

---

# Follow the Perturbed Leader: Optimism and Fast Parallel Algorithms for Smooth Minimax Games

---

**Arun Sai Suggala**  
Carnegie Mellon University  
asuggala@cs.cmu.edu

**Praneeth Netrapalli**  
Microsoft Research, India  
praneeth@microsoft.com

## Abstract

We consider the problem of online learning and its application to solving minimax games. For the online learning problem, Follow the Perturbed Leader (FTPL) is a widely studied algorithm which enjoys the optimal  $O(T^{1/2})$  *worst case* regret guarantee for both convex and nonconvex losses. In this work, we show that when the sequence of loss functions is *predictable*, a simple modification of FTPL which incorporates optimism can achieve better regret guarantees, while retaining the optimal worst case regret guarantee for unpredictable sequences. A key challenge in obtaining these tighter regret bounds is the stochasticity and optimism in the algorithm, which requires different analysis techniques than those commonly used in the analysis of FTPL. The key ingredient we utilize in our analysis is the dual view of perturbation as regularization. While our algorithm has several applications, we consider the specific application of minimax games. For solving smooth convex-concave games, our algorithm only requires access to a linear optimization oracle. For Lipschitz and smooth nonconvex-nonconcave games, our algorithm requires access to an optimization oracle which computes the perturbed best response. In both these settings, our algorithm solves the game up to an accuracy of  $O(T^{-1/2})$  using  $T$  calls to the optimization oracle. An important feature of our algorithm is that it is highly parallelizable and requires only  $O(T^{1/2})$  iterations, with each iteration making  $O(T^{1/2})$  parallel calls to the optimization oracle.

## 1 Introduction

In this work, we consider the problem of online learning, where in each iteration, the learner chooses an action and observes a loss function. The goal of the learner is to choose a sequence of actions which minimizes the cumulative loss suffered over the course of learning. The paradigm of online learning has many theoretical and practical applications and has been widely studied in a number of fields, including game theory and machine learning. One of the popular applications of online learning is in solving minimax games arising in various contexts such as boosting [1], robust optimization [2], Generative Adversarial Networks [3].

In recent years, a number of efficient algorithms have been developed for regret minimization. These algorithms fall into two broad categories, namely, Follow the Regularized Leader (FTRL) [4] and FTPL [5] style algorithms. When the sequence of loss functions encountered by the learner are convex, both these algorithms are known to achieve the optimal  $O(T^{1/2})$  worst case regret [6, 7]. While these algorithms have similar regret guarantees, they differ in computational aspects. Each iteration of FTRL involves optimization of a non-linear convex function over the action space (also called the projection step). In contrast, each step of FTPL involves solving a linear optimization problem, which can be implemented efficiently for many problems of interest [8, 9, 10]. For example, if the action space is an  $\ell_p$  ball for some  $p \notin \{1, 2, \infty\}$ , then projecting onto this set is much more computationally expensive than performing linear optimization over this set. As another example, consider the

scenario where the action space is the set of all positive semidefinite matrices. Then projecting onto this set requires performing expensive singular value decompositions. Whereas, linear optimization only requires computation of the leading eigenvector. This crucial difference between FTRL and FTPL makes the latter algorithm more attractive in practice. Even in the more general nonconvex setting, where the loss functions encountered by the learner can potentially be nonconvex, FTPL algorithms are attractive. In this setting, FTPL requires access to an offline optimization oracle which computes the perturbed best response, and achieves  $O(T^{1/2})$  worst case regret [11]. Furthermore, these optimization oracles can be efficiently implemented for many problems by leveraging the rich body of work on global optimization [12].

Despite its importance and popularity, FTPL has been mostly studied for the worst case setting, where the loss functions are assumed to be adversarially chosen. In a number of applications of online learning, the loss functions are actually benign and predictable [13]. In such scenarios, FTPL can not utilize the predictability of losses to achieve tighter regret bounds. While [11, 13] study variants of FTPL which can make use of predictability, these works either consider restricted settings or provide sub-optimal regret guarantees (see Section 2 for more details). This is unlike FTRL, where optimistic variants that can utilize the predictability of loss functions have been well understood [13, 14] and have been shown to provide faster convergence rates in applications such as minimax games. In this work, we aim to bridge this gap and study a variant of FTPL called Optimistic FTPL (OFTPL), which can achieve better regret bounds, while retaining the optimal worst case regret guarantee for unpredictable sequences. The main challenge in obtaining these tighter regret bounds is handling the stochasticity and optimism in the algorithm, which requires different analysis techniques to those commonly used in the analysis of FTPL. In this work, we rely on the dual view of perturbation as regularization to derive regret bounds of OFTPL.

To demonstrate the usefulness of OFTPL, we consider the problem of solving minimax games. A widely used approach for solving such games relies on online learning algorithms [6]. In this approach, both the minimization and the maximization players play a repeated game against each other and rely on online learning algorithms to choose their actions in each round of the game. In our algorithm for solving games, we let both the players use OFTPL to choose their actions. For solving smooth convex-concave games, our algorithm only requires access to a linear optimization oracle. For Lipschitz and smooth nonconvex-nonconcave games, our algorithm requires access to an optimization oracle which computes the perturbed best response. In both these settings, our algorithm solves the game up to an accuracy of  $O(T^{-1/2})$  using  $T$  calls to the optimization oracle. While there are prior algorithms that achieve these convergence rates [11, 15], an important feature of our algorithm is that it is highly parallelizable and requires only  $O(T^{1/2})$  iterations, with each iteration making  $O(T^{1/2})$  parallel calls to the optimization oracle. We note that such parallelizable algorithms are especially useful in large-scale machine learning applications such as training of GANs, adversarial training, which often involve huge datasets such as ImageNet [16].

## 2 Preliminaries and Background Material

**Online Learning.** The online learning framework can be seen as a repeated game between a learner and an adversary. In this framework, in each round  $t$ , the learner makes a prediction  $\mathbf{x}_t \in \mathcal{X} \subseteq \mathbb{R}^d$  for some compact set  $\mathcal{X}$ , and the adversary simultaneously chooses a loss function  $f_t : \mathcal{X} \rightarrow \mathbb{R}$  and observe each others actions. The goal of the learner is to choose a sequence of actions  $\{\mathbf{x}_t\}_{t=1}^T$  so that the following notion of regret is minimized:  $\sum_{t=1}^T f_t(\mathbf{x}_t) - \inf_{\mathbf{x} \in \mathcal{X}} \sum_{t=1}^T f_t(\mathbf{x})$ .

When the domain  $\mathcal{X}$  and loss functions  $f_t$  are convex, a number of efficient algorithms for regret minimization have been studied. Some of these include deterministic algorithms such as Online Mirror Descent, Follow the Regularized Leader (FTRL) [4, 7], and stochastic algorithms such as Follow the Perturbed Leader (FTPL) [5]. In FTRL, one predicts  $\mathbf{x}_t$  as  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \sum_{i=1}^{t-1} \langle \nabla_i, \mathbf{x} \rangle + R(\mathbf{x})$ , for some strongly convex regularizer  $R$ , where  $\nabla_i = \nabla f_i(\mathbf{x}_i)$ . FTRL is known to achieve the optimal  $O(T^{1/2})$  worst case regret in the convex setting [4]. In FTPL, one predicts  $\mathbf{x}_t$  as  $m^{-1} \sum_{j=1}^m \mathbf{x}_{t,j}$ , where  $\mathbf{x}_{t,j}$  is a minimizer of the following linear optimization problem:  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \left\langle \sum_{i=1}^{t-1} \nabla_i - \sigma_{t,j}, \mathbf{x} \right\rangle$ . Here,  $\{\sigma_{t,j}\}_{j=1}^m$  are independent random perturbations drawn from some appropriate probability distribution such as exponential distribution or uniform distribution in a hyper-cube. Various choices of perturbation distribution gives rise to various FTPL algorithms. When the loss functions are

linear, Kalai and Vempala [5] show that FTPL achieves  $O(T^{1/2})$  expected regret, irrespective of the choice of  $m$ . When the loss functions are convex, Hazan [7] showed that the deterministic version of FTPL (*i.e.*, as  $m \rightarrow \infty$ ) achieves  $O(T^{1/2})$  regret. While projection free methods for online convex learning have been studied since the early work of [17], surprisingly, regret bounds of FTPL for finite  $m$  have only been recently studied [10]. Hazan and Minasyan [10] show that for Lipschitz and convex functions, FTPL achieves  $O(T^{1/2} + m^{-1/2}T)$  expected regret, and for smooth convex functions, the algorithm achieves  $O(T^{1/2} + m^{-1}T)$  expected regret.

When either the domain  $\mathcal{X}$  or the loss functions  $f_t$  are non-convex, no deterministic algorithm can achieve  $o(T)$  regret [6, 11]. In such cases, one has to rely on randomized algorithms to achieve sub-linear regret. In randomized algorithms, in each round  $t$ , the learner samples the prediction  $\mathbf{x}_t$  from a distribution  $P_t \in \mathcal{P}$ , where  $\mathcal{P}$  is the set of all probability distributions supported on  $\mathcal{X}$ . The goal of the learner is to choose a sequence of distributions  $\{P_t\}_{t=1}^T$  to minimize the expected regret  $\sum_{t=1}^T \mathbb{E}_{\mathbf{x} \sim P_t} [f_t(\mathbf{x})] - \inf_{\mathbf{x} \in \mathcal{X}} \sum_{t=1}^T f_t(\mathbf{x})$ . A popular technique to minimize the expected regret is to consider a linearized problem in the space of probability distributions with losses  $\tilde{f}_t(P) = \mathbb{E}_{\mathbf{x} \sim P} [f_t(\mathbf{x})]$  and perform FTRL in this space. In such a technique,  $P_t$  is computed as:  $\operatorname{argmin}_{P \in \mathcal{P}} \sum_{i=1}^{t-1} \tilde{f}_i(P) + R(P)$ , for some strongly convex regularizer  $R(P)$ . When  $R(P)$  is the negative entropy of  $P$ , the algorithm is called entropic mirror descent or continuous exponential weights. This algorithm achieves  $O(T^{1/2})$  expected regret for bounded loss functions  $f_t$ . Another technique to minimize expected regret is to rely on FTPL [11, 18]. Here, the learner generates the random prediction  $\mathbf{x}_t$  by first sampling a random perturbation  $\sigma$  and then computing the perturbed best response, which is defined as  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \sum_{i=1}^{t-1} f_i(\mathbf{x}) - \langle \sigma, \mathbf{x} \rangle$ . In a recent work, Agarwal et al. [18] show that this algorithm achieves  $O(T^{2/3})$  expected regret, whenever the sequence of loss functions are Lipschitz. This was later improved to  $O(T^{1/2})$  by Suggala and Netrapalli [11]. We now briefly discuss the computational aspects of FTRL and FTPL. Each iteration of FTRL (with entropic regularizer) requires sampling from a non-logconcave distribution. In contrast, FTPL requires solving a nonconvex optimization problem to compute the perturbed best response. Of these, computing the perturbed best response seems significantly easier since standard algorithms such as gradient descent seem to be able to find approximate global optima reasonably fast, even for complicated tasks such as training deep neural networks.

**Online Learning with Optimism.** When the sequence of loss functions are convex and predictable, Rakhlin and Sridharan [13, 14] study optimistic variants of FTRL which can exploit the predictability to obtain better regret bounds. Let  $g_t$  be our guess of  $\nabla_t$  at the beginning of round  $t$ . Given  $g_t$ , we predict  $\mathbf{x}_t$  in Optimistic FTRL (OFTRL) as  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \left\langle \sum_{i=1}^{t-1} \nabla_i + g_t, \mathbf{x} \right\rangle + R(\mathbf{x})$ . Note that when  $g_t = 0$ , OFTRL is equivalent to FTRL. [13, 14] show that the regret bounds of OFTRL only depend on  $(g_t - \nabla_t)$ . Moreover, these works show that OFTRL provides faster convergence rates for solving smooth convex-concave games. In contrast to FTRL, the optimistic variants of FTPL have been less well understood. [13] studies OFTPL for linear loss functions. But they consider restrictive settings and their algorithms require the knowledge of sizes of deviations  $(g_t - \nabla_t)$ . [11] studies OFTPL for the more general nonconvex setting. The algorithm predicts  $\mathbf{x}_t$  as  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \sum_{i=1}^{t-1} f_i(\mathbf{x}) + g_t(\mathbf{x}) - \langle \sigma, \mathbf{x} \rangle$ , where  $g_t$  is our guess of  $f_t$ . However, the regret bounds of [11] are sub-optimal and weaker than the bounds we obtain in our work (see Theorem 4.2). Moreover, [11] does not provide any consequences of their results to minimax games. We note that their sub-optimal regret bounds translate to sub-optimal rates of convergence for solving smooth minimax games.

**Minimax Games.** Consider the following problem, which we refer to as minimax game:  $\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbf{y} \in \mathcal{Y}} f(\mathbf{x}, \mathbf{y})$ . In these games, we are often interested in finding a Nash Equilibrium (NE). A pair  $(P, Q)$ , where  $P$  is a probability distribution over  $\mathcal{X}$  and  $Q$  is a probability distribution over  $\mathcal{Y}$ , is called a NE if:  $\sup_{\mathbf{y} \in \mathcal{Y}} \mathbb{E}_{\mathbf{x} \sim P} [f(\mathbf{x}, \mathbf{y})] \leq \mathbb{E}_{\mathbf{x} \sim P, \mathbf{y} \sim Q} [f(\mathbf{x}, \mathbf{y})] \leq \inf_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\mathbf{y} \sim Q} [f(\mathbf{x}, \mathbf{y})]$ . A standard technique for finding a NE of the game is to rely on no-regret algorithms [6, 7]. Here, both  $\mathbf{x}$  and  $\mathbf{y}$  players play a repeated game against each other and use online learning algorithms to choose their actions. The average of the iterates generated via this repeated game can be shown to converge to a NE.

**Projection Free Learning.** Projection free learning algorithms are attractive as they only involve solving linear optimization problems. Two broad classes of projection free techniques have been considered for online convex learning and minimax games, namely, Frank-Wolfe (FW) methods

and FTPL based methods. Garber and Hazan [8] consider the problem of online learning when the action space  $\mathcal{X}$  is a *polytope*. They provide a FW method which achieves  $O(T^{1/2})$  regret using  $T$  calls to the linear optimization oracle. Hazan and Kale [17] provide a FW technique which achieves  $O(T^{3/4})$  regret for general online convex learning with Lipschitz losses and uses  $T$  calls to the linear optimization oracle. In a recent work, Hazan and Minasyan [10] show that FTPL achieves  $O(T^{2/3})$  regret for online convex learning with smooth losses, using  $T$  calls to the linear optimization oracle. This translates to  $O(T^{-1/3})$  rate of convergence for solving smooth convex-concave games. Note that, in contrast, our algorithm achieves  $O(T^{-1/2})$  convergence rate in the same setting. Gidel et al. [9] study FW methods for solving convex-concave games. When the constraint sets  $\mathcal{X}, \mathcal{Y}$  are *strongly convex*, the authors show geometric convergence of their algorithms. In a recent work, He and Harchaoui [15] propose a FW technique for solving smooth convex-concave games which converges at a rate of  $O(T^{-1/2})$  using  $T$  calls to the linear optimization oracle. We note that our simple OFTPL based algorithm achieves these rates, with the added advantage of parallelizability. That being said, He and Harchaoui [15] achieve dimension free convergence rates in the Euclidean setting, where the smoothness is measured w.r.t  $\|\cdot\|_2$  norm. In contrast, the rates of convergence of our algorithm depend on the dimension.

**Notation.**  $\|\cdot\|$  is a norm on some vector space, which is typically  $\mathbb{R}^d$  in our work.  $\|\cdot\|_*$  is the dual norm of  $\|\cdot\|$ , which is defined as  $\|\mathbf{x}\|_* = \sup\{\langle \mathbf{u}, \mathbf{x} \rangle : \mathbf{u} \in \mathbb{R}^d, \|\mathbf{u}\| \leq 1\}$ . We use  $\Psi_1, \Psi_2$  to denote norm compatibility constants of  $\|\cdot\|$ , which are defined as  $\Psi_1 = \sup_{\mathbf{x} \neq 0} \|\mathbf{x}\| / \|\mathbf{x}\|_2$ ,  $\Psi_2 = \sup_{\mathbf{x} \neq 0} \|\mathbf{x}\|_2 / \|\mathbf{x}\|$ .

We use the notation  $f_{1:t}$  to denote  $\sum_{i=1}^t f_i$  and  $\nabla_i$  to denote  $\nabla f_i(\mathbf{x}_i)$ . In some cases, when clear from context, we overload the notation  $f_{1:t}$  and use it to denote the set  $\{f_1, f_2 \dots f_t\}$ . For any convex function  $f$ ,  $\partial f(\mathbf{x})$  is the set of all subgradients of  $f$  at  $\mathbf{x}$ . For any function  $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ ,  $f(\cdot, \mathbf{y}), f(\mathbf{x}, \cdot)$  denote the functions  $\mathbf{x} \rightarrow f(\mathbf{x}, \mathbf{y}), \mathbf{y} \rightarrow f(\mathbf{x}, \mathbf{y})$ . For any function  $f : \mathcal{X} \rightarrow \mathbb{R}$  and any probability distribution  $P$ , we let  $f(P)$  denote  $\mathbb{E}_{\mathbf{x} \sim P} [f(\mathbf{x})]$ . Similarly, for any function  $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  and any two distributions  $P, Q$ , we let  $f(P, Q)$  denote  $\mathbb{E}_{\mathbf{x} \sim P, \mathbf{y} \sim Q} [f(\mathbf{x}, \mathbf{y})]$ . For any set of distributions  $\{P_j\}_{j=1}^m$ ,  $\frac{1}{m} \sum_{j=1}^m P_j$  is the mixture distribution which gives equal weights to its components. We use  $\text{Exp}(\eta)$  to denote the exponential distribution, whose CDF is given by  $P(Z \leq s) = 1 - \exp(-s/\eta)$ .

### 3 Dual view of Perturbation as Regularization

In this section, we present a key result which shows that when the sequence of loss functions are convex, every FTPL algorithm is an FTRL algorithm. Our analysis of OFTPL relies on this dual view to obtain tight regret bounds. This duality between FTPL and FTRL was originally studied by Hofbauer and Sandholm [19], where the authors show that any FTPL algorithm, with perturbation distribution admitting a strictly positive density on  $\mathbb{R}^d$ , is an FTRL algorithm w.r.t some convex regularizer. However, many popular perturbation distributions such as exponential and uniform distributions don't have a strictly positive density. In a recent work, Abernethy et al. [20] point out that the duality between FTPL and FTRL holds for very general perturbation distributions. However, the authors do not provide a formal theorem showing this result. Here, we provide a proposition formalizing the claim of [20].

