 holds. Then, $\sum_{t=1}^T \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] = \tilde{\mathcal{O}}(T^{1-q})$.*

Theorem 6 suggests that the convergence speed of $\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2$ under additive noise actually depends on q. Therefore, though the entire range of $q \in [0, 1/4]$ leads to $\mathcal{O}(\sqrt{T})$ regret, taking larger q may result in a more stabilized trajectory. This again goes against Propositions 1 and 2, which suggests smaller q leads to smaller regret in the face of adversarial opponents.

Finally, we also show last-iterate convergence of the trajectory of OG+ under additive noise.

Theorem 7. *Suppose that Assumptions 1–3 hold and all players run (OG+) with non-increasing learning rate sequences γ_t and η_t satisfying (4) and $\gamma_t = \Theta(1/(t^{\frac{1}{2}-q}\sqrt{\log t}))$, $\eta_t = \Theta(1/(\sqrt{t}\log t))$ for some $q \in (0, 1/4]$. Then, $\mathbf{X}_t$ converges almost surely to a Nash equilibrium. Moreover, if $\sup_{t\in\mathbb{N}} \mathbb{E}[\|\boldsymbol{\xi}_t\|^4] < +\infty$, then $\mathbf{X}_{t+\frac{1}{2}}$ converges almost surely to a Nash equilibrium.*

Theorem 7, in showing that the sequence $\mathbf{X}_t$ generated by OG+ converges under suitable learning rates, resolves an open question of [30]. However, compared to the said paper, the analysis of OG+ is much more involved due to the use of past feedback, as explained above for the proof of Theorem 4. Going further, in the second part of statement, we show that $\mathbf{X}_{t+\frac{1}{2}}$, the actual played action, also converges to a Nash equilibrium as long as the 4-th moment of the noise is bounded. Compared to OptDA+, it is possible to establish last-iterate convergence for OG+ under additive noise because we can use $\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2]$ (with $\mathbf{x}_\star \in \mathcal{X}_\star$) as a Lyapunov function here. The same strategy does not apply to OptDA+ due to summability issues. This is a common challenge shared by trajectory convergence analysis of any sort of dual averaging template under additive noise.

7 Concluding remarks

In this paper, we look into the fundamental problem of no-regret learning in games under uncertainty. We exhibited algorithms that enjoy constant regret under multiplicative noise. Building upon this encouraging result, we further studied an adaptive variant and proved trajectory convergence of the considered algorithms. A central element that is ubiquitous in our work is the trade-off between robustness in the fully adversarial setting and faster convergence in the game-theoretic case, as encoded by the exponent q. Whether this trade-off is inherent to the problem or an artifact of the algorithm design warrants further investigation.

Looking forward, there are many important problems that remain to be addressed. On the technical side, we would like to deepen our understanding on the convergence behavior of OptDA+ under additive noise. Extension of our results to learning in other type of games and/or under different types of uncertainty would also be highly valuable. Typical examples include learning in finite games with sampling- or bandit-based feedback.

Going one step further, we may want to reconsider the subject of the analysis itself. Departing from the dichotomy between the fully adversarial setup and the case where all players adhere to a prescribed algorithm, providing analysis for the situation where only a fraction of players deviate is practically relevant. Taking into account other type of regret that may be more suitable for the game-theoretic setting, such as policy regret [6], is yet another fruitful research direction to pursue.

Acknowledgement

This has been partially supported by MIAI@Grenoble Alpes, (ANR-19-P3IA-0003). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement n° 725594 - time-data) and from the Swiss National Science Foundation (SNSF) under grant number 200021_205011. P. Mertikopoulos is grateful for financial support by the French National Research Agency (ANR) in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02), the LabEx PERSYVAL (ANR-11-LABX-0025-01), MIAI@Grenoble Alpes (ANR-19-P3IA-0003), and the grant ALIAS (ANR-19-CE48-0018-01).

References

[1] Ioannis Anagnostides, Constantinos Daskalakis, Gabriele Farina, Maxwell Fishelson, Noah Golowich, and Tuomas Sandholm. Near-optimal no-regret learning for correlated equilibria in multi-player general-sum games. *arXiv preprint arXiv:2111.06008*, 2021.

[2] Ioannis Anagnostides, Gabriele Farina, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Tuomas Sandholm. Uncoupled learning dynamics with $o(logt)$ swap regret in multiplayer games. *arXiv preprint arXiv:2204.11417*, 2022.

[3] Kimon Antonakopoulos, Veronica Belmega, and Panayotis Mertikopoulos. An adaptive mirror-prox method for variational inequalities with singular operators. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

[4] Kimon Antonakopoulos, Veronica Belmega, and Panayotis Mertikopoulos. Adaptive extra-gradient methods for min-max optimization and games. In *International Conference on Learning Representations*, 2021.

[5] Kimon Antonakopoulos, Thomas Pethick, Ali Kavis, Panayotis Mertikopoulos, and Volkan Cevher. Sifting through the noise: Universal first-order methods for stochastic variational inequalities. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 13099–13111, 2021.

[6] Raman Arora, Michael Dinitz, Teodor Vanislavov Marinov, and Mehryar Mohri. Policy regret in repeated games. *Advances in Neural Information Processing Systems*, 31, 2018.

[7] Peter Auer, Nicolo Cesa-Bianchi, and Claudio Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64(1):48–75, 2002.

[8] Waïss Azizian, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. The last-iterate convergence rate of optimistic mirror descent in stochastic variational inequalities. In *Conference on Learning Theory*, pages 326–358. PMLR, 2021.

[9] Francis Bach and Kfir Y Levy. A universal algorithm for variational inequalities adaptive to smoothness and noise. In *Conference on Learning Theory*, pages 164–194. PMLR, 2019.

[10] David Blackwell. An analog of the minimax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6(1):1–8, 1956.

[11] Radu Ioan Boţ, Panayotis Mertikopoulos, Mathias Staudigl, and Phan Tu Vuong. Minibatch forward-backward-forward methods for solving stochastic variational inequalities. *Stochastic Systems*, 11(2):112–139, 2021.

[12] Xufeng Cai, Chaobing Song, Cristóbal Guzmán, and Jelena Diakonikolas. A stochastic halpern iteration with variance reduction for stochastic monotone inclusion problems. *arXiv preprint arXiv:2203.09436*, 2022.

[13] Tatjana Chavdarova, Gauthier Gidel, François Fleuret, and Simon Lacoste-Julien. Reducing noise in GAN training with variance reduced extragradient. In *Advances in Neural Information Processing Systems*, 2019.

[14] Chao-Kai Chiang, Tianbao Yang, Chia-Jung Lee, Mehrdad Mahdavi, Chi-Jen Lu, Rong Jin, and Shenghuo Zhu. Online optimization with gradual variations. In *Conference on Learning Theory*, 2012.

[15] Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.

[16] Constantinos Daskalakis, Alan Deckelbaum, and Anthony Kim. Near-optimal no-regret algorithms for zero-sum games. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 235–254. SIAM, 2011.

[17] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training GANs with optimism. In *International Conference on Learning Representations*, 2018.

[18] Constantinos Daskalakis, Maxwell Fishelson, and Noah Golowich. Near-optimal no-regret learning in general games. *Advances in Neural Information Processing Systems*, 34, 2021.

[19] Jelena Diakonikolas, Constantinos Daskalakis, and Michael I Jordan. Efficient methods for structured nonconvex-nonconcave min-max optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 2746–2754. PMLR, 2021.

[20] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research*, 12:2121–2159, 2011.

[21] Alina Ene and Huy L Nguyen. Adaptive and universal algorithms for variational inequalities with optimal convergence. In *36th AAAI Conference on Artificial Intelligence*, 2022.

[22] Gabriele Farina, Ioannis Anagnostides, Haipeng Luo, Chung-Wei Lee, Christian Kroer, and Tuomas Sandholm. Near-optimal no-regret learning for general convex games. *arXiv preprint arXiv:2206.08742*, 2022.

[23] Gabriele Farina, Chung-Wei Lee, Haipeng Luo, and Christian Kroer. Kernelized multiplicative weights for 0/1-polyhedral games: Bridging the gap between learning in extensive-form and normal-form games. *arXiv preprint arXiv:2202.00237*, 2022.

[24] Michail Fasoulakis, Evangelos Markakis, Yannis Pantazis, and Constantinos Varsos. Forward looking best-response multiplicative weights update methods. *arXiv preprint arXiv:2106.03579*, 2021.

[25] Genevieve E Flaspohler, Francesco Orabona, Judah Cohen, Soukayna Mouatadid, Miruna Oprescu, Paulo Orenstein, and Lester Mackey. Online learning with optimism and delay. In *International Conference on Machine Learning*, pages 3363–3373. PMLR, 2021.

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

[27] James Hannan. Approximation to bayes risk in repeated play. *Contributions to the Theory of Games*, 3(2):97–139, 1957.

[28] Elad Hazan. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.

[29] Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. On the convergence of single-call stochastic extra-gradient methods. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

[30] Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. Explore aggressively, update conservatively: Stochastic extragradient methods with variable stepsize scaling. In *Advances in Neural Information Processing Systems*, volume 33, pages 16223–16234, 2020.

[31] Yu-Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos. Adaptive learning in continuous games: Optimal regret bounds and convergence to nash equilibrium. In *Conference on Learning Theory*, 2021.

[32] Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. Multi-agent online optimization with delays: Asynchronicity, adaptivity, and optimism. *Journal of Machine Learning Research*, 2022.

[33] Alfredo N. Iusem, Alejandro Jofré, Roberto I. Oliveira, and Philip Thompson. Extragradient method with variance reduction for stochastic variational inequalities. *SIAM Journal on Optimization*, 27(2):686–724, 2017.

[34] Anatoli Juditsky, Arkadi Semen Nemirovski, and Claire Tauvel. Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58, 2011.

[35] Parameswaran Kamalaruban, Yu-Ting Huang, Ya-Ping Hsieh, Paul Rolland, Cheng Shi, and Volkan Cevher. Robust reinforcement learning via adversarial training with langevin dynamics. In *Advances in Neural Information Processing Systems*, volume 33, pages 8127–8138, 2020.

[36] Ehsan Asadi Kangarshahi, Ya-Ping Hsieh, Mehmet Fatih Sahin, and Volkan Cevher. Let's be honest: An optimal no-regret framework for zero-sum games. In *International Conference on Machine Learning*, pages 2488–2496, 2018.

[37] Aswin Kannan and Uday V Shanbhag. Optimal stochastic extragradient schemes for pseudomonotone stochastic variational inequality problems and their variants. *Computational Optimization and Applications*, 74(3):779–820, 2019.

[38] Ali Kavis, Kfir Y Levy, Francis Bach, and Volkan Cevher. Unixgrad: A universal, adaptive algorithm with optimal guarantees for constrained optimization. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

[39] G. M. Korpelevich. The extragradient method for finding saddle points and other problems. *Èkonom. i Mat. Metody*, 12:747–756, 1976.

[40] Jayash Koshal, Angelia Nedic, and Uday V Shanbhag. Regularized iterative stochastic approximation methods for stochastic variational inequality problems. *IEEE Transactions on Automatic Control*, 58(3):594–609, 2012.

[41] Sucheol Lee and Donghwan Kim. Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

[42] Tianyi Lin, Zhengyuan Zhou, Panayotis Mertikopoulos, and Michael I Jordan. Finite-time last-iterate convergence for multi-agent learning in games. In *International Conference on Machine Learning*, pages 6161–6171, 2020.

[43] Panayotis Mertikopoulos and Zhengyuan Zhou. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1-2):465–507, January 2019.

[44] Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *International Conference on Learning Representations*, 2019.

[45] Arkadi Semen Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

[46] Arkadi Semen Nemirovski, Anatoli Juditsky, Guanghui Lan, and Alexander Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009.

[47] Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Number 87 in Applied Optimization. Kluwer Academic Publishers, 2004.

[48] Yurii Nesterov. Dual extrapolation and its applications to solving variational inequalities and related problems. *Mathematical Programming*, 109(2):319–344, 2007.

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

[50] Francesco Orabona and Dávid Pál. Scale-free online learning. *Theoretical Computer Science*, 716:50–69, 2018.

[51] Thomas Pethick, Puya Latafat, Panagiotis Patrinos, Olivier Fercoq, and Volkan Cevhera. Escaping limit cycles: Global convergence for constrained nonconvex-nonconcave minimax problems. In *International Conference on Learning Representations*, 2022.

[52] Boris Teodorovich Polyak. *Introduction to Optimization*. Optimization Software, New York, NY, USA, 1987.

[53] Leonid Denisovich Popov. A modification of the Arrow–Hurwicz method for search of saddle points. *Mathematical Notes of the Academy of Sciences of the USSR*, 28(5):845–848, 1980.

[54] Alexander Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, 2013.

[55] Herbert Robbins and David Sigmund. A convergence theorem for nonnegative almost supermartingales and some applications. In J. S. Rustagi, editor, *Optimizing Methods in Statistics*, pages 233–257. Academic Press, New York, NY, 1971.

[56] Gesualdo Scutari, Daniel P Palomar, Francisco Facchinei, and Jong-Shi Pang. Convex optimization, game theory, and variational inequality theory. *IEEE Signal Processing Magazine*, 27(3): 35–49, 2010.

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

[58] Lin Xiao. Dual averaging methods for regularized stochastic learning and online optimization. *Journal of Machine Learning Research*, 11:2543–2596, October 2010.

[59] Guojun Zhang and Yaoliang Yu. Convergence of gradient methods on bilinear zero-sum games. In *International Conference on Learning Representations*, 2020.

[60] Martin Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *International Conference on Machine Learning*, pages 928–936, 2003.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [No]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

Appendix

Table of Contents

1 Introduction **1**

2 Problem Setup **3**

3 Optimistic gradient methods: Definitions, difficulties, and a test case **5**

4 Regret minimization with noisy feedback **5**

5 Adaptive learning rates **7**

6 Trajectory analysis **9**

7 Concluding remarks **10**

A Prelude **16**

B Further Related Work **16**

C Additional Figures **17**

D Technical Details and Notations **18**

E Preliminary Analysis for OG+ and OptDA+ **19**
- E.1 Generalized Schemes with Arbitrary Input Sequences 19
- E.2 Quasi-Descent Inequalities for OG+ and OptDA+ 20

F Regret Analysis with Predetermined Learning Rates **23**
- F.1 Bounds for OG+ 24
- F.2 Bounds for OptDA+ 28

G Regret Analysis with Adaptive Learning Rates **34**
- G.1 Preliminary Lemmas 35
- G.2 Robustness Against Adversarial Opponents 37
- G.3 Smaller Regret Against Opponents with Same Learning Algorithm 37

H Last-iterate Convergence **43**
- H.1 Lemmas on Stochastic Sequences 43
- H.2 Trajectory Convergence of OG+ under Additive Noise 45
- H.3 Trajectory Convergence of Non-Adaptive OptDA+ under Multiplicative Noise . . . 48
- H.4 Trajectory Convergence of Adaptive OptDA+ under Multiplicative Noise 49

A Prelude

The appendix is organized as follows. In Appendix B we complement our introduction with an overview on other related works. In Appendix C we expand on our plots for better visibility. We also provide some additional figures there. Subsequently, we build toward the proofs of our main results in Appendices D–H. Appendix D introduces the notations used in the proofs. Some technical details concerning the measurability of the noises and learning rates are discussed as well. Appendix E contains elementary energy inequalities that are repeatedly used through out our analysis. Appendices F and G are dedicated to the regret analysis of the non-adaptive and the adaptive variants. Bounds on the expectation of the sum of the squared operator norms $\sum_{t=1}^{T} \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2$ are also established in these two sections, as bounding this quantity often consists in an important step for bounding the regret. Finally, proofs on the trajectory convergence are presented in Appendix H.

Importantly, in the appendix we present our results in a way that fits better the analysis. Hence, both the organization and the ordering of the these results differ from those in the main paper. For the ease of the reader, we summarize below how the results in the appendix correspond to those in the main paper.

Results of main paper	Results of appendix
Lemma 1	Lemma 4; Lemma 6
Theorem 1	Theorem 9; Lemma 8
Theorem 2	Theorem 11; Lemma 8
Theorem 3	Theorem 13; Theorem 14; Lemma 8
Theorem 4	Theorem 10 (b); Theorem 17
Theorem 5	Theorem 18
Theorem 6	Theorem 8 (a); Theorem 10 (a); Theorem 12
Theorem 7	Theorem 15; Theorem 16
Proposition 1	Proposition 7; Lemma 8
Proposition 2	Proposition 8; Lemma 8

Table 2: Correspondence between results presented in the appendix and results presented in the main paper.

B Further Related Work

On the algorithmic side, both OG and EG have been extensively studied over the past decades in the contexts of, among others, variational inequalities [45, 48], online optimization [14], and learning in games [18, 54]. While the original design of these methods considered the use of the same learning rate for both the extrapolation and the update step, several recent works have shown the benefit of scale separation between the two steps. Our method is directly inspired by [29], which proposed a double step-size variant of EG for achieving last-iterate convergence in stochastic variationally stable games. Among the other uses of learning rate separation of optimistic gradient methods, we should mention here [24, 59] for faster convergence in bilinear games, [19, 41, 51] for performance guarantees under weaker assumptions, and [25, 32] for robustness against delays.

Concerning the last-iterate convergence of no-regret learning dynamics in games with noisy feedback, most existing results rely on the use of vanishing learning rates and are established under more restrictive assumptions such as strong monotonicity [8, 29, 37] or strict variational stability [43, 44]. Our work, in contrast, studies learning with potentially non-vanishing learning rates in variationally stable games. This is made possible thanks to a clear distinction between additive and multiplicative noise; the latter has only been formerly explored in the game-theoretic context by [4, 42] for the class of cocoercive games.[4] Relaxing the cocoercivity assumption is a nontrivial challenge, as testified by the few number of works that establish last-iterate convergence results of stochastic algorithms for

[4] In the said works they use the term absolute random noise and relative random noise for additive noise and multiplicative noise.

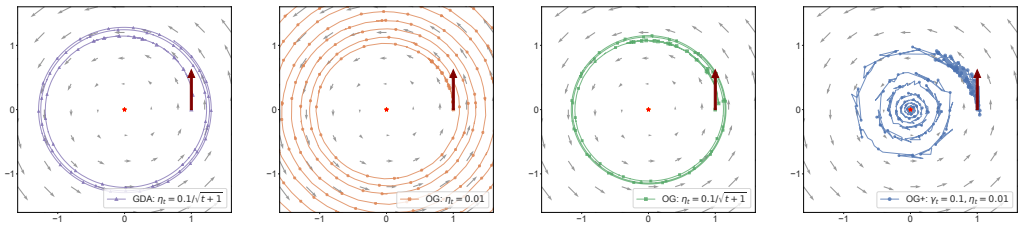
monotone games. Except for [30] mentioned above, this was achieved either through mini-batching [11, 33], Tikhonov regularization / Halpen iteration [40], or both [12].

C Additional Figures

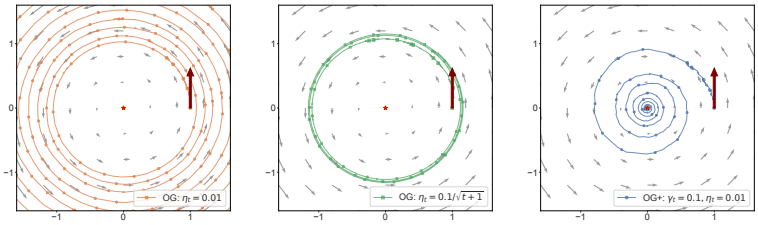
In this section we provide the complete version of Figure 1. In additional to the algorithms already considered in the said figure, we also present results for the case where the two players follow the vanilla gradient descent methods, which we mark as GDA (gradient descent/ascent).

To begin, we complement the leftmost plot of Figure 1 by Figure 2, where we present individual plots of the trajectories induced by different algorithms for better visibility. For optimistic algorithm, we present the trajectory both of the sequence of play $\mathbf{x}_t = \mathbf{X}_{t+\frac{1}{2}}$ and of the auxiliary iterate $\mathbf{X}_t$. The two algorithms GDA and OG have their iterates spiral out, indicating a divergence behavior, conformed to our previous discussions. For OG+ run with constant learning rate and adaptive OptDA+, we observe that the trajectory of $\mathbf{X}_t$ is much "smoother" than that of $\mathbf{x}_t = \mathbf{X}_{t+\frac{1}{2}}$. This is because the extrapolation step is taken with a larger learning rate. Finally, adaptive OptDA+ has its iterates go far away from the equilibrium in the first few iterations due to the initialization with large learning rates, but eventually finds the right learning rates itself and ends up with a convergence speed and regret that is competitive with carefully tuned OG+.

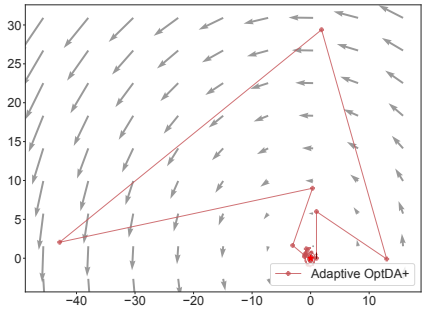
Next, In Figure 3, we expand on the right two plots of Figure 1 with additional curves for GDA. GDA and OG run with the same decreasing learning rate sequences $\eta_t = 0.1/\sqrt{t+1}$ turn out to have similar performance. This suggests that without learning rate separation, the benefit of the extrapolation step may be completely lost in the presence of noise.



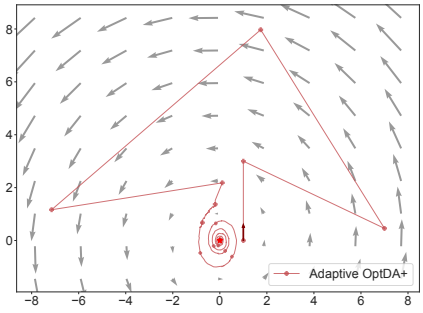
(a) Trajectory of $\mathbf{x}_t$ of different learning algorithms.



(b) Trajectory of $\mathbf{X}_t$ of different optimistic learning algorithms.



(c) Trajectory of $\mathbf{x}_t$ of adaptive OptDA+.



(d) Trajectory of $\mathbf{X}_t$ of adaptive OptDA+.

Figure 2: Trajectories induced by different learning algorithms on the model described in Example 1. We recall that for optimistic learning algorithms, the played point is $\mathbf{x}_t = \mathbf{X}_{t+\frac{1}{2}}$. We take $q = 1/4$ for adaptive OptDA+.

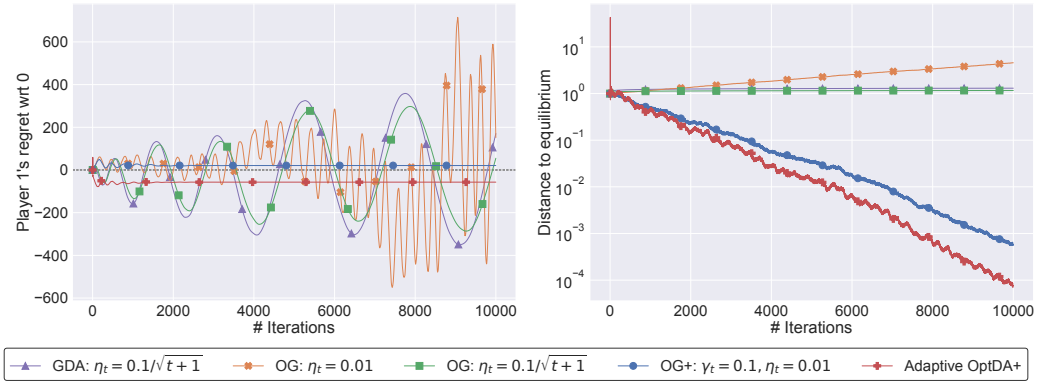


Figure 3: Player 1's regret and distance to equilibrium when both players follow a certain learning strategy in the model described in Example 1. We take $q = 1/4$ for adaptive OptDA+.

D Technical Details and Notations

In this section we introduce the necessary notations for our analysis and discuss some technical details omitted in the main text.

Noise, initialization, and measurability. Throughout our proof, to emphasize that $g_t^i = V^i(x_t^i) + \xi_t^i$ is a stochastic estimate of $V^i(X_{t+\frac{1}{2}}^i)$ in our algorithms, we use the notations $\hat{V}_{t+\frac{1}{2}}^i = g_t^i$ and $\xi_{t+\frac{1}{2}}^i = \xi_t^i$. For the update of $X_{3/2}^i$, we systematically take $g_0^i = \hat{V}_{1/2}^i = 0$. We also write $\xi_{1/2}^i = 0$.

A part of our analysis will be built on the fact that $\xi_{t-\frac{1}{2}}^i$ is $\mathcal{F}_t$-measurable. There is however no a priori reason for this to be true – as $(\mathcal{F}_t)_{t\in\mathbb{N}}$ is the natural filtration associated to $(\mathbf{x}_t)_{t\in\mathbb{N}}$, a sequence that can for example be taken constant independent of the feedback. To address this, we establish here that $\xi_{t-\frac{1}{2}}^i$ is indeed $\mathcal{F}_t$-measurable when player i uses OG+ or OptDA+ with learning rates satisfying a certain measurability assumption. To state it, we define $\mathcal{F}_t^i$ as the σ-algebra generated by $\{(\mathbf{x}_s)_{s=1}^t, (\xi_s^i)_{s=1}^{t-1}\}$.

Assumption 5. For all $t \in \mathbb{N}$, the learning rates γ_{t+1}^i and η_{t+1}^i are $\mathcal{F}_t^i$-measurable.