**Proposition 3.1.** *Consider the problem of online convex learning, where the sequence of loss functions  $\{f_t\}_{t=1}^T$  encountered by the learner are convex. Consider the deterministic version of FTPL algorithm, where the learner predicts  $\mathbf{x}_t$  as  $\mathbb{E}_\sigma [\arg\min_{\mathbf{x} \in \mathcal{X}} \langle \nabla_{1:t-1} - \sigma, \mathbf{x} \rangle]$ . Suppose the perturbation distribution is absolutely continuous w.r.t the Lebesgue measure. Then there exists a convex regularizer  $R : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{\infty\}$ , with domain  $\text{dom}(R) \subseteq \mathcal{X}$ , such that  $\mathbf{x}_t = \arg\min_{\mathbf{x} \in \mathcal{X}} \langle \nabla_{1:t-1}, \mathbf{x} \rangle + R(\mathbf{x})$ . Moreover,  $-\nabla_{1:t-1} \in \partial R(\mathbf{x}_t)$ , and  $\mathbf{x}_t = \partial R^{-1}(-\nabla_{1:t-1})$ , where  $\partial R^{-1}$  is the inverse of  $\partial R$  in the sense of multivalued mappings.*

---

**Algorithm 1** Convex OFTPL

---

1: **Input:** Perturbation Distribution  $P_{\text{PRTB}}$ , number of samples  $m$ , number of iterations  $T$   
2: Denote  $\nabla_0 = 0$   
3: **for**  $t = 1 \dots T$  **do**  
4:   Let  $g_t$  be the guess for  $\nabla_t$   
5:   **for**  $j = 1 \dots m$  **do**  
6:     Sample  $\sigma_{t,j} \sim P_{\text{PRTB}}$   
7:      $\mathbf{x}_{t,j} \in \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle \nabla_{0:t-1} + g_t - \sigma_{t,j}, \mathbf{x} \rangle$   
8:   **end for**  
9:   Let  $\mathbf{x}_t = \frac{1}{m} \sum_{j=1}^m \mathbf{x}_{t,j}$   
10:   Play  $\mathbf{x}_t$  and observe loss function  $f_t$   
11: **end for**

---

## 4 Online Learning with OFTPL

### 4.1 Online Convex Learning

In this section, we present the OFTPL algorithm for online convex learning and derive an upper bound on its regret. The algorithm we consider is similar to the OFTRL algorithm (see Algorithm 1). Let  $g_t[f_1 \dots f_{t-1}]$  be our guess for  $\nabla_t$  at the beginning of round  $t$ , with  $g_1 = 0$ . To simplify the notation, in the sequel, we suppress the dependence of  $g_t$  on  $\{f_i\}_{i=1}^{t-1}$ . Given  $g_t$ , we predict  $\mathbf{x}_t$  in OFTPL as follows. We sample independent perturbations  $\{\sigma_{t,j}\}_{j=1}^m$  from the perturbation distribution  $P_{\text{PRTB}}$  and compute  $\mathbf{x}_t$  as  $m^{-1} \sum_{j=1}^m \mathbf{x}_{t,j}$ , where  $\mathbf{x}_{t,j}$  is a minimizer of the following linear optimization problem

$$\mathbf{x}_{t,j} \in \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle \nabla_{1:t-1} + g_t - \sigma_{t,j}, \mathbf{x} \rangle.$$

We now present our main theorem which bounds the regret of OFTPL. A key quantity the regret depends on is the *stability* of predictions of the deterministic version of OFTPL. Intuitively, an algorithm is stable if its predictions in two consecutive iterations differ by a small quantity. To capture this notion, we first define function  $\nabla\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$  as:  $\nabla\Phi(g) = \mathbb{E}_\sigma [\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle g - \sigma, \mathbf{x} \rangle]$ . Observe that  $\nabla\Phi(\nabla_{1:t-1} + g_t)$  is the prediction of the deterministic version of OFTPL. We say the predictions of OFTPL are stable, if  $\nabla\Phi$  is a Lipschitz function.

**Definition 4.1** (Stability). The predictions of OFTPL are said to be  $\beta$ -stable w.r.t some norm  $\|\cdot\|$ , if

$$\forall g_1, g_2 \in \mathbb{R}^d \quad \|\nabla\Phi(g_1) - \nabla\Phi(g_2)\|_* \leq \beta \|g_1 - g_2\|.$$

**Theorem 4.1.** *Suppose the perturbation distribution  $P_{\text{PRTB}}$  is absolutely continuous w.r.t Lebesgue measure. Let  $D$  be the diameter of  $\mathcal{X}$  w.r.t  $\|\cdot\|$ , which is defined as  $D = \sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}} \|\mathbf{x}_1 - \mathbf{x}_2\|$ . Let  $\eta = \mathbb{E}_\sigma [\|\sigma\|_*]$ , and suppose the predictions of OFTPL are  $C\eta^{-1}$ -stable w.r.t  $\|\cdot\|_*$ , where  $C$  is a constant that depends on the set  $\mathcal{X}$ . Finally, suppose the sequence of loss functions  $\{f_t\}_{t=1}^T$  are convex, Holder smooth and satisfy*

$$\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \quad \|\nabla f_t(\mathbf{x}_1) - \nabla f_t(\mathbf{x}_2)\|_* \leq L \|\mathbf{x}_1 - \mathbf{x}_2\|^\alpha,$$

for some constant  $\alpha \in [0, 1]$ . Then the expected regret of Algorithm 1 satisfies

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E} \left[ \sum_{t=1}^T f_t(\mathbf{x}_t) - f_t(\mathbf{x}) \right] &\leq \eta D + \sum_{t=1}^T \frac{C}{2\eta} \mathbb{E} [\|\nabla_t - g_t\|_*^2] - \sum_{t=1}^T \frac{\eta}{2C} \mathbb{E} [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2] \\ &\quad + LT \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{1+\alpha}. \end{aligned}$$

where  $\mathbf{x}_t^\infty = \mathbb{E}[\mathbf{x}_t | g_t, f_{1:t-1}, \mathbf{x}_{1:t-1}]$  and  $\tilde{\mathbf{x}}_{t-1}^\infty = \mathbb{E}[\tilde{\mathbf{x}}_{t-1} | f_{1:t-1}, \mathbf{x}_{1:t-1}]$  and  $\tilde{\mathbf{x}}_{t-1}$  denotes the prediction in the  $t^{\text{th}}$  iteration of Algorithm 1, if guess  $g_t = 0$  was used. Here,  $\Psi_1, \Psi_2$  denote the norm compatibility constants of  $\|\cdot\|$ .

**Proof Sketch.** For any  $\mathbf{x} \in \mathcal{X}$ , we have

$$\sum_{t=1}^T \mathbb{E} [f_t(\mathbf{x}_t) - f_t(\mathbf{x})] \stackrel{(a)}{\leq} \sum_{t=1}^T \mathbb{E} [\langle \mathbf{x}_t - \mathbf{x}, \nabla_t \rangle] = \sum_{t=1}^T \mathbb{E} [\langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle] + \mathbb{E} [\langle \mathbf{x}_t^\infty - \mathbf{x}, \nabla_t \rangle],$$

---

**Algorithm 2** Nonconvex OFTPL

---

```
1: Input: Perturbation Distribution  $P_{\text{PRTB}}$ , number of samples  $m$ , number of iterations  $T$ 
2: Denote  $f_0 = 0$ 
3: for  $t = 1 \dots T$  do
4:   Let  $g_t$  be the guess for  $f_t$ 
5:   for  $j = 1 \dots m$  do
6:     Sample  $\sigma_{t,j} \sim P_{\text{PRTB}}$ 
7:      $\mathbf{x}_{t,j} \in \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} f_{0:t-1}(\mathbf{x}) + g_t(\mathbf{x}) - \sigma_{t,j}(\mathbf{x})$ 
8:   end for
9:   Let  $P_t$  be the empirical distribution over  $\{\mathbf{x}_{t,1}, \mathbf{x}_{t,2} \dots \mathbf{x}_{t,m}\}$ 
10:  Play  $\mathbf{x}_t$ , a random sample generated from  $P_t$ 
11:  Observe loss function  $f_t$ 
12: end for
```

---

where (a) follows from convexity of  $f_t$ . The first term in the RHS involves the difference between the iterates of stochastic and deterministic versions of OFTPL. To bound this term, we use the following two facts: (a) conditioned on the past randomness,  $\mathbf{x}_t - \mathbf{x}_t^\circ$  is the average of  $m$  i.i.d, bounded, mean 0 random variables and so its variance is  $O(D^2/m)$ , (b)  $f_t$  is Holder smooth. Using these two facts, we get  $\mathbb{E}[\langle \mathbf{x}_t - \mathbf{x}_t^\circ, \nabla_t \rangle] = O(D/\sqrt{m})^{1+\alpha}$ . The second term in the RHS is related to the regret of deterministic OFTPL. To bound this term, we rely on the duality between deterministic OFTPL and OFTRL (Proposition 3.1), and use a similar proof technique as the one used to derive regret bounds of OFTRL. One key distinction between OFTRL and OFTPL is that in OFTRL it is typically assumed that the regularizer  $R$  is differentiable. However, the regularizer corresponding to OFTPL need not be differentiable. As a result, the traditional Bregmann divergence used in the analysis of OFTRL is not well defined. So, our analysis instead relies on “pseudo” Bregmann divergence, which is obtained by replacing the gradient of  $R$  in Bregmann divergence with an appropriately chosen sub-gradient.  $\square$

Regret bounds that hold with high probability can be found in Appendix G. The above Theorem shows that the regret of OFTPL only depends on  $\|\nabla_t - g_t\|_*$ , which quantifies the accuracy of our guess  $g_t$ . In contrast, the regret of FTPL depends on  $\|\nabla_t\|_*$  [7]. This shows that for predictable sequences, with an appropriate choice of  $g_t$ , OFTPL can achieve better regret guarantees than FTPL. As we demonstrate in Section 5, this helps us design faster algorithms for solving minimax games.

Note that the above result is very general and holds for any absolutely continuous perturbation distribution. The key challenge in instantiating this result for any particular perturbation distribution is in showing the stability of predictions. Several past works have studied the stability of FTPL for various perturbation distributions such as uniform, exponential, Gumbel distributions [5, 7, 10]. Consequently, the above result can be used to derive tight regret bounds for all these perturbation distributions. As one particular instantiation of Theorem 4.1, we consider the special case of  $g_t = 0$  and derive regret bounds for FTPL, when the perturbation distribution is the uniform distribution over a ball centered at the origin.

**Corollary 4.1** (FTPL). *Suppose the perturbation distribution is equal to the uniform distribution over  $\{\mathbf{x} : \|\mathbf{x}\|_2 \leq (1 + d^{-1})\eta\}$ . Let  $D$  be the diameter of  $\mathcal{X}$  w.r.t  $\|\cdot\|_2$ . Then  $\mathbb{E}_\sigma[\|\sigma\|_2] = \eta$ , and the predictions of OFTPL are  $dD\eta^{-1}$ -stable w.r.t  $\|\cdot\|_2$ . Suppose, the sequence of loss functions  $\{f_t\}_{t=1}^T$  are  $G$ -Lipschitz and satisfy  $\sup_{\mathbf{x} \in \mathcal{X}} \|\nabla f_t(\mathbf{x})\|_2 \leq G$ . Moreover, suppose  $f_t$  satisfies the Holder smooth condition in Theorem 4.1 w.r.t  $\|\cdot\|_2$  norm. Then the expected regret of Algorithm 1 with guess  $g_t = 0$ , satisfies*

$$\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E} \left[ \sum_{t=1}^T f_t(\mathbf{x}_t) - f_t(\mathbf{x}) \right] \leq \eta D + \frac{dDG^2T}{2\eta} + LT \left( \frac{D}{\sqrt{m}} \right)^{1+\alpha}.$$

This recovers the regret bounds of FTPL for general convex loss functions derived by [10].

## 4.2 Online Nonconvex Learning

We now study OFTPL in the nonconvex setting. In this setting, we assume the sequence of loss functions belong to some function class  $\mathcal{F}$  containing real-valued measurable functions on  $\mathcal{X}$ . Some

popular choices for  $\mathcal{F}$  include the set of Lipschitz functions, the set of bounded functions. The OFTPL algorithm in this setting is described in Algorithm 2. Similar to the convex case, we first sample random perturbation functions  $\{\sigma_{t,j}\}_{j=1}^m$  from some distribution  $P_{\text{PRTB}}$ . Some examples of perturbation functions that have been considered in the past include  $\sigma_{t,j}(\mathbf{x}) = \langle \bar{\sigma}_{t,j}, \mathbf{x} \rangle$ , for some random vector  $\bar{\sigma}_{t,j}$  sampled from exponential or uniform distributions [11, 18]. Another popular choice for  $\sigma_{t,j}$  is the Gumbel process, which results in the continuous exponential weights algorithm [21]. Letting  $g_t$  be our guess of loss function  $f_t$  at the beginning of round  $t$ , the learner first computes  $\mathbf{x}_{t,j}$  as  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \sum_{i=1}^{t-1} f_i(\mathbf{x}) + g_t(\mathbf{x}) - \sigma_{t,j}(\mathbf{x})$ . We assume access to an optimization oracle which computes a minimizer of this problem. We often refer to this oracle as the *perturbed best response* oracle. Let  $P_t$  denote the empirical distribution of  $\{\mathbf{x}_{t,j}\}_{j=1}^m$ . The learner then plays an  $\mathbf{x}_t$  which is sampled from  $P_t$ . Algorithm 2 describes this procedure. We note that for the online learning problem,  $m = 1$  suffices, as the expected loss suffered by the learner in each round is independent of  $m$ ; that is  $\mathbb{E}[f_t(\mathbf{x}_t)] = \mathbb{E}[f_t(\mathbf{x}_{t,1})]$ . However, the choice of  $m$  affects the rate of convergence when Algorithm 2 is used for solving nonconvex nonconcave minimax games.

Before we present the regret bounds, we introduce the *dual space* associated with  $\mathcal{F}$ . Let  $\|\cdot\|_{\mathcal{F}}$  be a seminorm associated with  $\mathcal{F}$ . For example, when  $\mathcal{F}$  is the set of Lipschitz functions,  $\|\cdot\|_{\mathcal{F}}$  is the Lipschitz seminorm. Various choices of  $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$  induce various distance metrics on  $\mathcal{P}$ , the set of all probability distributions on  $\mathcal{X}$ . We let  $\gamma_{\mathcal{F}}$  denote the Integral Probability Metric (IPM) induced by  $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ , which is defined as

$$\gamma_{\mathcal{F}}(P, Q) = \sup_{f \in \mathcal{F}, \|f\|_{\mathcal{F}} \leq 1} \left| \mathbb{E}_{\mathbf{x} \sim P} [f(\mathbf{x})] - \mathbb{E}_{\mathbf{x} \sim Q} [f(\mathbf{x})] \right|.$$

We often refer to  $(\mathcal{P}, \gamma_{\mathcal{F}})$  as the dual space of  $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ . When  $\mathcal{F}$  is the set of Lipschitz functions and when  $\|\cdot\|_{\mathcal{F}}$  is the Lipschitz seminorm,  $\gamma_{\mathcal{F}}$  is the Wasserstein distance. Table 1 in Appendix E.1 presents examples of  $\gamma_{\mathcal{F}}$  induced by some popular function spaces. Similar to the convex case, the regret bounds in the nonconvex setting depend on the stability of predictions of OFTPL.

**Definition 4.2** (Stability). Suppose the perturbation function  $\sigma(\mathbf{x})$  is sampled from  $P_{\text{PRTB}}$ . For any  $f \in \mathcal{F}$ , define random variable  $\mathbf{x}_f(\sigma)$  as  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) - \sigma(\mathbf{x})$ . Let  $\nabla\Phi(f)$  denote the distribution of  $\mathbf{x}_f(\sigma)$ . The predictions of OFTPL are said to be  $\beta$ -stable w.r.t  $\|\cdot\|_{\mathcal{F}}$  if

$$\forall f, g \in \mathcal{F} \quad \gamma_{\mathcal{F}}(\nabla\Phi(f), \nabla\Phi(g)) \leq \beta \|f - g\|_{\mathcal{F}}.$$

**Theorem 4.2.** Suppose the sequence of loss functions  $\{f_t\}_{t=1}^T$  belong to  $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ . Suppose the perturbation distribution  $P_{\text{PRTB}}$  is such that  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) - \sigma(\mathbf{x})$  has a unique minimizer with probability one, for any  $f \in \mathcal{F}$ . Let  $\mathcal{P}$  be the set of probability distributions over  $\mathcal{X}$ . Define the diameter of  $\mathcal{P}$  as  $D = \sup_{P_1, P_2 \in \mathcal{P}} \gamma_{\mathcal{F}}(P_1, P_2)$ . Let  $\eta = \mathbb{E}[\|\sigma\|_{\mathcal{F}}]$ . Suppose the predictions of OFTPL are  $C\eta^{-1}$ -stable w.r.t  $\|\cdot\|_{\mathcal{F}}$ , for some constant  $C$  that depends on  $\mathcal{X}$ . Then the expected regret of Algorithm 2 satisfies

$$\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E} \left[ \sum_{t=1}^T f_t(\mathbf{x}_t) - f_t(\mathbf{x}) \right] \leq \eta D + \sum_{t=1}^T \frac{C}{2\eta} \mathbb{E} [\|f_t - g_t\|_{\mathcal{F}}^2] - \sum_{t=1}^T \frac{\eta}{2C} \mathbb{E} [\gamma_{\mathcal{F}}(P_t^{\infty}, \tilde{P}_{t-1}^{\infty})^2],$$

where  $P_t^{\infty} = \mathbb{E}[P_t | g_t, f_{1:t-1}, P_{1:t-1}]$ ,  $\tilde{P}_t^{\infty} = \mathbb{E}[\tilde{P}_{t-1} | f_{1:t-1}, P_{1:t-1}]$  and  $\tilde{P}_{t-1}$  is the empirical distribution computed in the  $t^{\text{th}}$  iteration of Algorithm 2, if guess  $g_t = 0$  was used.

**Proof Sketch.** The proof uses similar arguments as in the proof of Theorem 4.1. For any  $P \in \mathcal{P}$ , we have

$$\mathbb{E}[f_t(\mathbf{x}_t) - f_t(P)] = \mathbb{E}[f_t(P_t) - f_t(P_t^{\infty})] + \mathbb{E}[f_t(P_t^{\infty}) - f_t(P)] \stackrel{(a)}{=} \mathbb{E}[f_t(P_t^{\infty}) - f_t(P)],$$

where (a) follows from the fact that  $\mathbb{E}[f_t(P_t) - f_t(P_t^{\infty}) | g_t, f_{1:t-1}, P_{1:t-1}] = 0$ . The RHS is related to the regret of deterministic OFTPL. In the convex case, to bound this term, we relied on duality between OFTRL and OFTPL. However, in the nonconvex case, we can not take this route as there are no known analogs of Fenchel duality for infinite dimensional function spaces. As a result, more careful analysis is needed to obtain the regret bounds. Our analysis mimics the arguments made in the convex case, albeit without explicitly relying on duality theory.  $\square$

As in the convex case, the key challenge in instantiating the above result for any particular perturbation distribution is in showing the stability of predictions. In a recent work, [11] consider

linear perturbation functions  $\sigma(\mathbf{x}) = \langle \bar{\sigma}, \mathbf{x} \rangle$ , for  $\bar{\sigma}$  sampled from exponential distribution, and show stability of FTPL. We now instantiate the above Theorem for this setting.

**Corollary 4.2.** *Consider the setting of Theorem 4.2. Let  $\mathcal{F}$  be the set of Lipschitz functions and  $\|\cdot\|_{\mathcal{F}}$  be the Lipschitz seminorm, which is defined as  $\|f\|_{\mathcal{F}} = \sup_{\mathbf{x} \neq \mathbf{y} \text{ in } \mathcal{X}} |f(\mathbf{x}) - f(\mathbf{y})| / \|\mathbf{x} - \mathbf{y}\|_1$ . Suppose the perturbation function is such that  $\sigma(\mathbf{x}) = \langle \bar{\sigma}, \mathbf{x} \rangle$ , where  $\bar{\sigma} \in \mathbb{R}^d$  is a random vector whose entries are sampled independently from  $\text{Exp}(\eta)$ . Then  $\mathbb{E}_{\sigma} [\|\sigma\|_{\mathcal{F}}] = \eta \log d$ , and the predictions of OFTPL are  $O(d^2 D \eta^{-1})$ -stable w.r.t  $\|\cdot\|_{\mathcal{F}}$ . Moreover, the expected regret of Algorithm 2 is upper bounded by  $O\left(\eta D \log d + \sum_{t=1}^T \frac{d^2 D}{\eta} \mathbb{E} [\|f_t - g_t\|_{\mathcal{F}}^2] - \sum_{t=1}^T \frac{\eta}{d^2 D} \mathbb{E} \left[ \gamma_{\mathcal{F}}(P_t^{\circ}, \tilde{P}_{t-1}^{\circ})^2 \right]\right)$ .*

We note that the above regret bounds are tighter than the regret bounds of [11], where the authors show that the regret of OFTPL is bounded by  $O\left(\eta D \log d + \sum_{t=1}^T \frac{d^2 D}{\eta} \mathbb{E} [\|f_t - g_t\|_{\mathcal{F}}^2]\right)$ . These tighter bounds help us design faster algorithms for solving minimax games in the nonconvex setting (see Section 5 for a more detailed discussion).