The following lemma shows that whenever Assumption 5 holds, one can directly work with $(\mathcal{F}_t)_{t\in\mathbb{N}}$.

Lemma 2. *Let player i run* (OG+) *or* (OptDA+) *with learning rates satisfying Assumption 5. Then, for every $t \in \mathbb{N}$, it holds $\mathcal{F}_t^i = \mathcal{F}_t$. In other words, $\xi_{t-\frac{1}{2}}^i$ is $\mathcal{F}_t$-measurable.*

Proof. We prove the lemma by induction. For $t = 1$, this is true by definition. Now, fix $t \geq 2$ and assume that we have proven the statements for all $s \leq t-1$. To show that the statement is also true for t, we note that for both OG+ and OptDA+, $x_t^i = X_{t+\frac{1}{2}}^i$ is a linear combination of the vectors in $\{V^i(\mathbf{x}_s)\}_{s=1}^{t-1} \cup \{\xi_{s+\frac{1}{2}}^i\}_{s=1}^{t-1}$ with coefficients in $\{\eta_s^i\}_{s=1}^t \cup \{\gamma_t^i\}$. All the involved quantities except for $\xi_{t-\frac{1}{2}}^i$ is $\mathcal{F}_{t-1}$-measurable by the induction hypothesis. They are thus $\mathcal{F}_t$-measurable, and as X_t^i is $\mathcal{F}_t$-measurable by the definition of $\mathcal{F}_t$ we concludes that $\xi_{t-\frac{1}{2}}^i$ is also $\mathcal{F}_t$-measurable, which along with the induction hypothesis implies immediately $\mathcal{F}_t^i = \mathcal{F}_t$. □

An immediate consequence of Lemma 2 is the following.

Corollary 1. *Let player i run* (OG+) *or* (OptDA+) *with learning rates satisfying Assumption 5. Then for every $t \in \mathbb{N}$, γ_{t+1}^i and η_{t+1}^i are $\mathcal{F}_t$-measurable.*

Throughout the sequel, both Lemma 2 and Corollary 1 will be used implicitly. Our adaptive learning rates (Adapt) apparently satisfy Assumption 5. As for the non-adaptive case, for simplicity, we assume all their learning rates are predetermined, that is, they are $\mathcal{F}_1$-measurable; for more details on this point see Remark 1. $\mathcal{F}_0$ denotes the trivial σ-algebra.

As another technical detail, in our proofs we assume deterministic $\mathbf{X}_1$, but the entire analysis still goes through for random $\mathbf{X}_1$ under the following conditions

1. For non-adaptive algorithms, we require $\mathbb{E}[\|\mathbf{X}_1\|^2] < +\infty$.
2. For adaptive OptDA+, we require existence of $R \in \mathbb{R}_+$ such that $\|\mathbf{X}_1\| \leq R$ holds almost surely.

Notations related to the learning rates. For any $\mathbf{x} = (x^i)_{i\in\mathcal{N}} \in \mathcal{X} = \mathbb{R}^d$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^i)_{i\in\mathcal{N}} \in \mathbb{R}_+^N$, we write the weighted norm as $\|\mathbf{x}\|_{\boldsymbol{\alpha}} = \sqrt{\sum_{i=1}^N \alpha^i \|x^i\|^2}$. The weights $\boldsymbol{\alpha}$ will be taken as a function of the learning rates. It is thus convenient to write $\boldsymbol{\eta}_t = (\eta_t^i)_{i\in\mathcal{N}}$ and $\boldsymbol{\gamma}_t = (\gamma_t^i)_{i\in\mathcal{N}}$ for the joint learning rates. The arithmetic manipulation and the comparisons of these vectors should be taken elementwisely. For example, the element-wise division is $1/\boldsymbol{\eta}_t = (1/\eta_t^i)_{i\in\mathcal{N}}$. For ease of notation, we also write $\|\boldsymbol{\alpha}\|_1 = \sum_{i=1}^N \alpha^i$ and $\|\boldsymbol{\alpha}\|_\infty = \max_{i\in\mathcal{N}} \alpha^i$ respectively for the L1 norm and the L-infinity norm of an N-dimensional vector $\boldsymbol{\alpha}$.

E Preliminary Analysis for OG+ and OptDA+

In this section, we lay out the basis for the analysis of OG+ and OptDA+.

E.1 Generalized Schemes with Arbitrary Input Sequences

As a starting point, we derive elementary energy inequalities for the following two generalized schemes run with arbitrary vector sequences $(g_t)_{t\in\mathbb{N}}$ and $(g_{t+\frac{1}{2}})_{t\in\mathbb{N}}$.

- Generalized OG+ $\quad X_{t+\frac{1}{2}} = X_t - \gamma_t g_t, \;\; X_{t+1} = X_t - \eta_{t+1} g_{t+\frac{1}{2}}$
- Generalized OptDA+ $\quad X_{t+\frac{1}{2}} = X_t - \gamma_t g_t, \;\; X_{t+1} = X_1 - \eta_{t+1} \sum_{s=1}^t g_{s+\frac{1}{2}}$

In fact, Generalized OG+ with $g_{t+\frac{1}{2}} = \nabla f_t(X_{t+\frac{1}{2}})$ is nothing but the unconstrained, double step-size variant of the optimistic mirror descent method proposed in [54]. On the other hand, Generalized OptDA+ with single learning rate was introduced in [5] under the name of generalized extra-gradient. These two methods coincide when the learning rates are taken constant. In practice, $g_{t+\frac{1}{2}}$ is almost always an estimate of $\nabla f_t(X_{t+\frac{1}{2}})$ while g_t is an approximation of $g_{t+\frac{1}{2}}$. As a matter of fact, as we show in the following propositions, the dot product $\langle g_{t+\frac{1}{2}}, g_t\rangle$ appears with a negative sign in the energy inequalities, which results in a negative contribution when the two vectors are close.

We start with the energy inequality for Generalized OG+.

Proposition 3 (Energy inequality for Generalized OG+)**.** *Let $(X_t)_{t\in\mathbb{N}}$ and $(X_{t+\frac{1}{2}})_{t\in\mathbb{N}}$ be generated by Generalized OG+. It holds for any $p \in \mathcal{X}$ and $t \in \mathbb{N}$ that*

$$\|X_{t+1} - p\|^2 = \|X_t - p\|^2 - 2\eta_{t+1}\langle g_{t+\frac{1}{2}}, X_{t+\frac{1}{2}} - p\rangle - 2\gamma_t\eta_{t+1}\langle g_{t+\frac{1}{2}}, g_t\rangle + (\eta_{t+1})^2\|g_{t+\frac{1}{2}}\|^2.$$

Proof. We develop directly

$$\begin{aligned}\|X_{t+1} - p\|^2 &= \|X_t - \eta_{t+1} g_{t+\frac{1}{2}} - p\|^2 \\ &= \|X_t - p\|^2 - 2\langle g_{t+\frac{1}{2}}, X_t - p\rangle + (\eta_{t+1})^2\|g_{t+\frac{1}{2}}\|^2 \\ &= \|X_t - p\|^2 - 2\eta_{t+1}\langle g_{t+\frac{1}{2}}, X_{t+\frac{1}{2}} - p\rangle - 2\gamma_t\eta_{t+1}\langle g_{t+\frac{1}{2}}, g_t\rangle + (\eta_{t+1})^2\|g_{t+\frac{1}{2}}\|^2,\end{aligned}$$

where in the last equality we use the fact that $X_t = X_{t+\frac{1}{2}} + \gamma_t g_t$. □

For Generalized OptDA+ we have almost the same inequality but for squared distance weighted by $1/\eta_t$, with the notation $\eta_1 = \eta_2$.

Proposition 4 (Energy inequality for Generalized OptDA+)**.** *Let $(X_t)_{t\in\mathbb{N}}$ and $(X_{t+\frac{1}{2}})_{t\in\mathbb{N}}$ be generated by Generalized OptDA+. It holds for any $p \in \mathcal{X}$ and $t \in \mathbb{N}$ that*

$$\begin{aligned}\frac{\|X_{t+1} - p\|^2}{\eta_{t+1}} = \frac{\|X_t - p\|^2}{\eta_t} - \frac{\|X_t - X_{t+1}\|^2}{\eta_t} \\ + \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right)\|X_1 - p\|^2 - \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right)\|X_1 - X_{t+1}\|^2 \\ - 2\langle g_{t+\frac{1}{2}}, X_{t+\frac{1}{2}} - p\rangle - 2\gamma_t\langle g_{t+\frac{1}{2}}, g_t\rangle + \langle g_{t+\frac{1}{2}}, X_t - X_{t+1}\rangle.\end{aligned}$$

Proof. Using $g_{t+\frac{1}{2}} = (X_t - X_1)/\eta_t - (X_{t+1} - X_1)/\eta_{t+1}$, we can write

$$\begin{aligned}
\langle g_{t+\frac{1}{2}}, X_{t+1} - p\rangle &= \left\langle \frac{X_t - X_1}{\eta_t} - \frac{X_{t+1} - X_1}{\eta_{t+1}}, X_{t+1} - p \right\rangle \\
&= \frac{1}{\eta_t}\langle X_t - X_{t+1}, X_{t+1} - p\rangle + \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right)\langle X_1 - X_{t+1}, X_{t+1} - p\rangle \\
&= \frac{1}{2\eta_t}(\|X_t - p\|^2 - \|X_{t+1} - p\|^2 - \|X_t - X_{t+1}\|^2) \\
&\quad + \left(\frac{1}{2\eta_{t+1}} - \frac{1}{2\eta_t}\right)(\|X_1 - p\|^2 - \|X_{t+1} - p\|^2 - \|X_1 - X_{t+1}\|^2).
\end{aligned}$$

Multiplying the equality by 2 and rearranging, we get

$$\begin{aligned}
\frac{\|X_{t+1} - p\|^2}{\eta_{t+1}} &= \frac{\|X_t - p\|^2}{\eta_t} - \frac{\|X_t - X_{t+1}\|^2}{\eta_t} + \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right)\|X_1 - p\|^2 \\
&\quad - \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right)\|X_1 - X_{t+1}\|^2 - 2\langle g_{t+\frac{1}{2}}, X_{t+1} - p\rangle.
\end{aligned}$$

We conclude with the equality

$$\begin{aligned}
\langle g_{t+\frac{1}{2}}, X_{t+1} - p\rangle &= \langle g_{t+\frac{1}{2}}, X_{t+1} - X_t\rangle + \langle g_{t+\frac{1}{2}}, X_t - X_{t+\frac{1}{2}}\rangle + \langle g_{t+\frac{1}{2}}, X_{t+\frac{1}{2}} - p\rangle \\
&= \langle g_{t+\frac{1}{2}}, X_{t+1} - X_t\rangle + \gamma_t\langle g_{t+\frac{1}{2}}, g_t\rangle + \langle g_{t+\frac{1}{2}}, X_{t+\frac{1}{2}} - p\rangle,
\end{aligned}$$

where we have used $X_t = X_{t+\frac{1}{2}} + \gamma_t g_t$. □

Throughout our work, we assume the learning rate sequences to be non-increasing. This is essential for OptDA+, as it guarantees the following corollary.

Corollary 2. *Let $(X_t)_{t\in\mathbb{N}}$ and $(X_{t+\frac{1}{2}})_{t\in\mathbb{N}}$ be generated by Generalized OptDA+. For any $p \in \mathcal{X}$ and $t \in \mathbb{N}$, if $\eta_{t+1} \le \eta_t$, it holds that*

$$\begin{aligned}
\frac{\|X_{t+1} - p\|^2}{\eta_{t+1}} &\le \frac{\|X_t - p\|^2}{\eta_t} + \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right)\|X_1 - p\|^2 - 2\langle g_{t+\frac{1}{2}}, X_{t+\frac{1}{2}} - p\rangle \\
&\quad - 2\gamma_t\langle g_{t+\frac{1}{2}}, g_t\rangle + \eta_t^2\|g_{t+\frac{1}{2}}\|^2 + \min\left(\eta_t^2\|g_{t+\frac{1}{2}}\|^2 - \frac{\|X_t - X_{t+1}\|^2}{2\eta_t}, 0\right).
\end{aligned}$$

Proof. This is immediate from Proposition 4 by applying Young's inequality. More precisely, we use $(1/\eta_{t+1} - 1/\eta_t)\|X_1 - X_{t+1}\|^2 \ge 0$ and

$$2\langle g_{t+\frac{1}{2}}, X_{t+\frac{1}{2}} - p\rangle \le \min\left(\eta_t^2\|g_{t+\frac{1}{2}}\|^2 + \frac{\|X_t - X_{t+1}\|^2}{\eta_t}, 2\eta_t^2\|g_{t+\frac{1}{2}}\|^2 + \frac{\|X_t - X_{t+1}\|^2}{2\eta_t}\right).$$ □

E.2 Quasi-Descent Inequalities for OG+ and OptDA+

We now turn back to (OG+) and (OptDA+) introduced in Section 4. These are special cases of Generalized OG+ and Generalized OptDA+ with $g_t = g_{t-\frac{1}{2}} = \hat{V}^i_{t-\frac{1}{2}}$. The following lemma provides an upper bound on the conditional expectation of $\langle \hat{V}^i_{t+\frac{1}{2}}, \hat{V}^i_{t-\frac{1}{2}}\rangle$ when all the players follow one of the two strategies, and is essential for establishing our quasi-descent inequities.

Lemma 3. *Let Assumptions 1 and 3 hold and all players run either (OG+) or (OptDA+) with learning rates satisfying Assumption 5. Then, for all $i \in \mathcal{N}$ and $t \ge 2$, it holds*

$$\begin{aligned}
-2\,\mathbb{E}_{t-1}[\langle \hat{V}^i_{t+\frac{1}{2}}, \hat{V}^i_{t-\frac{1}{2}}\rangle] \le \mathbb{E}_{t-1}\Bigg[&- \|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2 - \|V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 \\
&+ \|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 \\
&+ L\left(\gamma_t^i\|\xi^i_{t-\frac{1}{2}}\|^2 + \sum_{j=1}^N \frac{(\eta_t^j + \gamma_t^j)^2\|\xi^j_{t-\frac{1}{2}}\|^2}{\gamma_t^i}\right)\Bigg]
\end{aligned}$$

Proof. Thanks to Lemma 2, we can apply the law of total expectation of the expectation to get

$$\begin{aligned}\mathbb{E}_{t-1}[\langle \hat{V}^i_{t+\frac{1}{2}}, \hat{V}^i_{t-\frac{1}{2}}\rangle] &= \mathbb{E}_{t-1}[\langle \mathbb{E}_t[\hat{V}^i_{t+\frac{1}{2}}], \hat{V}^i_{t-\frac{1}{2}}\rangle] \\ &= \mathbb{E}_{t-1}[\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), \hat{V}^i_{t-\frac{1}{2}}\rangle] \\ &= \mathbb{E}_{t-1}[\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), V^i(\mathbf{X}_{t-\frac{1}{2}})\rangle + \langle V^i(\mathbf{X}_{t+\frac{1}{2}}), \xi^i_{t-\frac{1}{2}}\rangle]. \end{aligned} \tag{7}$$

We rewrite the first term as

$$2\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), V^i(\mathbf{X}_{t-\frac{1}{2}})\rangle = \|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \|V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 - \|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2. \tag{8}$$

As for the second term, for all $j \in \mathcal{N}$, we define $\tilde{X}^j_{t+\frac{1}{2}} = X^j_{t+\frac{1}{2}} + (\eta^j_t + \gamma^j_t)\xi^j_{t-\frac{1}{2}}$ and as a surrogate for $X^j_{t+\frac{1}{2}}$ obtained by removing the noise of round $t-1$. For OG+ and OptDA+ we have respectively

$$\begin{aligned}\tilde{X}^j_{t+\frac{1}{2}} &= X^j_{t-1} - (\eta^j_t + \gamma^j_t)V^j(\mathbf{X}_{t-\frac{1}{2}}) \\ \tilde{X}^j_{t+\frac{1}{2}} &= X^i_1 - \eta^j_t \sum_{s=1}^{t-2} \hat{V}^j_{s+\frac{1}{2}} - (\eta^j_t + \gamma^j_t)V^j(\mathbf{X}_{t-\frac{1}{2}}).\end{aligned}$$

With Assumption 5 we then deduce that $\tilde{\mathbf{X}}_{t+\frac{1}{2}}$ is $\mathcal{F}_{t-1}$-measurable and hence

$$\mathbb{E}_{t-1}[\langle V^i(\tilde{\mathbf{X}}_{t+\frac{1}{2}}), \xi^i_{t-\frac{1}{2}}\rangle] = \langle V^i(\tilde{\mathbf{X}}_{t+\frac{1}{2}}), \mathbb{E}_{t-1}[\xi^i_{t-\frac{1}{2}}]\rangle = 0.$$

Moreover, by definition of $\tilde{\mathbf{X}}_{t+\frac{1}{2}}$ we have

$$\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_{t+\frac{1}{2}}\|^2 = \sum_{j=1}^N \|X^j_{t+\frac{1}{2}} - \tilde{X}^j_{t+\frac{1}{2}}\|^2 = \sum_{j=1}^N (\eta^j_t + \gamma^j_t)^2 \|\xi^j_{t-\frac{1}{2}}\|^2$$

It then follows from the Lipschitz continuity of V^i that

$$\begin{aligned}\mathbb{E}_{t-1}[-\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), \xi^i_{t-\frac{1}{2}}\rangle] &= \mathbb{E}_{t-1}[-\langle V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\tilde{\mathbf{X}}_{t+\frac{1}{2}}), \xi^i_{t-\frac{1}{2}}\rangle] \\ &\quad - \mathbb{E}_{t-1}[\langle V^i(\tilde{\mathbf{X}}_{t+\frac{1}{2}}), \xi^i_{t-\frac{1}{2}}\rangle] \\ &\le \mathbb{E}_{t-1}[L\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_{t+\frac{1}{2}}\|\|\xi^i_{t-\frac{1}{2}}\|] \\ &\le \mathbb{E}_{t-1}\left[L\left(\frac{\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_{t+\frac{1}{2}}\|^2}{2\gamma^i_t} + \frac{\gamma^i_t\|\xi^i_{t-\frac{1}{2}}\|^2}{2}\right)\right] \\ &= \mathbb{E}_{t-1}\left[L\left(\frac{\gamma^i_t\|\xi^i_{t-\frac{1}{2}}\|^2}{2} + \sum_{j=1}^N \frac{(\eta^j_t + \gamma^j_t)^2\|\xi^j_{t-\frac{1}{2}}\|^2}{2\gamma^i_t}\right)\right].\end{aligned} \tag{9}$$

Putting (7), (8), and (9) together gives the desired inequality. □

Quasi-Descent Inequalities for OG+. Below we establish respectively the individual and the global quasi-descent inequalities for OG+. In this part, all the players use the same learning rate sequences and we can thus drop the player index in the learning rates.

Lemma 4 (Individual quasi-descent inequality for OG+)**.** *Let Assumptions 1 and 3 hold and all players run* (OG+) *with the same predetermined learning rate sequences. Then, for all $i \in \mathcal{N}$, $t \ge 2$, and $p^i \in \mathcal{X}^i$, it holds*

$$\begin{aligned}\mathbb{E}_{t-1}[\|X^i_{t+1} - p^i\|^2] \le \mathbb{E}_{t-1}[&\|X^i_t - p^i\|^2 - 2\eta_{t+1}\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle \\ &- \gamma_t\eta_{t+1}(\|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \|V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2) \\ &+ \gamma_t\eta_{t+1}\|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 + \gamma_t^2\eta_{t+1}L\|\xi^i_{t-\frac{1}{2}}\|^2 \\ &+ \eta_{t+1}(\eta_t + \gamma_t)^2 L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2 + (\eta_{t+1})^2\|\hat{V}^i_{t+\frac{1}{2}}\|^2].\end{aligned} \tag{10}$$

Proof. We apply Proposition 3 to player i's update and $p \leftarrow p^i$. Since the inequality holds for any realization we can take expectation with respect to $\mathcal{F}_{t-1}$ to get

$$\mathbb{E}_{t-1}[\|X^i_{t+1} - p^i\|^2] = \mathbb{E}_{t-1}[\|X^i_t - p^i\|^2 - 2\eta_{t+1}\langle \hat{V}^i_{t+\frac{1}{2}}, X^i_{t+\frac{1}{2}} - p^i\rangle \\ - 2\gamma_t\eta_{t+1}\langle \hat{V}^i_{t+\frac{1}{2}}, \hat{V}^i_{t-\frac{1}{2}}\rangle + (\eta_{t+1})^2\|\hat{V}^i_{t+\frac{1}{2}}\|^2].$$

The learning rates γ_t and η_{t+1} being $\mathcal{F}_1$-measurable and in particular $\mathcal{F}_{t-1}$-measurable, we conclude immediately with Lemma 3 and the equality

$$\mathbb{E}_{t-1}[\eta_{t+1}\langle \hat{V}^i_{t+\frac{1}{2}}, X^i_{t+\frac{1}{2}} - p^i\rangle] = \eta_{t+1}\,\mathbb{E}_{t-1}[\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle]. \qquad \square$$

Remark 1. From the proof of Lemma 4 we see that the exact requirement concerning the measurability of the learning rates here is that both γ_t and η_{t+1} should be $\mathcal{F}_{t-1}$-measurable. For simplicity throughout our analysis for OG+ we simply say that all the learning rates are predetermined, i.e., $\mathcal{F}_1$-measurable. In contrast, for OptDA+ Assumption 5 is indeed sufficient. This is a technical detail that we have omitted in the main text.

Lemma 5 (Global quasi-descent inequality for OG+)**.** *Let Assumptions 1–3 hold and all players run* (OG+) *with the same predetermined learning rate sequences. Then, for all $t \geq 2$ and $\mathbf{x}_\star \in \mathcal{X}_\star$, we have*

$$\mathbb{E}_{t-1}[\|\mathbf{X}_{t+1} - \mathbf{x}_\star\|^2] \leq \mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2 - \gamma_t\eta_{t+1}(\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2) \\ + 3\gamma_t\eta_{t+1}NL^2((\eta_t^2 + \gamma_t^2)\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2 + (\gamma_{t-1})^2\|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2) \\ + (\gamma_t^2\eta_{t+1} + N\eta_{t+1}(\eta_t + \gamma_t)^2)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2 + (\eta_{t+1})^2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2].$$

Proof. We will apply Lemma 4 to $x^i_\star$. We first bound the variation $\|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2$ by

$$\|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 \leq 3\|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_t)\|^2 + 3\|V^i(\mathbf{X}_t) - V^i(\mathbf{X}_{t-1})\|^2 \\ + 3\|V^i(\mathbf{X}_{t-1}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 \\ \leq 3\gamma_t^2L^2\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2 + 3\eta_t^2L^2\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2 + 3(\gamma_{t-1})^2L^2\|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2. \quad (11)$$

In the second inequality, we have used the Lipschitz continuity of V^i and $\mathbf{X}_{t+\frac{1}{2}} = \mathbf{X}_t - \gamma_t\hat{\mathbf{V}}_{t-\frac{1}{2}}$ to obtain

$$\|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_t)\|^2 \leq L^2\|\mathbf{X}_{t+\frac{1}{2}} - \mathbf{X}_t\|^2 = 3\gamma_t^2L^2\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2.$$

The terms $\|V^i(\mathbf{X}_t) - V^i(\mathbf{X}_{t-1})\|^2$ and $\|V^i(\mathbf{X}_{t-1}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2$ were bounded in the same way. Applying Lemma 4 with $p^i \leftarrow x^i_\star$, plugging (11) into (10), and summing from $i = 1$ to N then yields

$$\mathbb{E}_{t-1}[\|\mathbf{X}_{t+1} - \mathbf{x}_\star\|^2] \leq \mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2 - \eta_{t+1}\langle \mathbf{V}(\mathbf{X}_{t+\frac{1}{2}}), \mathbf{X}_{t+\frac{1}{2}} - \mathbf{x}_\star\rangle \\ - \gamma_t\eta_{t+1}(\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2) \\ + 3\gamma_t\eta_{t+1}NL^2((\eta_t^2 + \gamma_t^2)\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2 + (\gamma_{t-1})^2\|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2) \\ + (\gamma_t^2\eta_{t+1} + N\eta_{t+1}(\eta_t + \gamma_t)^2)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2 + (\eta_{t+1})^2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2].$$

To conclude, we drop $-\eta_{t+1}\langle \mathbf{V}(\mathbf{X}_{t+\frac{1}{2}}), \mathbf{X}_{t+\frac{1}{2}} - \mathbf{x}_\star\rangle$ which is non-positive by Assumption 2. $\square$

Quasi-Descent Inequalities for OptDA+. Similarly, we establish quasi-descent inequalities for OptDA+ that will be used for both non-adaptive and adaptive analyses.