## 5 Minimax Games

We now consider the problem of solving minimax games of the following form

$$\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbf{y} \in \mathcal{Y}} f(\mathbf{x}, \mathbf{y}). \quad (1)$$

Nash equilibria of such games can be computed by playing two online learning algorithms against each other [6, 7]. In this work, we study the algorithm where both the players employ OFTPL to decide their actions in each round. For convex-concave games, both the players use the OFTPL algorithm described in Algorithm 1 (see Algorithm 3 in Appendix D). The following theorem derives the rate of convergence of this algorithm to a Nash equilibrium (NE).

**Theorem 5.1.** *Consider the minimax game in Equation (1). Suppose both the domains  $\mathcal{X}, \mathcal{Y}$  are compact subsets of  $\mathbb{R}^d$ , with diameter  $D = \max\{\sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}} \|\mathbf{x}_1 - \mathbf{x}_2\|_2, \sup_{\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{Y}} \|\mathbf{y}_1 - \mathbf{y}_2\|_2\}$ . Suppose  $f$  is convex in  $\mathbf{x}$ , concave in  $\mathbf{y}$  and is smooth w.r.t  $\|\cdot\|_2$*

$$\|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{x}} f(\mathbf{x}', \mathbf{y}')\|_2 + \|\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{y}} f(\mathbf{x}', \mathbf{y}')\|_2 \leq L \|\mathbf{x} - \mathbf{x}'\|_2 + L \|\mathbf{y} - \mathbf{y}'\|_2.$$

*Suppose Algorithm 3 is used to solve the minimax game. Suppose the perturbation distributions used by both the players are the same and equal to the uniform distribution over  $\{\mathbf{x} : \|\mathbf{x}\|_2 \leq (1 + d^{-1})\eta\}$ . Suppose the guesses used by  $\mathbf{x}, \mathbf{y}$  players in the  $t^{\text{th}}$  iteration are  $\nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1}), \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1})$ , where  $\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1}$  denote the predictions of  $\mathbf{x}, \mathbf{y}$  players in the  $t^{\text{th}}$  iteration, if guess  $g_t = 0$  was used. If Algorithm 3 is run with  $\eta = 6dD(L + 1), m = T$ , then the iterates  $\{(\mathbf{x}_t, \mathbf{y}_t)\}_{t=1}^T$  satisfy*

$$\sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ f \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t, \mathbf{y} \right) - f \left( \mathbf{x}, \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t \right) \right] = O \left( \frac{dD^2(L + 1)}{T} \right).$$

Rates of convergence which hold with high probability can be found in Appendix G. We note that Theorem 5.1 can be extended to more general noise distributions and settings where gradients of  $f$  are Holder smooth w.r.t non-Euclidean norms, and  $\mathcal{X}, \mathcal{Y}$  lie in spaces of different dimensions (see Theorem D.1 in Appendix). We now discuss the above result.

- Theorem 5.1 shows that for smooth convex-concave games, Algorithm 3 converges to a NE at  $O(T^{-1})$  rate using  $4T^2$  calls to the linear optimization oracle. Moreover, the algorithm runs in  $T$  iterations, with each iteration making  $4T$  parallel calls to the optimization oracle. In contrast, FTPL makes  $2T^3$  calls to the linear optimization oracle to achieve  $O(T^{-1})$  rates of convergence and runs for  $T^2$  iterations, with each iteration making  $2T$  parallel calls to the optimization oracle. This can be obtained by setting  $m = \sqrt{T}, \alpha = 1$ , and  $\eta = O(\sqrt{T})$  in Corollary 4.1.
- The Frank-Wolfe technique of He and Harchaoui [15] achieves the same convergence rates as our algorithm; that is, it achieves  $O(T^{-1})$  rates using  $T^2$  calls to the linear optimization oracle. However, unlike [15], our algorithm is parallelizable and can be run in  $T$  iterations.
- He and Harchaoui [15] achieve dimension free convergence rates in the Euclidean setting, where the smoothness is measured w.r.t  $\|\cdot\|_2$  norm. In contrast, the rates of convergence of our algorithm depend on the dimension. We believe the dimension dependence in the rates can be removed by appropriately choosing the perturbation distributions based on domains  $\mathcal{X}, \mathcal{Y}$  (see Appendix F).

- Note that OFTRL also achieves  $O(T^{-1})$  rates of convergence after  $T$  iterations. However, each iteration of OFTRL involves optimization of a non-linear convex function over the domains  $\mathcal{X}, \mathcal{Y}$ , which can be quite expensive in practice.

We now consider the more general nonconvex-nonconcave games. In this case, both the players use the nonconvex OFTPL algorithm described in Algorithm 2 to choose their actions. Instead of generating a single sample from the empirical distribution  $P_t$  computed in  $t^{\text{th}}$  iteration of Algorithm 2, the players now play the entire distribution  $P_t$  (see Algorithm 4 in Appendix E). Letting  $\{P_t\}_{t=1}^T, \{Q_t\}_{t=1}^T$ , be the sequence of iterates generated by the  $\mathbf{x}$  and  $\mathbf{y}$  players, the following theorem shows that  $\left(\frac{1}{T} \sum_{t=1}^T P_t, \frac{1}{T} \sum_{t=1}^T Q_t\right)$  converges to a NE.

**Theorem 5.2.** *Consider the minimax game in Equation (1). Suppose the domains  $\mathcal{X}, \mathcal{Y}$  are compact subsets of  $\mathbb{R}^d$  with diameter  $D = \max\{\sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}} \|\mathbf{x}_1 - \mathbf{x}_2\|_1, \sup_{\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{Y}} \|\mathbf{y}_1 - \mathbf{y}_2\|_1\}$ . Suppose  $f$  is Lipschitz w.r.t  $\|\cdot\|_1$  and satisfies*

$$\max \left\{ \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})\|_{\infty}, \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \|\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})\|_{\infty} \right\} \leq G.$$

Moreover, suppose  $f$  satisfies the following smoothness property

$$\|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{x}} f(\mathbf{x}', \mathbf{y}')\|_{\infty} + \|\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{y}} f(\mathbf{x}', \mathbf{y}')\|_{\infty} \leq L\|\mathbf{x} - \mathbf{x}'\|_1 + L\|\mathbf{y} - \mathbf{y}'\|_1.$$

Suppose both  $\mathbf{x}$  and  $\mathbf{y}$  players use Algorithm 4 to solve the game with linear perturbation functions  $\sigma(\mathbf{z}) = \langle \bar{\sigma}, \mathbf{z} \rangle$ , where  $\bar{\sigma} \in \mathbb{R}^d$  is such that each of its entries is sampled independently from  $\text{Exp}(\eta)$ . Suppose the guesses used by  $\mathbf{x}$  and  $\mathbf{y}$  players in the  $t^{\text{th}}$  iteration are  $f(\cdot, \tilde{Q}_{t-1}), f(\tilde{P}_{t-1}, \cdot)$ , where  $\tilde{P}_{t-1}, \tilde{Q}_{t-1}$  denote the predictions of  $\mathbf{x}, \mathbf{y}$  players in the  $t^{\text{th}}$  iteration, if guess  $g_t = 0$  was used. If Algorithm 4 is run with  $\eta = 10d^2D(L+1), m = T$ , then the iterates  $\{(P_t, Q_t)\}_{t=1}^T$  satisfy

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ f \left( \frac{1}{T} \sum_{t=1}^T P_t, \mathbf{y} \right) - f \left( \mathbf{x}, \frac{1}{T} \sum_{t=1}^T Q_t \right) \right] &= O \left( \frac{d^2 D^2 (L+1) \log d}{T} \right) \\ &+ O \left( \min \left\{ D^2 L, \frac{d^2 G^2 \log T}{LT} \right\} \right). \end{aligned}$$

More general versions of the Theorem, which consider other function classes and general perturbation distributions, can be found in Appendix E. We now discuss the above result.

- Theorem 5.2 shows that Algorithm 4 converges to a NE at  $\tilde{O}(T^{-1})$  rate using  $T^2$  calls to the perturbed best response oracle. This matches the rates of convergence of FTPL [11]. However, the key advantage of our algorithm is that it is highly parallelizable and runs in  $T$  iterations, in contrast to FTPL, which runs in  $T^2$  iterations.
- As previously stated, [11] also study OFTPL for non-convex losses and upper bound its regret as  $O\left(\eta D \log d + \sum_{t=1}^T \frac{d^2 D}{\eta} \mathbb{E}[\|f_t - g_t\|_{\mathcal{F}}^2]\right)$ . However, by relying on this regret bound, we can only obtain  $O(T^{-3/4})$  convergence rates, even if we set  $m = \infty$  and run the algorithm for  $T$  iterations (see Appendix E.4).

## 6 Conclusion

We studied an optimistic variant of FTPL which achieves better regret guarantees when the sequence of loss functions is predictable. As one specific application of our algorithm, we considered the problem of solving minimax games. For solving convex-concave games, our algorithm requires access to a linear optimization oracle and for nonconvex-nonconcave games our algorithm requires access to a more powerful perturbed best response oracle. In both these settings, our algorithm achieves  $O(T^{-1/2})$  convergence rates using  $T$  calls to the oracles. Moreover, our algorithm runs in  $O(T^{1/2})$  iterations, with each iteration making  $O(T^{1/2})$  parallel calls to the optimization oracle. We believe our improved algorithms for solving minimax games are useful in a number of modern machine learning applications such as training of GANs, adversarial training, which involve solving nonconvex-nonconcave minimax games and often deal with huge datasets.

## Broader Impact

Many problems in machine learning and statistics have a game theoretic component to them. Two popular modern applications that illustrate this are adversarial training and density estimation using IPMs. These applications often involve solving large-scale minimax games. In many cases these games are nonconvex-nonconcave, which makes it even more harder to find a NE. Existing approaches for solving these games have mostly relied on algorithms from online convex learning. However, such algorithms are not guaranteed to converge to a NE of nonconvex-nonconcave games. As a result, there is a need for faster algorithms for provably solving large-scale nonconvex-nonconcave games. Our work takes a first step towards this goal by proposing fast parallelizable algorithms which provably converge to a NE in both convex and nonconvex settings.

## References

- [1] Yoav Freund and Robert E Schapire. Game theory, on-line prediction and boosting. In *COLT*, volume 96, pages 325–332. Citeseer, 1996.
- [2] Robert S Chen, Brendan Lucier, Yaron Singer, and Vasilis Syrgkanis. Robust optimization for non-convex objectives. In *Advances in Neural Information Processing Systems*, pages 4705–4714, 2017.
- [3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [4] H Brendan McMahan. A survey of algorithms and analysis for adaptive online learning. *The Journal of Machine Learning Research*, 18(1):3117–3166, 2017.
- [5] Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.
- [6] Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.
- [7] Elad Hazan. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.
- [8] Dan Garber and Elad Hazan. Playing non-linear games with linear oracles. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 420–428. IEEE, 2013.
- [9] Gauthier Gidel, Tony Jebara, and Simon Lacoste-Julien. Frank-wolfe algorithms for saddle point problems. *arXiv preprint arXiv:1610.07797*, 2016.
- [10] Elad Hazan and Edgar Minasyan. Faster projection-free online learning. *CoRR*, abs/2001.11568, 2020. URL <https://arxiv.org/abs/2001.11568>.
- [11] Arun Sai Suggala and Praneeth Netrapalli. Online non-convex learning: Following the perturbed leader is optimal. In Aryeh Kontorovich and Gergely Neu, editors, *Proceedings of the 31st International Conference on Algorithmic Learning Theory*, volume 117 of *Proceedings of Machine Learning Research*, pages 845–861, San Diego, California, USA, 08 Feb–11 Feb 2020. PMLR. URL <http://proceedings.mlr.press/v117/suggala20a.html>.
- [12] Reiner Horst and Panos M Pardalos. *Handbook of global optimization*, volume 2. Springer Science & Business Media, 2013.
- [13] Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. *arXiv preprint arXiv:1208.3728*, 2012.
- [14] Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. In *Advances in Neural Information Processing Systems*, pages 3066–3074, 2013.
- [15] Niao He and Zaid Harchaoui. Semi-proximal mirror-prox for nonsmooth composite minimization. In *Advances in Neural Information Processing Systems*, pages 3411–3419, 2015.
- [16] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.
- [17] Elad Hazan and Satyen Kale. Projection-free online learning. *arXiv preprint arXiv:1206.4657*, 2012.

- [18] Naman Agarwal, Alon Gonen, and Elad Hazan. Learning in non-convex games with an optimization oracle. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 18–29, Phoenix, USA, 25–28 Jun 2019. PMLR. URL <http://proceedings.mlr.press/v99/agarwal19a.html>.
- [19] Josef Hofbauer and William H Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002.
- [20] Jacob Abernethy, Chansoo Lee, and Ambuj Tewari. Perturbation techniques in online learning and optimization. *Perturbations, Optimization, and Statistics*, page 233, 2016.
- [21] Chris J Maddison, Daniel Tarlow, and Tom Minka. A\* sampling. In *Advances in Neural Information Processing Systems*, pages 3086–3094, 2014.
- [22] R Tyrrell Rockafellar. *Convex analysis*. Number 28. Princeton university press, 1970.
- [23] Dimitri P Bertsekas. Stochastic optimization problems with nondifferentiable cost functionals. *Journal of Optimization Theory and Applications*, 12(2):218–231, 1973.
- [24] Shai Shalev-Shwartz. Thesis submitted for the degree of “doctor of philosophy”. 2007.
- [25] Eduardo H Zarantonello. Dense single-valuedness of monotone operators. *Israel Journal of Mathematics*, 15(2):158–166, 1973.
- [26] Daniel Hsu, Sham Kakade, Tong Zhang, et al. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17, 2012.
- [27] Chi Jin, Praneeth Netrapalli, Rong Ge, Sham M Kakade, and Michael I Jordan. A short note on concentration inequalities for random vectors with subgaussian norm. *arXiv preprint arXiv:1902.03736*, 2019.
- [28] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.
- [29] Sham Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. On the duality of strong convexity and strong smoothness: Learning applications and matrix regularization. *Unpublished Manuscript*, <http://ttic.uchicago.edu/shai/papers/KakadeShalevTewari09.pdf>, 2(1), 2009.

## Contents

<b>A Dual view of Perturbations as Regularization</b>	<b>13</b>
A.1 Proof of Theorem 3.1 . . . . .	13
<b>B Online Convex Learning</b>	<b>14</b>
B.1 Proof of Theorem 4.1 . . . . .	14
B.1.1 Notation . . . . .	14
B.1.2 Main Argument . . . . .	14
B.2 Proof of Corollary 4.1 . . . . .	16
<b>C Online Nonconvex Learning</b>	<b>17</b>
C.1 Proof of Theorem 4.2 . . . . .	17
C.1.1 Notation . . . . .	17
C.1.2 Intermediate Results . . . . .	17
C.1.3 Main Argument . . . . .	20
C.2 Proof of Corollary 4.2 . . . . .	22
C.2.1 Intermediate Results . . . . .	22
C.2.2 Main Argument . . . . .	22
<b>D Convex-Concave Games</b>	<b>23</b>
D.1 General Result . . . . .	23
D.2 Proof of Theorem 5.1 . . . . .	26
<b>E Nonconvex-Nonconcave Games</b>	<b>26</b>
E.1 Primal Dual Spaces . . . . .	26
E.2 General Result . . . . .	27
E.3 Proof of Theorem 5.2 . . . . .	30
E.4 Regret Bounds of OFTPL in [11] . . . . .	31
<b>F Choice of Perturbation Distributions</b>	<b>32</b>
<b>G High Probability Bounds</b>	<b>33</b>
G.1 Online Convex Learning . . . . .	33
G.2 Convex-Concave Games . . . . .	35
G.3 Nonconvex-Nonconcave Games . . . . .	37
<b>H Background on Convex Analysis</b>	<b>40</b>

## A Dual view of Perturbations as Regularization

### A.1 Proof of Theorem 3.1

We first define a convex function  $\Psi : \mathbb{R}^d \rightarrow \mathbb{R}$  as

$$\Psi(f) = \mathbb{E}_\sigma \left[ \sup_{\mathbf{x} \in \mathcal{X}} \langle f + \sigma, \mathbf{x} \rangle \right] = \mathbb{E}_\sigma \left[ \sup_{\mathbf{x} \in \mathcal{X}} \langle f + \sigma, \mathbf{x} \rangle \right],$$

where perturbation  $\sigma$  follows probability distribution  $P_{\text{PRTB}}$  which is absolutely continuous w.r.t the Lebesgue measure. For our choice of  $P_{\text{PRTB}}$ , we now show that  $\Psi$  is differentiable. Consider the function  $\psi(g) = \sup_{\mathbf{x} \in \mathcal{X}} \langle g, \mathbf{x} \rangle$ . Since  $\psi(g)$  is a proper convex function, we know that it is differentiable almost everywhere, except on a set of Lebesgue measure 0 [see Theorem 25.5 of 22]. Moreover, it is easy to verify that  $\operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} \langle g, \mathbf{x} \rangle \in \partial\psi(g)$ . These two observations, together with the fact that  $P_{\text{PRTB}}$  is absolutely continuous, show that the sup expression inside the expectation of  $\Psi$  has a unique maximizer with probability one.

Since the sup expression inside the expectation has a unique maximizer with probability 1, we can swap the expectation and gradient to obtain [see Proposition 2.2 of 23]

$$\nabla\Psi(f) = \mathbb{E}_\sigma \left[ \operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} \langle f + \sigma, \mathbf{x} \rangle \right]. \quad (2)$$

Note that  $\nabla\Psi$  is related to the prediction of deterministic version of FTPL. Specifically,  $\nabla\Psi(-\nabla_{1:t-1})$  is the prediction of deterministic FTPL in the  $t^{\text{th}}$  iteration. We now show that  $\nabla\Psi(f) = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle -f, \mathbf{x} \rangle + R(\mathbf{x})$ , for some convex function  $R$ .

Since all differentiable functions are closed,  $\Psi(f)$  is a proper, closed and differentiable convex function over  $\mathbb{R}^d$ . Let  $R(\mathbf{x})$  denote the Fenchel conjugate of  $\Psi(f)$

$$R(\mathbf{x}) = \sup_{f \in \operatorname{dom}(\Psi)} \langle \mathbf{x}, f \rangle - \Psi(f),$$

where  $\operatorname{dom}(\Psi)$  denotes the domain of  $\Psi$ . Following Theorem H.1 (see Appendix H),  $\Psi(f)$  is the Fenchel conjugate of  $R(\mathbf{x})$

$$\Psi(f) = \sup_{\mathbf{x} \in \operatorname{dom}(R)} \langle f, \mathbf{x} \rangle - R(\mathbf{x}).$$

Furthermore, from Theorem H.2 we have

$$\nabla\Psi(f) = \operatorname{argmax}_{\mathbf{x} \in \operatorname{dom}(R)} \langle f, \mathbf{x} \rangle - R(\mathbf{x}).$$

We now show that the domain of  $R$  is a subset of  $\mathcal{X}$ . This, together with the previous two equations, would then immediately imply

$$\Psi(f) = \sup_{\mathbf{x} \in \mathcal{X}} \langle f, \mathbf{x} \rangle - R(\mathbf{x}), \quad (3)$$

$$\nabla\Psi(f) = \operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} \langle f, \mathbf{x} \rangle - R(\mathbf{x}). \quad (4)$$

From Theorem H.4, we know that the domain of  $R$  satisfies

$$\operatorname{ri}(\operatorname{dom}(R)) \subseteq \operatorname{range}\nabla\Psi \subseteq \operatorname{dom}(R),$$

where  $\operatorname{ri}(A)$  denotes the relative interior of a set  $A$ . Moreover, from the definition of  $\nabla\Psi(f)$  in Equation (2), we have  $\operatorname{range}\nabla\Psi \subseteq \mathcal{X}$ . Combining these two properties, we can show that one of the following statements is true

$$\begin{aligned} \operatorname{ri}(\operatorname{dom}(R)) \subseteq \operatorname{range}\nabla\Psi \subseteq \mathcal{X} \subseteq \operatorname{dom}(R), \\ \operatorname{ri}(\operatorname{dom}(R)) \subseteq \operatorname{range}\nabla\Psi \subseteq \operatorname{dom}(R) \subseteq \mathcal{X}. \end{aligned}$$

Suppose the first statement is true. Since  $\mathcal{X}$  is a compact set, it is easy to see that  $\mathcal{X} = \operatorname{dom}(R)$ . If the second statement is true, then  $\operatorname{dom}(R) \subseteq \mathcal{X}$ . Together, these two statements imply  $\operatorname{dom}(R) \subseteq \mathcal{X}$ .