Lemma 6 (Individual quasi-descent inequality for OptDA+)**.** *Let Assumptions 1 and 3 hold and all players run* (OptDA+) *with non-increasing learning rates satisfying Assumption 5. Then, for all*

$i \in \mathcal{N}$, $t \geq 2$, *and* $p^i \in \mathcal{X}^i$, *it holds*

$$\begin{aligned}\mathbb{E}_{t-1}\left[\frac{\|X^i_{t+1}-p^i\|^2}{\eta^i_{t+1}}\right] \leq \mathbb{E}_{t-1}\Bigg[&\frac{\|X^i_t-p^i\|^2}{\eta^i_t}+\left(\frac{1}{\eta^i_{t+1}}-\frac{1}{\eta^i_t}\right)\|X^i_1-p^i\|^2\\ &-2\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}}-p^i\rangle\\ &-\gamma^i_t(\|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2+\|V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2)\\ &+\gamma^i_t\|V^i(\mathbf{X}_{t+\frac{1}{2}})-V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2\\ &+\min\left(-\frac{\|X^i_t-X^i_{t+1}\|^2}{2\eta^i_t}+\eta^i_t\|\hat{V}^i_{t+\frac{1}{2}}\|^2,\,0\right)\\ &+(\gamma^i_t)^2L\|\xi^i_{t-\frac{1}{2}}\|^2+L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{(\boldsymbol{\eta}_t+\boldsymbol{\gamma}_t)^2}+\eta^i_t\|\hat{V}^i_{t+\frac{1}{2}}\|^2\Bigg]. \qquad (12)\end{aligned}$$

Proof. This is an immediate by combining Corollary 2 and Lemma 3. We just notice that as γ^i_t is $\mathcal{F}_{t-1}$-measurable, we have $\mathbb{E}_{t-1}[\gamma^i_t\langle\hat{V}^i_{t+\frac{1}{2}},\hat{V}^i_{t-\frac{1}{2}}\rangle]=\gamma^i_t\,\mathbb{E}_{t-1}[\langle\hat{V}^i_{t+\frac{1}{2}},\hat{V}^i_{t-\frac{1}{2}}\rangle]$. □

Lemma 7 (Global quasi-descent inequality for OptDA+)**.** *Let Assumptions 1–3 hold and all players run* (OptDA+) *with non-increasing learning rates satisfying Assumption 5. Then, for all* $t \geq 2$ *and* $\mathbf{x}_\star \in \mathcal{X}_\star$, *if* $\boldsymbol{\eta}_t \leq \boldsymbol{\gamma}_t$, *we have*

$$\begin{aligned}\mathbb{E}_{t-1}[\|\mathbf{X}_{t+1}-\mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_{t+1}}] \leq \mathbb{E}_{t-1}[&\|\mathbf{X}_t-\mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_t}+\|\mathbf{X}_1-\mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_{t+1}-1/\boldsymbol{\eta}_t}\\ &-\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t}-\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t}\\ &-\|\mathbf{X}_t-\mathbf{X}_{t+1}\|^2_{1/(2\boldsymbol{\eta}_t)}+3\|\mathbf{V}(\mathbf{X}_t)-\mathbf{V}(\mathbf{X}_{t-1})\|^2_{\boldsymbol{\gamma}_t}\\ &+3L^2(\|\boldsymbol{\gamma}_t\|_1\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2}+\|\boldsymbol{\gamma}_{t-1}\|_1\|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2_{(\boldsymbol{\gamma}_{t-1})^2})\\ &+(4N+1)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2}+2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2_{\boldsymbol{\eta}_t}]. \qquad (13)\end{aligned}$$

Proof. The result is proved in the same way as Lemma 5 but instead of Lemma 4 we make use of Lemma 6 with

$$\min\left(-\frac{\|X^i_t-X^i_{t+1}\|^2}{2\eta^i_t}+\eta^i_t\|\hat{V}^i_{t+\frac{1}{2}}\|^2,\,0\right) \leq -\frac{\|X^i_t-X^i_{t+1}\|^2}{2\eta^i_t}+\eta^i_t\|\hat{V}^i_{t+\frac{1}{2}}\|^2.$$

Moreover, as there is not a simple expression for $\|\mathbf{X}_t-\mathbf{X}_{t+1}\|$, in the place of (11) we use

$$\|V^i(\mathbf{X}_{t+\frac{1}{2}})-V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 \leq 3L^2\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2}+3L^2\|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2_{(\boldsymbol{\gamma}_{t-1})^2}+3\|V^i(\mathbf{X}_t)-V^i(\mathbf{X}_{t-1})\|^2. \qquad (14)$$

To obtain (13), we further use $\boldsymbol{\eta}_t \leq \boldsymbol{\gamma}_t$ and $\|\boldsymbol{\gamma}_t\|_1 \leq \|\boldsymbol{\gamma}_{t-1}\|_1$. □

Remark 2. The players can take different learning rates in OptDA+ because in the quasi-descent inequality (12), there is no learning rate in front of $\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}}-p^i\rangle$. Take $p^i \leftarrow x^i_\star$ and summing from $i=1$ to N we get directly $\langle\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}}),\mathbf{X}_{t+\frac{1}{2}}-\mathbf{x}_\star\rangle$ which is non-negative according to Assumption 2. While it is also possible to put (10) in the form of Lemma 1, we are not able to control the sum of $(1/\eta^i_{t+1}-1/\eta^i_t)\|X^i_t-p^i\|^2$ as explained in Section 4.

F Regret Analysis with Predetermined Learning Rates

In this section, we tackle the regret analysis of OG+ and OptDA+ run with non-adaptive learning rates. We prove bounds on the pseudo-regret $\max_{p^i\in\mathcal{K}^i}\mathbb{E}[\mathrm{Reg}^i_T(p^i)]$ and on the sum of the expected magnitude of the noiseless feedback $\sum_{t=1}^T\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2]$. In fact, in our analysis, building bounds on $\sum_{t=1}^T\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2]$ is a crucial step for deriving bounds on the pseudo-regret.

Moreover, as the loss functions are convex in their respective player's action parameter, a player's regret can be bounded by its linearized counterpart, as stated in the following lemma.

Lemma 8. *Let Assumption 1 hold. Then, for all $i \in \mathcal{N}$, any sequence of actions $(\mathbf{x}_t)_{t\in\mathbb{N}}$, and all reference point $p^i \in \mathcal{X}^i$, we have*

$$\mathrm{Reg}_T^i(p^i) \leq \sum_{t=1}^{T} \langle V^i(\mathbf{x}_t), x_t^i - p^i \rangle$$

We therefore focus exclusively on bounding the linearized regret in the sequel.

F.1 Bounds for OG+

In this part we will simply assume the learning rates to be $\mathcal{F}_1$-measurable, a technical detailed that we ignored in the main text. The global quasi-descent inequality of OG+ introduced in Lemma 5 indeed allows us to bound several important quantities, as shown below.

Proposition 5 (Bound on sum of squared norms)**.** *Let Assumptions 1–3 hold and all players run* (OG+) *with learning rates described in Theorem 1. Then, for all $T \in \mathbb{N}$ and $\mathbf{x}_\star \in \mathcal{X}_\star$, we have*

$$\begin{aligned}
&\mathbb{E}[\|\mathbf{X}_{t+1} - \mathbf{x}_\star\|^2] + \frac{1}{2}\sum_{t=1}^{T} \gamma_t \eta_{t+1}\, \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] \\
&\quad \leq \|\mathbf{X}_1 - \mathbf{x}_\star\|^2 + \gamma_1\eta_2\|V(\mathbf{X}_1)\|^2 + \sum_{t=1}^{T} \left(9\gamma_t^3\eta_{t+1}NL^2 + \gamma_t^2\eta_{t+1}(4N+1)L + (\eta_{t+1})^2\right) N\sigma_A^2.
\end{aligned}$$

Accordingly, $\sum_{t=1}^{\infty} \gamma_t\eta_{t+1}\, \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] < \infty$.

Proof. Since $\hat{\mathbf{V}}_{1/2} = 0$, we have $\mathbf{X}_{3/2} = \mathbf{X}_1$ and with $\mathbf{X}_2 = \mathbf{X}_1 - \eta_2\hat{\mathbf{V}}_{3/2}$ we obtain

$$\|\mathbf{X}_2 - \mathbf{x}_\star\|^2 = \|\mathbf{X}_1 - \mathbf{x}_\star\|^2 - 2\eta_2\langle \hat{\mathbf{V}}_{3/2}, \mathbf{X}_{3/2} - \mathbf{x}_\star\rangle + \eta_2^2\|\hat{\mathbf{V}}_{3/2}\|^2.$$

Taking expectation then gives

$$\begin{aligned}
\mathbb{E}[\|\mathbf{X}_2 - \mathbf{x}_\star\|^2] &= \mathbb{E}[\|\mathbf{X}_1 - \mathbf{x}_\star\|^2 - 2\eta_2\langle \mathbf{V}(\mathbf{X}_{3/2}), \mathbf{X}_{3/2} - \mathbf{x}_\star\rangle + \eta_2^2\|\hat{\mathbf{V}}_{3/2}\|^2] \\
&\leq \mathbb{E}[\|\mathbf{X}_1 - \mathbf{x}_\star\|^2 + \eta_2^2\|\hat{\mathbf{V}}_{3/2}\|^2],
\end{aligned} \tag{15}$$

where we have used Assumption 2 to deduce that $\langle \mathbf{V}(\mathbf{X}_{3/2}), \mathbf{X}_{3/2} - \mathbf{x}_\star\rangle \geq 0$. Taking total expectation of the inequality of Lemma 5, summing from $t = 2$ to T, and further adding (15) gives

$$\begin{aligned}
&\underbrace{\mathbb{E}\left[\|\mathbf{X}_{T+1} - \mathbf{x}_\star\|^2 + \sum_{t=2}^{T}\gamma_t\eta_{t+1}(\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2)\right]}_{(A)} \\
&\quad \leq \mathbb{E}\Bigg[\|\mathbf{X}_1 - \mathbf{x}_\star\|^2 + \sum_{t=1}^{T}(\eta_{t+1})^2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2 + \sum_{t=2}^{T} 3\gamma_t\eta_{t+1}(\eta_t^2 + \gamma_t^2)NL^2\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2 \\
&\qquad + \sum_{t=2}^{T} 3\gamma_t\eta_{t+1}(\gamma_{t-1})^2NL^2\|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2 + \sum_{t=2}^{T}\left(\gamma_t^2\eta_{t+1} + N\eta_{t+1}(\eta_t + \gamma_t)^2\right)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2\Bigg].
\end{aligned}$$

We use Assumption 3 to bound the noise terms. For example, we have

$$\mathbb{E}[\|\boldsymbol{\xi}_{t+\frac{1}{2}}\|^2] = \sum_{i=1}^{N}\mathbb{E}[\|\xi_{t+\frac{1}{2}}^i\|^2] \leq \sum_{i=1}^{N}(\sigma_A^2 + \sigma_M^2\|V^i(\mathbf{x}_t)\|^2) = \sigma_M^2\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + N\sigma_A^2. \tag{16}$$

Subsequently,

$$\begin{aligned}
\mathbb{E}[(\eta_{t+1})^2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2] &= (\eta_{t+1})^2\,\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \|\boldsymbol{\xi}_{t+\frac{1}{2}}\|^2] \\
&\leq (\eta_{t+1})^2\left(\mathbb{E}[(1+\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] + N\sigma_A^2\right).
\end{aligned}$$

Along with the fact that the learning rates are non-increasing and $\eta_s \leq \gamma_s$ for all $s \in \mathbb{N}$, we get

$$
\begin{aligned}
(A) \leq \mathbb{E}\Bigg[& \|\mathbf{X}_1 - \mathbf{x}_\star\|^2 + \sum_{t=1}^{T} (\eta_{t+1})^2 \left((1+\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + N\sigma_A^2 \right) \\
& + \sum_{t=2}^{T} \left(6\gamma_t^3 \eta_{t+1} N L^2 (1+\sigma_M^2) + \gamma_t^2 \eta_{t+1} (4N+1) L \sigma_M^2 \right) \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2 \\
& + \sum_{t=2}^{T} \left(6\gamma_t^3 \eta_{t+1} N L^2 + \gamma_t^2 \eta_{t+1} (4N+1) L \right) N\sigma_A^2 \\
& + \sum_{t=3}^{T} 3(\gamma_{t-1})^3 \eta_t N L^2 \left((1+\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{t-\frac{3}{2}})\|^2 + N\sigma_A^2 \right) \Bigg].
\end{aligned} \tag{17}
$$

Re-indexing the summations and adding positive terms to the right-hand side (RHS) of the inequality, we deduce

$$
\begin{aligned}
(A) \leq \mathbb{E}\Bigg[& \|\mathbf{X}_1 - \mathbf{x}_\star\|^2 + \sum_{t=2}^{T} \left(9\gamma_t^3 \eta_{t+1} N L^2 (1+\sigma_M^2) + \gamma_t^2 \eta_{t+1} (4N+1) L \sigma_M^2 \right) \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2 \\
& + \sum_{t=1}^{T} (\eta_{t+1})^2 (1+\sigma_M^2) \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 \\
& + \sum_{t=1}^{T} \left((\eta_{t+1})^2 + 9\gamma_t^3 \eta_{t+1} N L^2 + \gamma_t^2 \eta_{t+1} (4N+1) L \right) N\sigma_A^2 \Bigg].
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
(A) = & \|\mathbf{X}_{T+1} - \mathbf{x}_\star\|^2 - \gamma_1 \eta_2 \|\mathbf{V}(\mathbf{X}_{3/2})\|^2 \\
& + \sum_{t=1}^{T} \gamma_t \eta_{t+1} \, \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] + \sum_{t=2}^{T} \gamma_t \eta_{t+1} \, \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2].
\end{aligned}
$$

Combining the above two (in)equalities, rearranging, and using $\mathbf{X}_{3/2} = \mathbf{X}_1$ leads to

$$
\begin{aligned}
& \mathbb{E}[\|\mathbf{X}_{T+1} - \mathbf{x}_\star\|^2] + \sum_{t=1}^{T} \gamma_t \eta_{t+1} \left(1 - \frac{(1+\sigma_M^2)\eta_{t+1}}{\gamma_t} \right) \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] \\
& \quad + \sum_{t=2}^{T} \gamma_t \eta_{t+1} (1 - a_t(1+\sigma_M^2) - b_t \sigma_M^2) \, \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2] \\
& \leq \|\mathbf{X}_1 - \mathbf{x}_\star\|^2 + \gamma_1 \eta_2 \|V(\mathbf{X}_1)\|^2 + \sum_{t=1}^{T} \gamma_t \eta_{t+1} \left(\frac{\eta_{t+1}}{\gamma_t} + a_t + b_t \right) N\sigma_A^2,
\end{aligned}
$$

where $a_t = 9\gamma_t^2 N L^2$ and $b_t = \gamma_t (4N+1) L$. To conclude, we notice that with the learning rate choices of Theorem 1, it always holds $1 - (1+\sigma_M^2)(\eta_{t+1}/\gamma_t) \geq 1/2$, $1 - a_t(1+\sigma_M^2) - b_t\sigma_M^2 \geq 0$, and $\sum_{t=1}^{+\infty} \gamma_t \eta_{t+1} (\eta_{t+1}/\gamma_t + a_t + b_t) N\sigma_A^2 < +\infty$. □

From Proposition 5 we obtain immediately the bounds on $\sum_{t=1}^{T} \mathbb{E}[\|\mathbf{V}(X_{t+\frac{1}{2}})\|^2]$ of OG+ as claimed in Section 6.

Theorem 8. *Let Assumptions 1–3 hold and all players run (OG+) with non-increasing learning rate sequences $(\gamma_t)_{t\in\mathbb{N}}$ and $(\eta_t)_{t\in\mathbb{N}}$ satisfying (4). We have*

(a) If there exists $q \in [0, 1/4]$ such that $\gamma_t = \mathcal{O}(1/(t^{\frac{1}{4}}\sqrt{\log t}))$, $\gamma_t = \Omega(1/t^{\frac{1}{2}-q})$, and $\eta_t = \Theta(1/(\sqrt{t}\log t))$, then

$$
\sum_{t=1}^{T} \mathbb{E}[\|\mathbf{V}(X_{t+\frac{1}{2}})\|^2] = \tilde{\mathcal{O}}\left(T^{1-q}\right)
$$

(b) *If the noise is multiplicative (i.e., $\sigma_A = 0$) and the learning rates are constant $\gamma_t \equiv \gamma$, $\eta_t \equiv \eta$, then*

$$\sum_{t=1}^{T} \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] \leq \frac{2\,\mathrm{dist}(\mathbf{X}_1, \mathcal{X}_\star)^2}{\gamma\eta} + 2\|\mathbf{V}(\mathbf{X}_1)\|^2.$$

In particular, if the equalities hold in (4), then the above is in $\mathcal{O}(N^3L^2(1+\sigma_M^2)^3)$.

Proof. Let $\mathbf{x}_\star = \Pi_{\mathcal{X}_\star}(\mathbf{X}_1)$. By the choice of our learning rates, the constant

$$C := \|\mathbf{X}_1 - \mathbf{x}_\star\|^2 + \gamma_1\eta_2\|V(\mathbf{X}_1)\|^2 + \sum_{t=1}^{+\infty}(9\gamma_t^3\eta_{t+1}NL^2 + \gamma_t^2\eta_{t+1}(4N+1)L + (\eta_{t+1})^2)N\sigma_A^2.$$

is finite. In addition, from Proposition 5 we know tat

$$\sum_{t=1}^{T} \gamma_t\eta_{t+1}\,\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] \leq 2C.$$

On the other hand since the learning rates are non-increasing, it holds

$$\sum_{t=1}^{T} \gamma_t\eta_{t+1}\,\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] \geq \gamma_{T+1}\eta_{T+1}\sum_{t=1}^{T}\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2].$$

As a consequence,

$$\sum_{t=1}^{T} \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] \leq \frac{2C}{\gamma_{T+1}\eta_{T+1}}. \tag{18}$$

The results are then immediate from our choice of learning rates. □

Remark 3. In the estimation of (b) we use $\mathrm{dist}(\mathbf{X}_1, \mathcal{X}_\star)^2 = \mathcal{O}(N)$ and $1/\gamma = \mathcal{O}(NL(1+\sigma_M^2))$. We can get improved dependence on N if the noises of the players are supposed to be mutually independent conditioned on the past. In fact, in this case we only require $\gamma \leq \min\left(\frac{1}{3L\sqrt{2N(1+\sigma_M^2)}}, \frac{1}{8L\sigma_M^2}\right)$.

Bounding Linearized Regret. We proceed to bound the linearized regret. The following lemma is a direct consequence of the individual quasi-descent inequality of Lemma 4.

Lemma 9 (Bound on linearized regret)**.** *Let Assumptions 1–3 hold and all players run* (OG+) *with learning rates described in Theorem 1. Then, for all $i \in \mathcal{N}$, $T \in \mathbb{N}$, and $p^i \in \mathcal{X}^i$, we have*

$$\sum_{t=1}^{T} \mathbb{E}[\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle] \leq \mathbb{E}\left[\frac{\|X_1^i - p^i\|^2}{2\eta_2} + \sum_{t=2}^{T}\left(\frac{1}{2\eta_{t+1}} - \frac{1}{2\eta_t}\right)\|X_t^i - p^i\|^2 + \sum_{t=1}^{T}\left(\frac{3\gamma_t}{4}\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \frac{a_t\sigma_A^2}{2}\right)\right],$$

where $a_t = 9\gamma_t^3L^2N + \gamma_t^2(4N+1)L + \eta_{t+1}$.

Proof. Applying Lemma 4, dividing both sides of (10) by η_{t+1}, rearranging, taking total expectation, and using Assumption 3, we get

$$\begin{aligned}
&\mathbb{E}[2\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle] \\
&\leq \mathbb{E}\Bigg[\frac{\|X_t^i - p^i\|^2}{\eta_{t+1}} - \frac{\|X_{t+1}^i - p^i\|^2}{\eta_{t+1}} \\
&\quad - \gamma_t(\|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \|V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2) + \gamma_t\|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 \\
&\quad + \gamma_t^2L\sigma_M^2\|V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 + (\eta_t+\gamma_t)^2L\sigma_M^2\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2 + \eta_{t+1}(1+\sigma_M^2)\|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2 \\
&\quad + \gamma_t^2L\sigma_A^2 + (\eta_t+\gamma_t)^2LN\sigma_A^2 + \eta_{t+1}\sigma_A^2\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{E}\Bigg[\frac{\|X_t^i - p^i\|^2}{\eta_{t+1}} - \frac{\|X_{t+1}^i - p^i\|^2}{\eta_{t+1}} \\
&\quad + 3\gamma_t^3 L^2 \|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2 + 3\gamma_t \eta_t^2 L^2 \|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2 + 3\gamma_t(\gamma_{t-1})^2 L^2 \|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2 \\
&\quad + 5\gamma_t^2 L \sigma_M^2 \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2 + \eta_{t+1}(1+\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \gamma_t^2(4N+1)L\sigma_A^2 + \eta_{t+1}\sigma_A^2\Bigg] \\
&\leq \mathbb{E}\Bigg[\frac{\|X_t^i - p^i\|^2}{\eta_t} + \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t}\right)\|X_t^i - p^i\|^2 - \frac{\|X_{t+1}^i - p^i\|^2}{\eta_{t+1}} \\
&\quad + \frac{\gamma_t}{2}\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \frac{5\gamma_t}{6}\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2 + 3(\gamma_{t-1})^3 L^2 \|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2 \\
&\quad + 6\gamma_t^3 L^2 N \sigma_A^2 + \gamma_t^2(4N+1)L\sigma_A^2 + \eta_{t+1}\sigma_A^2\Bigg].
\end{aligned}
$$

In the last inequality we have used $\eta_t^2 \leq \gamma_t^2 \leq 1/(18L^2(1+\sigma_M^2))$, $5\gamma_t L\sigma_M^2 \leq \gamma_1(4N+1)L\sigma_M^2 \leq 1/2$ and $\eta_{t+1}(1+\sigma_M^2) \leq \eta_t(1+\sigma_M^2) \leq \gamma_t/2$. As for the $\|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2$ term, we recall that $\hat{\mathbf{V}}_{1/2} = 0$ and otherwise its expectation can again be bounded using Assumption 3. Summing the above inequality from $t = 2$ to T and dividing both sides by 2, we then obtain

$$
\begin{aligned}
\sum_{t=2}^{T} \mathbb{E}[\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X_{t+\frac{1}{2}}^i - p^i\rangle] \leq \mathbb{E}\Bigg[&\frac{\|X_2^i - p^i\|^2}{2\eta_2} + \sum_{t=2}^{T}\left(\frac{1}{2\eta_{t+1}} - \frac{1}{2\eta_t}\right)\|X_t^i - p^i\|^2 \\
&+ \sum_{t=2}^{T}\frac{1}{4}\Bigg(\gamma_t\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + 2\gamma_t\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2 \\
&+ 18\gamma_t^3 L^2 N\sigma_A^2 + \gamma_t^2(8N+2)L\sigma_A^2 + 2\eta_{t+1}\sigma_A^2\Bigg)\Bigg].
\end{aligned} \tag{19}
$$

For $t = 1$, since $X_{3/2}^i = X_1^i$ and $X_2^i = X_1^i - \eta_2 \hat{V}_{3/2}^i$, we have

$$
\|X_2^i - p^i\|^2 = \|X_t^1 - p^i\|^2 - 2\eta_2\langle \hat{V}_{3/2}^i, X_{3/2}^i - p^i\rangle + \eta_2^2\|\hat{V}_{3/2}^i\|^2.
$$

Taking expectation then gives

$$
\mathbb{E}[\|X_2^i - p^i\|^2] \leq \mathbb{E}[\|X_1^i - p^i\|^2 - 2\eta_2\langle V^i(\mathbf{X}_{3/2}), X_{3/2}^i - p^i\rangle + \eta_2^2(1+\sigma_M^2)\|V^i(\mathbf{X}_{3/2})\|^2 + \eta_2^2\sigma_A^2]. \tag{20}
$$

Combining (19) and (20) and bounding $\eta_2(1+\sigma_M^2)\|V^i(\mathbf{X}_{3/2})\|^2 \leq (\gamma_t/2)\|\mathbf{V}(\mathbf{X}_{3/2})\|^2$, we get the desired inequality. □

With Lemma 9 and Proposition 5, we are now ready to prove our result concerning the regret of OG+. The main difficulty here consists in controlling the sum of $(1/(2\eta_{t+1}) - 1/(2\eta_t))\,\mathbb{E}[\|X_t^i - p^i\|^2]$ when the learning rates are not constant.

Theorem 9. *Let Assumptions 1–3 hold and all players run (OG+) with non-increasing learning rate sequences $(\gamma_t)_{t\in\mathbb{N}}$ and $(\eta_t)_{t\in\mathbb{N}}$ satisfying (4). For any $i \in \mathcal{N}$ and bounded set $\mathcal{K}^i \subset \mathcal{X}^i$ with $R \geq \sup_{p^i}\|X_1^i - p^i\|$, we have:*

(a) If $\gamma_t = \mathcal{O}(1/(t^{\frac{1}{4}}\sqrt{\log t}))$ and $\eta_t = \Theta(1/(\sqrt{t}\log t))$, then

$$
\max_{p^i\in\mathcal{K}^i} \mathbb{E}\left[\sum_{t=1}^{T}\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X_{t+\frac{1}{2}}^i - p^i\rangle\right] = \tilde{\mathcal{O}}\left(\sqrt{T}\right).
$$

(b) If the noise is multiplicative (i.e., $\sigma_A = 0$) and the learning rates are constant $\gamma_t \equiv \gamma$, $\eta_t \equiv \eta$, then

$$
\max_{p^i\in\mathcal{K}^i} \mathbb{E}\left[\sum_{t=1}^{T}\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X_{t+\frac{1}{2}}^i - p^i\rangle\right] \leq \frac{R^2}{2\eta} + \frac{2}{\eta}(\operatorname{dist}(\mathbf{X}_1, \mathcal{X}_\star)^2 + \gamma\eta\|\mathbf{V}(\mathbf{X}_1)\|^2).
$$

In particular, if the equalities hold in (4), the above is in $\mathcal{O}(N^2L(1+\sigma_M^2)^2)$.

Proof. Let $\mathbf{x}_\star = \Pi_{\mathcal{X}_\star}(\mathbf{X}_1)$ be the projection of $\mathbf{X}_1$ onto the solution set. For any $p^i \in \mathcal{K}^i$, it holds

$$\begin{aligned}
&\sum_{t=2}^{T}\left(\frac{1}{2\eta_{t+1}}-\frac{1}{2\eta_t}\right)\|X_t^i-p^i\|^2\\
&\le \sum_{t=2}^{T}\left(\frac{1}{\eta_{t+1}}-\frac{1}{\eta_t}\right)\left(\|X_t^i-x_\star^i\|^2+\|x_\star^i-p^i\|^2\right)\\
&\le \sum_{t=2}^{T}\left(\frac{1}{\eta_{t+1}}-\frac{1}{\eta_t}\right)\left(\|\mathbf{X}_t-\mathbf{x}_\star\|^2+\|x_\star^i-X_1^i+X_1^i-p^i\|^2\right)\\
&\le \left(\frac{1}{\eta_{T+1}}-\frac{1}{\eta_2}\right)\left(2\|X_1^i-x_\star^i\|^2+2R^2\right)+\sum_{t=2}^{T}\left(\frac{1}{\eta_{t+1}}-\frac{1}{\eta_t}\right)\|\mathbf{X}_t-\mathbf{x}_\star\|^2. \qquad (21)
\end{aligned}$$

To proceed, with Proposition 5, we know that for C defined in the proof of Theorem 8, we have for all $t \in \mathbb{N}$

$$\mathbb{E}[\|\mathbf{X}_{t+1}-\mathbf{x}_\star\|^2]+\sum_{s=1}^{t}\frac{\gamma_s\eta_{s+1}}{2}\,\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{s+\frac{1}{2}})\|^2]\le C \qquad (22)$$

We can therefore write

$$\sum_{t=2}^{T}\left(\frac{1}{\eta_{t+1}}-\frac{1}{\eta_t}\right)\mathbb{E}[\|\mathbf{X}_t-\mathbf{x}_\star\|^2]\le\sum_{t=2}^{T}\left(\frac{1}{\eta_{t+1}}-\frac{1}{\eta_t}\right)C\le\frac{C}{\eta_{T+1}}. \qquad (23)$$

Since $\eta_{T+1}\le\eta_{t+1}$ for all $t\le T$. From (22) we also deduce

$$\sum_{t=1}^{T}\gamma_t\,\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2]\le\frac{1}{\eta_{T+1}}\sum_{t=1}^{T}\gamma_t\eta_{t+1}\,\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2]\le\frac{2C}{\eta_{T+1}}. \qquad (24)$$

Plugging (21), (23), and (24) into Lemma 9, we obtain

$$\sum_{t=1}^{T}\mathbb{E}[\langle V^i(\mathbf{X}_{t+\frac{1}{2}}),X_{t+\frac{1}{2}}^i-p^i\rangle]\le\mathbb{E}\left[\frac{2\operatorname{dist}(\mathbf{X}_1,\mathcal{X}_\star)^2+2R^2}{\eta_{T+1}}+\frac{3C}{\eta_{T+1}}+\sum_{t=1}^{T}\frac{a_t\sigma_A^2}{2}\right].$$

The result is now immediate from $\gamma_t=\mathcal{O}(1/(t^{\frac{1}{4}}\sqrt{\log t}))$ and $\eta_t=\Theta(1/(\sqrt{t}\log t))$.

(b) Let $p^i\in\mathcal{K}^i$. With $\sigma_A^2=0$, constant learning rates, and $\|X_1^i-p^i\|\le R^2$, Lemma 9 gives

$$\sum_{t=1}^{T}\mathbb{E}[\langle V^i(\mathbf{X}_{t+\frac{1}{2}}),X_{t+\frac{1}{2}}^i-p^i\rangle]\le\mathbb{E}\left[\frac{R^2}{2\eta}+\sum_{t=1}^{T}\frac{3\gamma}{4}\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2\right],$$

We conclude immediately with the help of Theorem 8(b). □

F.2 Bounds for OptDA+

For the analysis of OptDA+, we first establish two preliminary bounds respectively for the linearized regret and for the sum of the squared operator norms. These bounds are used later for deriving more refined bounds in the non-adaptive and the adaptive case. We use the notation $\eta_1^i=\eta_2^i$.

Lemma 10 (Bound on linearized regret)**.** *Let Assumptions 1 and 3 hold and all players run* (OptDA+) *with non-increasing learning rates satisfying Assumption 5 and $\boldsymbol{\eta}_t\le\boldsymbol{\gamma}_t$ for all $t\in\mathbb{N}$. Then, for all $i\in\mathcal{N}$, $T\in\mathbb{N}$, and $p^i\in\mathcal{X}^i$, we have*

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}\langle V^i(\mathbf{X}_{t+\frac{1}{2}}),X_{t+\frac{1}{2}}^i-p^i\rangle\right]\le\mathbb{E}\Bigg[&\frac{\|X_1^i-p^i\|^2}{2\eta_{T+1}^i}+\frac{1}{2}\sum_{t=1}^{T}\eta_t^i\|\hat{V}_{t+\frac{1}{2}}^i\|^2\\
&+\sum_{t=2}^{T}\gamma_t^iL^2\left(3\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2+\frac{3}{2}\|\mathbf{X}_t-\mathbf{X}_{t-1}\|^2\right)\\
&+\frac{1}{2}\sum_{t=2}^{T}((\gamma_t^i)^2L\|\xi_{t-\frac{1}{2}}^i\|^2+4L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2)\Bigg]. \qquad (25)
\end{aligned}$$

Proof. Applying Lemma 6, dropping non-positive terms on the RHS of (12), using

$$\min\left(-\frac{\|X_t^i - X_{t+1}^i\|^2}{2\eta_t^i} + \eta_t^i\|\hat{V}_{t+\frac{1}{2}}^i\|^2,\, 0\right) \le 0$$

and taking total expectation gives

$$\begin{aligned}\mathbb{E}\left[\frac{\|X_{t+1}^i - p^i\|^2}{\eta_{t+1}^i}\right] \le \mathbb{E}\Bigg[&\frac{\|X_t^i - p^i\|^2}{\eta_t^i} + \left(\frac{1}{\eta_{t+1}^i} - \frac{1}{\eta_t^i}\right)\|X_1^i - p^i\|^2 \\ &- 2\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X_{t+\frac{1}{2}}^i - p^i\rangle + \gamma_t^i\|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 \\ &+ (\gamma_t^i)^2 L\|\xi_{t-\frac{1}{2}}^i\|^2 + L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{(\boldsymbol{\eta}_t+\boldsymbol{\gamma}_t)^2} + \eta_t^i\|\hat{V}_{t+\frac{1}{2}}^i\|^2\Bigg]. \end{aligned} \tag{26}$$

The above inequality holds for $t \ge 2$. As for $t = 1$, we notice that with $X_2^i = X_1^i - \eta_2^i\hat{V}_{3/2}^i$, we have in fact

$$\|X_2^i - p^i\|^2 = \|X_1^i - p^i\|^2 - 2\eta_2^i\langle\hat{V}_{3/2}^i, X_1^i - p^i\rangle + (\eta_2^i)^2\|\hat{V}_{3/2}^i\|^2.$$

As $X_{3/2}^i = X_1^i = 0$ and $\eta_1^i = \eta_2^i$, the above implies

$$\mathbb{E}\left[\langle V^i(X_{3/2}^i), X_{3/2}^i - p^i\rangle\right] = \mathbb{E}\left[\frac{\|X_1^i - p^i\|^2}{2\eta_2^i} - \frac{\|X_2^i - p^i\|^2}{2\eta_2^i} + \frac{\eta_1^i\|\hat{V}_{3/2}^i\|^2}{2}\right]. \tag{27}$$

Summing (26) from $t = 2$ to T, dividing by 2, adding (27), and using $\boldsymbol{\eta}_t \le \boldsymbol{\gamma}_t$ leads to

$$\begin{aligned}\sum_{t=1}^T \mathbb{E}[\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X_{t+\frac{1}{2}}^i - p^i\rangle] \le \frac{1}{2}\mathbb{E}\Bigg[&\frac{\|X_1^i - p^i\|^2}{\eta_{T+1}^i} + \sum_{t=1}^T \eta_t^i\|\hat{V}_{t+\frac{1}{2}}^i\|^2 \\ &+ \sum_{t=2}^T \gamma_t^i\|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 \\ &+ \sum_{t=2}^T((\gamma_t^i)^2 L\|\xi_{t-\frac{1}{2}}^i\|^2 + 4L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2})\Bigg].\end{aligned}$$

Similar to (11), we can bound the difference term by

$$\|V^i(\mathbf{X}_{t+\frac{1}{2}}) - V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 \le 3L^2\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2} + 3L^2\|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2_{(\boldsymbol{\gamma}_{t-1})^2} + 3L^2\|\mathbf{X}_t - \mathbf{X}_{t-1}\|^2.$$

Combining the above two inequalities and using $\hat{\mathbf{V}}_{1/2} = 0$ gives the desired inequality. □

Lemma 11 (Bound on sum of squared norms)**.** *Let Assumptions 1–3 hold and all players run* (OptDA+) *with non-increasing learning rates satisfying Assumption 5 and $\boldsymbol{\eta}_t \le \boldsymbol{\gamma}_t$ for all $t \in \mathbb{N}$. Then, for all $T \in \mathbb{N}$ and $\mathbf{x}_\star \in \mathcal{X}_\star$, we have*

$$\begin{aligned}&\sum_{t=2}^T \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t} + \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t}] \\ &\le \mathbb{E}\Bigg[\|\mathbf{X}_1 - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_{T+1}} + \sum_{t=1}^T\left(3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|^2_{\boldsymbol{\gamma}_t} - \|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2_{1/(2\boldsymbol{\eta}_t)}\right) \\ &\quad + \sum_{t=2}^T 6\|\boldsymbol{\gamma}_t\|_1 L^2\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2} + \sum_{t=2}^T(4N+1)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2} + \sum_{t=1}^T 2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2_{\boldsymbol{\eta}_t}\Bigg].\end{aligned} \tag{28}$$

Proof. This is a direct consequence of Lemma 7. In fact, taking total expectation of (13) and summing from $t = 2$ to T gives already

$$\begin{aligned}
&\sum_{t=2}^{T} \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2 + \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2] \\
&\leq \mathbb{E}\Bigg[\|\mathbf{X}_2 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_2}^2 + \|\mathbf{X}_1 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_{T+1} - 1/\boldsymbol{\eta}_2}^2 \\
&\qquad + \sum_{t=2}^{T} (3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t-1})\|_{\boldsymbol{\gamma}_t}^2 - \|\mathbf{X}_t - \mathbf{X}_{t+1}\|_{1/(2\boldsymbol{\eta}_t)}^2) \\
&\qquad + \sum_{t=2}^{T} 6\|\boldsymbol{\gamma}_t\|_1 L^2 \|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2 + \sum_{t=2}^{T} (4N+1)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2 + \sum_{t=2}^{T} 2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|_{\boldsymbol{\eta}_t}^2\Bigg]. \qquad (29)
\end{aligned}$$

We have in particular used $\hat{\mathbf{V}}_{1/2} = 0$ to bound

$$\begin{aligned}
&\sum_{t=2}^{T} 3L^2(\|\boldsymbol{\gamma}_t\|_1 \|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2 + \|\boldsymbol{\gamma}_{t-1}\|_1 \|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|_{(\boldsymbol{\gamma}_{t-1})^2}^2) \\
&= \sum_{t=2}^{T} 3\|\boldsymbol{\gamma}_t\|_1 L^2 \|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2 + \sum_{t=3}^{T} \|\boldsymbol{\gamma}_t\|_1 3NL^2 \|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2 \\
&\leq \sum_{t=2}^{T} 6\|\boldsymbol{\gamma}_t\|_1 L^2 \|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2.
\end{aligned}$$

To obtain (28), we further bound

$$\sum_{t=2}^{T} 3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t-1})\|_{\boldsymbol{\gamma}_t}^2 = \sum_{t=1}^{T-1} 3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|_{\boldsymbol{\gamma}_{t+1}}^2 \leq \sum_{t=1}^{T} 3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|_{\boldsymbol{\gamma}_t}^2 \qquad (30)$$

For $t = 1$, we use (27) with $p^i \leftarrow x_\star^i$; that is

$$\frac{\|X_2^i - x_\star^i\|^2}{\eta_2^i} = \frac{\|X_1^i - x_\star^i\|^2}{\eta_2^i} - 2\langle V^i(\mathbf{X}_{3/2}) + \xi_{3/2}^i, X_1^i - x_\star^i\rangle + \eta_1^i \|\hat{V}_{3/2}^i\|^2.$$

Since $\mathbf{X}_{3/2} = \mathbf{X}_1$, summing the above inequality from $i = 1$ to N leads to

$$\|\mathbf{X}_2 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_2}^2 = \|\mathbf{X}_1 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_2}^2 - 2\langle \mathbf{V}(\mathbf{X}_{3/2}) + \boldsymbol{\xi}_{3/2}, \mathbf{X}_{3/2} - \mathbf{x}_\star\rangle + \|\hat{\mathbf{V}}_{3/2}\|_{\boldsymbol{\eta}_1}^2. \qquad (31)$$

Assumptions 2 and 3 together ensure

$$\mathbb{E}[\langle \mathbf{V}(\mathbf{X}_{3/2}) + \boldsymbol{\xi}_{3/2}, \mathbf{X}_{3/2} - \mathbf{x}_\star\rangle] = \langle \mathbf{V}(\mathbf{X}_{3/2}), \mathbf{X}_{3/2} - \mathbf{x}_\star\rangle \geq 0.$$

Subsequently,

$$\begin{aligned}
\mathbb{E}[\|\mathbf{X}_2 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_2}^2] &\leq \mathbb{E}[\|\mathbf{X}_1 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_2}^2 + \|\hat{V}_{3/2}^i\|_{\boldsymbol{\eta}_2}^2] \\
&\leq \mathbb{E}[\|\mathbf{X}_1 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_2}^2 + 2\|\hat{V}_{3/2}^i\|_{\boldsymbol{\eta}_1}^2 - \|\mathbf{X}_1 - \mathbf{X}_2\|_{(1/2\boldsymbol{\eta}_1)}^2]. \qquad (32)
\end{aligned}$$

Combining (29), (30), and (32) gives exactly (28). □

F.2.1 Dedicated Analysis for Non-Adaptive OptDA+

In this part, we show how the non-adaptive learning rates suggested in Theorem 2 helps to achieve small regret and lead to fast convergence of the norms of the payoff gradients.

Proposition 6 (Bound on sum of squared norms)**.** *Let Assumptions 1–3 hold and all players run* (OptDA+) *with learning rates described in Theorem 2. Then, for all $T \in \mathbb{N}$ and $\mathbf{x}_\star \in \mathcal{X}_\star$, we have*

$$\begin{aligned}
&\frac{1}{2}\sum_{s=1}^{T} \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2] + \sum_{t=1}^{T} 21\|\boldsymbol{\gamma}_1\|_\infty NL^2\, \mathbb{E}[\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2] \\
&\leq \|\mathbf{X}_1 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_{T+1}}^2 + \|\mathbf{V}(\mathbf{X}_1)\|_{\boldsymbol{\gamma}_1}^2 + \sum_{t=1}^{T} \left(6\|\boldsymbol{\gamma}_t\|_\infty^3 NL^2 + \|\boldsymbol{\gamma}_t\|_\infty^2 (4N+1)L + 2\|\boldsymbol{\eta}_t\|_\infty\right) N\sigma_A^2
\end{aligned}$$

Proof. We first apply Lemma 11 to obtain (28). We bound the expectations of the following three terms separately.

$$A_t = 3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|^2_{\boldsymbol{\gamma}_t} - \|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2_{1/(2\boldsymbol{\eta}_t)},$$
$$B_t = 6\|\boldsymbol{\gamma}_t\|_1 L^2 \|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2} + (4N+1)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2}, \quad C_t = 2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2_{\boldsymbol{\eta}_t}.$$

To bound A_t, we first use $\boldsymbol{\eta}_t \leq \boldsymbol{\gamma}_t/(4(1+\sigma_M^2)) \leq \|\boldsymbol{\gamma}_1\|_\infty/(4(1+\sigma_M^2))$ to get

$$\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2_{1/(2\boldsymbol{\eta}_t)} \geq \frac{2(1+\sigma_M^2)}{\|\boldsymbol{\gamma}_1\|_\infty}\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2.$$

Moreover, with $\|\boldsymbol{\gamma}_1\|^2_\infty \leq 1/(12NL^2(1+\sigma_M^2))$ we indeed have

$$\frac{2(1+\sigma_M^2)}{\|\boldsymbol{\gamma}_1\|_\infty} \geq 24NL^2(1+\sigma_M^2)^2\|\boldsymbol{\gamma}_1\|_\infty \geq 24NL^2\|\boldsymbol{\gamma}_1\|_\infty.$$

On the other hand, with the Lipschitz continuity of $(V^i)_{i\in\mathcal{N}}$ it holds

$$3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|^2_{\boldsymbol{\gamma}_t} \leq \sum_{i=1}^N 3\gamma_t^i L^2\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2 \leq 3\|\boldsymbol{\gamma}_1\|_\infty NL^2\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2.$$

Combining the above inequalities we deduce that $A_t \leq -21\|\boldsymbol{\gamma}_1\|_\infty NL^2\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2$ and accordingly

$$\mathbb{E}[A_t] \leq \mathbb{E}[-21\|\boldsymbol{\gamma}_1\|_\infty NL^2\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2]. \tag{33}$$

We proceed to bound $\mathbb{E}[B_t]$. The exploration learning rates $\boldsymbol{\gamma}_t$ being $\mathcal{F}_{t-1}$-measurable, using Assumption 3 and the law of total expectation, we get

$$\begin{aligned}
\mathbb{E}[B_t] &= \mathbb{E}\left[\mathbb{E}_{t-1}[6\|\boldsymbol{\gamma}_t\|_1 L^2\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2} + (4N+1)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t}]\right] \\
&= \mathbb{E}\left[\sum_{i=1}^N \left(6\|\boldsymbol{\gamma}_t\|_1(\gamma_t^i)^2L^2\,\mathbb{E}_{t-1}[\|\hat{V}^i_{t-\frac{1}{2}}\|^2] + (\gamma_t^i)^2(4N+1)L\,\mathbb{E}_{t-1}[\|\xi^i_{t-\frac{1}{2}}\|^2]\right)\right] \\
&\leq \mathbb{E}\Bigg[6\|\boldsymbol{\gamma}_t\|^2_\infty NL^2(1+\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t} + \|\boldsymbol{\gamma}_t\|_\infty(4N+1)L\sigma_M^2\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t} \\
&\qquad + (6\|\boldsymbol{\gamma}_t\|^3_\infty NL^2 + \|\boldsymbol{\gamma}_t\|^2_\infty(4N+1)L)N\sigma_A^2\Bigg].
\end{aligned} \tag{34}$$

Similarly, $\boldsymbol{\eta}_{t+1}$ being deterministic and in particular $\mathcal{F}_t$-measurable, we have

$$\mathbb{E}[C_t] = \mathbb{E}[\mathbb{E}_t[2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2_{\boldsymbol{\eta}_t}]] \leq \mathbb{E}\left[2(1+\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\eta}_t} + 2\|\boldsymbol{\eta}_t\|_\infty N\sigma_A^2\right]. \tag{35}$$

Putting together (28), (33), (34), and (35), we get

$$\begin{aligned}
&\sum_{t=2}^T \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t - 2(1+\sigma_M^2)\boldsymbol{\eta}_t} + (1 - a_t(1+\sigma_M^2) - b_t\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t}] \\
&\leq \mathbb{E}\Bigg[\|\mathbf{X}_1 - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_{T+1}} + 2(1+\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{3/2})\|^2_{\boldsymbol{\eta}_1} - \sum_{t=1}^T 21\|\boldsymbol{\gamma}_1\|_\infty NL^2\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2 \\
&\qquad + \sum_{t=1}^T (a_t + b_t + 2\|\boldsymbol{\eta}_t\|_\infty)N\sigma_A^2\Bigg],
\end{aligned}$$

where $a_t = 6\|\boldsymbol{\gamma}_t\|^3_\infty NL^2$ and $b_t = \|\boldsymbol{\gamma}_t\|^2_\infty(4N+1)L$. We conclude by using $\mathbf{X}_{3/2} = \mathbf{X}_1$ and noticing that under our learning rate requirement it is always true that $1 - 6\|\boldsymbol{\gamma}_t\|^2_\infty NL^2(1+\sigma_M^2) - \|\boldsymbol{\gamma}_t\|_\infty(4N+1)L\sigma_M^2 \geq 0$ and $\boldsymbol{\gamma}_t - 2(1+\sigma_M^2)\boldsymbol{\eta}_t \geq \boldsymbol{\gamma}_t/2$. □

Remark 4. We notice that in the analysis, we can replace the common Lipschitz constant by the ones that are proper to each player (i.e., V^i is L^i-Lipschitz continuous) when bounding B_t. This is however hot the case for our bound on A_t, unless we bound directly $\gamma_t^i (L^i)^2$ by a constant.

Again, from Proposition 6 we obtain immediately the bounds on $\sum_{t=1}^T \mathbb{E}[\|\mathbf{V}(X_{t+\frac{1}{2}})\|^2]$ of non-adaptive OptDA+ as claimed in Section 6.

Theorem 10. *Let Assumptions 1–3 hold and all players run (OptDA+) with non-increasing learning rate sequences $(\gamma_t^i)_{t\in\mathbb{N}}$ and $(\eta_t^i)_{t\in\mathbb{N}}$ satisfying (5). We have*

(a) If there exists $q \in [0, 1/4]$ such that $\gamma_t^j = \mathcal{O}(1/t^{\frac{1}{4}})$, $\gamma_t^j = \Omega(1/t^{\frac{1}{2}-q})$, and $\eta_t^j = \Theta(1/\sqrt{t})$ for all $j \in \mathcal{N}$, then

$$\sum_{t=1}^T \mathbb{E}[\|\mathbf{V}(X_{t+\frac{1}{2}})\|^2] = \mathcal{O}\left(T^{1-q}\right)$$

(b) If the noise is multiplicative (i.e., $\sigma_A = 0$) and the learning rates are constant $\gamma_t \equiv \gamma$, $\eta_t \equiv \eta$, then

$$\sum_{t=1}^T \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] \leq \frac{2}{\min_{i\in\mathcal{N}} \gamma^i} \left(\operatorname{dist}_{1/\boldsymbol{\eta}}(\mathbf{X}_1, \mathcal{X}_\star)^2 + \|\mathbf{V}(\mathbf{X}_1)\|_{\boldsymbol{\gamma}_t}^2\right)$$

In particular, if the equalities hold in (5), then the above is in $\mathcal{O}(N^3 L^2 (1+\sigma_M^2)^3)$.

Proof. Let us define $a_t = 6\|\boldsymbol{\gamma}_t\|_\infty^3 N L^2 + \|\boldsymbol{\gamma}_t\|_\infty^2 (4N+1)L + 2\|\boldsymbol{\eta}_t\|_\infty$. From Proposition 6 we know that for all $\mathbf{x}_\star \in \mathcal{X}_\star$, it holds

$$\sum_{s=1}^T \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\gamma}_t/2}^2]] \leq \|\mathbf{X}_1 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_{T+1}}^2 + \|\mathbf{V}(\mathbf{X}_1)\|_{\boldsymbol{\gamma}_1}^2 + \sum_{t=1}^T a_t N \sigma_A^2,$$

Since the learning rates are decreasing, we can lower bound $\boldsymbol{\gamma}_t$ by $\boldsymbol{\gamma}_t \geq \boldsymbol{\gamma}_T \geq \min_{i\in\mathcal{N}} \gamma_T^i$. Accordingly,

$$\sum_{s=1}^T \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2]] \leq \frac{2}{\min_{i\in\mathcal{N}} \gamma_T^i} \left(\|\mathbf{X}_1 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_{T+1}}^2 + \|\mathbf{V}(\mathbf{X}_1)\|_{\boldsymbol{\gamma}_1}^2 + \sum_{t=1}^T a_t N \sigma_A^2\right), \quad (36)$$

The result then follows immediately from our learning rate choices. For (a), we observe that with $\|\boldsymbol{\gamma}_t\|_\infty = \mathcal{O}(1/t^{\frac{1}{4}})$ and $\|\boldsymbol{\eta}_t\|_\infty = \mathcal{O}(1/\sqrt{t})$, we have $\sum_{t=1}^T a_t = \mathcal{O}(\sqrt{T})$, while $\gamma_t^j = \Omega(1/t^{\frac{1}{2}-q})$, and $\eta_t^j = \Omega(1/\sqrt{t})$ guarantees $1/\min_{i\in\mathcal{N}} \gamma_t^T = \mathcal{O}(T^{\frac{1}{2}-q})$ and $1/\min_{i\in\mathcal{N}} \eta_{t+1}^T = \mathcal{O}(\sqrt{T})$. For (b), we take $\mathbf{x}_\star = \arg\min_{\mathbf{x}\in\mathcal{X}_\star} \|\mathbf{X}_1 - \mathbf{x}\|_{1/\boldsymbol{\eta}}$. □

Bounding Linearized Regret. To bound the linearized regret, we refine Lemma 10 as follows.