**Connecting back to FTPL.** We now connect the above results to FTPL. From Equation (2), we know that the prediction at iteration  $t$  of deterministic FTPL is equal to  $\nabla\Psi(-\nabla_{1:t-1})$ . From Equation (4),  $\nabla\Psi(-\nabla_{1:t-1})$  is defined as

$$\mathbf{x}_t = \nabla\Psi(-\nabla_{1:t-1}) = \operatorname{argmax}_{\mathbf{x} \in \mathcal{X}} \langle -\nabla_{1:t-1}, \mathbf{x} \rangle - R(\mathbf{x}).$$

This shows that

$$\mathbf{x}_t = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle \nabla_{1:t-1}, \mathbf{x} \rangle + R(\mathbf{x}).$$

So the prediction of FTPL can also be obtained using FTRL for some convex regularizer  $R(\mathbf{x})$ . Finally, to show that  $-\nabla_{1:t-1} \in \partial R(\mathbf{x}_t)$ ,  $\mathbf{x}_t = \partial R^{-1}(-\nabla_{1:t-1})$ , we rely on Theorem H.3. Since  $\mathbf{x}_t = \nabla\Psi(-\nabla_{1:t-1})$ , from Theorem H.3, we have

$$-\nabla_{1:t-1} \in \partial R(\mathbf{x}_t), \quad \mathbf{x}_t = \nabla\Psi(-\nabla_{1:t-1}) = \partial R^{-1}(-\nabla_{1:t-1}),$$

where  $\partial R^{-1}$  is the inverse of  $\partial R$  in the sense of multivalued mappings. Note that, even though  $\partial R$  can be a multivalued mapping, its inverse  $\partial R^{-1} = \nabla\Psi$  is a singlevalued mapping (this follows from differentiability of  $\Psi$ ). This finishes the proof of the Theorem.

## B Online Convex Learning

### B.1 Proof of Theorem 4.1

Before presenting the proof of the Theorem, we introduce some notation.

#### B.1.1 Notation

We define functions  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $R : \mathbb{R}^d \rightarrow \mathbb{R}$  as follows

$$\Phi(f) = \mathbb{E}_\sigma \left[ \inf_{\mathbf{x} \in \mathcal{X}} \langle f - \sigma, \mathbf{x} \rangle \right], \quad R(\mathbf{x}) = \sup_{f \in \mathbb{R}^d} \langle f, \mathbf{x} \rangle + \Phi(-f).$$

Note that  $\Phi$  is related to the function  $\Psi$  defined in the proof of Proposition 3.1. To be precise,  $\Psi(f) = -\Phi(-f)$ . Moreover,  $R(\mathbf{x})$  is the Fenchel conjugate of  $\Psi$ . For our choice of perturbation distribution,  $\Psi$  is differentiable (see proof of Proposition 3.1). This implies  $\Phi$  is also differentiable with gradient  $\nabla\Phi$  defined as

$$\nabla\Phi(f) = \mathbb{E}_\sigma \left[ \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle f - \sigma, \mathbf{x} \rangle \right].$$

Note that  $\nabla\Phi$  is the prediction of deterministic version of FTPL. In Proposition 3.1 we showed that

$$\nabla\Phi(f) = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle f, \mathbf{x} \rangle + R(\mathbf{x}).$$

#### B.1.2 Main Argument

Since  $\mathbf{x}_t^\infty$  is the prediction of deterministic version of FTPL, following FTPL-FTRL duality proved in Proposition 3.1,  $\mathbf{x}_t^\infty$  can equivalently be written as

$$\mathbf{x}_t^\infty = \nabla\Phi(\nabla_{1:t-1} + g_t) = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle \nabla_{1:t-1} + g_t, \mathbf{x} \rangle + R(\mathbf{x}).$$

Similarly,  $\tilde{\mathbf{x}}_t^\infty$  can be written as

$$\tilde{\mathbf{x}}_t^\infty = \nabla\Phi(\nabla_{1:t}) = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle \nabla_{1:t}, \mathbf{x} \rangle + R(\mathbf{x}).$$

We use the notation  $\nabla_{1:0} = 0$ . So  $\tilde{\mathbf{x}}_0^\infty, \mathbf{x}_1^\infty$  are equal to  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} R(\mathbf{x})$ . From the first order optimality conditions, we have

$$-\nabla_{1:t-1} - g_t \in \partial R(\mathbf{x}_t^\infty), \quad -\nabla_{1:t} \in \partial R(\tilde{\mathbf{x}}_t^\infty).$$

Define functions  $B(\cdot, \mathbf{x}_t^\infty), B(\cdot, \tilde{\mathbf{x}}_t^\infty)$  for any  $t \in [T]$  as

$$\begin{aligned} B(\mathbf{x}, \mathbf{x}_t^\infty) &= R(\mathbf{x}) - R(\mathbf{x}_t^\infty) + \langle \nabla_{1:t-1} + g_t, \mathbf{x} - \mathbf{x}_t^\infty \rangle, \\ B(\mathbf{x}, \tilde{\mathbf{x}}_t^\infty) &= R(\mathbf{x}) - R(\tilde{\mathbf{x}}_t^\infty) + \langle \nabla_{1:t}, \mathbf{x} - \tilde{\mathbf{x}}_t^\infty \rangle. \end{aligned}$$

From the stability of predictions of OFTPL we know that:  $\|\nabla\Phi(g_1) - \nabla\Phi(g_2)\| \leq C\eta^{-1}\|g_1 - g_2\|_*$ . Following our connection between  $\Psi, \Phi$ , this implies  $\|\nabla\Psi(g_1) - \nabla\Psi(g_2)\| \leq C\eta^{-1}\|g_1 - g_2\|_*$ . This implies the following smoothness condition on  $\Psi$  [see Lemma 15 of 24]

$$\Psi(g_2) \leq \Psi(g_1) + \langle \nabla\Psi(g_1), g_2 - g_1 \rangle + \frac{C\eta^{-1}}{2}\|g_1 - g_2\|_*^2.$$

Since  $\Psi$  is  $C\eta^{-1}$ -smooth w.r.t  $\|\cdot\|_*$ , following duality between strong convexity and strong smoothness properties (see Theorem H.5), we can infer that  $R$  is  $C^{-1}\eta$ -strongly convex w.r.t  $\|\cdot\|$  norm and satisfies

$$B(\mathbf{x}, \mathbf{x}_t^\infty) \geq \frac{\eta}{2C}\|\mathbf{x} - \mathbf{x}_t^\infty\|^2, \quad B(\mathbf{x}, \tilde{\mathbf{x}}_t^\infty) \geq \frac{\eta}{2C}\|\mathbf{x} - \tilde{\mathbf{x}}_t^\infty\|^2.$$

We now go ahead and bound the regret of the learner. For any  $\mathbf{x} \in \mathcal{X}$ , we have

$$\begin{aligned} f_t(\mathbf{x}_t) - f_t(\mathbf{x}) &\stackrel{(a)}{\leq} \langle \mathbf{x}_t - \mathbf{x}, \nabla_t \rangle = \langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle + \langle \mathbf{x}_t^\infty - \mathbf{x}, \nabla_t \rangle \\ &= \langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle + \langle \mathbf{x}_t^\infty - \tilde{\mathbf{x}}_t^\infty, \nabla_t - g_t \rangle + \langle \mathbf{x}_t^\infty - \tilde{\mathbf{x}}_t^\infty, g_t \rangle \\ &\quad + \langle \tilde{\mathbf{x}}_t^\infty - \mathbf{x}, \nabla_t \rangle \\ &\leq \langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle + \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_t^\infty\| \|\nabla_t - g_t\|_* + \langle \mathbf{x}_t^\infty - \tilde{\mathbf{x}}_t^\infty, g_t \rangle \\ &\quad + \langle \tilde{\mathbf{x}}_t^\infty - \mathbf{x}, \nabla_t \rangle, \end{aligned}$$

where (a) follows from convexity of  $f$ . Next, a simple calculation shows that

$$\begin{aligned} \langle \mathbf{x}_t^\infty - \tilde{\mathbf{x}}_t^\infty, g_t \rangle &= B(\tilde{\mathbf{x}}_t^\infty, \tilde{\mathbf{x}}_{t-1}^\infty) - B(\tilde{\mathbf{x}}_t^\infty, \mathbf{x}_t^\infty) - B(\mathbf{x}_t^\infty, \tilde{\mathbf{x}}_{t-1}^\infty) \\ \langle \tilde{\mathbf{x}}_t^\infty - \mathbf{x}, \nabla_t \rangle &= B(\mathbf{x}, \tilde{\mathbf{x}}_{t-1}^\infty) - B(\mathbf{x}, \tilde{\mathbf{x}}_t^\infty) - B(\tilde{\mathbf{x}}_t^\infty, \tilde{\mathbf{x}}_{t-1}^\infty). \end{aligned}$$

Substituting this in the previous inequality gives us

$$\begin{aligned} f_t(\mathbf{x}_t) - f_t(\mathbf{x}) &\leq \langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle + \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_t^\infty\| \|\nabla_t - g_t\|_* \\ &\quad + B(\tilde{\mathbf{x}}_t^\infty, \tilde{\mathbf{x}}_{t-1}^\infty) - B(\tilde{\mathbf{x}}_t^\infty, \mathbf{x}_t^\infty) - B(\mathbf{x}_t^\infty, \tilde{\mathbf{x}}_{t-1}^\infty) \\ &\quad + B(\mathbf{x}, \tilde{\mathbf{x}}_{t-1}^\infty) - B(\mathbf{x}, \tilde{\mathbf{x}}_t^\infty) - B(\tilde{\mathbf{x}}_t^\infty, \tilde{\mathbf{x}}_{t-1}^\infty) \\ &= \langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle + \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_t^\infty\| \|\nabla_t - g_t\|_* \\ &\quad + B(\mathbf{x}, \tilde{\mathbf{x}}_{t-1}^\infty) - B(\mathbf{x}, \tilde{\mathbf{x}}_t^\infty) - B(\tilde{\mathbf{x}}_t^\infty, \mathbf{x}_t^\infty) - B(\mathbf{x}_t^\infty, \tilde{\mathbf{x}}_{t-1}^\infty) \\ &\stackrel{(a)}{\leq} \langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle + \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_t^\infty\| \|\nabla_t - g_t\|_* \\ &\quad + B(\mathbf{x}, \tilde{\mathbf{x}}_{t-1}^\infty) - B(\mathbf{x}, \tilde{\mathbf{x}}_t^\infty) - \frac{\eta\|\tilde{\mathbf{x}}_t^\infty - \mathbf{x}_t^\infty\|^2}{2C} - \frac{\eta\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2}{2C}, \end{aligned}$$

where (a) follows from strongly convexity of  $R$ . Summing over  $t = 1, \dots, T$ , gives us

$$\begin{aligned} \sum_{t=1}^T f_t(\mathbf{x}_t) - f_t(\mathbf{x}) &\leq \sum_{t=1}^T \langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle + \underbrace{B(\mathbf{x}, \tilde{\mathbf{x}}_0^\infty) - B(\mathbf{x}, \tilde{\mathbf{x}}_T^\infty)}_{S_1} \\ &\quad + \sum_{t=1}^T \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_t^\infty\| \|\nabla_t - g_t\|_* \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T (\|\tilde{\mathbf{x}}_t^\infty - \mathbf{x}_t^\infty\|^2 + \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2). \end{aligned}$$

**Bounding  $S_1$ .** We now bound  $B(\mathbf{x}, \tilde{\mathbf{x}}_0^\infty) - B(\mathbf{x}, \tilde{\mathbf{x}}_T^\infty)$ . From the definition of  $B$ , we have

$$B(\mathbf{x}, \tilde{\mathbf{x}}_0^\infty) - B(\mathbf{x}, \tilde{\mathbf{x}}_T^\infty) = R(\tilde{\mathbf{x}}_T^\infty) - \langle \nabla_{1:T}, \mathbf{x} - \tilde{\mathbf{x}}_T^\infty \rangle - R(\tilde{\mathbf{x}}_0^\infty) + \langle \nabla_{1:0}, \mathbf{x} - \tilde{\mathbf{x}}_0^\infty \rangle.$$

Note that  $\nabla_{1:0} = 0$ . This gives us

$$B(\mathbf{x}, \tilde{\mathbf{x}}_0^\infty) - B(\mathbf{x}, \tilde{\mathbf{x}}_T^\infty) = R(\tilde{\mathbf{x}}_T^\infty) - \langle \nabla_{1:T}, \mathbf{x} - \tilde{\mathbf{x}}_T^\infty \rangle - R(\tilde{\mathbf{x}}_0^\infty).$$

We now use duality to convert the RHS of the above equation, which is currently in terms of  $R$ , into a quantity which depends on  $\Phi$ . From Proposition 3.1 we have

$$\Phi(g) = -\Psi(-g) = \inf_{\mathbf{x} \in \mathcal{X}} \langle g, \mathbf{x} \rangle + R(\mathbf{x}).$$

Since  $\tilde{\mathbf{x}}_T^\infty$  is the minimizer of  $\langle \nabla_{1:T}, \mathbf{x} \rangle + R(\mathbf{x})$ , we have  $\Phi(\nabla_{1:T}) = \langle \nabla_{1:T}, \tilde{\mathbf{x}}_T^\infty \rangle + R(\tilde{\mathbf{x}}_T^\infty)$ . Similarly,  $\Phi(0) = R(\tilde{\mathbf{x}}_0^\infty)$ . Substituting these in the previous equation gives us

$$\begin{aligned}
B(\mathbf{x}, \tilde{\mathbf{x}}_T^\infty) - B(\mathbf{x}, \tilde{\mathbf{x}}_T^\infty) &= \Phi(\nabla_{1:T}) - \langle \nabla_{1:T}, \mathbf{x} \rangle - \Phi(0) \\
&= \mathbb{E}_\sigma \left[ \inf_{\mathbf{x}' \in \mathcal{X}} \langle \nabla_{1:T} - \sigma, \mathbf{x}' \rangle \right] - \langle \nabla_{1:T}, \mathbf{x} \rangle - \mathbb{E}_\sigma \left[ \inf_{\mathbf{x}' \in \mathcal{X}} \langle -\sigma, \mathbf{x}' \rangle \right] \\
&\leq \mathbb{E}_\sigma [\langle \nabla_{1:T} - \sigma, \mathbf{x} \rangle] - \langle \nabla_{1:T}, \mathbf{x} \rangle - \mathbb{E}_\sigma \left[ \inf_{\mathbf{x}' \in \mathcal{X}} \langle -\sigma, \mathbf{x}' \rangle \right] \\
&= \mathbb{E}_\sigma \left[ \inf_{\mathbf{x}' \in \mathcal{X}} \langle \sigma, \mathbf{x}' \rangle \right] - \mathbb{E}_\sigma [\langle \sigma, \mathbf{x} \rangle] \\
&\leq D \mathbb{E}_\sigma [\|\sigma\|_*] = \eta D
\end{aligned}$$

**Bounding Regret.** Substituting this in our regret bound and taking expectation on both sides gives us

$$\begin{aligned}
\mathbb{E} \left[ \sum_{t=1}^T f_t(\mathbf{x}_t) - f_t(\mathbf{x}) \right] &\leq \sum_{t=1}^T \mathbb{E} [\langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle] + \eta D + \sum_{t=1}^T \mathbb{E} [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_t^\infty\| \|\nabla_t - g_t\|_*] \\
&\quad - \frac{\eta}{2C} \sum_{t=1}^T (\mathbb{E} [\|\tilde{\mathbf{x}}_t^\infty - \mathbf{x}_t^\infty\|^2] + \mathbb{E} [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2]) \\
&\leq \sum_{t=1}^T \mathbb{E} [\langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle] + \eta D + \sum_{t=1}^T \frac{C}{2\eta} \mathbb{E} [\|\nabla_t - g_t\|_*^2] \\
&\quad - \frac{\eta}{2C} \sum_{t=1}^T \mathbb{E} [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2]
\end{aligned}$$

To finish the proof, we make use of the Holder's smoothness assumption on  $f_t$  to bound the first term in the RHS above. From Holder's smoothness assumption, we have

$$\langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t - \nabla f_t(\mathbf{x}_t^\infty) \rangle \leq L \|\mathbf{x}_t - \mathbf{x}_t^\infty\|^{1+\alpha}.$$

Using this, we get

$$\begin{aligned}
\mathbb{E} [\langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle | g_t, \mathbf{x}_{1:t-1}, f_{1:t}] &\leq \mathbb{E} [\langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla f_t(\mathbf{x}_t^\infty) \rangle] + L \|\mathbf{x}_t - \mathbf{x}_t^\infty\|^{1+\alpha} | g_t, \mathbf{x}_{1:t-1}, f_{1:t}] \\
&\stackrel{(a)}{\leq} L \mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_t^\infty\|^{1+\alpha} | g_t, \mathbf{x}_{1:t-1}, f_{1:t}] \\
&\stackrel{(b)}{\leq} \Psi_1^{1+\alpha} L \mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_t^\infty\|_2^{1+\alpha} | g_t, \mathbf{x}_{1:t-1}, f_{1:t}] \\
&\stackrel{(c)}{\leq} \Psi_1^{1+\alpha} L \mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_t^\infty\|_2^2 | g_t, \mathbf{x}_{1:t-1}, f_{1:t}]^{(1+\alpha)/2} \\
&\stackrel{(d)}{\leq} L \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{1+\alpha},
\end{aligned}$$

where (a) follows from the fact that  $\mathbb{E} [\langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla f_t(\mathbf{x}_t^\infty) \rangle | g_t, \mathbf{x}_{1:t-1}, f_{1:t}] = 0$ , (b) follows from the definition of norm compatibility constant  $\Psi_1$ , (c) follows from Holders inequality and (d) uses the fact that conditioned on  $\{g_t, \mathbf{x}_{1:t-1}, f_{1:t}\}$ ,  $\mathbf{x}_t - \mathbf{x}_t^\infty$  is the average of  $m$  i.i.d bounded mean 0 random variables, the variance of which scales as  $O(D^2/m)$ . Substituting this in the above regret bound gives us the required result.

## B.2 Proof of Corollary 4.1

We first bound  $\mathbb{E}_\sigma [\|\sigma\|_2]$ . Relying on spherical symmetry of the perturbation distribution and the fact that the density of  $P_{\text{PRTB}}$  on the spherical shell of radius  $r$  is proportional to  $r^{d-1}$ , we get

$$\mathbb{E}_\sigma [\|\sigma\|_2] = \frac{\int_{r=0}^{(1+d^{-1})\eta} r \times r^{d-1} dr}{\int_{r=0}^{(1+d^{-1})\eta} r^{d-1} dr} = \eta.$$

We now bound the stability of predictions of OFTPL. Our technique for bounding the stability uses similar arguments as Hazan and Minasyan [10] (see Lemma 4.2 of [10]). Recall, to bound stability, we need to show that  $\Phi(g) = \mathbb{E}_\sigma [\inf_{\mathbf{x} \in \mathcal{X}} \langle g - \sigma, \mathbf{x} \rangle]$  is smooth. Let  $\phi_0(g) = \inf_{\mathbf{x} \in \mathcal{X}} \langle g, \mathbf{x} - \mathbf{x}_{00} \rangle$ , where  $\mathbf{x}_{00}$  is an arbitrary point in  $\mathcal{X}$ . We can rewrite  $\Phi(g)$  as

$$\Phi(g) = \mathbb{E}_\sigma [\phi_0(g - \sigma)] + \langle g, \mathbf{x}_{00} \rangle.$$

Since the second term in the RHS above is linear in  $g$ , any upper bound on the smoothness of  $\mathbb{E}_\sigma [\phi_0(g - \sigma)]$  is also a bound on the smoothness of  $\Phi(g)$ . So we focus on bounding the smoothness of  $\mathbb{E}_\sigma [\phi_0(g - \sigma)]$ .

First note that  $\phi_0(g)$  is  $D$  Lipschitz and satisfies the following for any  $g_1, g_2 \in \mathbb{R}^d$

$$\begin{aligned} \phi_0(g_1) - \phi_0(g_2) &= \inf_{\mathbf{x} \in \mathcal{X}} \langle -g_2, \mathbf{x} - \mathbf{x}_{00} \rangle - \inf_{\mathbf{x} \in \mathcal{X}} \langle -g_1, \mathbf{x} - \mathbf{x}_{00} \rangle \\ &\leq \sup_{\mathbf{x} \in \mathcal{X}} \langle g_1 - g_2, \mathbf{x} - \mathbf{x}_{00} \rangle \\ &\leq D \|g_1 - g_2\|_2. \end{aligned}$$

Letting  $\Phi_0(g) = \mathbb{E}_\sigma [\phi_0(g - \sigma)]$ , Lemma 4.2 of Hazan and Minasyan [10] shows that  $\Phi_0(g)$  is smooth and satisfies

$$\|\nabla \Phi_0(g_1) - \nabla \Phi_0(g_2)\|_2 \leq dD\eta^{-1} \|g_1 - g_2\|_2.$$

This shows that the predictions of OFTPL are  $dD\eta^{-1}$  stable. The rest of the proof involves substituting  $C = dD$  in the regret bound of Theorem 4.1 and setting  $g_t = 0$  and using the fact that  $\|\nabla_t\|_2 \leq G$ .

## C Online Nonconvex Learning

### C.1 Proof of Theorem 4.2

Before we present the proof of the Theorem, we introduce some notation and present some useful intermediate results. We note that unlike the convex case, there are no known Fenchel duality theorems for infinite dimensional setting. So more careful arguments are needed to obtain tight regret bounds. Our proof mimics the proof of Theorem 4.1.

#### C.1.1 Notation

Let  $\mathcal{P}$  be the set of all probability measures on  $\mathcal{X}$ . We define functions  $\Phi : \mathcal{F} \rightarrow \mathbb{R}$ ,  $R : \mathcal{P} \rightarrow \mathbb{R}$  as follows

$$\begin{aligned} \Phi(f) &= \mathbb{E}_\sigma \left[ \inf_{P \in \mathcal{P}} \mathbb{E}_{\mathbf{x} \sim P} [f(\mathbf{x}) - \sigma(\mathbf{x})] \right], \\ R(P) &= \sup_{f \in \mathcal{F}} -\mathbb{E}_{\mathbf{x} \sim P} [f(\mathbf{x})] + \Phi(f). \end{aligned}$$

Also, note that the function  $\nabla \Phi : \mathcal{F} \rightarrow \mathcal{P}$  defined in Section 4.2 can be written as

$$\nabla \Phi(f) = \mathbb{E}_\sigma \left[ \operatorname{argmin}_{P \in \mathcal{P}} \mathbb{E}_{\mathbf{x} \sim P} [f(\mathbf{x}) - \sigma(\mathbf{x})] \right].$$

Note that,  $\nabla \Phi(f)$  is well defined because from our assumption on the perturbation distribution, the minimization problem inside the expectation has a unique minimizer with probability one. To simplify the notation, in the sequel, we use the shorthand notation  $\langle P, f \rangle$  to denote  $\mathbb{E}_{\mathbf{x} \sim P} [f(\mathbf{x})]$ , for any  $P \in \mathcal{P}$  and  $f \in \mathcal{F}$ . Similarly, for any  $P_1, P_2 \in \mathcal{P}$  and  $f \in \mathcal{F}$ , we use the notation  $\langle P_1 - P_2, f \rangle$  to denote  $\mathbb{E}_{\mathbf{x} \sim P_1} [f(\mathbf{x})] - \mathbb{E}_{\mathbf{x} \sim P_2} [f(\mathbf{x})]$ .

#### C.1.2 Intermediate Results

**Lemma C.1.** For any  $g \in \mathcal{F}$ ,  $R(\nabla \Phi(g)) = -\langle \nabla \Phi(g), g \rangle + \Phi(g)$ .

*Proof.* Define  $P_{g,\sigma}$  as

$$P_{g,\sigma} = \operatorname{argmin}_{P \in \mathcal{P}} \mathbb{E}_{\mathbf{x} \sim P} [g(\mathbf{x}) - \sigma(\mathbf{x})].$$

Note that  $\nabla\Phi(g) = \mathbb{E}_\sigma [P_{g,\sigma}]$ . For any  $g, h \in \mathcal{F}$ , we have

$$\begin{aligned}\Phi(h) &= \mathbb{E}_\sigma \left[ \inf_{P \in \mathcal{P}} \langle P, h - \sigma \rangle \right] \\ &\leq \mathbb{E}_\sigma [\langle P_{g,\sigma}, h - \sigma \rangle] \\ &= \mathbb{E}_\sigma [\langle P_{g,\sigma}, g - \sigma \rangle] + \mathbb{E}_\sigma [\langle P_{g,\sigma}, h - g \rangle] \\ &= \Phi(g) + \langle \nabla\Phi(g), h - g \rangle.\end{aligned}$$

This shows that for any  $g, h \in \mathcal{F}$

$$\Phi(h) - \langle \nabla\Phi(g), h \rangle \leq \Phi(g) - \langle \nabla\Phi(g), g \rangle. \quad (5)$$

Taking supremum over  $h$  of the LHS quantity gives us

$$R(\nabla\Phi(g)) = \sup_{h \in \mathcal{F}} \Phi(h) - \langle \nabla\Phi(g), h \rangle = \Phi(g) - \langle \nabla\Phi(g), g \rangle.$$

□

**Lemma C.2** (Strong Smoothness). *The function  $-\Phi$  is convex and strongly smooth and satisfies the following inequality for any  $g_1, g_2 \in \mathcal{F}$*

$$-\Phi(g_2) \leq -\Phi(g_1) - \langle \nabla\Phi(g_1), g_2 - g_1 \rangle + \frac{C}{2\eta} \|g_2 - g_1\|_{\mathcal{F}}^2.$$

*Proof.* Let  $g_1, g_2 \in \mathcal{F}$  and  $\alpha \in [0, 1]$ . Then

$$\begin{aligned}\Phi(\alpha g_1 + (1 - \alpha)g_2) &= \mathbb{E}_\sigma \left[ \inf_{P \in \mathcal{P}} \langle P, \alpha g_1 + (1 - \alpha)g_2 - \sigma \rangle \right] \\ &\geq \alpha \mathbb{E}_\sigma \left[ \inf_{P \in \mathcal{P}} \langle P, g_1 - \sigma \rangle \right] + (1 - \alpha) \mathbb{E}_\sigma \left[ \inf_{P \in \mathcal{P}} \langle P, g_2 - \sigma \rangle \right] \\ &= \alpha \Phi(g_1) + (1 - \alpha) \Phi(g_2).\end{aligned}$$

This shows that  $-\Phi$  is convex. To show smoothness, we rely on the following stability property

$$\forall g_1, g_2 \in \mathcal{F} \quad \gamma_{\mathcal{F}}(\nabla\Phi(g_1), \nabla\Phi(g_2)) \leq \frac{C}{\eta} \|g_1 - g_2\|_{\mathcal{F}}.$$

Let  $T$  be an arbitrary positive integer and for  $t \in \{0, 1, \dots, T\}$ , define  $\alpha_t = t/T$ . Let  $h = g_2 - g_1$ . We have

$$\begin{aligned}\Phi(g_1) - \Phi(g_2) &= \Phi(g_1 + \alpha_0 h) - \Phi(g_1 + \alpha_T h) \\ &= \sum_{t=0}^{T-1} (\Phi(g_1 + \alpha_t h) - \Phi(g_1 + \alpha_{t+1} h))\end{aligned}$$

Since  $-\Phi$  is convex and satisfies Equation (5), we have

$$\begin{aligned}\Phi(g_1) - \Phi(g_2) &= \sum_{t=0}^{T-1} (\Phi(g_1 + \alpha_t h) - \Phi(g_1 + \alpha_{t+1} h)) \\ &\leq - \sum_{t=0}^{T-1} \frac{1}{T} \langle \nabla\Phi(g_1 + \alpha_{t+1} h), h \rangle\end{aligned}$$

Using stability, we get

$$\begin{aligned}
\Phi(g_1) - \Phi(g_2) &\leq - \sum_{t=0}^{T-1} \frac{1}{T} \langle \nabla \Phi(g_1 + \alpha_{t+1}h), h \rangle \\
&= \sum_{t=0}^{T-1} \frac{1}{T} (\langle \nabla \Phi(g_1) - \nabla \Phi(g_1 + \alpha_{t+1}h), h \rangle - \langle \nabla \Phi(g_1), h \rangle) \\
&\stackrel{(a)}{\leq} - \langle \nabla \Phi(g_1), h \rangle + \sum_{t=0}^{T-1} \frac{1}{T} \gamma_{\mathcal{F}}(\nabla \Phi(g_1), \nabla \Phi(g_1 + \alpha_{t+1}h)) \|h\|_{\mathcal{F}} \\
&\stackrel{(b)}{\leq} - \langle \nabla \Phi(g_1), h \rangle + \sum_{t=0}^{T-1} \frac{C}{T\eta} \|\alpha_{t+1}h\|_{\mathcal{F}} \|h\|_{\mathcal{F}} \\
&= - \langle \nabla \Phi(g_1), h \rangle + \sum_{t=0}^{T-1} \frac{C\alpha_{t+1}}{T\eta} \|h\|_{\mathcal{F}}^2 \\
&= - \langle \nabla \Phi(g_1), h \rangle + \frac{C}{\eta} \frac{T+1}{2T} \|h\|_{\mathcal{F}}^2,
\end{aligned}$$

where (a) follows from the definition of  $\gamma_{\mathcal{F}}$  and (b) follows from the stability assumption. Taking  $T \rightarrow \infty$ , we get

$$-\Phi(g_2) \leq -\Phi(g_1) - \langle \nabla \Phi(g_1), g_2 - g_1 \rangle + \frac{C}{2\eta} \|g_2 - g_1\|_{\mathcal{F}}^2.$$

□

**Lemma C.3 (Strong Convexity).** For any  $P \in \mathcal{P}$  and  $g \in \mathcal{F}$ ,  $R$  satisfies the following inequality

$$R(P) \geq R(\nabla \Phi(g)) + \langle \nabla \Phi(g) - P, g \rangle + \frac{\eta}{2C} \gamma_{\mathcal{F}}(P, \nabla \Phi(g))^2.$$

*Proof.* From Lemma C.2 we know that the following holds for any  $g, h \in \mathcal{F}$

$$\Phi(g) \geq \underbrace{\Phi(h) + \langle \nabla \Phi(h), g - h \rangle - \frac{C}{2\eta} \|g - h\|_{\mathcal{F}}^2}_{\Phi_{\text{lb},h}(g)}.$$

Define  $R_{\text{lb},h}(P)$  as

$$R_{\text{lb}}(P) = \sup_{g \in \mathcal{F}} -\langle P, g \rangle + \Phi_{\text{lb},h}(g).$$

Since  $\Phi(g) \geq \Phi_{\text{lb},h}(g)$  for all  $g \in \mathcal{F}$ ,  $R(P) \geq R_{\text{lb},h}(P)$  for all  $P$ . We now derive an expression for  $R_{\text{lb},h}(P)$ . Note that from Lemma C.1 we have  $R(\nabla \Phi(h)) = -\langle \nabla \Phi(h), h \rangle + \Phi(h)$ . Using this, we get

$$\begin{aligned}
R_{\text{lb},h}(P) &= \sup_{g \in \mathcal{F}} -\langle P, g \rangle + \Phi_{\text{lb},h}(g) \\
&\stackrel{(a)}{=} \sup_{g \in \mathcal{F}} \left( -\langle P, g \rangle + \Phi(h) + \langle \nabla \Phi(h), g - h \rangle - \frac{C}{2\eta} \|g - h\|_{\mathcal{F}}^2 \right) \\
&\stackrel{(b)}{=} R(\nabla \Phi(h)) + \sup_{g \in \mathcal{F}} \left( \langle \nabla \Phi(h) - P, g \rangle - \frac{C}{2\eta} \|g - h\|_{\mathcal{F}}^2 \right),
\end{aligned}$$

where (a) follows from the definition of  $\Phi_{\text{lb},h}(g)$  and (b) follows from Lemma C.1. We now do a change of variables in the supremum of the above expression. Substituting  $g' = g - h$ , we get

$$R_{\text{lb},h}(P) = R(\nabla \Phi(h)) + \langle \nabla \Phi(h) - P, h \rangle + \sup_{g' \in \mathcal{F}} \left( \langle \nabla \Phi(h) - P, g' \rangle - \frac{C}{2\eta} \|g'\|_{\mathcal{F}}^2 \right).$$

We now show that

$$\sup_{g' \in \mathcal{F}} \left( \langle \nabla \Phi(h) - P, g' \rangle - \frac{C}{2\eta} \|g'\|_{\mathcal{F}}^2 \right) \geq \frac{\eta}{2C} \gamma_{\mathcal{F}}(P, \nabla \Phi(h))^2.$$

To this end, we choose a  $g'' \in \mathcal{F}$  such that

$$\|g''\|_{\mathcal{F}} = \frac{\eta}{C} \gamma_{\mathcal{F}}(P, \nabla\Phi(h)), \quad \langle \nabla\Phi(h) - P, g'' \rangle = \frac{\eta}{C} \gamma_{\mathcal{F}}(P, \nabla\Phi(h))^2. \quad (6)$$

If such a  $g''$  can be found, we have

$$\begin{aligned} \sup_{g' \in \mathcal{F}} \left( \langle \nabla\Phi(h) - P, g' \rangle - \frac{C}{2\eta} \|g'\|_{\mathcal{F}}^2 \right) &\geq \langle \nabla\Phi(h) - P, g'' \rangle - \frac{C}{2\eta} \|g''\|_{\mathcal{F}}^2 \\ &= \frac{\eta}{2C} \gamma_{\mathcal{F}}(P, \nabla\Phi(h))^2. \end{aligned}$$

This would then imply the main claim of the Lemma.

$$R(P) \geq R_{\text{lb},h}(P) \geq R(\nabla\Phi(h)) + \langle \nabla\Phi(h) - P, h \rangle + \frac{\eta}{2C} \gamma_{\mathcal{F}}(P, \nabla\Phi(h))^2.$$

**Finding  $g''$ .** We now construct a  $g''$  which satisfies Equation (6). From the definition of  $\gamma_{\mathcal{F}}$  we know that

$$\gamma_{\mathcal{F}}(P, \nabla\Phi(h)) = \sup_{\|g'\|_{\mathcal{F}} \leq 1} |\langle \nabla\Phi(h) - P, g' \rangle|$$

Suppose the supremum is achieved at  $g^*$ . Define  $g''$  as  $\frac{\eta s}{C} \gamma_{\mathcal{F}}(P, \nabla\Phi(h)) g^*$ , where  $s = \text{sign}(\langle \nabla\Phi(h) - P, g^* \rangle)$ . It can be easily verified that  $g''$  satisfies Equation (6).

If the supremum is never achieved, the same argument as above can still be made using a sequence of functions  $\{g_n\}_{n=1}^{\infty}$  such that

$$\|g_n\|_{\mathcal{F}} \leq 1, \quad \lim_{n \rightarrow \infty} |\langle \nabla\Phi(h) - P, g_n \rangle| = \gamma_{\mathcal{F}}(P, \nabla\Phi(h)).$$

Define  $g''_n$  as  $\frac{\eta s_n}{C} \gamma_{\mathcal{F}}(P, \nabla\Phi(h)) g_n$ , where  $s_n = \text{sign}(\langle \nabla\Phi(h) - P, g_n \rangle)$ . Since  $\lim_{n \rightarrow \infty} \|g_n\|_{\mathcal{F}} = 1$ , we have  $\lim_{n \rightarrow \infty} \|g''_n\|_{\mathcal{F}} = \frac{\eta}{C} \gamma_{\mathcal{F}}(P, \nabla\Phi(h))$ . Moreover,

$$\lim_{n \rightarrow \infty} \langle \nabla\Phi(h) - P, g''_n \rangle = \lim_{n \rightarrow \infty} \frac{\eta}{C} \gamma_{\mathcal{F}}(P, \nabla\Phi(h)) |\langle \nabla\Phi(h) - P, g_n \rangle| = \frac{\eta}{C} \gamma_{\mathcal{F}}(P, \nabla\Phi(h))^2.$$

This shows that

$$\begin{aligned} \sup_{g' \in \mathcal{F}} \left( \langle \nabla\Phi(h) - P, g' \rangle - \frac{C}{2\eta} \|g'\|_{\mathcal{F}}^2 \right) &\geq \lim_{n \rightarrow \infty} \left( \langle \nabla\Phi(h) - P, g''_n \rangle - \frac{C}{2\eta} \|g''_n\|_{\mathcal{F}}^2 \right) \\ &= \frac{\eta}{2C} \gamma_{\mathcal{F}}(P, \nabla\Phi(h))^2. \end{aligned}$$

This finishes the proof of the Lemma.  $\square$

### C.1.3 Main Argument

We are now ready to prove Theorem 4.2. Our proof relies on Lemma C.3 and uses similar arguments as used in the proof of Theorem 4.1. We first rewrite  $P_t, \tilde{P}_t$  as

$$\begin{aligned} P_t &= \frac{1}{m} \sum_{j=1}^m \underset{P \in \mathcal{P}}{\text{argmin}} \mathbb{E}_{\mathbf{x} \sim P} \left[ \sum_{i=1}^{t-1} f_i(\mathbf{x}) + g_t(\mathbf{x}) - \sigma_{t,j}(\mathbf{x}) \right], \\ \tilde{P}_t &= \frac{1}{m} \sum_{j=1}^m \underset{P \in \mathcal{P}}{\text{argmin}} \mathbb{E}_{\mathbf{x} \sim P} \left[ \sum_{i=1}^t f_i(\mathbf{x}) - \sigma'_{t,j}(\mathbf{x}) \right]. \end{aligned}$$

Note that

$$\begin{aligned} P_t^{\infty} &= \mathbb{E} [P_t | g_t, f_{1:t-1}, P_{1:t-1}] = \nabla\Phi(f_{1:t-1} + g_t), \\ \tilde{P}_t^{\infty} &= \mathbb{E} [\tilde{P}_t | f_{1:t-1}, P_{1:t-1}] = \nabla\Phi(f_{1:t}), \end{aligned}$$

with  $P_1^{\infty} = \tilde{P}_0^{\infty} = \nabla\Phi(0)$ . Define functions  $B(\cdot, P_t^{\infty}), B(\cdot, \tilde{P}_t^{\infty})$  as

$$\begin{aligned} B(P, P_t^{\infty}) &= R(P) - R(P_t^{\infty}) + \langle P - P_t^{\infty}, f_{1:t-1} + g_t \rangle, \\ B(P, \tilde{P}_t^{\infty}) &= R(P) - R(\tilde{P}_t^{\infty}) + \langle P - \tilde{P}_t^{\infty}, f_{1:t} \rangle. \end{aligned}$$

From Lemma C.3, we have

$$B(P, P_t^\infty) \geq \frac{\eta}{2C} \gamma_{\mathcal{F}}(P, P_t^\infty)^2, \quad B(P, \tilde{P}_t^\infty) \geq \frac{\eta}{2C} \gamma_{\mathcal{F}}(P, \tilde{P}_t^\infty)^2.$$