Lemma 12 (Bound on linearized regret)**.** *Let Assumptions 1–3 hold and all players run (OptDA+) with learning rates described in Theorem 2. Then, for all $i \in \mathcal{N}$, $T \in \mathbb{N}$, and $p^i \in \mathcal{X}^i$, we have*

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^T \langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X_{t+\frac{1}{2}}^i - p^i\rangle\right] \leq \mathbb{E}\Bigg[&\frac{\|X_1^i - p^i\|^2}{2\eta_{T+1}^i} + \sum_{t=1}^T \frac{5}{8}\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2 \\ &+ \sum_{t=1}^{T-1} \frac{3\|\boldsymbol{\gamma}_1\|_\infty L^2}{2}\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2 \\ &+ \frac{1}{2}\sum_{t=1}^T \left(6\|\boldsymbol{\gamma}_t\|_\infty^3 N L^2 + \|\boldsymbol{\gamma}_t\|_\infty^2 (4N+1)L + \eta_t^i\right)\sigma_A^2\Bigg].\end{aligned}$$

Proof. Thanks to Lemma 10 and Assumption 3, we can bound

$$\begin{aligned}
&\left[\sum_{t=1}^{T}\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle\right] \\
&\leq \mathbb{E}\left[\frac{\|X^i_1 - p^i\|^2}{2\eta^i_{T+1}} + \sum_{t=2}^{T}\gamma^i_t L^2\left(3(1+\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2_{\boldsymbol{\gamma}^2_t} + 3\|\boldsymbol{\gamma}^2_t\|_1\sigma_A^2 + \frac{3}{2}\|\mathbf{X}_t - \mathbf{X}_{t-1}\|^2\right)\right. \\
&\qquad + \frac{1}{2}\sum_{t=2}^{T}\left((\gamma^i_t)^2 L(\sigma_M^2\|V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 + \sigma_A^2) + 4L(\sigma_M^2\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2_{\boldsymbol{\gamma}^2_t} + \|\boldsymbol{\gamma}^2_t\|_1\sigma_A^2)\right) \\
&\qquad \left. + \frac{1}{2}\sum_{t=1}^{T}\eta^i_t\left((1+\sigma_M^2)\|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \sigma_A^2\right)\right].
\end{aligned}$$

In the following, we further bound the above inequality using *i)* $\eta^i_t \leq \gamma^i_t/(4(1+\sigma_M^2))$, *ii)* $\boldsymbol{\gamma}_{t+1} \leq \boldsymbol{\gamma}_t$, *iii)* $\alpha^i_t\|V^i(\mathbf{x})\|^2 \leq \|\mathbf{V}(\mathbf{x})\|^2_{\boldsymbol{\alpha}}$ for any $\boldsymbol{\alpha} \in \mathbb{R}^N_+$ and $\mathbf{x} \in \mathcal{X}$, and *iv)* $\|\boldsymbol{\alpha}\|_\infty = \max_{i\in\mathcal{N}}\alpha^i$ and in particular $\|\boldsymbol{\alpha}^2\|_1 \leq N\|\boldsymbol{\alpha}\|^2_\infty$ for $\boldsymbol{\alpha} \in \mathbb{R}^N_+$.

$$\begin{aligned}
&\left[\sum_{t=1}^{T}\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle\right] \\
&\leq \mathbb{E}\left[\frac{\|X^i_1 - p^i\|^2}{2\eta^i_{T+1}} + \sum_{t=2}^{T}3\|\boldsymbol{\gamma}_t\|^2_\infty L^2\left((1+\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t} + \|\boldsymbol{\gamma}_t\|_\infty N\sigma_A^2\right)\right. \\
&\qquad + \frac{1}{2}\sum_{t=2}^{T}\left(\|\boldsymbol{\gamma}_t\|_\infty L\sigma_M^2(\gamma^i_t\|V^i(\mathbf{X}_{t-\frac{1}{2}})\|^2 + 4\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t}) + \|\boldsymbol{\gamma}_t\|^2_\infty(4N+1)L\sigma_A^2\right) \\
&\qquad \left. + \sum_{t=2}^{T}\frac{3\|\boldsymbol{\gamma}_t\|_\infty L^2}{2}\|\mathbf{X}_t - \mathbf{X}_{t-1}\|^2 + \frac{1}{2}\sum_{t=1}^{T}\left(\frac{\gamma^i_t}{4}\|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \eta^i_t\sigma_A^2\right)\right] \\
&\leq \mathbb{E}\left[\frac{\|X^i_1 - p^i\|^2}{2\eta^i_{T+1}} + \sum_{t=1}^{T}\left(3\|\boldsymbol{\gamma}_t\|^2_\infty L^2(1+\sigma_M^2) + \frac{5\|\boldsymbol{\gamma}_t\|_\infty L\sigma_M^2}{2} + \frac{1}{8}\right)\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t}\right. \\
&\qquad + \sum_{t=1}^{T-1}\frac{3\|\boldsymbol{\gamma}_1\|_\infty L^2}{2}\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2 \\
&\qquad \left. + \frac{1}{2}\sum_{t=1}^{T}\left(6\|\boldsymbol{\gamma}_t\|^3_\infty NL^2\sigma_A^2 + \|\boldsymbol{\gamma}_t\|^2_\infty(4N+1)L\sigma_A^2 + \eta^i_t\sigma_A^2\right)\right].
\end{aligned}$$

To conclude, we notice that under that our learning rate requirements it holds that $3\|\boldsymbol{\gamma}_t\|^2_\infty L^2(1+\sigma_M^2) + 5\|\boldsymbol{\gamma}_t\|_\infty L\sigma_M^2/2 \leq 1/2$. □

Our main regret guarantees of non-adaptive OptDA+ follows from the combination of Lemma 12 and Proposition 6.

Theorem 11. *Let Assumptions 1–3 hold and all players run (OptDA+) with non-increasing learning rate sequences $(\gamma^i_t)_{t\in\mathbb{N}}$ and $(\eta^i_t)_{t\in\mathbb{N}}$ satisfying (5). For any $i \in \mathcal{N}$ and bounded set $\mathcal{K}^i \subset \mathcal{X}^i$ with $R \geq \sup_{p^i}\|X^i_1 - p^i\|$, we have:*

(a) If $\gamma^j_t = \mathcal{O}(1/t^{\frac{1}{4}})$ and $\eta^j_t = \Theta(1/\sqrt{t})$ for all $j \in \mathcal{N}$, then

$$\max_{p^i\in\mathcal{K}^i}\mathbb{E}\left[\sum_{t=1}^{T}\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle\right] = \mathcal{O}\left(\sqrt{T}\right).$$

(b) If the noise is multiplicative (i.e., $\sigma_A = 0$) and the learning rates are constant $\gamma_t \equiv \gamma$, $\eta_t \equiv \eta$, then

$$\max_{p^i\in\mathcal{K}^i}\mathbb{E}\left[\sum_{t=1}^{T}\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle\right] \leq \frac{R^2}{2\eta^i} + \frac{5}{4}\left(\mathrm{dist}_{1/\boldsymbol{\eta}}(\mathbf{X}_1, \mathcal{X}_\star)^2 + \|\mathbf{V}(\mathbf{X}_1)\|^2_{\boldsymbol{\gamma}}\right).$$

In particular, if the equalities hold in (5), the above is in $\mathcal{O}(N^2L(1+\sigma_M^2)^2)$.

Proof. Let $p^i \in \mathcal{K}^i$ and $\mathbf{x}_\star = \arg\min_{\mathbf{x}\in\mathcal{X}_\star} \|\mathbf{X}_1 - \mathbf{x}\|_{1/\boldsymbol{\eta}}$. We define $a_t = 6\|\boldsymbol{\gamma}_t\|_\infty^3 NL^2 + \|\boldsymbol{\gamma}_t\|_\infty^2(4N+1)L + 2\|\boldsymbol{\eta}_t\|_\infty$ as in the proof of Theorem 10. Combining Proposition 6 and Lemma 12, we know that

$$\mathbb{E}\left[\sum_{t=1}^T \langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle\right]$$
$$\leq \mathbb{E}\left[\frac{R^2}{2\eta^i_{T+1}} + \frac{1}{2}\sum_{t=1}^T a_t\sigma_A^2 + \frac{5}{4}\left(\|\mathbf{X}_1 - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_{T+1}} + \|\mathbf{V}(\mathbf{X}_1)\|^2_{\boldsymbol{\gamma}_1} + \sum_{t=1}^T a_t N\sigma_A^2\right)\right].$$

The claims of the theorem follow immediately. □

To close this section, we bound the regret of non-adaptive OptDA+ when played against arbitrary opponents.

Proposition 7. *Let Assumption 3 hold and player i run (OptDA+) with non-increasing learning rates $\gamma_t^i = \Theta(1/t^{\frac{1}{2}-q})$ and $\eta_t^i = \Theta(1/\sqrt{t})$ for some $q \in [0, 1/4]$. Then, if there exists $G \in \mathbb{R}_+$ such that $\sup_{x^i\in\mathcal{X}^i}\|V^i(x^i)\| \leq G$, it holds for any bounded set $\mathcal{K}^i$ with $R \geq \sup_{p^i\in\mathcal{K}^i}\|X_1^i - p^i\|$ that*

$$\max_{p^i\in\mathcal{K}^i} \mathbb{E}\left[\sum_{t=1}^T \langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle\right] = \mathcal{O}\left(R^2\sqrt{T} + ((1+\sigma_M^2)G^2 + \sigma_A^2)T^{\frac{1}{2}+q}\right).$$

Proof. Let $p^i \in \mathcal{K}^i$. From Corollary 2 and Young's inequality we get

$$\begin{aligned}
\langle \hat{V}^i_{t+\frac{1}{2}}, X^i_{t+\frac{1}{2}} - p^i\rangle &\leq \frac{\|X_t^i - p^i\|^2}{2\eta_t^i} - \frac{\|X_{t+1}^i - p^i\|^2}{2\eta_{t+1}^i} - \frac{\|X_t^i - X_{t+1}^i\|^2}{2\eta_t^i} \\
&\quad + \left(\frac{1}{2\eta_{t+1}^i} - \frac{1}{2\eta_t^i}\right)\|X_1^i - p^i\|^2 - \gamma_t^i\langle \hat{V}^i_{t+\frac{1}{2}}, \hat{V}^i_{t-\frac{1}{2}}\rangle + \eta_t^i\|\hat{V}^i_{t+\frac{1}{2}}\|^2 \\
&\leq \frac{R^2}{2\eta_t^i} - \frac{\|X_{t+1}^i - p^i\|^2}{2\eta_{t+1}^i} - \frac{\|X_t^i - X_{t+1}^i\|^2}{2\eta_t^i} \\
&\quad + \left(\frac{1}{2\eta_{t+1}^i} - \frac{1}{2\eta_t^i}\right)\|X_1^i - p^i\|^2 + \frac{\gamma_t^i}{2}(\|\hat{V}^i_{t+\frac{1}{2}}\|^2 + \|\hat{V}^i_{t-\frac{1}{2}}\|^2) + \eta_t^i\|\hat{V}^i_{t+\frac{1}{2}}\|^2
\end{aligned}$$

As $\mathbf{V}^i_{1/2} = 0$ and $\eta_1^i = \eta_2^i$, summing the above from $t = 1$ to T gives

$$\sum_{t=1}^T \langle \hat{V}^i_{t+\frac{1}{2}}, X^i_{t+\frac{1}{2}} - p^i\rangle \leq \frac{\|X_1^i - p^i\|^2}{2\eta^i_{T+1}} - \sum_{t=1}^T \frac{\|X_t^i - X_{t+1}^i\|^2}{2\eta_t^i} + \sum_{t=1}^T(\gamma_t^i + \eta_t^i)\|\hat{V}^i_{t+\frac{1}{2}}\|^2. \tag{37}$$

Dropping the negative term and taking expectation leads to

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^T \langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i\rangle\right] &\leq \mathbb{E}\left[\frac{R^2}{2\eta^i_{T+1}} + \sum_{t=1}^T(\gamma_t^i + \eta_t^i)((1+\sigma_M^2)\|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \sigma_A^2)\right] \\
&\leq \frac{R^2}{2\eta^i_{T+1}} + \sum_{t=1}^T(\gamma_t^i + \eta_t^i)((1+\sigma_M^2)G^2 + \sigma_A^2)
\end{aligned}$$

The claim then follows immediately from the choice of the learning rates. □

G Regret Analysis with Adaptive Learning Rates

In this section, we tackle the regret analysis of OptDA+ run with adaptive learning rates. For ease of notation, we introduce the following quantities[5]

$$\lambda_t^i = \sum_{s=1}^t \|\hat{V}^i_{s+\frac{1}{2}}\|^2, \quad \mu_t^i = \sum_{s=1}^t \|X_s^i - X_{s+1}^i\|^2.$$

[5]For $t \leq 0$, $\lambda_t^i = \mu_t^i = 0$.

Clearly, our adaptive learning rates (Adapt) correspond to $\gamma_t^i = 1/(1+\lambda_{t-2}^i)^{\frac{1}{2}-q}$ and $\eta_t^i = 1/\sqrt{1+\lambda_{t-2}^i+\mu_{t-2}^i}$. As Assumption 4 assumes the noise to be bounded *almost surely*, whenever this assumption is used, the stated inequalities only hold almost surely. To avoid repetition, we will not mention this explicitly in the following. Finally, Assumption 5 is obviously satisfied by the learning rates given by (Adapt); therefore, Lemmas 10 and 11 can be effectively applied.

G.1 Preliminary Lemmas

We start by establishing several basic lemmas that will be used repeatedly in the analysis. We first state the apparent fact that λ_t^i grows at most linearly under Assumption 4.

Lemma 13. *Let Assumption 4 hold. Then, for all $i \in \mathcal{N}$ and $T \in \mathbb{N}$, we have*

$$\lambda_T^i \leq 2(G^2+\bar{\sigma}^2)T.$$

Proof. Using Assumption 4, we deduce that

$$\|\hat{V}_{t+\frac{1}{2}}^i\|^2 \leq 2\|V^i(\mathbf{X}_{t+\frac{1}{2}})\|^2 + 2\|\xi_{t+\frac{1}{2}}^i\|^2 \leq 2G^2 + 2\bar{\sigma}^2,$$

The claimed inequality is then immediate from the definition of λ_T^i. □

The next lemma is a slight generalization of the AdaGrad lemma [7, Lemma 3.5].

Lemma 14. *Let $T \in \mathbb{N}$, $\varepsilon > 0$, and $q \in [0,1)$. For any sequence of non-negative real numbers $a_1, \dots, a_T$, it holds*

$$\sum_{t=1}^T \frac{a_t}{\left(\varepsilon + \sum_{s=1}^t a_s\right)^q} \leq \frac{1}{1-q}\left(\sum_{t=1}^T a_t\right)^{1-q}. \tag{38}$$

Proof. The function $y \in \mathbb{R}_+ \mapsto y^{1-q}$ is concave and has derivative $y \mapsto (1-q)/y^q$. Therefore, it holds for every $y, z > 0$ that

$$z^{1-q} \leq y^{1-q} + \frac{1-q}{y^q}(z-y).$$

For $\varepsilon' \in (0,\varepsilon)$, we apply the above inequality to $y = \varepsilon' + \sum_{s=1}^t a_s$ and $z = \varepsilon' + \sum_{s=1}^{t-1} a_s$. This gives

$$\begin{aligned}\frac{1}{1-q}\left(\varepsilon' + \sum_{s=1}^{t-1} a_s\right)^{1-q} &\leq \frac{1}{1-q}\left(\varepsilon' + \sum_{s=1}^{t} a_s\right)^{1-q} - \frac{a_t}{\left(\varepsilon' + \sum_{s=1}^t a_s\right)^q} \\ &\leq \frac{1}{1-q}\left(\varepsilon' + \sum_{s=1}^{t} a_s\right)^{1-q} - \frac{a_t}{\left(\varepsilon + \sum_{s=1}^t a_s\right)^q}. \end{aligned} \tag{39}$$

Moreover, at $t = 1$ we have

$$\frac{a_1}{(\varepsilon + a_1)^q} \leq (\varepsilon' + a_1)^{1-q} \leq \frac{1}{1-q}(\varepsilon' + a_1)^{1-q}. \tag{40}$$

Summing (39) from $t = 2$ to T, adding (40), and rearranging leads to

$$\sum_{t=1}^T \frac{a_t}{\left(\varepsilon + \sum_{s=1}^t a_s\right)^q} \leq \frac{1}{1-q}\left(\epsilon' + \sum_{t=1}^T a_t\right)^{1-q}.$$

Provided that the above inequality holds for any $\varepsilon' \in (0,\varepsilon)$, we obtain (38) by taking $\varepsilon' \to 0$. □

The above two lemmas together provide us with the following bound on the sum of the weighted squared norms of feedback.

Lemma 15. *Let Assumption 4 hold, $s \in \mathbb{N}_0$, and $r \in [0,1)$. Then, for all $i \in \mathcal{N}$ and $T \in \mathbb{N}$, we have*

$$\sum_{t=1}^T \frac{\|\hat{V}_{t+\frac{1}{2}}^i\|^2}{(1+\lambda_{t-s}^i)^r} \leq \frac{(\lambda_T^i)^{1-r}}{1-r} + 2s(G^2+\bar{\sigma}^2).$$

Proof. Since $1/(1+\lambda_t^i)^r \le 1/(1+\lambda_{t-s}^i)^r$ and $\|\hat{V}_{t+\frac{1}{2}}^i\|^2 \le 2G^2 + 2\bar{\sigma}^2$, we have

$$\left(\frac{1}{(1+\lambda_{t-s}^i)^r} - \frac{1}{(1+\lambda_t^i)^r}\right)\|\hat{V}_{t+\frac{1}{2}}^i\|^2 \le \left(\frac{1}{(1+\lambda_{t-s}^i)^r} - \frac{1}{(1+\lambda_t^i)^r}\right)2(G^2+\bar{\sigma}^2).$$

Subsequently, it follows from Lemma 14 that

$$\begin{aligned}
\sum_{t=1}^T \frac{\|\hat{V}_{t+\frac{1}{2}}^i\|^2}{(1+\lambda_{t-s}^i)^r} &= \sum_{t=1}^T \left(\frac{\|\hat{V}_{t+\frac{1}{2}}^i\|^2}{(1+\lambda_t^i)^r} + \left(\frac{1}{(1+\lambda_{t-s}^i)^r} - \frac{1}{(1+\lambda_t^i)^r}\right)\|\hat{V}_{t+\frac{1}{2}}^i\|^2\right) \\
&\le \sum_{t=1}^T \frac{\|\hat{V}_{t+\frac{1}{2}}^i\|^2}{(1+\lambda_t^i)^r} + \sum_{t=1}^T \left(\frac{1}{(1+\lambda_{t-s}^i)^r} - \frac{1}{(\lambda_t^i)^r}\right)2(G^2+\bar{\sigma}^2) \\
&\le \frac{(\lambda_T^i)^{1-r}}{1-r} + \sum_{t=-s+1}^0 \frac{2(G^2+\bar{\sigma}^2)}{(1+\lambda_t^i)^r} \\
&= \frac{(\lambda_T^i)^{1-r}}{1-r} + 2s(G^2+\bar{\sigma}^2).
\end{aligned}$$

□

We also state a variant of the above result that takes into account the feedback of all players.

Lemma 16. *Let Assumption 4 hold, $s \in \mathbb{N}_0$, $r \in [0,1)$, and $(\boldsymbol{\alpha}_t)_{t\in\mathbb{N}}$ be a sequence of non-negative N-dimensional vectors such that $\alpha_t^i \le 1/(1+\lambda_{t-s}^i)^r$. Then, for all $T \in \mathbb{N}$, we have*

$$\sum_{t=1}^T \|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|_{\boldsymbol{\alpha}_t}^2 \le 2Ns(G^2+\bar{\sigma}^2) + \sum_{i=1}^N \frac{(\lambda_T^i)^{1-r}}{1-r}.$$

Proof. This is immediate from Lemma 15. □

Both Lemma 15 and Lemma 16 are essential for our analysis as they allow us to express the sums appearing in our analysis as a power of λ_t^i plus a constant. We end up with a technical lemma for bounding the inverse of η_t^i.

Lemma 17. *Let the learning rates be defined as in* (Adapt)*. For any $i \in \mathcal{N}$, $T \in \mathbb{N}$, and $a, b \in \mathbb{R}_+$, we have*

$$\frac{a}{\eta_{T+1}^i} - b\sum_{t=1}^T \frac{\|X_t^i - X_{t+1}^i\|^2}{\eta_t^i} \le a\sqrt{1+\lambda_{T-1}^i} + \frac{a^2}{4b}.$$

Proof. On one hand, we have

$$\frac{a}{\eta_{T+1}^i} = a\sqrt{1+\lambda_{T-1}^i+\mu_{T-1}^i} \le a\sqrt{1+\lambda_{T-1}^i} + a\sqrt{\mu_{T-1}^i}.$$

On the other hand, with $\eta_t^i \le 1$, it holds

$$b\sum_{t=1}^T \frac{\|X_t^i - X_{t+1}^i\|^2}{\eta_t^i} \ge b\sum_{t=1}^T \|X_t^i - X_{t+1}^i\|^2 \ge b\mu_{T-1}^i.$$

Let us define the function $f\colon y \in \mathbb{R} \mapsto -by^2 + ay$. Then

$$a\sqrt{\mu_{T-1}^i} - b\mu_{T-1}^i \le \max_{y\in\mathbb{R}} f(y) = \frac{a^2}{4b}.$$

Combining the above inequalities gives the desired result. □

G.2 Robustness Against Adversarial Opponents

In this part, we derive regret bounds for adaptive OptDA+ when played against adversarial opponents.

Proposition 8. *Let Assumption 4 hold and player i run (OptDA+) with learning rates (Adapt). Then, for any bounded set $\mathcal{K}^i$ with $R \geq \sup_{p^i \in \mathcal{K}^i} \|X_1^i - p^i\|$, it holds*

$$\max_{p^i \in \mathcal{K}^i} \mathbb{E}\left[\sum_{t=1}^{T} \langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X_{t+\frac{1}{2}}^i - p^i \rangle\right] = \mathcal{O}\left(((G^2+\bar{\sigma}^2)T)^{\frac{1}{2}+q} + R^2(G+\bar{\sigma})\sqrt{T} + R^4 + G^2 + \bar{\sigma}^2\right).$$

Proof. To begin, we notice that inequality (37) that we established in the proof of Proposition 7 still holds here for any $p^i \in \mathcal{K}^i$. Furthermore, applying Lemma 17 with $a \leftarrow R^2/2$, $b \leftarrow 1/2$ leads to

$$\frac{R^2}{2\eta_{T+1}^i} - \sum_{t=1}^{T} \frac{\|X_t^i - X_{t+1}^i\|^2}{2\eta_t^i} \leq \frac{R^2\sqrt{1+\lambda_{T-1}^i}}{2} + \frac{R^4}{8}.$$

On the other hand, invoking Lemma 15 with either $r \leftarrow 1/4 + q$ or $r \leftarrow 1/2$ guarantees that

$$\sum_{t=1}^{T} (\gamma_t^i + \eta_t^i)\|\hat{V}_{t+\frac{1}{2}}^i\|^2 \leq \frac{4(\lambda_T^i)^{3/4-q}}{3-4q} + 2\sqrt{\lambda_T^i} + 8(G^2+\bar{\sigma}^2).$$

Putting the above inequalities together, we obtain

$$\max_{p^i \in \mathcal{K}^i} \mathbb{E}\left[\sum_{t=1}^{T} \langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X_{t+\frac{1}{2}}^i - p^i \rangle\right] \leq \mathbb{E}\left[\frac{R^2\sqrt{1+\lambda_{T-1}^i}}{2} + \frac{4(\lambda_T^i)^{3/4-q}}{3-4q} + 2\sqrt{\lambda_T^i}\right] + \frac{R^4}{8} + 8(G^2+\bar{\sigma}^2).$$

We conclude with the help of Lemma 13. □

G.3 Smaller Regret Against Opponents with Same Learning Algorithm

We now address the more challenging part of the analysis: fast regret minimization when all players adopt adaptive OptDA+. For this, we need to control the different terms appearing in Lemmas 10 and 11. For the latter we build the following lemma to control the sum of some differences. As argued in Section 5, this is the reason that we include $\|X_s^i - X_{s+1}^i\|^2$ in the definition of η_t^i.

Lemma 18. *Let Assumptions 1 and 4 hold and the learning rates be defined as in (Adapt), then for all $T \in \mathbb{N}$, we have*

$$\sum_{t=1}^{T} \left(3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|_{\boldsymbol{\gamma}_t}^2 - \|\mathbf{X}_t - \mathbf{X}_{t+1}\|_{1/(4\boldsymbol{\eta}_t)}^2\right) \leq 432N^3L^6 + 24N^2G^2.$$

Proof. For all $i \in \mathcal{N}$, let us define

$$\bar{t}^i := \max\left\{s \in \{0, ..., T\} : \eta_t^i \geq \frac{1}{12NL^2}\right\},$$

where we set $\eta_0^i = 1/(12NL^2)$ to ensure that $\bar{t}^i$ is always well-defined. By the definition of η_t^i, the inequality $\eta_{\bar{t}^i}^i \geq 1/(12NL^2)$ implies $\mu_{\bar{t}^i-2}^i \leq 144N^2L^2$. We next define the sets

$$\mathcal{T} := \bigcup_{i \in \mathcal{N}} \{\bar{t}^i - 1, \bar{t}^i\} \cap \{1, ..., T\}$$

Clearly, $\operatorname{card}(\mathcal{T}) \leq 2N$. With $\boldsymbol{\gamma}_t \leq 1$, Assumptions 1 and 4, we obtain

$$\begin{aligned}
&\sum_{t=1}^{T} 3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|_{\boldsymbol{\gamma}_t}^2 \\
&\quad \leq \sum_{t=1}^{T} 3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|^2 \\
&\quad = \sum_{t\in[T]\setminus\mathcal{T}} 3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|^2 + \sum_{t\in\mathcal{T}} 3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|^2 \\
&\quad \leq \sum_{t\in[T]\setminus\mathcal{T}} 3NL^2\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2 + \sum_{t\in\mathcal{T}} 6\left(\|\mathbf{V}(\mathbf{X}_t)\|^2 + \|\mathbf{V}(\mathbf{X}_{t+1})\|^2\right) \\
&\quad \leq \sum_{i=1}^{N} \sum_{t\in[T]\setminus\mathcal{T}} 3NL^2\|X_t^i - X_{t+1}^i\|^2 + \sum_{t\in\mathcal{T}} 12NG^2 \\
&\quad \leq \sum_{i=1}^{N} 3NL^2 \Bigg(\underbrace{\sum_{t=1}^{\bar{t}^i-2} \|X_t^i - X_{t+1}^i\|^2}_{\mu_{\bar{t}^i-2}^i \leq 144N^2L^2} + \sum_{t=\bar{t}^i+1}^{T} \|X_t^i - X_{t+1}^i\|^2 \Bigg) + 24N^2G^2 \qquad (41)
\end{aligned}$$

On the other hand, by the choice of $\bar{t}^i$ we know that $1/\eta_t^i \geq 12NL^2$ for all $t \geq \bar{t}^i + 1$; hence

$$\begin{aligned}
\sum_{t=1}^{T} \|\mathbf{X}_t - \mathbf{X}_{t+1}\|_{1/(4\boldsymbol{\eta}_t)}^2 &= \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{\|X_t^i - X_{t+1}^i\|^2}{4\eta_t^i} \\
&\geq \sum_{i=1}^{N} \sum_{t=\bar{t}^i+1}^{T} \frac{\|X_t^i - X_{t+1}^i\|^2}{4\eta_t^i} \\
&\geq \sum_{i=1}^{N} \sum_{t=\bar{t}^i+1}^{T} 3NL^2\|X_t^i - X_{t+1}^i\|^2. \qquad (42)
\end{aligned}$$

Combining (41) and (42) gives the desired result. □

With Lemma 18 and the lemmas introduced in Appendix G.1, we are in a position to provide a bound on the expectation of the sum of the weighted squared operators norms plus the second-order path length. The next lemma is a fundamental building block for showing faster rates of adaptive OptDA+.

Lemma 19 (Bound on sum of squared norms)**.** *Let Assumptions 1, 2 and 4 hold and all players run OptDA+ with adaptive learning rates* (Adapt)*. Then, for all $T \in \mathbb{N}$ we have*

$$\sum_{t=1}^{T} \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2] + \frac{1}{8}\sum_{t=1}^{T} \mathbb{E}[\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2] \leq c_1 \sum_{i=1}^{N} \mathbb{E}\left[\sqrt{\lambda_T^i}\right] + c_2,$$

where

$$\begin{aligned}
\rho &= \min_{\mathbf{x}_\star \in \mathcal{X}_\star} \max_{i\in\mathcal{N}} \|X_1^i - x_\star^i\|, \\
c_1 &= 12NL^2 + 8NL + 2L + \rho^2 + 4, \\
c_2 &= 432N^3L^6 + 24N^2G^2 + (12NL^2 + 8NL + 2L + 8)(NG^2 + N\bar{\sigma}^2) + N\rho^2 + 2N\rho^4.
\end{aligned}$$

Proof. As in the proof of Proposition 6, and proceed to bound in expectation the sum of the following quantities

$$\begin{aligned}
A_t &= 3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|_{\boldsymbol{\gamma}_t}^2 - \|\mathbf{X}_t - \mathbf{X}_{t+1}\|_{1/(2\boldsymbol{\eta}_t)}^2, \\
B_t &= 6\|\boldsymbol{\gamma}_t\|_1 L^2 \|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2 + (4N+1)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2, \quad C_t = 2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|_{\boldsymbol{\eta}_t}^2. \qquad (43)
\end{aligned}$$

Thanks to Lemma 18, we know that the sum of A_t can be bounded directly without taking expectation by

$$\sum_{t=1}^{T} A_t = \sum_{t=1}^{T} \left(3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t+1})\|_{\boldsymbol{\gamma}_t}^2 - \|\mathbf{X}_t - \mathbf{X}_{t+1}\|_{1/(4\boldsymbol{\eta}_t)}^2 - \|\mathbf{X}_t - \mathbf{X}_{t+1}\|_{1/(4\boldsymbol{\eta}_t)}^2 \right)$$
$$\leq 432N^3L^6 + 24N^2G^2 - \sum_{t=1}^{T} \|\mathbf{X}_t - \mathbf{X}_{t+1}\|_{1/(4\boldsymbol{\eta}_t)}^2.$$

To obtain the above inequality we have also used $\boldsymbol{\eta}_t \leq 1$. To bound $\mathbb{E}[B_t]$, we use $\mathbb{E}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2] \leq \mathbb{E}[\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2]$ shown in (49), $\|\boldsymbol{\gamma}_t\|_1 \leq N$, and Lemma 16 (as $(\gamma_{t+1}^i)^2 \leq 1/\sqrt{1+\lambda_{t-1}^i}$) to obtain

$$\sum_{t=2}^{T} \mathbb{E}[B_t] \leq \mathbb{E}\left[\sum_{t=2}^{T} (6NL^2 + (4N+1)L) \|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}_t^2}^2 \right]$$
$$= \mathbb{E}\left[\sum_{t=1}^{T-1} (6NL^2 + (4N+1)L) \|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|_{(\boldsymbol{\gamma}_{t+1})^2}^2 \right]$$
$$\leq (6NL^2 + (4N+1)L) \left(2N(G^2 + \bar{\sigma}^2) + \sum_{i=1}^{N} 2\,\mathbb{E}\left[\sqrt{\lambda_{T-1}^i}\right] \right). \tag{44}$$

Similarly, the sum of C_t can be bounded in expectation by

$$\sum_{t=1}^{T} \mathbb{E}[C_t] \leq 8N(G^2 + \bar{\sigma}^2) + \sum_{i=1}^{N} 4\,\mathbb{E}\left[\sqrt{\lambda_T^i}\right]. \tag{45}$$

Let us choose $\mathbf{x}_\star = \arg\min_{\mathbf{x}\in\mathcal{X}_\star} \max_{i\in\mathcal{N}} \|X_1^i - x^i\|$ so that $\rho = \max_{i\in\mathcal{N}} \|X_1^i - x_\star^i\|$. Plugging (43), (44), and (45) into (28) of Lemma 11, we get readily

$$\sum_{t=2}^{T} \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2 + \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2] + \sum_{t=1}^{T} \mathbb{E}[\|\mathbf{X}_t - \mathbf{X}_{t+1}\|_{1/(8\boldsymbol{\eta}_t)}^2]$$
$$\leq \mathbb{E}[\|\mathbf{X}_1 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_{T+1}}^2] - \sum_{t=1}^{T} \|\mathbf{X}_t - \mathbf{X}_{t+1}\|_{1/(8\boldsymbol{\eta}_t)}^2 + (12NL^2 + 8NL + 2L + 4) \sum_{i=1}^{N} \mathbb{E}\left[\sqrt{\lambda_T^i}\right]$$
$$+ 432N^3L^6 + 24N^2G^2 + (12NL^2 + 8NL + 2L + 8)(NG^2 + N\bar{\sigma}^2) \tag{46}$$

Using Lemma 17, we can then further bound the RHS of (46) with

$$\|\mathbf{X}_1 - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_{T+1}}^2 - \sum_{t=1}^{T} \|\mathbf{X}_t - \mathbf{X}_{t+1}\|_{1/(8\boldsymbol{\eta}_t)}^2 = \sum_{i=1}^{N} \left(\frac{\|X_1^i - x_\star^i\|^2}{\eta_{T+1}^i} - \sum_{t=1}^{T} \frac{\|X_t^i - X_{t+1}^i\|^2}{8\eta_t^i} \right)$$
$$\leq \sum_{i=1}^{N} \left(\|X_1^i - x_\star^i\|^2 \sqrt{1+\lambda_{T-1}^i} + 2\|x_\star^i\|^4 \right)$$
$$\leq N\rho^2 + 2N\rho^4 + \sum_{i=1}^{N} \rho^2 \sqrt{\lambda_{T-1}^i}. \tag{47}$$

Finally, using $\boldsymbol{\eta}_t \leq 1$, $\mathbf{X}_{3/2} = \mathbf{X}_1$, and $\boldsymbol{\gamma}_2 = \boldsymbol{\gamma}_1$, the left-hand side (LHS) of (46) can be bounded from below by

$$\sum_{t=2}^{T} \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2 + \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2] + \sum_{t=1}^{T} \mathbb{E}[\|\mathbf{X}_t - \mathbf{X}_{t+1}\|_{1/(8\boldsymbol{\eta}_t)}^2]$$
$$\geq \sum_{t=1}^{T} \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2] + \frac{1}{8} \sum_{t=1}^{T} \mathbb{E}[\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2]. \tag{48}$$

Combining (46), (47), and (48) gives the desired result. □

We also refine Lemma 10 for the case of adaptive learning rates. The next lemma suggests the terms that need to be bounded in expectation in order to control the regret.

Lemma 20 (Bound on linearized regret)**.** *Let Assumptions 1, 2 and 4 hold and all players run OptDA+ with adaptive learning rates* (Adapt)*. Then, for all $i \in \mathcal{N}$, $T \in \mathbb{N}$, and bounded set $\mathcal{K}^i \subset \mathcal{X}^i$ with $R \geq \sup_{p^i \in \mathcal{K}^i} \|X_1^i - p^i\|$, it holds that*

$$\max_{p^i \in \mathcal{K}^i} \mathbb{E}\left[\sum_{t=1}^{T} \langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i \rangle\right] \leq \mathbb{E}\Bigg[\left(\frac{R^2}{2} + \frac{L+1}{2}\right)\sqrt{\lambda_T^i} + (6L^2+4L)\sum_{j=1}^{N}\sqrt{\lambda_{T-1}^j} + \frac{R^2\sqrt{\mu_{T-1}^i}}{2} + \frac{3L^2}{2}\sum_{t=1}^{T-1}\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2 + \frac{R^2}{2} + (6NL^2+4NL+L+2)(G^2+\bar{\sigma}^2)\Bigg].$$

Proof. We will derive inequality (49) from Lemma 10. To begin, by Assumption 3(a) the noises are conditionally unbiased and we can thus write

$$\mathbb{E}_{t-1}[\|\hat{V}^i_{t-\frac{1}{2}}\|^2] = \|V^i(X^i_{t-\frac{1}{2}})\|^2 + \mathbb{E}_{t-1}[\|\xi^i_{t-\frac{1}{2}}\|^2] \geq \mathbb{E}_{t-1}[\|\xi^i_{t-\frac{1}{2}}\|^2].$$

Subsequently, $\boldsymbol{\gamma}_t$ being $\mathcal{F}_{t-1}$-measurable, applying the law of total expectation gives

$$\begin{aligned}\mathbb{E}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2}] &= \mathbb{E}\left[\sum_{i=1}^{N}(\gamma_t^i)^2\,\mathbb{E}_{t-1}[\|\xi^i_{t-\frac{1}{2}}\|^2]\right] \\ &\leq \mathbb{E}\left[\sum_{i=1}^{N}(\gamma_t^i)^2\,\mathbb{E}_{t-1}[\|\hat{V}^i_{t-\frac{1}{2}}\|^2]\right] = \mathbb{E}[\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2}].\end{aligned} \tag{49}$$

Plugging the above two inequalities into the inequality of Lemma 10 and using $\gamma_t^i \leq 1$ results in

$$\begin{aligned}&\max_{p^i \in \mathcal{K}^i} \mathbb{E}\left[\sum_{t=1}^{T} \langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}} - p^i \rangle\right] \\ &\leq \mathbb{E}\Bigg[\frac{R^2}{2\eta^i_{T+1}} + \sum_{t=2}^{T}\gamma_t^i L^2\left(3\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2} + \frac{3}{2}\|\mathbf{X}_t - \mathbf{X}_{t-1}\|^2\right) \\ &\quad + \frac{1}{2}\sum_{t=2}^{T}((\gamma_t^i)^2 L\|\hat{V}^i_{t-\frac{1}{2}}\|^2 + 4L\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2}) + \frac{1}{2}\sum_{t=1}^{T}\eta_t^i\|\hat{V}^i_{t+\frac{1}{2}}\|^2\Bigg] \\ &\leq \mathbb{E}\Bigg[\frac{R^2\sqrt{1+\lambda^i_{T-1}+\mu^i_{T-1}}}{2} + \sum_{t=1}^{T-1}(3L^2+2L)\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2_{(\boldsymbol{\gamma}_{t+1})^2} + \frac{3L^2}{2}\sum_{t=1}^{T-1}\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2 \\ &\quad + \frac{1}{2}\sum_{t=1}^{T-1}(\gamma^i_{t+1})^2 L\|\hat{V}^i_{t+\frac{1}{2}}\|^2 + \frac{1}{2}\sum_{t=1}^{T}\eta_t^i\|\hat{V}^i_{t+\frac{1}{2}}\|^2\Bigg].\end{aligned}$$

Since we have both $(\gamma^i_{t+1})^2 \leq 1/\sqrt{1+\lambda^i_{t-1}}$ and $\eta_t^i \leq 1/\sqrt{1+\lambda^i_{t-2}}$, applying Lemma 15 leads to

$$\sum_{t=1}^{T-1}(\gamma^i_{t+1})^2 L\|\hat{V}^i_{t+\frac{1}{2}}\|^2 + \sum_{t=1}^{T}\eta_t^i\|\hat{V}^i_{t+\frac{1}{2}}\|^2 \leq L\left(\sqrt{\lambda^i_{T-1}} + 2(G^2+\bar{\sigma}^2)\right) + \sqrt{\lambda^i_T} + 4(G^2+\bar{\sigma}^2).$$

Similarly, using Lemma 16 we deduce

$$\sum_{t=1}^{T-1}(3L^2+2L)\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2_{(\boldsymbol{\gamma}_{t+1})^2} \leq (3L^2+2L)\left(2N(G^2+\bar{\sigma}^2) + \sum_{j=1}^{N}2\sqrt{\lambda^j_{T-1}}\right)$$

Putting the above inequalities together and using $\sqrt{1+\lambda^i_{T-1}+\mu^i_{T-1}} \leq 1 + \sqrt{\lambda^i_{T-1}} + \sqrt{\mu^i_{T-1}}$ gives the desired result. □

G.3.1 The Case of Additive Noise

From Lemma 19 and Lemma 20 we can readily derive our main results for the case of additive noise.

Theorem 12. *Let Assumptions 1, 2 and 4 hold and all players run OptDA+ with adaptive learning rates* (Adapt)*. Then,*

$$\sum_{t=1}^{T} \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})]\|^2 = \mathcal{O}\left(T^{1-q}\right).$$

Proof. With Lemma 13, for $t \in \{1, ..., T\}$, we can lower bound the learning rate γ_t^i by

$$\gamma_t^i = \frac{1}{(1+\lambda_{t-2}^i)^{\frac{1}{2}-q}} \geq \frac{1}{(1+2\max(t-2,0)(G^2+\bar{\sigma}^2))^{\frac{1}{2}-q}} \geq \frac{1}{(1+2T(G^2+\bar{\sigma}^2))^{\frac{1}{2}-q}}.$$

Lemma 19 thus guarantees

$$\frac{\sum_{t=1}^{T} \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2]}{(1+2T(G^2+\bar{\sigma}^2))^{\frac{1}{2}-q}} \leq c_1 \sum_{i=1}^{N} \mathbb{E}\left[\sqrt{\lambda_t^i}\right] + c_2.$$

We conclude by using again Lemma 13. □

Theorem 13. *Let Assumptions 1, 2 and 4 hold and all players run OptDA+ with adaptive learning rates* (Adapt)*. Then, for any $i \in \mathcal{N}$ and bounded set $\mathcal{K}^i \subset \mathcal{X}^i$, we have*

$$\max_{p^i \in \mathcal{K}^i} \mathbb{E}\left[\sum_{t=1}^{T} \langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X_{t+\frac{1}{2}}^i - p^i \rangle\right] = \mathcal{O}\left(\sqrt{T}\right).$$

Proof. This follows from Lemma 20. To begin, with Lemma 13, we have clearly

$$\mathbb{E}\left[\left(\frac{R^2}{2} + \frac{L+1}{2}\right)\sqrt{\lambda_t^i} + (6L^2+4L)\sum_{j=1}^{N}\sqrt{\lambda_{T-1}^j}\right] = \mathcal{O}\left(\sqrt{T}\right).$$

Next, thanks to Lemma 19 we can bound

$$\begin{aligned}\mathbb{E}\left[\frac{R^2\sqrt{\mu_{T-1}^i}}{2} + \frac{3L^2}{2}\sum_{t=1}^{T-1}\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2\right] &\leq \mathbb{E}\left[\left(\frac{R^2}{2} + \frac{3L^2}{2}\right)\sum_{t=1}^{T-1}\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2\right] \\ &\leq (4R^2+12L^2)\left(c_1\sum_{i=1}^{N}\mathbb{E}\left[\sqrt{\lambda_t^i}\right] + c_2\right).\end{aligned}$$

This is again in $\mathcal{O}(\sqrt{T})$. Plugging the above into Lemma 20 concludes the proof. □

G.3.2 The Case of Multiplicative Noise

The case of multiplicative noise is more delicate. As explained in Section 5, the main step is to establish an inequality in the form of (6). This is achieved in Lemma 22 by using Lemma 19. Before that, we derive a lemma to show how inequality (6) implies boundedness of the relevant quantities.

Lemma 21. *Let $p, r, c \in \mathbb{R}_+$ such that $p > r$, $c \in \mathbb{R}_+$, and $(a^1, \ldots, a^N)$ be a collection of N non-negative real-valued random variables. If*

$$\sum_{i=1}^{N} \mathbb{E}[(a^i)^p] \leq c \sum_{i=1}^{N} \mathbb{E}[(a^i)^r], \tag{50}$$

Then $\sum_{i=1}^{N} \mathbb{E}[(a^i)^p] \leq N c^{\frac{p}{p-r}}$ and $\sum_{i=1}^{N} \mathbb{E}[(a^i)^r] \leq N c^{\frac{r}{p-r}}$.

Proof. Since $p > r$, the function $y \in \mathbb{R}_+ \cup \{0\} \mapsto y^{\frac{r}{p}}$ is concave. Applying Jensen's inequality for the expectation gives $\mathbb{E}[(a^i)^r] \leq \mathbb{E}[(a^i)^p]^{\frac{r}{p}}$. Next, we apply Jensen's inequality for the average to obtain

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[(a^i)^p]^{\frac{r}{p}} \leq \left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[(a^i)^p]\right)^{\frac{r}{p}}. \tag{51}$$

Along with inequality (50) we then get

$$\sum_{i=1}^{N}\mathbb{E}[(a^i)^p] \leq c\sum_{i=1}^{N}\mathbb{E}[(a^i)^r] \leq cN^{1-\frac{r}{p}}\left(\sum_{i=1}^{N}\mathbb{E}[(a^i)^p]\right)^{\frac{r}{p}}. \tag{52}$$

In other words

$$\left(\sum_{i=1}^{N}\mathbb{E}[(a^i)^p]\right)^{1-\frac{r}{p}} \leq cN^{1-\frac{r}{p}}.$$

Taking both sides of the inequality to the power of $p/(p-r)$, we obtain effectively

$$\sum_{i=1}^{N}\mathbb{E}[(a^i)^p] \leq Nc^{\frac{p}{p-r}}$$

The second inequality combines the above with second part of (52). □

In the next lemma we build inequality (6), and combined with Lemma 21 we obtain the boundedness of various quantities. This is also where the factor $1/q$ shows up.

Lemma 22. *Let Assumptions 1, 2 and 4 hold and all players run OptDA+ with adaptive learning rates* (Adapt). *Assume additionally Assumption 3 with $\sigma_A = 0$. Then, for any $T \in \mathbb{N}$, we have*

$$\sum_{i=1}^{N}\mathbb{E}\left[(1+\lambda_T^i)^{\frac{1}{2}+q}\right] \leq N\left((1+\sigma_M^2)c_1 + \frac{(1+\sigma_M^2)c_2+1}{N}\right)^{1+\frac{1}{2q}}, \tag{53}$$

$$\sum_{i=1}^{N}\mathbb{E}\left[\sqrt{1+\lambda_T^i}\right] \leq N\left((1+\sigma_M^2)c_1 + \frac{(1+\sigma_M^2)c_2+1}{N}\right)^{\frac{1}{2q}}, \tag{54}$$

$$\sum_{i=1}^{N}\mathbb{E}[\mu_T^i] \leq 8Nc_1\left((1+\sigma_M^2)c_1 + \frac{(1+\sigma_M^2)c_2+1}{N}\right)^{\frac{1}{2q}} + 8c_2. \tag{55}$$

Proof. From Lemma 19 we know that

$$\sum_{t=1}^{T}\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2] \leq c_1\sum_{i=1}^{N}\mathbb{E}\left[\sqrt{\lambda_T^i}\right] + c_2,$$

Since $\boldsymbol{\gamma}_t$ is $\mathcal{F}_t$-measurable (it is even $\mathcal{F}_{t-1}$-measurable), using the relative noise assumption and the law of total expectation we get

$$\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\gamma}_t}^2] = \sum_{i=1}^{N}\mathbb{E}[\gamma_t^i\,\mathbb{E}_t[\|V^i(\mathbf{X}_{t+\frac{1}{2}})]\|^2] \geq \sum_{i=1}^{N}\mathbb{E}\left[\gamma_t^i\,\mathbb{E}_t\left[\frac{\|\hat{V}_{t+\frac{1}{2}}^i\|^2}{1+\sigma_M^2}\right]\right] = \frac{\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|_{\boldsymbol{\gamma}_t}^2}{1+\sigma_M^2}.$$

The learning rates $\boldsymbol{\gamma}_t$ being non-increasing, we can then bound from below the sum of $\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t}]$ by

$$\begin{aligned}
\sum_{t=1}^{T}\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t}] &\geq \frac{1}{1+\sigma_M^2}\sum_{t=1}^{T}\mathbb{E}[\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t}] \\
&\geq \frac{1}{1+\sigma_M^2}\sum_{t=1}^{T}\mathbb{E}[\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_{T+2}}] \\
&= \frac{1}{1+\sigma_M^2}\sum_{i=1}^{N}\mathbb{E}\left[\frac{\sum_{t=1}^{T}\|\hat{V}^i_{t+\frac{1}{2}}\|^2}{(1+\lambda^i_T)^{\frac{1}{2}-q}}\right] \\
&= \frac{1}{1+\sigma_M^2}\sum_{i=1}^{N}\mathbb{E}\left[\frac{\lambda^i_T+1-1}{(1+\lambda^i_T)^{\frac{1}{2}-q}}\right] \\
&\geq -\frac{1}{1+\sigma_M^2}+\frac{1}{1+\sigma_M^2}\sum_{i=1}^{N}\mathbb{E}\left[(1+\lambda^i_T)^{\frac{1}{2}+q}\right].
\end{aligned}$$

As a consequence, we have shown that

$$\sum_{i=1}^{N}\mathbb{E}\left[(1+\lambda^i_T)^{\frac{1}{2}+q}\right] \leq (1+\sigma_M^2)c_1\sum_{i=1}^{N}\mathbb{E}\left[\sqrt{\lambda^i_T}\right]+(1+\sigma_M^2)c_2+1,$$

Subsequently,

$$\sum_{i=1}^{N}\mathbb{E}\left[(1+\lambda^i_T)^{\frac{1}{2}+q}\right] \leq \left((1+\sigma_M^2)c_1+\frac{(1+\sigma_M^2)c_2+1}{N}\right)\sum_{i=1}^{N}\mathbb{E}\left[\sqrt{1+\lambda^i_T}\right].$$

We deduce (53) and (54) with the help of Lemma 21 taking $p \leftarrow 3/4-q$, $r \leftarrow 1/2$, $c \leftarrow (1+\sigma_M^2)c_1+((1+\sigma_M^2)c_2+1)/N$, and $a^i \leftarrow 1+\lambda^i_T$. Plugging (54) into Lemma 19 gives (55). □

Now, as an immediate consequence of all our previous results, we obtain the constant regret bound of adaptive OptDA+ under multiplicative noise.

Theorem 14. *Let Assumptions 1, 2 and 4 hold and all players run OptDA+ with adaptive learning rates* (Adapt)*. Assume additionally Assumption 3 with $\sigma_A = 0$. Then, for any $i \in \mathcal{N}$ and bounded set $\mathcal{K}^i$, we have*

$$\mathbb{E}\left[\max_{p^i\in\mathcal{K}^i}\sum_{t=1}^{T}\langle V^i(\mathbf{X}_{t+\frac{1}{2}}), X^i_{t+\frac{1}{2}}-p^i\rangle\right] = \mathcal{O}\left(\exp\left(\frac{1}{2q}\right)\right).$$

Proof. Combining Lemma 20 and Lemma 22 gives the desired result. □

H Last-iterate Convergence

We close our appendix with proofs on almost-sure last-iterate convergence of the trajectories. The global proof schema was sketched in Section 6 for the particular case of Theorem 4 (which corresponds to the upcoming Theorem 17). To prove last-iterate convergence we make heavy use of the different results that we derived in previous sections.