For any  $P \in \mathcal{P}$ , we have

$$\begin{aligned} \mathbb{E}[f_t(\mathbf{x}_t) - f_t(P)] &= \mathbb{E}[f_t(P_t) - f_t(P)] \\ &= \mathbb{E}[\langle P_t - P, f_t \rangle] \\ &= \mathbb{E}[\langle P_t - P_t^\infty, f_t \rangle] + \mathbb{E}[\langle P_t^\infty - P, f_t \rangle] \\ &= \mathbb{E}[\langle P_t - P_t^\infty, f_t \rangle] + \mathbb{E}[\langle P_t^\infty - \tilde{P}_t^\infty, f_t - g_t \rangle] \\ &\quad + \mathbb{E}[\langle P_t^\infty - \tilde{P}_t^\infty, g_t \rangle] + \mathbb{E}[\langle \tilde{P}_t^\infty - P, f_t \rangle] \\ &\stackrel{(a)}{\leq} \mathbb{E}[\gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_t^\infty) \|f_t - g_t\|_{\mathcal{F}}] + \mathbb{E}[\langle P_t^\infty - \tilde{P}_t^\infty, g_t \rangle] \\ &\quad + \mathbb{E}[\langle \tilde{P}_t^\infty - P, f_t \rangle], \end{aligned}$$

where (a) follows from the fact that  $\mathbb{E}[\langle P_t - P_t^\infty, f_t \rangle | g_t, f_{1:t-1}, P_{1:t-1}] = 0$  and as a result  $\mathbb{E}[\langle P_t - P_t^\infty, f_t \rangle] = 0$ . Next, a simple calculation shows that

$$\begin{aligned} \langle P_t^\infty - \tilde{P}_t^\infty, g_t \rangle &= B(\tilde{P}_t^\infty, \tilde{P}_{t-1}^\infty) - B(\tilde{P}_t^\infty, P_t^\infty) - B(P_t^\infty, \tilde{P}_{t-1}^\infty) \\ \langle \tilde{P}_t^\infty - P, f_t \rangle &= B(P, \tilde{P}_{t-1}^\infty) - B(P, \tilde{P}_t^\infty) - B(\tilde{P}_t^\infty, \tilde{P}_{t-1}^\infty). \end{aligned}$$

Substituting this in the previous regret bound gives us

$$\begin{aligned} \mathbb{E}[f_t(\mathbf{x}_t) - f_t(P)] &\leq \mathbb{E}[\gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_t^\infty) \|f_t - g_t\|_{\mathcal{F}}] + \mathbb{E}[B(\tilde{P}_t^\infty, \tilde{P}_{t-1}^\infty) - B(\tilde{P}_t^\infty, P_t^\infty) - B(P_t^\infty, \tilde{P}_{t-1}^\infty)] \\ &\quad + \mathbb{E}[B(P, \tilde{P}_{t-1}^\infty) - B(P, \tilde{P}_t^\infty) - B(\tilde{P}_t^\infty, \tilde{P}_{t-1}^\infty)] \\ &= \mathbb{E}[\gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_t^\infty) \|f_t - g_t\|_{\mathcal{F}}] \\ &\quad + \mathbb{E}[B(P, \tilde{P}_{t-1}^\infty) - B(P, \tilde{P}_t^\infty) - B(\tilde{P}_t^\infty, P_t^\infty) - B(P_t^\infty, \tilde{P}_{t-1}^\infty)] \\ &\stackrel{(a)}{\leq} \mathbb{E}[\gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_t^\infty) \|f_t - g_t\|_{\mathcal{F}}] \\ &\quad + \mathbb{E}[B(P, \tilde{P}_{t-1}^\infty) - B(P, \tilde{P}_t^\infty)] - \mathbb{E}\left[\frac{\eta}{2C} \gamma_{\mathcal{F}}(\tilde{P}_t^\infty, P_t^\infty)^2 + \frac{\eta}{2C} \gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_{t-1}^\infty)^2\right] \\ &\stackrel{(b)}{\leq} \frac{C}{2\eta} \mathbb{E}[\|f_t - g_t\|_{\mathcal{F}}^2] + \mathbb{E}[B(P, \tilde{P}_{t-1}^\infty) - B(P, \tilde{P}_t^\infty)] - \mathbb{E}\left[\frac{\eta}{2C} \gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_{t-1}^\infty)^2\right] \end{aligned}$$

where (a) follows from Lemma C.3, and (b) uses the fact that  $|xy| \leq \frac{1}{2c}|x|^2 + \frac{c}{2}|y|^2$ , for any  $x, y$ ,  $c > 0$ . Summing over  $t = 1, \dots, T$  gives us

$$\begin{aligned} \sum_{t=1}^T \mathbb{E}[f_t(\mathbf{x}_t) - f_t(P)] &\leq \underbrace{\mathbb{E}[B(P, \tilde{P}_0^\infty) - B(P, \tilde{P}_T^\infty)]}_{S_1} + \sum_{t=1}^T \frac{C}{2\eta} \mathbb{E}[\|f_t - g_t\|_{\mathcal{F}}^2] \\ &\quad - \sum_{t=1}^T \frac{\eta}{2C} \mathbb{E}[\gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_{t-1}^\infty)^2] \end{aligned}$$

To finish the proof of the Theorem, we need to bound  $S_1$ .

**Bounding  $S_1$ .** From the definition of  $B$ , we have

$$B(P, \tilde{P}_0^\infty) - B(P, \tilde{P}_T^\infty) = R(\tilde{P}_T^\infty) - \langle P - \tilde{P}_T^\infty, f_{1:T} \rangle - R(\tilde{\mathbf{x}}_0^\infty),$$

where we used the fact that  $f_{1:0} = 0$ . We now rely on Lemma C.1 to convert the above equation, which is currently in terms of  $R$ , into a quantity which depends on  $\Phi$ . Using Lemma C.1, we get

$$B(P, \tilde{P}_0^\infty) - B(P, \tilde{P}_T^\infty) = \Phi(f_{1:T}) - \langle P, f_{1:T} \rangle - \Phi(0).$$

From the definition of  $\Phi$  we have

$$\begin{aligned}
B(P, \tilde{P}_0^\infty) - B(P, \tilde{P}_T^\infty) &= \Phi(f_{1:T}) - \langle P, f_{1:T} \rangle - \Phi(0) \\
&= \mathbb{E}_\sigma \left[ \inf_{P' \in \mathcal{P}} \langle P', f_{1:T} - \sigma \rangle \right] - \langle P, f_{1:T} \rangle - \mathbb{E}_\sigma \left[ \inf_{P' \in \mathcal{P}} \langle P', -\sigma \rangle \right] \\
&\leq \mathbb{E}_\sigma [\langle P, f_{1:T} - \sigma \rangle] - \langle P, f_{1:T} \rangle - \mathbb{E}_\sigma \left[ \inf_{P' \in \mathcal{P}} \langle P', -\sigma \rangle \right] \\
&= \mathbb{E}_\sigma \left[ \sup_{P' \in \mathcal{P}} \langle P', \sigma \rangle \right] - \mathbb{E}_\sigma [\langle P, \sigma \rangle] \\
&\leq D \mathbb{E}_\sigma [\|\sigma\|_{\mathcal{F}}] = \eta D,
\end{aligned}$$

where the last inequality follows from our bound on the diameter of  $\mathcal{P}$ . Substituting this in the above regret bound gives us the required result.

## C.2 Proof of Corollary 4.2

To prove the corollary we first show that for our choice of perturbation distribution,  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) - \sigma(\mathbf{x})$  has a unique minimizer with probability one, for any  $f \in \mathcal{F}$ . Next, we show that the predictions of OFTPL are stable.

### C.2.1 Intermediate Results

**Lemma C.4** (Unique Minimizers). *Suppose the perturbation function is such that  $\sigma(\mathbf{x}) = \langle \bar{\sigma}, \mathbf{x} \rangle$ , where  $\bar{\sigma} \in \mathbb{R}^d$  is a random vector whose entries are sampled independently from  $\operatorname{Exp}(\eta)$ . Then, for any  $f \in \mathcal{F}$ ,  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) - \sigma(\mathbf{x})$  has a unique minimizer with probability one.*

*Proof.* Define  $\mathbf{x}_f(\sigma)$  as

$$\mathbf{x}_f(\bar{\sigma}) \in \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) - \langle \bar{\sigma}, \mathbf{x} \rangle.$$

For any  $\bar{\sigma}_1, \bar{\sigma}_2$  we now show that  $\mathbf{x}_f(\bar{\sigma})$  satisfies the following monotonicity property

$$\langle \mathbf{x}_f(\bar{\sigma}_1) - \mathbf{x}_f(\bar{\sigma}_2), \bar{\sigma}_1 - \bar{\sigma}_2 \rangle \geq 0.$$

From the optimality of  $\mathbf{x}_f(\bar{\sigma}_1), \mathbf{x}_f(\bar{\sigma}_2)$  we have

$$\begin{aligned}
f(\mathbf{x}_f(\bar{\sigma}_1)) - \langle \bar{\sigma}_1, \mathbf{x}_f(\bar{\sigma}_1) \rangle &\leq f(\mathbf{x}_f(\bar{\sigma}_2)) - \langle \bar{\sigma}_1, \mathbf{x}_f(\bar{\sigma}_2) \rangle \\
&= f(\mathbf{x}_f(\bar{\sigma}_2)) - \langle \bar{\sigma}_2, \mathbf{x}_f(\bar{\sigma}_2) \rangle + \langle \bar{\sigma}_2 - \bar{\sigma}_1, \mathbf{x}_f(\bar{\sigma}_2) \rangle \\
&\leq f(\mathbf{x}_f(\bar{\sigma}_1)) - \langle \bar{\sigma}_2, \mathbf{x}_f(\bar{\sigma}_1) \rangle + \langle \bar{\sigma}_2 - \bar{\sigma}_1, \mathbf{x}_f(\bar{\sigma}_2) \rangle.
\end{aligned}$$

This shows that  $\langle \bar{\sigma}_2 - \bar{\sigma}_1, \mathbf{x}_f(\bar{\sigma}_2) - \mathbf{x}_f(\bar{\sigma}_1) \rangle \geq 0$ . To finish the proof of Lemma, we rely on Theorem 1 of Zarantonello [25], which shows that the set of points for which a monotone operator is not single-valued has Lebesgue measure zero. Since the distribution of  $\bar{\sigma}$  is absolutely continuous w.r.t Lebesgue measure, this shows that  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) - \sigma(\mathbf{x})$  has a unique minimizer with probability one.  $\square$

### C.2.2 Main Argument

For our choice of perturbation distribution,  $\mathbb{E}_\sigma [\|\sigma\|_{\mathcal{F}}] = \mathbb{E}_{\bar{\sigma}} [\|\bar{\sigma}\|_\infty] = \eta \log d$ . We now bound the stability of predictions of OFTPL. First note that for our choice of primal space  $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ ,  $\gamma_{\mathcal{F}}$  is the Wasserstein-1 metric, which is defined as

$$\gamma_{\mathcal{F}}(P_1, P_2) = \sup_{f \in \mathcal{F}, \|f\|_{\mathcal{F}} \leq 1} \left| \mathbb{E}_{\mathbf{x} \sim P_1} [f(\mathbf{x})] - \mathbb{E}_{\mathbf{x} \sim P_2} [f(\mathbf{x})] \right| = \inf_{Q \in \Gamma(P_1, P_2)} \mathbb{E}_{(\mathbf{x}_1, \mathbf{x}_2) \sim Q} [\|\mathbf{x}_1 - \mathbf{x}_2\|_1],$$

where  $\Gamma(P_1, P_2)$  is the set of all probability measures on  $\mathcal{X} \times \mathcal{X}$  with marginals  $P_1, P_2$  on the first and second factors respectively. Define  $\mathbf{x}_f(\bar{\sigma})$  as

$$\mathbf{x}_f(\bar{\sigma}) \in \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) - \langle \bar{\sigma}, \mathbf{x} \rangle.$$

---

**Algorithm 3** OFTPL for convex-concave games
 

---

```

1: Input: Perturbation Distributions  $P_{\text{PRTB}}^1, P_{\text{PRTB}}^2$  of  $\mathbf{x}, \mathbf{y}$  players, number of samples  $m$ , iterations  $T$ 
2: for  $t = 1 \dots T$  do
3:   if  $t = 1$  then
4:     Sample  $\{\sigma_{1,j}^1\}_{j=1}^m, \{\sigma_{1,j}^2\}_{j=1}^m$  from  $P_{\text{PRTB}}^1, P_{\text{PRTB}}^2$ 
5:      $\mathbf{x}_1 = \frac{1}{m} \sum_{j=1}^m [\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle -\sigma_{1,j}^1, \mathbf{x} \rangle], \mathbf{y}_1 = \frac{1}{m} [\sum_{j=1}^m \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \langle \sigma_{1,j}^2, \mathbf{y} \rangle]$ 
6:     continue
7:   end if
8:   //Compute guesses
9:   for  $j = 1 \dots m$  do
10:    Sample  $\sigma_{t,j}^1 \sim P_{\text{PRTB}}^1, \sigma_{t,j}^2 \sim P_{\text{PRTB}}^2$ 
11:     $\tilde{\mathbf{x}}_{t-1,j} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle \sum_{i=1}^{t-1} \nabla_{\mathbf{x}} f(\mathbf{x}_i, \mathbf{y}_i) - \sigma_{t,j}^1, \mathbf{x} \rangle$ 
12:     $\tilde{\mathbf{y}}_{t-1,j} = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \langle \sum_{i=1}^{t-1} \nabla_{\mathbf{y}} f(\mathbf{x}_i, \mathbf{y}_i) + \sigma_{t,j}^2, \mathbf{y} \rangle$ 
13:   end for
14:    $\tilde{\mathbf{x}}_{t-1} = \frac{1}{m} \sum_{j=1}^m \tilde{\mathbf{x}}_{t-1,j}, \tilde{\mathbf{y}}_{t-1} = \frac{1}{m} \sum_{j=1}^m \tilde{\mathbf{y}}_{t-1,j}$ 
15:   //Use the guesses to compute the next action
16:   for  $j = 1 \dots m$  do
17:    Sample  $\sigma_{t,j}^1 \sim P_{\text{PRTB}}^1, \sigma_{t,j}^2 \sim P_{\text{PRTB}}^2$ 
18:     $\mathbf{x}_{t,j} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \langle \sum_{i=1}^{t-1} \nabla_{\mathbf{x}} f(\mathbf{x}_i, \mathbf{y}_i) + \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1}) - \sigma_{t,j}^1, \mathbf{x} \rangle$ 
19:     $\mathbf{y}_{t,j} = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \langle \sum_{i=1}^{t-1} \nabla_{\mathbf{y}} f(\mathbf{x}_i, \mathbf{y}_i) + \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1}) + \sigma_{t,j}^2, \mathbf{y} \rangle$ 
20:   end for
21:    $\mathbf{x}_t = \frac{1}{m} \sum_{j=1}^m \mathbf{x}_{t,j}, \mathbf{y}_t = \frac{1}{m} \sum_{j=1}^m \mathbf{y}_{t,j}$ 
22: end for
23: return  $\{(\mathbf{x}_t, \mathbf{y}_t)\}_{t=1}^T$ 

```

---

Note that  $\nabla \Phi(f)$  is the distribution of random variable  $\mathbf{x}_f(\bar{\sigma})$ . Suggala and Netrapalli [11] show that for any  $f, g \in \mathcal{F}$

$$\mathbb{E}_{\bar{\sigma}} [\|\mathbf{x}_f(\bar{\sigma}) - \mathbf{x}_g(\bar{\sigma})\|_1] \leq \frac{125d^2D}{\eta} \|f - g\|_{\mathcal{F}}.$$

Since  $\gamma_{\mathcal{F}}(\nabla \Phi(f), \nabla \Phi(g)) \leq \mathbb{E}_{\bar{\sigma}} [\|\mathbf{x}_f(\bar{\sigma}) - \mathbf{x}_g(\bar{\sigma})\|_1]$ , this shows that OFTPL is  $O(d^2D\eta^{-1})$  stable w.r.t  $\|\cdot\|_{\mathcal{F}}$ . Substituting the stability bound in the regret bound of Theorem 4.2 shows that

$$\begin{aligned} \sup_{P \in \mathcal{P}} \mathbb{E} \left[ \sum_{t=1}^T f_t(\mathbf{x}_t) - f_t(P) \right] &= \eta D \log d \\ &+ O \left( \sum_{t=1}^T \frac{d^2D}{\eta} \mathbb{E} [\|f_t - g_t\|_{\mathcal{F}}^2] - \sum_{t=1}^T \frac{\eta}{d^2D} \mathbb{E} [\gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_{t-1}^\infty)^2] \right). \end{aligned}$$

## D Convex-Concave Games

Our algorithm for convex-concave games is presented in Algorithm 3. Before presenting the proof of Theorem 5.1, we first present a more general result in Section D.1. Theorem 5.1 immediately follows from our general result by instantiating it for the uniform noise distribution.

### D.1 General Result

**Theorem D.1.** Consider the minimax game in Equation (1). Suppose  $f$  is convex in  $\mathbf{x}$ , concave in  $\mathbf{y}$  and is Holder smooth w.r.t some norm  $\|\cdot\|$

$$\begin{aligned} \|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{x}} f(\mathbf{x}', \mathbf{y}')\|_* &\leq L_1 \|\mathbf{x} - \mathbf{x}'\|^\alpha + L_2 \|\mathbf{y} - \mathbf{y}'\|^\alpha, \\ \|\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{y}} f(\mathbf{x}', \mathbf{y}')\|_* &\leq L_2 \|\mathbf{x} - \mathbf{x}'\|^\alpha + L_1 \|\mathbf{y} - \mathbf{y}'\|^\alpha. \end{aligned}$$

Define diameter of sets  $\mathcal{X}, \mathcal{Y}$  as  $D = \max\{\sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}} \|\mathbf{x}_1 - \mathbf{x}_2\|, \sup_{\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{Y}} \|\mathbf{y}_1 - \mathbf{y}_2\|\}$ . Let  $L = \{L_1, L_2\}$ . Suppose both  $\mathbf{x}$  and  $\mathbf{y}$  players use Algorithm 1 to solve the minimax game. Suppose the perturbation distributions  $P_{\text{PRTB}}^1, P_{\text{PRTB}}^2$ , used by  $\mathbf{x}, \mathbf{y}$  players are absolutely continuous

and satisfy  $\mathbb{E}_{\sigma \sim P_{PRTB}^1} [\|\sigma\|_*] = \mathbb{E}_{\sigma \sim P_{PRTB}^2} [\|\sigma\|_*] = \eta$ . Suppose the predictions of both the players are  $C\eta^{-1}$ -stable w.r.t  $\|\cdot\|_*$ . Suppose the guesses used by  $\mathbf{x}, \mathbf{y}$  players in the  $t^{\text{th}}$  iteration are  $\nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1}), \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1})$ , where  $\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1}$  denote the predictions of  $\mathbf{x}, \mathbf{y}$  players in the  $t^{\text{th}}$  iteration, if guess  $g_t = 0$  was used in that iteration. Then the iterates  $\{(\mathbf{x}_t, \mathbf{y}_t)\}_{t=1}^T$  generated by the OFTPL based algorithm satisfy

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ f \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t, \mathbf{y} \right) - f \left( \mathbf{x}, \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t \right) \right] &\leq 2L_1 \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{1+\alpha} + \frac{2\eta D}{T} \\ &\quad + \frac{20CL^2}{\eta} \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{2\alpha} + 10L \left( \frac{5CL}{\eta} \right)^{\frac{1+\alpha}{1-\alpha}} \end{aligned}$$

*Proof.* Since both the players are responding to each others actions using OFTPL, using Theorem 4.1, we get the following regret bounds for the players

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}_t) - f(\mathbf{x}, \mathbf{y}_t) \right] &\leq L_1 T \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{1+\alpha} + \eta D \\ &\quad + \frac{C}{2\eta} \sum_{t=1}^T \mathbb{E} [\|\nabla_{\mathbf{x}} f(\mathbf{x}_t, \mathbf{y}_t) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1})\|_*^2] \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T \mathbb{E} [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2]. \\ \sup_{\mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}) - f(\mathbf{x}_t, \mathbf{y}_t) \right] &\leq L_1 T \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{1+\alpha} + \eta D \\ &\quad + \frac{C}{2\eta} \sum_{t=1}^T \mathbb{E} [\|\nabla_{\mathbf{y}} f(\mathbf{x}_t, \mathbf{y}_t) - \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1})\|_*^2] \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T \mathbb{E} [\|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|^2]. \end{aligned}$$