H.1 Lemmas on Stochastic Sequences

To begin, we state several basic lemmas concerning stochastic sequences. The first one translates a bound of expectation into almost sure boundedness and convergence. It is a special case of Doob's martingale convergence theorem [26], but we also provide another elementary proof below. For simplicity, throughout the sequel, we use the term finite random variable to refer to those random variables which are finite almost surely.

Lemma 23. *Let $(U_t)_{t\in\mathbb{N}}$ be a sequence of non-decreasing and non-negative real-valued random variables. If there exists constant $C \in \mathbb{R}$ such that*

$$\forall\, t \in \mathbb{N}, \quad \mathbb{E}[U_t] \leq C.$$

Then $(U_t)_{t\in\mathbb{N}}$ converges almost surely to a finite random variable. In particular, for any sequence of non-negative real-valued random variables $(\chi_t)_{t\in\mathbb{N}}$, the fact that $\sum_{t=1}^{+\infty} \mathbb{E}[\chi_t] < +\infty$ implies $\sum_{t=1}^{+\infty} \chi_t < +\infty$ almost surely, and accordingly $\lim_{t\to+\infty} \chi_t = 0$ almost surely.

Proof. Let U_∞ be the pointwise limit of $(U_t)_{t\in\mathbb{N}}$. Applying Beppo Levi's lemma we deduce that U_∞ is also measurable and $\lim_{t\to+\infty} \mathbb{E}[U_t] = \mathbb{E}[U_\infty]$. Accordingly, $\mathbb{E}[U_\infty] \leq C$. The random variable U_∞ being non-negative, $\mathbb{E}[U_\infty] \leq C < +\infty$ implies that U_∞ is finite almost surely, which concludes the first statement of the lemma. The second statement is derived from the first statement by setting $U_t = \sum_{s=1}^{t} \chi_s$. □

The next lemma is essential for building almost sure last-iterate converge in the case of vanishing learning rates, as it allows to extract a convergent subsequence.

Lemma 24. *Let $(U_t)_{t\in\mathbb{N}}$ be a sequence of non-negative real-valued random variables such that*

$$\liminf_{t\to+\infty} \mathbb{E}[U_t] = 0.$$

Then, i) there exists a subsequence $(U_{\omega(t)})_{t\in\mathbb{N}}$ of $(U_t)_{t\in\mathbb{N}}$ that converges to 0 almost surely;[6] and accordingly ii) it holds almost surely that $\liminf_{t\to+\infty} U_t = 0$.

Proof. Since $\liminf_{t\to+\infty} \mathbb{E}[U_t] = 0$, we can extract a subsequence $(U_{\omega(t)})_{t\in\mathbb{N}}$ such that for all $t \in \mathbb{N}$, $\mathbb{E}[U_{\omega(t)}] \leq 2^{-t}$. This gives $\sum_{t=1}^{+\infty} \mathbb{E}[U_{\omega(t)}] < +\infty$ and invoking Lemma 23 we then know that $\sum_{t=1}^{+\infty} U_{\omega(t)} < +\infty$ almost surely, which in turn implies that $U_{\omega(t)}$ converges to 0 almost surely. To prove (*ii*), we just notice that for any realization such that $\lim_{t\to+\infty} U_{\omega(t)} = 0$, we have $0 = \lim_{t\to+\infty} U_{\omega(t)} \geq \liminf_{t\to+\infty} U_t \geq 0$ and thus the equalities must hold, i.e., $\liminf_{t\to+\infty} U_t = 0$. □

Another important building block is Robbins-Sigmunds's theorem that allows us to show almost sure convergence of the Lyapunov function to a finite random variable.

Lemma 25 (Robbins and Sigmund [55])**.** *Consider a filtration $(\mathcal{G}_t)_{t\in\mathbb{N}}$ and four non-negative real-valued $(\mathcal{G}_t)_{t\in\mathbb{N}}$-adapted processes $(U_t)_{t\in\mathbb{N}}$, $(\alpha_t)_{t\in\mathbb{N}}$, $(\chi_t)_{t\in\mathbb{N}}$, $(\zeta_t)_{t\in\mathbb{N}}$ such that $\mathbb{E}[U_1] < +\infty$, $\sum_{t=1}^{+\infty} \mathbb{E}[\alpha_t] < \infty$, $\sum_{t=1}^{+\infty} \mathbb{E}[\chi_t] < \infty$, and for all $t \in \mathbb{N}$,*

$$\mathbb{E}[U_{t+1} \mid \mathcal{G}_t] \leq (1+\alpha_t)U_t + \chi_t - \zeta_t.$$

Then $(U_t)_{t\in\mathbb{N}}$ converges almost surely to a finite random variable and $\sum_{t=1}^{+\infty} \zeta_t < \infty$ almost surely.

Finally, since the solution may not be unique, we need a to translate the result with respect to a single point to the one that applies to the entire set. This is achieved through the following lemma.

Lemma 26. *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a closed set, $(\mathbf{u}_t)_{t\in\mathbb{N}}$ be a sequence of $\mathbb{R}^d$-valued random variable, and $(\boldsymbol{\alpha}_t)_{t\in\mathbb{N}}$ be a sequence of $\mathbb{R}^N$-valued random variable such that*

- *(a) For all $i \in \mathcal{N}$, $\alpha_1^i \geq 1$, $(\alpha_t^i)_{t\in\mathbb{N}}$ is non-decreasing and converges to a finite constant almost surely.*
- *(b) For all $\mathbf{x} \in \mathcal{K}$, $\|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}_t}$ converges almost surely.*

Then, with probability 1, the vector $\boldsymbol{\alpha}_\infty = \lim_{t\to+\infty} \boldsymbol{\alpha}_t$ is well-defined, finite, and $\|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty}$ converges for all $\mathbf{x} \in \mathcal{K}$.

Proof. As $\mathbb{R}^d$ is a separable metric space, $\mathcal{K}$ is also separable and we can find a countable set $\mathcal{Z}$ such that $\mathcal{K} = \mathrm{cl}(\mathcal{Z})$. Let us define the event

$$\mathcal{E} \coloneqq \{\boldsymbol{\alpha}_\infty = \lim_{t\to+\infty} \boldsymbol{\alpha}_t \text{ is well-defined and finite;} \quad \|\mathbf{u}_t - \mathbf{z}\|_{\boldsymbol{\alpha}_t} \text{ converges for all } \mathbf{z} \in \mathcal{Z}.\} \tag{56}$$

[6] We remark that the choice of the subsequence does not depend on the realization but only the distribution of the random variables.

The set $\mathcal{Z}$ being countable, from (a) and (b) we then know that $\mathbb{P}(\mathcal{E}) = 1$. In the following, we show that $\|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty}$ converges for all $\mathbf{x} \in \mathcal{K}$ whenever $\mathcal{E}$ happens, which concludes our proof.

Let us now consider a realization of $\mathcal{E}$. We first establish the convergence of $\|\mathbf{u}_t - \mathbf{z}\|_{\boldsymbol{\alpha}_\infty}$ for any $\mathbf{z} \in \mathcal{Z}$. To begin, the convergence of $\|\mathbf{u}_t - \mathbf{z}\|_{\boldsymbol{\alpha}_t}$ implies the boundedness of this sequence, from which we deduce immediately the boundedness of $\|\mathbf{u}_t - \mathbf{z}\|$ as $\|\mathbf{u}_t - \mathbf{z}\| \leq \|\mathbf{u}_t - \mathbf{z}\|_{\boldsymbol{\alpha}_t}$ by $\boldsymbol{\alpha}_t \geq \boldsymbol{\alpha}_1 \geq 1$. In other words, $C = \sup_{t\in\mathbb{N}} \|\mathbf{u}_t - \mathbf{z}\|$ is finite. Furthermore, we have

$$0 \leq \|\mathbf{u}_t - \mathbf{z}\|^2_{\boldsymbol{\alpha}_\infty} - \|\mathbf{u}_t - \mathbf{z}\|^2_{\boldsymbol{\alpha}_t} = \sum_{i=1}^{N} (\alpha^i_\infty - \alpha^i_t)\|u^i_t - z^i\|^2 \leq \sum_{i=1}^{N} (\alpha^i_\infty - \alpha^i_t) C^2. \tag{57}$$

Since $\alpha^i_\infty - \alpha^i_t$ converges to 0 when t goes to infinity, from (57) we get immediately $\lim_{t\to+\infty}(\|\mathbf{u}_t - \mathbf{z}\|^2_{\boldsymbol{\alpha}_\infty} - \|\mathbf{u}_t - \mathbf{z}\|^2_{\boldsymbol{\alpha}_t}) = 0$. This shows that $\|\mathbf{u}_t - \mathbf{z}\|^2_{\boldsymbol{\alpha}_\infty}$ converges to $\lim_{t\to+\infty} \|\mathbf{u}_t - \mathbf{z}\|^2_{\boldsymbol{\alpha}_t}$, which exists by definition of $\mathcal{E}$. We have thus shown the convergence of $\|\mathbf{u}_t - \mathbf{z}\|_{\boldsymbol{\alpha}_\infty}$.

To conclude, we need to show that $\|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty}$ in fact converges for all $\mathbf{x} \in \mathcal{K}$. Let $\mathbf{x} \in \mathcal{K}$. As $\mathcal{Z}$ is dense in $\mathcal{K}$, there exists a sequence of points $(\mathbf{z}_k)_{k\in\mathbb{N}}$ with $\mathbf{z}_k \in \mathcal{Z}$ for all $k \in \mathbb{N}$ such that $\lim_{k\to+\infty} \mathbf{z}_k = \mathbf{x}$. For any $t, k \in \mathbb{N}$, the triangular inequality implies

$$-\|\mathbf{z}_k - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty} \leq \|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty} - \|\mathbf{u}_t - \mathbf{z}_k\|_{\boldsymbol{\alpha}_\infty} \leq \|\mathbf{z}_k - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty}.$$

Since $\mathbf{z}_k \in \mathcal{Z}$, we have shown that $\lim_{t\to+\infty} \|\mathbf{u}_t - \mathbf{z}_k\|$ exists. Subsequently, we get

$$\begin{aligned}
-\|\mathbf{z}_k - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty} &\leq \liminf_{t\to+\infty} \|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty} - \lim_{t\to+\infty} \|\mathbf{u}_t - \mathbf{z}_k\|_{\boldsymbol{\alpha}_\infty} \\
&\leq \limsup_{t\to+\infty} \|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty} - \lim_{t\to+\infty} \|\mathbf{u}_t - \mathbf{z}_k\|_{\boldsymbol{\alpha}_\infty} \\
&\leq \|\mathbf{z}_k - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty}.
\end{aligned}$$

Taking the limit as $k \to +\infty$, we deduce that $\lim_{k\to+\infty} \lim_{t\to+\infty} \|\mathbf{u}_t - \mathbf{z}_k\|_{\boldsymbol{\alpha}_\infty}$ exists and

$$\liminf_{t\to+\infty} \|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty} = \lim_{k\to+\infty} \lim_{t\to+\infty} \|\mathbf{u}_t - \mathbf{z}_k\|_{\boldsymbol{\alpha}_\infty} = \limsup_{t\to+\infty} \|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty}.$$

This shows the convergence of $\|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}_\infty}$. □

Corollary 3. *Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a closed set, $(\mathbf{u}_t)_{t\in\mathbb{N}}$ be a sequence of $\mathbb{R}^d$-valued random variable, and $\boldsymbol{\alpha} \in \mathbb{R}^N$ such that $\alpha^i \geq 1$ for all $i \in \mathcal{N}$, and for all $\mathbf{x} \in \mathcal{K}$, $\|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}}$ converges almost surely. Then, with probability 1, $\|\mathbf{u}_t - \mathbf{x}\|_{\boldsymbol{\alpha}}$ converges for all $\mathbf{x} \in \mathcal{K}$.*

H.2 Trajectory Convergence of OG+ under Additive Noise

We start by proving the almost sure last-iterate convergence of OG+ under additive noise. This proof is, in a sense, the most technical once the results of the previous sections are established. This is because with vanishing learning rates, we cannot show that every cluster point of $(\mathbf{X}_{t+\frac{1}{2}})_{t\in\mathbb{N}}$ is a Nash equilibrium with probability 1. Instead we need to work with subsequences.

We will prove the convergence of $\mathbf{X}_t$ and $\mathbf{X}_{t+\frac{1}{2}}$ separately, and under relaxed learning rate requirements. For the convergence of $\mathbf{X}_t$, a learning rate condition introduced in [30] for double step-size EG is considered.

Theorem 15. *Let Assumptions 1–3 hold and all players run (OG+) with non-increasing learning rate sequences $(\gamma_t)_{t\in\mathbb{N}}$ and $(\eta_t)_{t\in\mathbb{N}}$ satisfying (4) and*

$$\sum_{t=1}^{+\infty} \gamma_t \eta_{t+1} = \infty, \quad \sum_{t=1}^{+\infty} \gamma_t^2 \eta_{t+1} < \infty, \quad \sum_{t=1}^{+\infty} \eta_t^2 < \infty. \tag{58}$$

Then, $\mathbf{X}_t$ converges almost surely to a Nash equilibrium.

Proof. Our proof is divided into four steps. To begin, let us define $\tilde{\mathbf{X}}_1 = \mathbf{X}_1$ and for all $t \geq 2$,

$$\tilde{\mathbf{X}}_t = \mathbf{X}_t + \eta_t \boldsymbol{\xi}_{t-\frac{1}{2}} = \mathbf{X}_{t-1} - \eta_t \mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})$$

Notice that $\tilde{\mathbf{X}}_t$ is $\mathcal{F}_{t-1}$-measurable. This surrogate of $\mathbf{X}_t$ plays an important role in the subsequent analysis.

(1) *With probability* 1, $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|$ *converges for all* $\mathbf{x}_\star \in \mathcal{X}_\star$. Let $x_\star \in \mathcal{X}_\star$. We would like to apply Robbins-Siegmund's theorem (Lemma 25) to the inequality of Lemma 5 with

$$\begin{aligned}
&\mathcal{G}_t \leftarrow \mathcal{F}_{t-1}, \quad U_t \leftarrow \mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2], \quad \alpha_t \leftarrow 0,\\
&\zeta_t \leftarrow \gamma_t\eta_{t+1}(\mathbb{E}_{t-1}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] + \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2),\\
&\chi_t \leftarrow \mathbb{E}_{t-1}[3\gamma_t\eta_{t+1}NL^2((\eta_t^2+\gamma_t^2)\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2 + (\gamma_{t-1})^2\|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2)\\
&\qquad\qquad + (\gamma_t^2\eta_{t+1} + N\eta_{t+1}(\eta_t+\gamma_t)^2)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2 + (\eta_{t+1})^2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2].
\end{aligned}$$

As Lemma 5 only applies to $t \geq 2$, for $t = 1$ we use inequality (15). We thus choose $\zeta_1 = 0$ and $\chi_1 = \eta_2^2\|\hat{\mathbf{V}}_{3/2}\|^2$.

We claim that $\sum_{t=1}^{+\infty}\mathbb{E}[\chi_t] < +\infty$. In fact, following the proof of Proposition 5, we can deduce

$$\sum_{t=1}^{+\infty}\mathbb{E}[\chi_t] \leq \underbrace{\sum_{t=1}^{+\infty}\gamma_t\eta_{t+1}(a_t(1+\sigma_M^2)+b_t\sigma_M^2)\,\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2]}_{(A)} + \underbrace{\sum_{t=1}^{+\infty}\gamma_t\eta_{t+1}(a_t+b_t)N\sigma_A^2}_{(B)},$$

for $a_t = \eta_{t+1}/\gamma_t + 9\gamma_t^2NL^2$ and $b_t = \gamma_t(4N+1)L$. With our learning rate requirements it is true that $a_t(1+\sigma_M^2) + b_t\sigma_M^2 \leq 3/2$, so Proposition 5 implies (A) is finite. On the other hand, from $\sum_{t=1}^{+\infty}\gamma_t^2\eta_{t+1} < +\infty$ and $\sum_{t=1}^{+\infty}\eta_t^2 < +\infty$ we deduce that (B) is also finite. We then conclude that it is effectively true that $\sum_{t=1}^{+\infty}\mathbb{E}[\chi_t] < +\infty$.

As a consequence, applying Robbins-Siegmund theorem gives the almost sure convergence of $\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2]$ to a finite random variable U_∞. To proceed, we use the equality

$$\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2] = \mathbb{E}_{t-1}[\|\tilde{\mathbf{X}}_t - \eta_t\boldsymbol{\xi}_{t-\frac{1}{2}} - \mathbf{x}_\star\|^2] = \|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|^2 + \eta_t^2\,\mathbb{E}_{t-1}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2].$$

Accordingly,

$$\begin{aligned}
\sum_{t=2}^{+\infty}\mathbb{E}[\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2] - \|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|^2] &= \sum_{t=2}^{+\infty}\mathbb{E}[\eta_t^2\,\mathbb{E}_{t-1}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2]]\\
&\leq \sum_{t=2}^{+\infty}\eta_t^2\,\mathbb{E}[\sigma_M^2\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2 + N\sigma_A^2]\\
&\leq \sum_{t=1}^{+\infty}(\gamma_t\eta_{t+1}\sigma_M^2\,\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] + (\eta_{t+1})^2N\sigma_A^2)\\
&< +\infty.
\end{aligned} \tag{59}$$

To obtain the last inequality we have applied *i*) Proposition 5; and *ii*) the summability of $(\eta_t^2)_{t\in\mathbb{N}}$. Invoking Lemma 23, we deduce that $\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2] - \|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|^2$ converges to 0 almost surely. This together with the almost sure convergence of $\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2]$ to U_∞ we obtain the almost sure convergence of $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|^2$ to U_∞.

To summarize, we have shown that for all $\mathbf{x}_\star \in \mathcal{X}_\star$, the distance $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|$ almost surely converges. Applying Corollary 3, we conclude that the event $\{\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|$ converges for all $\mathbf{x}_\star \in \mathcal{X}_\star\}$ happens with probability 1.

(2) *There exists and increasing function* $\omega\colon \mathbb{N} \to \mathbb{N}$ *such that* $\|\mathbf{V}(\mathbf{X}_{\omega(t)+\frac{1}{2}})\|^2 + \|\mathbf{X}_{\omega(t)+\frac{1}{2}} - \tilde{\mathbf{X}}_{\omega(t)}\|^2$ *converges to* 0 *almost surely.* From Lemma 24, we know it is sufficient to show that

$$\liminf_{t\to+\infty}\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\|^2] = 0.$$

Since $\sum_{t=1}^{+\infty}\gamma_t\eta_{t+1} = +\infty$ in all the cases, the above is implied by

$$\sum_{t=2}^{+\infty}\gamma_t\eta_{t+1}\,\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\|^2] < +\infty. \tag{60}$$

Using Assumption 3 and $\eta_t < \gamma_t$, we have

$$\begin{aligned}\mathbb{E}[\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\|^2] &= \mathbb{E}[\|\gamma_t \mathbf{V}(\mathbf{X}_{t-\frac{1}{2}}) + (\eta_t + \gamma_t)\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2] \\ &= \gamma_t^2\, \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2] + (\eta_t + \gamma_t)^2\, \mathbb{E}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2] \\ &\leq \gamma_t^2 (1 + 4\sigma_M^2)\, \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2] + 4\gamma_t^2 N \sigma_A^2.\end{aligned}$$

Subsequently, with Proposition 5, the summability of $(\gamma_t^2 \eta_{t+1})_{t\in\mathbb{N}}$ and the fact that the learning rates are non-increasing, we obtain

$$\begin{aligned}\sum_{t=2}^{+\infty} \gamma_t \eta_{t+1} \|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\|^2] &\leq \sum_{t=2}^{+\infty} \gamma_t \eta_{t+1}\, \mathbb{E}[\gamma_t^2 (1 + 4\sigma_M^2) \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2 + 4\gamma_t^2 N \sigma_A^2] \\ &\leq \sum_{t=1}^{+\infty} \gamma_1^2 \gamma_t \eta_{t+1} (1 + 4\sigma_M^2) \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 + \sum_{t=1}^{+\infty} 4\gamma_1 \gamma_t^2 \eta_{t+1} N \sigma_A^2 \\ &< +\infty.\end{aligned}$$

Invoking Proposition 5 again gives $\sum_{t=2}^{+\infty} \gamma_t \eta_{t+1}\, \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2] < +\infty$ and thus we have effectively (60). This concludes the proof of this step.

(3) $(\tilde{\mathbf{X}}_t)_{t\in\mathbb{N}}$ *converges to a point in* $\mathcal{X}_\star$ *almost surely.* Let us define the event

$$\mathcal{E} = \{\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\| \text{ converges for all } \mathbf{x}_\star \in \mathcal{X}_\star;\ \|\mathbf{V}(\mathbf{X}_{\omega(t)+\frac{1}{2}})\|^2 + \|\mathbf{X}_{\omega(t)+\frac{1}{2}} - \tilde{\mathbf{X}}_{\omega(t)}\|^2 \text{ converges to } 0\}$$

Combining the aforementioned two points we know that $\mathbb{P}(\mathcal{E}) = 1$. It is thus sufficient to show that $(\tilde{\mathbf{X}}_t)_{t\in\mathbb{N}}$ converges to a point in $\mathcal{X}_\star$ for any realization $\mathcal{E}$.

Let us consider a realization of $\mathcal{E}$. The set $\mathcal{X}_\star$ being non-empty, the convergence of $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|$ for a $\mathbf{x}_\star \in \mathcal{X}_\star$ implies the boundedness of $(\tilde{\mathbf{X}}_t)_{t\in\mathbb{N}}$. Therefore, we can extract a subsequence of $(\tilde{\mathbf{X}}_{\omega(t)})_t$, which we denote by $(\tilde{\mathbf{X}}_{\omega(\psi(t))})_t$ that converges to a point $\mathbf{x}_\infty \in \mathcal{X}$. As $\lim_{t\to+\infty} \|\mathbf{X}_{\omega(\psi(t))+\frac{1}{2}} - \tilde{\mathbf{X}}_{\omega(\psi(t))}\|^2 = 0$, we deduce that $(\mathbf{X}_{\omega(\psi(t)+\frac{1}{2}})_t$ also converges to $\mathbf{x}_\infty \in \mathcal{X}$. Moreover, we also have $\lim_{t\to+\infty} \|\mathbf{V}(\mathbf{X}_{\omega(\psi(t))+\frac{1}{2}})\|^2 = 0$. By continuity of $\mathbf{V}$ we then know that $\mathbf{V}(\mathbf{x}_\infty) = 0$, i.e., $\mathbf{x}_\infty \in \mathcal{X}_\star$. By definition of $\mathcal{E}$, this implies the convergence of $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\infty\|$. The limit $\lim_{t\to+\infty} \|\tilde{\mathbf{X}}_t - \mathbf{x}_\infty\|$ is thus well defined and $\lim_{t\to+\infty} \|\tilde{\mathbf{X}}_t - \mathbf{x}_\infty\| = \lim_{t\to+\infty} \|\tilde{\mathbf{X}}_{\omega(\psi(t))} - \mathbf{x}_\infty\|$. However, $\lim_{t\to+\infty} \|\tilde{\mathbf{X}}_{\omega(\psi(t))} - \mathbf{x}_\infty\| = 0$ by the choice of $\mathbf{x}_\infty$. We have therefore $\lim_{t\to+\infty} \|\tilde{\mathbf{X}}_t - \mathbf{x}_\infty\| = 0$. Recalling that $\mathbf{x}_\infty \in \mathcal{X}_\star$, we have indeed shown that $(\tilde{X}_t)_{t\in\mathbb{N}}$ converges to a point in $\mathcal{X}_\star$.

(4) *Conclude:* $(\mathbf{X}_t)_{t\in\mathbb{N}}$ *converges to a point in* $\mathcal{X}_\star$ *almost surely* . We claim that $\|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|$ converges to 0. In fact, similar to (59), it holds that

$$\sum_{t=1}^{+\infty} \mathbb{E}[\|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|^2] = \sum_{t=2}^{+\infty} \eta_t^2\, \mathbb{E}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2] < +\infty.$$

Invoking Lemma 23 we get almost sure convergence of $\|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|$ to 0. Moreover, we have shown in the previous point that $(\tilde{\mathbf{X}}_t)_{t\in\mathbb{N}}$ converges to a point in $\mathcal{X}_\star$ almost surely. Combining the above two arguments we obtain the almost sure convergence of $(\mathbf{X}_t)_{t\in\mathbb{N}}$ to a point in $\mathcal{X}_\star$. □

Provided that the players use larger extrapolation steps, the convergence of $\mathbf{X}_t$ does not necessarily imply the convergence of $\mathbf{X}_{t+\frac{1}{2}}$. The next theorem derives sufficient condition for the latter to hold.

Theorem 16. *Let Assumptions 1–3 hold and all players run* (OG+) *with non-increasing learning rate sequences* $(\gamma_t)_{t\in\mathbb{N}}$ *and* $(\eta_t)_{t\in\mathbb{N}}$ *satisfying* (4) *and* (58)*. Assume further that* $\gamma_t^3 = \mathcal{O}(\eta_t)$ *and there exists* $r \in (2, 4]$ *and* $\sigma > 0$ *such that* $\mathbb{E}[\|\boldsymbol{\xi}_t\|^r] \leq \sigma^r$ *for all* t *and* $\sum_{t=1}^{+\infty} \gamma^r < \infty$*. Then, the actual point of play* $\mathbf{X}_{t+\frac{1}{2}}$ *converges almost surely to a Nash equilibrium.*

Proof. Since we already know that $(\mathbf{X}_t)_{t\in\mathbb{N}}$ converges to a point in $\mathcal{X}_\star$ almost surely, it is sufficient to show that $\lim_{t\to+\infty} \|\mathbf{X}_t - \mathbf{X}_{t+\frac{1}{2}}\| = 0$ almost surely. By the update rule of OG+, we have, for

$t \geq 2$, $\mathbf{X}_t - \mathbf{X}_{t+\frac{1}{2}} = \gamma_t \mathbf{V}(\mathbf{X}_{t-\frac{1}{2}}) + \gamma_t \boldsymbol{\xi}_{t-\frac{1}{2}}$. We will deal with the two terms separately. For the noise term, we notice that under the additional assumptions we have

$$\sum_{t=2}^{+\infty} \mathbb{E}[\|\gamma_t \boldsymbol{\xi}_{t-\frac{1}{2}}\|^r] \leq \sum_{t=2}^{+\infty} \gamma_t^r \sigma^r < +\infty.$$

Therefore, applying Lemma 23 gives the almost sure convergence of $\|\gamma_t \boldsymbol{\xi}_{t-\frac{1}{2}}\|$ to 0. As for the operator term, for $t \geq 3$ we bound

$$\|\gamma_t \mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\| \leq \gamma_t \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}}) - \mathbf{V}(\mathbf{X}_{t-1})\| + \gamma_t \|\mathbf{V}(\mathbf{X}_{t-1})\|.$$

On one hand, as $(\mathbf{X}_t)_{t\in\mathbb{N}}$ converges to a point in $\mathcal{X}_\star$ almost surely, the term $\gamma_t \|\mathbf{V}(\mathbf{X}_{t-1})\|$ converges to 0 almost surely by continuity of $\mathbf{V}$. On the other hand, by Lipschitz continuity of $\mathbf{V}$ we have

$$\begin{aligned}\sum_{t=2}^{+\infty} \mathbb{E}[(\gamma_{t+1})^2 \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}}) - \mathbf{V}(\mathbf{X}_t)\|^2] &\leq \sum_{t=2}^{+\infty} (\gamma_{t+1})^2 \gamma_t^2 N L^2 \, \mathbb{E}[\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2] \\ &\leq \sum_{t=2}^{+\infty} \gamma_t^4 N L^2 \, \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2] + \sum_{t=2}^{+\infty} \gamma_t^4 N L^2 \, \mathbb{E}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2].\end{aligned} \tag{61}$$

Since $\gamma_t^3 = \mathcal{O}(\eta_t)$, there exists $C \in \mathbb{R}_+$ such that $\gamma_t^3 \leq C\eta_t$ for all $t \in \mathbb{N}$. Along with Proposition 5 we get

$$\sum_{t=2}^{+\infty} \gamma_t^4 N L^2 \, \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2] \leq \sum_{t=2}^{+\infty} \gamma_{t-1} \eta_t C N L^2 \, \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2] < +\infty. \tag{62}$$

Since $r > 2$, by Jensen's inequality $\mathbb{E}[\|\boldsymbol{\xi}_t\|^r] \leq \sigma^r$ implies $\mathbb{E}[\|\boldsymbol{\xi}_t\|^2] \leq \sigma^2$. Along with $r \leq 4$ and $\sum_{t=1}^{+\infty} \gamma_t^r < +\infty$ we deduce

$$\sum_{t=2}^{+\infty} \gamma_t^4 N L^2 \, \mathbb{E}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2] \leq \sum_{t=2}^{+\infty} \gamma_t^r \gamma_1^{4-r} N L^2 \sigma^2 < +\infty. \tag{63}$$

Combining (61), (62), and (63) we obtain $\sum_{t=2}^{+\infty} \mathbb{E}[(\gamma_{t+1})^2 \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}}) - \mathbf{V}(\mathbf{X}_t)\|^2] < +\infty$, which implies $\lim_{t\to+\infty} \gamma_{t+1} \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}}) - \mathbf{V}(\mathbf{X}_t)\| = 0$ using Lemma 23. In summary, we have shown the three sequences $(\gamma_t \|\boldsymbol{\xi}_{t-\frac{1}{2}}\|)_{t\in\mathbb{N}}$, $(\gamma_t \|\mathbf{V}(\mathbf{X}_{t-1})\|)_{t\in\mathbb{N}}$, and $(\gamma_t \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}}) - \mathbf{V}(\mathbf{X}_{t-1})\|)_{t\in\mathbb{N}}$ converge almost surely to 0. As we have

$$\|\mathbf{X}_t - \mathbf{X}_{t+\frac{1}{2}}\| = \|\gamma_t \mathbf{V}(\mathbf{X}_{t-\frac{1}{2}}) + \gamma_t \boldsymbol{\xi}_{t-\frac{1}{2}}\| \leq \gamma_t \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}}) - \mathbf{V}(\mathbf{X}_{t-1})\| + \gamma_t \|\mathbf{V}(\mathbf{X}_{t-1})\| + \gamma_t \|\boldsymbol{\xi}_{t-\frac{1}{2}}\|,$$

we can indeed conclude that $\lim_{t\to+\infty} \|\mathbf{X}_t - \mathbf{X}_{t+\frac{1}{2}}\| = 0$ almost surely. □

H.3 Trajectory Convergence of Non-Adaptive OptDA+ under Multiplicative Noise

We now turn to the case of multiplicative noise and prove almost sure last-iterate convergence with constant learning rates.

Theorem 17. *Let Assumptions 1–3 hold with $\sigma_A = 0$ and all players run (OG+) / (OptDA+) with learning rates given in Theorem 2(b). Then, both $\mathbf{X}_t$ and $\mathbf{X}_{t+\frac{1}{2}}$ converge almost surely to a Nash equilibrium.*

Proof. As in the proof Theorem 15, we define $\tilde{\mathbf{X}}_1 = \mathbf{X}_1$ and for all $i \in \mathcal{N}$, $t \geq 2$,

$$\tilde{X}_t^i = X_t^i + \eta^i \xi_{t-\frac{1}{2}}^i = -\eta^i \sum_{s=1}^{t-2} \hat{V}_{t+\frac{1}{2}}^i - \eta^i V^i(\mathbf{X}_{t-\frac{1}{2}}).$$

$\tilde{\mathbf{X}}_t$ serves a surrogate for $\mathbf{X}_t$ and is $\mathcal{F}_{t-1}$-measurable. Our first step is to show that

With probability 1, $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}}$ *converges for all* $\mathbf{x}_\star \in \mathcal{X}_\star$.

For this, we fix $\mathbf{x}_\star \in \mathcal{X}_\star$ and apply Robbins-Siegmund's theorem (Lemma 25) to inequality (13) of Lemma 7 with

$$\mathcal{G}_t \leftarrow \mathcal{F}_{t-1}, \quad U_t \leftarrow \mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}}^2], \quad \alpha_t \leftarrow 0, \quad \zeta_t \leftarrow \mathbb{E}_{t-1}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\gamma}}^2] + \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|_{\boldsymbol{\gamma}}^2,$$
$$\chi_t \leftarrow \mathbb{E}_{t-1}[3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t-1})\|_{\boldsymbol{\gamma}}^2 + (4N+1)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}^2}^2 + 3L^2(\|\boldsymbol{\gamma}\|_1\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|_{\boldsymbol{\gamma}^2}^2 + \|\boldsymbol{\gamma}\|_1\|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|_{\boldsymbol{\gamma}^2}^2) + 2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|_{\boldsymbol{\eta}}^2].$$

For $t = 1$ we use (31); thus $\zeta_t = 0$ and $\chi_t = \|\hat{\mathbf{V}}_{3/2}\|_{\boldsymbol{\eta}}^2$. To see that Robbins-Siegmund's theorem is effectively applicable, we use Assumptions 1 and 3 with $\sigma_A = 0$ to establish[7]

$$\begin{aligned}\mathbb{E}[\chi_t] \leq \mathbb{E}[&3\|\boldsymbol{\gamma}\|_\infty L^2\|\mathbf{X}_t - \mathbf{X}_{t-1}\|^2 \\ &+ (\|\boldsymbol{\gamma}\|_\infty(4N+1)L\sigma_M^2 + 3\|\boldsymbol{\gamma}\|_\infty^2 NL^2(1+\sigma_M^2))\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|_{\boldsymbol{\gamma}}^2] \\ &+ 3\|\boldsymbol{\gamma}\|_\infty^2 NL^2(1+\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{t-\frac{3}{2}})\|_{\boldsymbol{\gamma}}^2 + 2(1+\sigma_M^2)\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|_{\boldsymbol{\eta}}^2].\end{aligned}$$

With $2(1+\sigma_M^2)\boldsymbol{\eta} \leq \boldsymbol{\gamma}$, it follows immediately from Proposition 6 that $\sum_{t=1}^{+\infty}\mathbb{E}[\chi_t] < +\infty$. Robbins-Siegmund's theorem thus ensures the almost sure convergence of $\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2]$ to a finite random variable. By definition of $\tilde{X}_t^i$, we have

$$\mathbb{E}_{t-1}[\|X_t^i - x_\star^i\|^2] = \mathbb{E}_{t-1}[\|\tilde{X}_t^i - \eta^i\xi_{t-\frac{1}{2}}^i - x_\star^i\|^2] = \|\tilde{X}_t^i - x_\star^i\|^2 + (\eta^i)^2\,\mathbb{E}_{t-1}[\|\xi_{t-\frac{1}{2}}^i\|^2].$$

Subsequently

$$\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}}^2] = \|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}}^2 + \mathbb{E}_{t-1}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|_{\boldsymbol{\eta}}^2].$$

Therefore, by Assumption 3 with $\sigma_A = 0$ and Proposition 6 we get

$$\sum_{t=2}^{+\infty}\mathbb{E}[\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}}^2] - \|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}}^2] = \sum_{t=2}^{+\infty}\mathbb{E}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|_{\boldsymbol{\eta}}^2] \leq \sum_{t=2}^{+\infty}\sigma_M^2\,\mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|_{\boldsymbol{\eta}}^2] < +\infty.$$

Following the proof of Theorem 15, we deduce with the help of Lemma 23 and Corollary 3 that the claimed argument is effectively true, i.e., with probability 1, $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}}$ converges for all $\mathbf{x}_\star \in \mathcal{X}_\star$.

Since $\|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|^2 = \|\boldsymbol{\xi}_{t-\frac{1}{2}}\|_{\boldsymbol{\eta}^2}^2$ and $\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\|^2 = \sum_{i=1}^N\|\gamma^i\hat{V}_{t-\frac{1}{2}}^i + \eta^i\xi_{t-\frac{1}{2}}^i\|^2$ (for $t \geq 2$), applying the relative noise assumption, Proposition 6 (or simply point (a)), and Lemma 23 we deduce that both $\|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|$ and $\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\|$ converge to 0 almost surely. Moreover, point (a) along with Lemma 23 also implies the almost sure convergence of $\|\mathbf{V}(X_{t+\frac{1}{2}})\|$ to 0. In summary, we have shown that the event

$$\mathcal{E} := \left\{\begin{array}{c}\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}} \text{ converges for all } \mathbf{x}_\star \in \mathcal{X}_\star, \\ \lim_{t\to+\infty}\|\mathbf{X}_t - \tilde{\mathbf{X}}_t\| = 0, \quad \lim_{t\to+\infty}\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\| = 0, \quad \lim_{t\to+\infty}\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\| = 0\end{array}\right\}$$

happens almost surely. To conclude, we just need to show that $\mathbf{X}_t$ and $\mathbf{X}_{t+\frac{1}{2}}$ converge to a point in $\mathcal{X}_\star$ whenever $\mathcal{E}$ happens. The convergence of $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}}$ for a point $\mathbf{x}_\star$ in particular implies the boundedness of $(\tilde{\mathbf{X}}_t)_{t\in\mathbb{N}}$. Therefore, $(\tilde{\mathbf{X}}_t)_{t\in\mathbb{N}}$ has at least a cluster point, which we denote by $\mathbf{x}_\infty$. Provided that $\lim_{t\to+\infty}\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\| = 0$, the point $\mathbf{x}_\infty$ is clearly also a cluster point of $(\mathbf{X}_{t+\frac{1}{2}})_{t\in\mathbb{N}}$. By $\lim_{t\to+\infty}\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\| = 0$ and the continuity of $\mathbf{V}$ we then have $\mathbf{V}(\mathbf{x}_\infty) = 0$, i.e., $\mathbf{x}_\infty \in \mathcal{X}_\star$. This in turn implies that $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\infty\|_{1/\boldsymbol{\eta}}$ converges, so this limit can only be 0. In other words, $(\tilde{\mathbf{X}}_t)_{t\in\mathbb{N}}$ converges to $\mathbf{x}_\infty$; we conclude by $\lim_{t\to+\infty}\|\mathbf{X}_t - \tilde{\mathbf{X}}_t\| = 0$ and $\lim_{t\to+\infty}\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\| = 0$. □

H.4 Trajectory Convergence of Adaptive OptDA+ under Multiplicative Noise

In closing, we prove the almost sure last-iterate convergence of adaptive OptDA+ under multiplicative noise. As claimed in Section 6, we first show that the learning rates almost surely converge to positive constant. This intuitively means that the analysis of the last section should apply as well.

[7] For $t = 1$ and $t = 2$, we remove the terms that involve either $\mathbf{X}_{1/2}$, $\mathbf{X}_0$, or $\mathbf{X}_{-1/2}$.

Lemma 27. *Let Assumptions 1–4 hold with $\sigma_A = 0$ and all players run OptDA+ with adaptive learning rates (Adapt). Then,*

(a) With probability 1, for all $i \in \mathcal{N}$, $(\lambda_t^i)_{t\in\mathbb{N}}$ and $(\mu_t^i)_{t\in\mathbb{N}}$ converge to finite constant.

(b) With probability 1, for all $i \in \mathcal{N}$, the learning rates $(\gamma_t^i)_{t\in\mathbb{N}}$ and $(\eta_t^i)_{t\in\mathbb{N}}$ converge to positive constants.

Proof. We notice that (b) is a direct consequence of (a) so we will only show (a) below. For this, we make use of Lemma 22 and Lemma 23. In fact, $(\sqrt{\lambda_t^i})_{t\in\mathbb{N}}$ is clearly non-decreasing and by Lemma 22, $\sup_{t\in\mathbb{N}} \mathbb{E}[\sqrt{\lambda_t^i}] < +\infty$. Therefore, Lemma 23 ensures the almost sure convergence of $(\sqrt{\lambda_t^i})_{t\in\mathbb{N}}$ to a finite random variable, which in turn implies that $(\lambda_t^i)_{t\in\mathbb{N}}$ converges to a finite constant almost surely. Similarly, $(\mu_t^i)_{t\in\mathbb{N}}$ is non-decreasing and $\sup_{t\in\mathbb{N}} \mathbb{E}[\mu_t^i] < +\infty$ by Lemma 22. We thus deduce by Lemma 23 that $(\mu_t^i)_{t\in\mathbb{N}}$ converges to finite constant almost surely. □

We now adapt the proof of Theorem 17 to the case of adaptive learning rates. Note that the fact that the learning rates are not constant also causes some additional challenges.

Theorem 18. *Let Assumptions 1–4 hold with $\sigma_A = 0$ and all players run (OptDA+) with adaptive learning rates (Adapt). Then,*

(a) It holds almost surely that $\sum_{t=1}^{+\infty} \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 < +\infty$.

(b) Both $(\mathbf{X}_t)_{t\in\mathbb{N}}$ and $(\mathbf{X}_{t+\frac{1}{2}})_{t\in\mathbb{N}}$ converge to a Nash equilibrium almost surely.

Proof. In the following, we define $\boldsymbol{\gamma}_\infty = \lim_{t\to+\infty} \boldsymbol{\gamma}_t$ and $\boldsymbol{\eta}_\infty = \lim_{t\to+\infty} \boldsymbol{\eta}_t$ as the limits of the learning rate sequences. Since for each $i \in \mathcal{N}$, $(\gamma_t^i)_{t\in\mathbb{N}}$ and $(\eta_t^i)_{t\in\mathbb{N}}$ are non-negative non-increasing sequences, both $\boldsymbol{\gamma}_\infty$ and $\boldsymbol{\eta}_\infty$ are well-defined. Moreover, by Lemma 27 we know that $\boldsymbol{\gamma}_\infty$ and $\boldsymbol{\eta}_\infty$ are positive almost surely.

(a) Combining Lemma 19 and Lemma 22 we get immediately $\sum_{t=1}^{+\infty} \mathbb{E}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t}] < +\infty$. Therefore, using Lemma 23 we deduce that $\sum_{t=1}^{+\infty} \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t} < +\infty$ almost surely. By definition of $\boldsymbol{\gamma}_\infty$ we have

$$\sum_{t=1}^{+\infty} \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t} \geq \sum_{t=1}^{+\infty} \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_\infty} \geq \min_{i\in\mathcal{N}} \gamma_\infty^i \sum_{t=1}^{+\infty} \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2$$

As a consequence, whenever *i)* $C := \sum_{t=1}^{+\infty} \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t}$ is finite; and *ii)* $\min_{i\in\mathcal{N}} \gamma_\infty^i > 0$, we have

$$\sum_{t=1}^{+\infty} \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 \leq \frac{C}{\min_{i\in\mathcal{N}} \gamma_\infty^i} < +\infty.$$

As both *i)* and *ii)* hold almost surely, we have indeed shown that $\sum_{t=1}^{+\infty} \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2 < +\infty$ almost surely.

(b) To prove this point, we follow closely the proof of Theorem 17. To begin, we fix $\mathbf{x}_\star \in \mathcal{X}_\star$ and show that we can always apply Robbins-Siegmund's theorem (Lemma 25) to inequality (13) of Lemma 7 (or inequality (31) for $t = 1$). This gives, for $t \geq 2$,

$$\mathcal{G}_t = \mathcal{F}_{t-1}, \quad U_t = \mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_t}], \quad \alpha_t = 0, \quad \zeta_t = \mathbb{E}_{t-1}[\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t}] + \|\mathbf{V}(\mathbf{X}_{t-\frac{1}{2}})\|^2_{\boldsymbol{\gamma}_t},$$
$$\chi_t = \mathbb{E}_{t-1}[3\|\mathbf{V}(\mathbf{X}_t) - \mathbf{V}(\mathbf{X}_{t-1})\|^2_{\boldsymbol{\gamma}_t} + \|\mathbf{X}_1 - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_{t+1}-1/\boldsymbol{\eta}_t} + (4N+1)L\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2} + 3L^2$$
$$(\|\boldsymbol{\gamma}_t\|_1 \|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\gamma}_t^2} + \|\boldsymbol{\gamma}_{t-1}\|_1 \|\hat{\mathbf{V}}_{t-\frac{3}{2}}\|^2_{(\boldsymbol{\gamma}_{t-1})^2}) + 2\|\hat{\mathbf{V}}_{t+\frac{1}{2}}\|^2_{\boldsymbol{\eta}_t}].$$

As for $t = 1$, we replace the above with $\zeta_t = 0$ and $\chi_t = \|\hat{\mathbf{V}}_{3/2}\|^2_{\boldsymbol{\eta}_1}$. Using Assumption 1, (44), and (45), we can bound the sum of the expectation of χ_t by

$$\sum_{t=1}^{T} \mathbb{E}[\chi_t] \leq \sum_{t=1}^{T-1} 3L^2\, \mathbb{E}[\|\mathbf{X}_t - \mathbf{X}_{t+1}\|^2] + \sum_{i=1}^{N} \left(\|X_1^i - x_\star^i\|^2\, \mathbb{E}\left[\sqrt{1 + \lambda_{T-1}^i + \mu_{T-1}^i}\right]\right)$$

$$+ (6NL^2 + (4N+1)L)\left(2N(G^2+\bar{\sigma}^2) + \sum_{i=1}^{N} 2\,\mathbb{E}\left[\sqrt{\lambda_{T-1}^i}\right]\right) +$$
$$+ 8N(G^2+\bar{\sigma}^2) + \sum_{i=1}^{N} 4\,\mathbb{E}\left[\sqrt{\lambda_T^i}\right]$$

It then follows immediately from Lemma 22 that $\sum_{t=1}^{+\infty}\mathbb{E}[\chi_t] < +\infty$. With Robbins-Siegmund's theorem we deduce that $\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_t}]$ converges almost surely to a finite random variable.

As in the proof of Theorems 15 and 17, we next define $\tilde{\mathbf{X}}_1 = \mathbf{X}_1$ and for all $i \in \mathcal{N}$, $t \geq 2$,

$$\tilde{X}_t^i = X_t^i + \eta_t^i \xi_{t-\frac{1}{2}}^i = -\eta_t^i \sum_{s=1}^{t-2} \hat{V}_{t+\frac{1}{2}}^i - \eta_t^i V^i(\mathbf{X}_{t-\frac{1}{2}}).$$

Then,

$$\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_t}] = \|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_t} + \mathbb{E}_{t-1}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\eta}_t}].$$

Using $\mathbb{E}_{t-1}[\|\xi_{t-\frac{1}{2}}^i\|^2] \leq \mathbb{E}_{t-1}[\|\hat{V}_{t-\frac{1}{2}}^i\|^2]$, the law of total expectation, the fact that $\boldsymbol{\eta}_t$ is $\mathcal{F}_{t-1}$-measurable, Lemma 16, and Lemma 22, we then get

$$\begin{aligned}
\sum_{t=2}^{+\infty} \mathbb{E}[\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_t}] - \|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_t}] &= \sum_{t=2}^{+\infty} \mathbb{E}[\|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\eta}_t}] \\
&\leq \sum_{t=2}^{+\infty} \mathbb{E}[\|\hat{\mathbf{V}}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\eta}_t}] \\
&\leq 2N(G^2+\bar{\sigma}^2) + \sup_{t\in\mathbb{N}} \sum_{i=1}^{N} 2\,\mathbb{E}\left[\sqrt{\lambda_t^i}\right] \\
&< +\infty. \qquad (64)
\end{aligned}$$

Invoking Lemma 23 we deduce that $\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_t}] - \|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_t}$ almost surely converges to 0. Since we have shown $\mathbb{E}_{t-1}[\|\mathbf{X}_t - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_t}]$ almost surely converges to a finite random variable, we now know that $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|^2_{1/\boldsymbol{\eta}_t}$ almost surely converges to this finite random variable as well. To summarize, we have shown that for any $\mathbf{x}_\star \in \mathcal{X}_\star$, $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_t}$ converges almost surely. Moreover, we also know that $(1/\boldsymbol{\eta}_\infty)$, the limit of $(1/\boldsymbol{\eta}_t)_{t\in\mathbb{N}}$ is finite almost surely. Therefore, applying Lemma 26 with $\mathcal{K} \leftarrow \mathcal{X}_\star$, $\mathbf{u}_t \leftarrow \tilde{\mathbf{X}}_t$, and $\boldsymbol{\alpha}_t \leftarrow 1/\boldsymbol{\eta}_t$, we deduce that with probability 1, the vector $1/\boldsymbol{\eta}_\infty$ is finite and $\|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_\infty}$ converges for all $\mathbf{x}_\star \in \mathcal{X}_\star$.

Next, with $\|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|^2 = \|\boldsymbol{\xi}_{t-\frac{1}{2}}\|^2_{\boldsymbol{\eta}_t^2}$ and $\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\|^2 = \sum_{i=1}^N \|\gamma_t^i \hat{V}_{t-\frac{1}{2}}^i + \eta_t^i \xi_{t-\frac{1}{2}}^i\|^2$ (for $t \geq 2$), following the reasoning of (64), we get both $\sum_{t=1}^{+\infty}\mathbb{E}[\|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|^2] < +\infty$ and $\sum_{t=1}^{+\infty}\mathbb{E}[\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\|^2] < +\infty$. By Lemma 23 we then know that $\|\mathbf{X}_t - \tilde{\mathbf{X}}_t\|$ and $\|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\|$ converge to 0 almost surely. Finally, from point (a) we know that $\|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\|$ converges to 0 almost surely. To conclude, let us define the event

$$\mathcal{E} := \left\{ \begin{array}{c} 1/\boldsymbol{\eta}_\infty \text{ is finite and } \|\tilde{\mathbf{X}}_t - \mathbf{x}_\star\|_{1/\boldsymbol{\eta}_\infty} \text{ converges for all } \mathbf{x}_\star \in \mathcal{X}_\star, \\ \lim_{t\to+\infty} \|\mathbf{X}_t - \tilde{\mathbf{X}}_t\| = 0, \quad \lim_{t\to+\infty} \|\mathbf{X}_{t+\frac{1}{2}} - \tilde{\mathbf{X}}_t\| = 0, \quad \lim_{t\to+\infty} \|\mathbf{V}(\mathbf{X}_{t+\frac{1}{2}})\| = 0 \end{array} \right\}$$

We have shown that $\mathbb{P}(\mathcal{E}) = 1$. Moreover, with the arguments of Theorem 17 we deduce that whenever $\mathcal{E}$ happens both $(\mathbf{X}_t)_{t\in\mathbb{N}}$ and $(\mathbf{X}_{t+\frac{1}{2}})_{t\in\mathbb{N}}$ converge to a point in $\mathcal{X}_\star$, and this ends the proof. □