First, consider the regret of the  $\mathbf{x}$  player. Since  $\|a_1 + \dots + a_5\|^2 \leq 5(\|a_1\|^2 + \dots + \|a_5\|^2)$ , we have

$$\begin{aligned} \|\nabla_{\mathbf{x}} f(\mathbf{x}_t, \mathbf{y}_t) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1})\|_*^2 &\leq 5 \|\nabla_{\mathbf{x}} f(\mathbf{x}_t, \mathbf{y}_t) - \nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t)\|_*^2 \\ &\quad + 5 \|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t) - \nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty)\|_*^2 \\ &\quad + 5 \|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_*^2 \\ &\quad + 5 \|\nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1})\|_*^2 \\ &\quad + 5 \|\nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1})\|_*^2 \\ &\stackrel{(a)}{\leq} 5L_1^2 \|\mathbf{x}_t - \mathbf{x}_t^\infty\|^{2\alpha} + 5L_1^2 \|\tilde{\mathbf{x}}_{t-1} - \tilde{\mathbf{x}}_{t-1}^\infty\|^{2\alpha} \\ &\quad + 5L_2^2 \|\mathbf{y}_t - \mathbf{y}_t^\infty\|^{2\alpha} + 5L_2^2 \|\tilde{\mathbf{y}}_{t-1} - \tilde{\mathbf{y}}_{t-1}^\infty\|^{2\alpha} \\ &\quad + 5 \|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_*^2. \end{aligned}$$

where (a) follows from the Holder's smoothness of  $f$ . Using a similar technique as in the proof of Theorem 4.1, relying on Holders inequality, we get

$$\begin{aligned} \mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_t^\infty\|^{2\alpha} | \tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1}, \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}] &\leq \mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_t^\infty\|^2 | \tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1}, \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}]^\alpha \\ &\leq \Psi_1^{2\alpha} \mathbb{E} [\|\mathbf{x}_t - \mathbf{x}_t^\infty\|_2^2 | \tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1}, \mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}]^\alpha \\ &\stackrel{(a)}{\leq} \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{2\alpha}, \end{aligned}$$

where (a) follows from the fact that conditioned on past randomness,  $\mathbf{x}_t - \mathbf{x}_t^\infty$  is the average of  $m$  i.i.d bounded mean 0 random variables, the variance of which scales as  $O(D^2/m)$ . A similar bound

holds for the expectation of other quantities appearing in the RHS of the above equation. Using this, the regret of  $\mathbf{x}$  player can be upper bounded as

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}_t) - f(\mathbf{x}, \mathbf{y}_t) \right] &\leq L_1 T \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{1+\alpha} + \eta D + \frac{10CL^2 T}{\eta} \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{2\alpha} \\ &\quad + \frac{5C}{2\eta} \sum_{t=1}^T \mathbb{E} [\|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_*^2] \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T \mathbb{E} [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2]. \end{aligned}$$

Similarly, the regret of  $\mathbf{y}$  player can be bounded as

$$\begin{aligned} \sup_{\mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}) - f(\mathbf{x}_t, \mathbf{y}_t) \right] &\leq L_1 T \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{1+\alpha} + \eta D + \frac{10CL^2 T}{\eta} \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{2\alpha} \\ &\quad + \frac{5C}{2\eta} \sum_{t=1}^T \mathbb{E} [\|\nabla_{\mathbf{y}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_*^2] \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T \mathbb{E} [\|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|^2]. \end{aligned}$$

Summing the above two inequalities, we get

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}) - f(\mathbf{x}, \mathbf{y}_t) \right] &\leq 2L_1 T \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{1+\alpha} + 2\eta D + \frac{20CL^2 T}{\eta} \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{2\alpha} \\ &\quad + \frac{5C}{2\eta} \sum_{t=1}^T \mathbb{E} [\|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_*^2] \\ &\quad + \frac{5C}{2\eta} \sum_{t=1}^T \mathbb{E} [\|\nabla_{\mathbf{y}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_*^2] \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T (\mathbb{E} [\|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|^2] + \mathbb{E} [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2]). \end{aligned}$$

From Holder's smoothness assumption on  $f$ , we have

$$\begin{aligned} \mathbb{E} [\|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_*^2] &\leq 2\mathbb{E} [\|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_*^2] \\ &\quad + 2\mathbb{E} [\|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \tilde{\mathbf{y}}_{t-1}^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_*^2] \\ &\stackrel{(a)}{\leq} 2L^2 \mathbb{E} [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^{2\alpha}] + 2L^2 \mathbb{E} [\|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|^{2\alpha}], \end{aligned}$$

Using a similar argument, we get

$$\mathbb{E} [\|\nabla_{\mathbf{y}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_*^2] \leq 2L^2 \mathbb{E} [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^{2\alpha}] + 2L^2 \mathbb{E} [\|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|^{2\alpha}].$$

Plugging this in the previous bound, we get

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}) - f(\mathbf{x}, \mathbf{y}_t) \right] &\leq 2L_1 T \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{1+\alpha} + 2\eta D + \frac{20CL^2 T}{\eta} \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{2\alpha} \\ &\quad + \frac{10CL^2}{\eta} \sum_{t=1}^T (\mathbb{E} [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^{2\alpha}] + \mathbb{E} [\|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|^{2\alpha}]) \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T (\mathbb{E} [\|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|^2] + \mathbb{E} [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2]). \end{aligned}$$

**Case  $\alpha = 1$ .** We first consider the case of  $\alpha = 1$ . In this case, choosing  $\eta > \sqrt{20}CL$ , we get

$$\sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}) - f(\mathbf{x}, \mathbf{y}_t) \right] \leq 2L_1 T \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{1+\alpha} + 2\eta D + \frac{20CL^2 T}{\eta} \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{2\alpha}.$$

**General  $\alpha$ .** The more general case relies on AM-GM inequality. Consider the following

$$\begin{aligned} \frac{10CL^2}{\eta} \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^{2\alpha} &= \left( (2\alpha C)^{\frac{\alpha}{1-\alpha}} \eta^{-\frac{1+\alpha}{1-\alpha}} (10CL^2)^{\frac{1}{1-\alpha}} \right)^{1-\alpha} \left( \frac{\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2}{2\alpha C \eta^{-1}} \right)^\alpha \\ &\stackrel{(a)}{\leq} (1-\alpha) \left( (2\alpha C)^{\frac{\alpha}{1-\alpha}} \eta^{-\frac{1+\alpha}{1-\alpha}} (10CL^2)^{\frac{1}{1-\alpha}} \right) + \frac{\eta}{2C} \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2 \\ &= \sqrt{20}L \left( \frac{\sqrt{20}CL}{\eta} \right)^{\frac{1+\alpha}{1-\alpha}} + \frac{\eta}{2C} \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2 \end{aligned}$$

where (a) follows from AM-GM inequality. Plugging this in the previous bound, we get

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}) - f(\mathbf{x}, \mathbf{y}_t) \right] &\leq 2L_1 T \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{1+\alpha} + 2\eta D \\ &\quad + \frac{20CL^2 T}{\eta} \left( \frac{\Psi_1 \Psi_2 D}{\sqrt{m}} \right)^{2\alpha} + 4\sqrt{5}LT \left( \frac{\sqrt{20}CL}{\eta} \right)^{\frac{1+\alpha}{1-\alpha}}. \end{aligned}$$

The claim of the theorem then follows from the observation that

$$\mathbb{E} \left[ f \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t, \mathbf{y} \right) - f \left( \mathbf{x}, \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t \right) \right] \leq \frac{1}{T} \mathbb{E} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}) - f(\mathbf{x}, \mathbf{y}_t) \right].$$

□

## D.2 Proof of Theorem 5.1

To prove the Theorem, we instantiate Theorem D.1 for the uniform noise distribution. As shown in Corollary 4.1, the predictions of OFTPL are  $dD\eta^{-1}$ -stable in this case. Plugging this in the bound of Theorem D.1 and using the fact that  $\Psi_1 = \Psi_2 = 1$  and  $\alpha = 1$  gives us

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ f \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t, \mathbf{y} \right) - f \left( \mathbf{x}, \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t \right) \right] &\leq 2L \left( \frac{D}{\sqrt{m}} \right)^2 + \frac{2\eta D}{T} \\ &\quad + \frac{20dDL^2}{\eta} \left( \frac{D}{\sqrt{m}} \right)^2 + 10L \left( \frac{5dDL}{\eta} \right)^\infty. \end{aligned}$$

Plugging in  $\eta = 6dD(L+1)$ ,  $m = T$  in the above bound gives us

$$\sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ f \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t, \mathbf{y} \right) - f \left( \mathbf{x}, \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t \right) \right] \leq O \left( \frac{dD^2(L+1)}{T} \right).$$

## E Nonconvex-Nonconcave Games

Our algorithm for nonconvex-nonconcave games is presented in Algorithm 4. Note that in each iteration of this game, both the players play empirical distributions  $(P_t, Q_t)$ . Before presenting the proof of Theorem 5.2, we first present a more general result in Section E.2. Theorem 5.2 immediately follows from our general result by instantiating it for exponential noise distribution.

### E.1 Primal Dual Spaces

In this section, we present some integral probability metrics induced by popular choices of functions spaces  $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ .

---

**Algorithm 4** OFTPL for nonconvex-nonconcave games

---

```

1: Input: Perturbation Distributions  $P_{\text{PRTB}}^1, P_{\text{PRTB}}^2$  of  $\mathbf{x}, \mathbf{y}$  players, number of samples  $m$ , iterations  $T$ 
2: for  $t = 1 \dots T$  do
3:   if  $t = 1$  then
4:     for  $j = 1 \dots m$  do
5:       Sample  $\sigma_{t,j}^1 \sim P_{\text{PRTB}}^1, \sigma_{t,j}^2 \sim P_{\text{PRTB}}^2$ 
6:        $\mathbf{x}_{1,j} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} -\sigma_{1,j}^1(\mathbf{x})$ 
7:        $\mathbf{y}_{1,j} = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \sigma_{1,j}^2(\mathbf{y})$ 
8:     end for
9:     Let  $P_1, Q_1$  be the empirical distributions over  $\{\mathbf{x}_{1,j}\}_{j=1}^m, \{\mathbf{y}_{1,j}\}_{j=1}^m$ 
10:    continue
11:  end if
12:  //Compute guesses
13:  for  $j = 1 \dots m$  do
14:    Sample  $\sigma_{t,j}^1 \sim P_{\text{PRTB}}^1, \sigma_{t,j}^2 \sim P_{\text{PRTB}}^2$ 
15:     $\tilde{\mathbf{x}}_{t-1,j} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \sum_{i=1}^{t-1} f(\mathbf{x}, Q_i) - \sigma_{t,j}^1(\mathbf{x})$ 
16:     $\tilde{\mathbf{y}}_{t-1,j} = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \sum_{i=1}^{t-1} f(P_i, \mathbf{y}) + \sigma_{t,j}^2(\mathbf{y})$ 
17:  end for
18:  Let  $\tilde{P}_{t-1}, \tilde{Q}_{t-1}$  be the empirical distributions over  $\{\tilde{\mathbf{x}}_{t-1,j}\}_{j=1}^m, \{\tilde{\mathbf{y}}_{t-1,j}\}_{j=1}^m$ 
19:  //Use the guesses to compute the next action
20:  for  $j = 1 \dots m$  do
21:    Sample  $\sigma_{t,j}^1 \sim P_{\text{PRTB}}^1, \sigma_{t,j}^2 \sim P_{\text{PRTB}}^2$ 
22:     $\mathbf{x}_{t,j} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} \sum_{i=1}^{t-1} f(\mathbf{x}, Q_i) + f(\mathbf{x}, \tilde{Q}_{t-1}) - \sigma_{t,j}^1(\mathbf{x})$ 
23:     $\mathbf{y}_{t,j} = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} \sum_{i=1}^{t-1} f(P_i, \mathbf{y}) + f(\tilde{P}_{t-1}, \mathbf{y}) + \sigma_{t,j}^2(\mathbf{y})$ 
24:  end for
25:  Let  $P_t, Q_t$  be the empirical distributions over  $\{\mathbf{x}_{t,j}\}_{j=1}^m, \{\mathbf{y}_{t,j}\}_{j=1}^m$ 
26: end for
27: return  $\{(P_t, Q_t)\}_{t=1}^T$ 

```

---

$\gamma_{\mathcal{F}}(P, Q)$	$\ f\ _{\mathcal{F}}$	$\mathcal{F}$
Dudley Metric	$\operatorname{Lip}(f) + \ f\ _{\infty}$	$\{f : \operatorname{Lip}(f) + \ f\ _{\infty} < \infty\}$
Kantorovich Metric (or) Wasserstein-1 Metric	$\operatorname{Lip}(f)$	$\{f : \operatorname{Lip}(f) < \infty\}$
Total Variation (TV) Distance	$\ f\ _{\infty}$	$\{f : \ f\ _{\infty} < \infty\}$
Maximum Mean Discrepancy (MMD) for RKHS $\mathcal{H}$	$\ f\ _{\mathcal{H}}$	$\{f : \ f\ _{\mathcal{H}} < \infty\}$

Table 1: Table showing some popular Integral Probability Metrics. Here  $\operatorname{Lip}(f)$  is the Lipschitz constant of  $f$  which is defined as  $\sup_{\mathbf{x}, \mathbf{y} \in \mathcal{X}} |f(\mathbf{x}) - f(\mathbf{y})| / \|\mathbf{x} - \mathbf{y}\|$  and  $\|f\|_{\infty}$  is the supremum norm of  $f$ .

## E.2 General Result

**Theorem E.1.** Consider the minimax game in Equation (1). Suppose the domains  $\mathcal{X}, \mathcal{Y}$  are compact subsets of  $\mathbb{R}^d$ . Let  $\mathcal{F}, \mathcal{F}'$  be the set of Lipschitz functions over  $\mathcal{X}, \mathcal{Y}$ , and  $\|g_1\|_{\mathcal{F}}, \|g_2\|_{\mathcal{F}'}$  be the Lipschitz constants of functions  $g_1 : \mathcal{X} \rightarrow \mathbb{R}, g_2 : \mathcal{Y} \rightarrow \mathbb{R}$  w.r.t some norm  $\|\cdot\|$ . Suppose  $f$  is such that  $\max\{\sup_{\mathbf{x} \in \mathcal{X}} \|f(\cdot, \mathbf{y})\|_{\mathcal{F}}, \sup_{\mathbf{y} \in \mathcal{Y}} \|f(\mathbf{x}, \cdot)\|_{\mathcal{F}'}\} \leq G$  and satisfies the following smoothness property

$$\begin{aligned} \|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{x}} f(\mathbf{x}', \mathbf{y}')\|_* &\leq L\|\mathbf{x} - \mathbf{x}'\| + L\|\mathbf{y} - \mathbf{y}'\|, \\ \|\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{y}} f(\mathbf{x}', \mathbf{y}')\|_* &\leq L\|\mathbf{x} - \mathbf{x}'\| + L\|\mathbf{y} - \mathbf{y}'\|. \end{aligned}$$

Let  $\mathcal{P}, \mathcal{Q}$  be the set of probability distributions over  $\mathcal{X}, \mathcal{Y}$ . Define diameter of  $\mathcal{P}, \mathcal{Q}$  as  $D = \max\{\sup_{P_1, P_2 \in \mathcal{P}} \gamma_{\mathcal{F}}(P_1, P_2), \sup_{Q_1, Q_2 \in \mathcal{Q}} \gamma_{\mathcal{F}'}(Q_1, Q_2)\}$ . Suppose both  $\mathbf{x}, \mathbf{y}$  players use Algorithm 2 to solve the game. Suppose the perturbation distributions  $P_{\text{PRTB}}^1, P_{\text{PRTB}}^2$ , used by  $\mathbf{x}, \mathbf{y}$  players are such that  $\operatorname{argmin}_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) - \sigma(\mathbf{x}), \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}} f(\mathbf{y}) + \sigma(\mathbf{y})$  have unique optimizers with probability one, for any  $f$  in  $\mathcal{F}, \mathcal{F}'$  respectively. Moreover, suppose  $\mathbb{E}_{\sigma \sim P_{\text{PRTB}}^1} [\|\sigma\|_{\mathcal{F}}] = \mathbb{E}_{\sigma \sim P_{\text{PRTB}}^2} [\|\sigma\|_{\mathcal{F}'}] = \eta$  and predictions of both the players are  $C\eta^{-1}$ -stable w.r.t norms  $\|\cdot\|_{\mathcal{F}}, \|\cdot\|_{\mathcal{F}'}$ . Suppose the guesses

used by  $\mathbf{x}, \mathbf{y}$  players in the  $t^{\text{th}}$  iteration are  $f(\cdot, \tilde{Q}_{t-1}), f(\tilde{P}_{t-1}, \cdot)$ , where  $\tilde{P}_{t-1}, \tilde{Q}_{t-1}$  denote the predictions of  $\mathbf{x}, \mathbf{y}$  players in the  $t^{\text{th}}$  iteration, if guess  $g_t = 0$  was used. Then the iterates  $\{(P_t, Q_t)\}_{t=1}^T$  generated by the Algorithm 3 satisfy the following, for  $\eta > \sqrt{3}CL$

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ f \left( \frac{1}{T} \sum_{t=1}^T P_t, \mathbf{y} \right) - f \left( \mathbf{x}, \frac{1}{T} \sum_{t=1}^T Q_t \right) \right] &= O \left( \frac{\eta D}{T} + \frac{CD^2L^2}{\eta m} \right) \\ &+ O \left( \min \left\{ \frac{dC\Psi_1^2\Psi_2^2G^2 \log(2m)}{\eta m}, \frac{CD^2L^2}{\eta} \right\} \right). \end{aligned}$$

*Proof.* The proof of this Theorem uses similar arguments as Theorem D.1. Since both the players are responding to each others actions using OFTPL, using Theorem 4.2, we get the following regret bounds for the players

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E} \left[ \sum_{t=1}^T f(P_t, Q_t) - f(\mathbf{x}, Q_t) \right] &\leq \eta D + \sum_{t=1}^T \frac{C}{2\eta} \mathbb{E} \left[ \|f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1})\|_{\mathcal{F}}^2 \right] \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T \mathbb{E} \left[ \gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_{t-1}^\infty)^2 \right], \\ \sup_{\mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ \sum_{t=1}^T f(P_t, \mathbf{y}) - f(P_t, Q_t) \right] &\leq \eta D + \sum_{t=1}^T \frac{C}{2\eta} \mathbb{E} \left[ \|f(P_t, \cdot) - f(\tilde{P}_{t-1}, \cdot)\|_{\mathcal{F}'}^2 \right] \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T \mathbb{E} \left[ \gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty)^2 \right], \end{aligned}$$

where  $P_t^\infty, \tilde{P}_{t-1}^\infty, Q_t^\infty, \tilde{Q}_{t-1}^\infty$  are as defined in Theorem 4.2. First, consider the regret of the  $\mathbf{x}$  player. We upper bound  $\|f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1})\|_{\mathcal{F}}^2$  as

$$\begin{aligned} \|f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1})\|_{\mathcal{F}}^2 &\leq 3\|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}}^2 \\ &\quad + 3\|f(\cdot, Q_t^\infty) - f(\cdot, \tilde{Q}_{t-1}^\infty)\|_{\mathcal{F}}^2 \\ &\quad + 3\|f(\cdot, \tilde{Q}_{t-1}^\infty) - f(\cdot, \tilde{Q}_{t-1})\|_{\mathcal{F}}^2. \end{aligned}$$

We now show that  $\mathbb{E} \left[ \|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}}^2 | \tilde{P}_{t-1}, \tilde{Q}_{t-1}, P_{1:t-1}, Q_{1:t-1} \right]$  is  $O(1/m)$ . To simplify the notation, we let  $\zeta_t = \{\tilde{P}_{t-1}, \tilde{Q}_{t-1}, P_{1:t-1}, Q_{1:t-1}\}$ . Let  $\mathcal{N}_\epsilon$  be the  $\epsilon$ -net of  $\mathcal{X}$  w.r.t  $\|\cdot\|$ . Then

$$\begin{aligned} \|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}} &\stackrel{(a)}{\leq} \sup_{\mathbf{x} \in \mathcal{X}} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_* \\ &\stackrel{(b)}{\leq} \sup_{\mathbf{x} \in \mathcal{N}_\epsilon} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_* + 2L\epsilon, \end{aligned}$$

where (a) follows from the definition of Lipschitz constant and (b) follows from our smoothness assumption on  $f$ . Using this, we get

$$\mathbb{E} \left[ \|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}}^2 | \zeta_t \right] \leq 2\mathbb{E} \left[ \sup_{\mathbf{x} \in \mathcal{N}_\epsilon} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_*^2 | \zeta_t \right] + 8L^2\epsilon^2,$$

Since  $f$  is Lipschitz,  $\|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})\|_*$  is bounded by  $G$ . So  $\|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_*$  is bounded by  $2G$  and  $\|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_2$  is bounded by  $2\Psi_1 G$ . Moreover, conditioned on past randomness  $(\zeta_t)$ ,  $\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)$  is a sub-Gaussian random vector and satisfies the following bound

$$\mathbb{E} [\langle \mathbf{u}, \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty) \rangle | \zeta_t] \leq \exp(2\Psi_1^2 G^2 \|\mathbf{u}\|_2^2 / m).$$

From tail bounds of sub-Gaussian random vectors [26], we have

$$\mathbb{P} \left( \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_2^2 > \frac{4\Psi_1^2 G^2}{m} (d + 2\sqrt{ds} + 2s) | \zeta_t \right) \leq e^{-s},$$

for any  $s > 0$ . Using union bound, and the fact that  $\log |\mathcal{N}_\epsilon|$  is upper bounded by  $d \log(1 + 2D/\epsilon)$ , we get

$$\mathbb{P} \left( \sup_{\mathbf{x} \in \mathcal{N}_\epsilon} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_2^2 > \frac{4\Psi_1^2 G^2}{m} (d + 2\sqrt{ds} + 2s) \Big| \zeta_t \right) \leq e^{-s + d \log(1 + 2D/\epsilon)}.$$

Let  $Z = \sup_{\mathbf{x} \in \mathcal{N}_\epsilon} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_2^2$ . The expectation of  $Z$  can be bounded as follows

$$\begin{aligned} \mathbb{E}[Z|\zeta_t] &= \mathbb{P}(Z \leq a|\zeta_t) \mathbb{E}[Z|\zeta_t, Z \leq a] + \mathbb{P}(Z > a|\zeta_t) \mathbb{E}[Z|\zeta_t, Z > a] \\ &\leq a + 4\Psi_1^2 G^2 \mathbb{P}(Z > a|\zeta_t). \end{aligned}$$

Choosing  $\epsilon = Dm^{-1/2}$ ,  $s = 3d \log(1 + 2m^{1/2})$ , and  $a = \frac{44d\Psi_1^2 G^2 \log(1 + 2m^{1/2})}{m}$ , we get

$$\mathbb{E}[Z|\zeta_t] \leq \frac{48d\Psi_1^2 G^2 \log(1 + 2m^{1/2})}{m}.$$

This shows that  $\mathbb{E}[\|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}}^2|\zeta_t] \leq \frac{96d\Psi_1^2 \Psi_2^2 G^2 \log(1 + 2m^{1/2})}{m} + \frac{8D^2 L^2}{m}$ . Note that another trivial upper bound for  $\|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}}$  is  $DL$ , which can be obtained as follows

$$\begin{aligned} \|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}} &= \sup_{\mathbf{x} \in \mathcal{X}} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_* \\ &= \|\mathbb{E}_{\mathbf{y}_1 \sim Q_t, \mathbf{y}_2 \sim Q_t^\infty} [\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}_1) - \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}_2)]\|_* \\ &\stackrel{(a)}{\leq} LD, \end{aligned}$$

where (a) follows from the smoothness assumption on  $f$  and the fact that the diameter of  $\mathcal{X}$  is  $D$ . When  $L$  is close to 0, this bound can be much better than the above bound. So we have

$$\mathbb{E}[\|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}}^2|\zeta_t] \leq \min \left( \frac{96d\Psi_1^2 \Psi_2^2 G^2 \log(1 + 2m^{1/2})}{m} + \frac{8D^2 L^2}{m}, L^2 D^2 \right).$$

Using this, the regret of the  $\mathbf{x}$  player can be bounded as follows

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E} \left[ \sum_{t=1}^T f(P_t, Q_t) - f(\mathbf{x}, Q_t) \right] &\leq \eta D + \frac{24CD^2 L^2 T}{\eta m} \\ &\quad + \min \left( \frac{288dC\Psi_1^2 \Psi_2^2 G^2 T \log(1 + 2m^{1/2})}{\eta m}, \frac{3CD^2 L^2 T}{\eta} \right) \\ &\quad + \sum_{t=1}^T \frac{3C}{2\eta} \mathbb{E} \left[ \|f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1}^\infty)\|_{\mathcal{F}}^2 \right] \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T \mathbb{E} \left[ \gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_{t-1}^\infty)^2 \right]. \end{aligned}$$

A similar analysis shows that the regret of  $\mathbf{y}$  player can be bounded as

$$\begin{aligned} \sup_{\mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ \sum_{t=1}^T f(P_t, \mathbf{y}) - f(P_t, Q_t) \right] &\leq \eta D + \frac{24CD^2 L^2 T}{\eta m} \\ &\quad + \min \left( \frac{288dC\Psi_1^2 \Psi_2^2 G^2 T \log(1 + 2m^{1/2})}{\eta m}, \frac{3CD^2 L^2 T}{\eta} \right) \\ &\quad + \sum_{t=1}^T \frac{3C}{2\eta} \mathbb{E} \left[ \|f(P_t^\infty, \cdot) - f(\tilde{P}_{t-1}^\infty, \cdot)\|_{\mathcal{F}'}^2 \right] \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T \mathbb{E} \left[ \gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty)^2 \right], \end{aligned}$$

Summing the above two inequalities, we get

$$\begin{aligned}
\sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ \sum_{t=1}^T f(P_t, \mathbf{y}) - f(P, Q_t) \right] &\leq 2\eta D + \frac{48CD^2L^2T}{\eta m} \\
&+ \min \left( \frac{576dC\Psi_1^2\Psi_2^2G^2T \log(1+2m^{1/2})}{\eta m}, \frac{6CD^2L^2T}{\eta} \right) \\
&+ \sum_{t=1}^T \frac{3C}{2\eta} \mathbb{E} \left[ \|f(\cdot, Q_t^\infty) - f(\cdot, \tilde{Q}_{t-1}^\infty)\|_{\mathcal{F}}^2 \right] \\
&+ \sum_{t=1}^T \frac{3C}{2\eta} \mathbb{E} \left[ \|f(P_t^\infty, \cdot) - f(\tilde{P}_{t-1}^\infty, \cdot)\|_{\mathcal{F}'}^2 \right] \\
&- \frac{\eta}{2C} \sum_{t=1}^T \left( \mathbb{E} \left[ \gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_{t-1}^\infty)^2 \right] + \mathbb{E} \left[ \gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty)^2 \right] \right).
\end{aligned}$$

From our assumption on smoothness of  $f$ , we have

$$\|f(\cdot, Q_t^\infty) - f(\cdot, \tilde{Q}_{t-1}^\infty)\|_{\mathcal{F}} \leq L\gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty), \quad \|f(P_t^\infty, \cdot) - f(\tilde{P}_{t-1}^\infty, \cdot)\|_{\mathcal{F}'} \leq L\gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_{t-1}^\infty).$$

To see this, consider the following

$$\begin{aligned}
\|f(\cdot, Q_t^\infty) - f(\cdot, \tilde{Q}_{t-1}^\infty)\|_{\mathcal{F}} &= \sup_{\mathbf{x} \in \mathcal{X}} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty) - \nabla_{\mathbf{x}} f(\mathbf{x}, \tilde{Q}_{t-1}^\infty)\|_* \\
&= \sup_{\mathbf{x} \in \mathcal{X}, \|\mathbf{u}\| \leq 1} \left\langle \mathbf{u}, \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty) - \nabla_{\mathbf{x}} f(\mathbf{x}, \tilde{Q}_{t-1}^\infty) \right\rangle \\
&= \sup_{\mathbf{x} \in \mathcal{X}, \|\mathbf{u}\| \leq 1} \mathbb{E}_{\mathbf{y} \sim Q_t^\infty} [\langle \mathbf{u}, \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) \rangle] - \mathbb{E}_{\mathbf{y} \sim \tilde{Q}_{t-1}^\infty} [\langle \mathbf{u}, \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) \rangle] \\
&\leq \gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty) \sup_{\mathbf{x} \in \mathcal{X}, \|\mathbf{u}\| \leq 1} \|\langle \mathbf{u}, \nabla_{\mathbf{x}} f(\mathbf{x}, \cdot) \rangle\|_{\mathcal{F}'} \\
&= \gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty) \sup_{\mathbf{x} \in \mathcal{X}, \|\mathbf{u}\| \leq 1} \left( \sup_{\mathbf{y}_1 \neq \mathbf{y}_2 \in \mathcal{Y}} \frac{|\langle \mathbf{u}, \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}_1) \rangle - \langle \mathbf{u}, \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}_2) \rangle|}{\|\mathbf{y}_1 - \mathbf{y}_2\|} \right) \\
&\leq \gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty) \sup_{\mathbf{x} \in \mathcal{X}} \left( \sup_{\mathbf{y}_1 \neq \mathbf{y}_2 \in \mathcal{Y}} \frac{\|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}_1) - \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}_2)\|_*}{\|\mathbf{y}_1 - \mathbf{y}_2\|} \right) \\
&\stackrel{(a)}{\leq} L\gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty),
\end{aligned}$$

where (a) follows from smoothness of  $f$ . Substituting this in the previous equation, and choosing  $\eta > \sqrt{3}CL$ , we get

$$\begin{aligned}
\sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ \sum_{t=1}^T f(P_t, \mathbf{y}) - f(P, Q_t) \right] &\leq 2\eta D + \frac{48CD^2L^2T}{\eta m} \\
&+ \min \left( \frac{576dC\Psi_1^2\Psi_2^2G^2T \log(1+2m^{1/2})}{\eta m}, \frac{6CD^2L^2T}{\eta} \right)
\end{aligned}$$

This finishes the proof of the Theorem.  $\square$

**Remark E.1.** We note that a similar result can be obtained for other choice of function classes such as the set of all bounded and Lipschitz functions. The only difference between proving such a result vs. proving Theorem E.1 is in bounding  $\|f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1})\|_{\mathcal{F}}$ .

### E.3 Proof of Theorem 5.2

To prove the Theorem, we instantiate Theorem E.1 for exponential noise distribution. Recall, in Corollary 4.2, we showed that  $\mathbb{E}_\sigma [\|\sigma\|_{\mathcal{F}}] = \eta \log d$  and OFTPL is  $O(d^2D\eta^{-1})$  stable w.r.t  $\|\cdot\|_{\mathcal{F}}$ ,

for this choice of perturbation distribution (similar results hold for  $(\mathcal{F}', \|\cdot\|_{\mathcal{F}'})$ ). Substituting this in the bounds of Theorem E.1 and using the fact that  $\Psi_1 = \sqrt{d}$ ,  $\Psi_2 = 1$ , we get

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ f \left( \frac{1}{T} \sum_{t=1}^T P_t, \mathbf{y} \right) - f \left( \mathbf{x}, \frac{1}{T} \sum_{t=1}^T Q_t \right) \right] &= O \left( \frac{\eta D \log d}{T} + \frac{d^2 D^3 L^2}{\eta m} \right) \\ &+ O \left( \min \left\{ \frac{d^4 D G^2 \log(2m)}{\eta m}, \frac{d^2 D^3 L^2}{\eta} \right\} \right). \end{aligned}$$

Choosing  $\eta = 10d^2 D(L+1)$ ,  $m = T$ , we get

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ f \left( \frac{1}{T} \sum_{t=1}^T P_t, \mathbf{y} \right) - f \left( \mathbf{x}, \frac{1}{T} \sum_{t=1}^T Q_t \right) \right] &= O \left( \frac{d^2 D^2 (L+1) \log d}{T} \right) \\ &+ O \left( \min \left\{ \frac{d^2 G^2 \log(T)}{LT}, D^2 L \right\} \right). \end{aligned}$$

#### E.4 Regret Bounds of OFTPL in [11]

In this section, we rely on the OFTPL regret bounds of [11] to derive the rate of convergence of Algorithm 4 to a NE of nonconvex-nonconcave games. Our goal is to show that even if we set  $m = \infty$ , the OFTPL regret bounds of [11] only give us  $O(T^{-3/4})$  convergence rates. In contrast, using the tighter regret bounds in Theorem 4.2, we obtain  $O(T^{-1})$  convergence rates.

**Theorem E.2.** *Consider the setting of Theorem 5.2. Suppose OFTPL is run for  $T$  iterations with  $\eta = (d^3 DGL)^{1/2} T^{1/4}$ ,  $m = \infty$ . then the iterates  $\{(P_t, Q_t)\}_{t=1}^T$  satisfy*

$$\sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} f \left( \frac{1}{T} \sum_{t=1}^T P_t, \mathbf{y} \right) - f \left( \mathbf{x}, \frac{1}{T} \sum_{t=1}^T Q_t \right) = O \left( \frac{(d^3 D^3 GL)^{1/2} \log d}{T^{3/4}} \right).$$

*Proof.* From the regret bounds of OFTPL derived by [11] we know that the regret of  $\mathbf{x}$  player can be bounded as

$$\sup_{\mathbf{x} \in \mathcal{X}} \sum_{t=1}^T f(P_t, Q_t) - f(\mathbf{x}, Q_t) \leq O \left( \eta D \log d + \sum_{t=1}^T \frac{d^2 D}{\eta} \|f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1})\|_{\mathcal{F}}^2 \right),$$

where  $\|f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1})\|_{\mathcal{F}}$  is the Lipschitz constant of  $f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1})$ , which can be upper bounded as follows

$$\begin{aligned} \|f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1})\|_{\mathcal{F}} &\leq L \mathbb{E}_{\sigma} [\|\mathbf{y}_t(\sigma) - \tilde{\mathbf{y}}_{t-1}(\sigma)\|_1] \quad (\text{smoothness of } F) \\ &\leq O \left( \frac{LGd^2 D}{\eta} \right) \quad (\text{stability}), \end{aligned}$$

where the last inequality follows from the stability of FTPL proved in Theorem 1 of Suggala and Netrapalli [11]. Plugging this in the regret bound of the  $\mathbf{x}$  player, we get

$$\sup_{\mathbf{x} \in \mathcal{X}} \left( \sum_{t=1}^T f(P_t, Q_t) - f(\mathbf{x}, Q_t) \right) \leq O \left( \frac{d^6 D^3 G^2 L^2 T}{\eta^3} + \eta D \log d \right).$$

A similar argument shows that the regret of  $\mathbf{y}$  player can be bounded as

$$\sup_{\mathbf{y} \in \mathcal{Y}} \left( \sum_{t=1}^T f(P_t, \mathbf{y}) - f(P_t, Q_t) \right) \leq O \left( \frac{d^6 D^3 G^2 L^2 T}{\eta^3} + \eta D \log d \right).$$

Summing the above two regret bounds and substituting  $\eta = (d^3 DGL)^{1/2} T^{1/4}$  gives us the required bound.  $\square$

## F Choice of Perturbation Distributions

**Regularization of some Perturbation Distributions.** We first study the regularization effect of various perturbation distributions. Table 2 presents the regularizer  $R$  corresponding to some commonly used perturbation distributions, when the action space  $\mathcal{X}$  is  $\ell_\infty$  ball of radius 1 centered at origin.

Perturbation Distribution $P_{\text{PTB}}$	Regularizer
Uniform over $[0, \eta]^d$	$\eta \ \mathbf{x} - 1\ _2^2$
Exponential $P(\sigma > t) = \exp(-t/\eta)$	$\sum_i \eta(\mathbf{x}_i + 1) [\log(\mathbf{x}_i + 1) - (1 + \log 2)]$
Gaussian $P(\sigma = t) \propto e^{-t^2/2\eta^2}$	$\sum_i \sup_{u \in \mathbb{R}} u [\mathbf{x}_i - 1 + 2F(-u/\eta)]$

Table 2: Regularizers corresponding to various perturbation distributions used in FTPL when the action space  $\mathcal{X}$  is  $\ell_\infty$  ball of radius 1 centered at origin. Here,  $F$  is the CDF of a standard normal random variable.

**Dimension independent rates.** Recall, the OFTPL algorithm described in Algorithm 3 converges at  $O(d/T)$  rate to a Nash equilibrium of smooth convex-concave games (see Theorem 5.1). We now show that for certain constraint sets  $\mathcal{X}, \mathcal{Y}$ , by choosing the perturbation distributions appropriately, the dimension dependence in the rates can *potentially* be removed.

Suppose the action set is  $\mathcal{X} = \{\mathbf{x} : \|\mathbf{x}\|_2 \leq 1\}$ . Suppose the perturbation distribution  $P_{\text{PTB}}$  is the multivariate Gaussian distribution with mean 0 and covariance  $\eta^2 I_{d \times d}$ , where  $I_{d \times d}$  is the identity matrix. We now try to explicitly compute the regularizer corresponding to this perturbation distribution and action set. Define function  $\Psi$  as

$$\Psi(f) = \mathbb{E}_\sigma \left[ \max_{\mathbf{x} \in \mathcal{X}} \langle f + \sigma, \mathbf{x} \rangle \right] = \mathbb{E}_\sigma [\|f + \sigma\|_2].$$

As shown in Proposition 3.1, the regularizer  $R$  corresponding to any perturbation distribution is given by the Fenchel conjugate of  $\Psi$

$$R(\mathbf{x}) = \sup_f \langle f, \mathbf{x} \rangle - \Psi(f).$$

Since getting an exact expression for  $R$  is a non-trivial task, we only compute an *approximate expression* for  $R$ . Consider the high dimensional setting (*i.e.*, very large  $d$ ). In this setting,  $\|f + \sigma\|_2$ , for  $\sigma$  drawn from  $\mathcal{N}(0, \eta^2 I_{d \times d})$ , can be approximated as follows

$$\begin{aligned} \|f + \sigma\|_2 &= \sqrt{\|f\|_2^2 + \|\sigma\|_2^2 + 2\langle f, \sigma \rangle} \\ &\stackrel{(a)}{\approx} \sqrt{\|f\|_2^2 + \eta^2 d + 2\langle f, \sigma \rangle} \\ &\stackrel{(b)}{\approx} \sqrt{\|f\|_2^2 + \eta^2 d} \end{aligned}$$

where (a) follows from the fact that  $\|\sigma\|_2^2$  is highly concentrated around  $\eta^2 d$  [26]. To be precise

$$\mathbb{P}(\|\sigma\|_2^2 \geq \eta^2(d + 2\sqrt{dt} + 2t)) \leq e^{-t}.$$

A similar bound holds for the lower tail. Approximation (b) follows from the fact that  $\langle f, \sigma \rangle$  is a Gaussian random variable with mean 0 and variance  $\eta^2 \|f\|_2^2$ , and with high probability its magnitude is upper bounded by  $\tilde{O}(\eta \|f\|_2)$ . Since  $\eta \|f\|_2 \ll \sqrt{d} \eta \|f\|_2 \leq \|f\|_2^2 + \eta^2 d$ , approximation (b) holds. This shows that  $\Psi(f)$  can be approximated as

$$\Psi(f) \approx \sqrt{\|f\|_2^2 + \eta^2 d}.$$

Using this approximation, we now compute the regularizer corresponding to the perturbation distribution

$$R(\mathbf{x}) = \sup_f \langle f, \mathbf{x} \rangle - \Psi(f) \approx \sup_f \langle f, \mathbf{x} \rangle - \sqrt{\|f\|_2^2 + \eta^2 d} = -\eta \sqrt{d} \sqrt{1 - \|\mathbf{x}\|_2^2}.$$

This shows that  $R$  is  $\eta\sqrt{d}$ -strongly convex w.r.t  $\|\cdot\|_2$  norm. Following duality between strong convexity and strong smoothness,  $\Psi(f)$  is  $(\eta^2 d)^{-1/2}$  strongly smooth w.r.t  $\|\cdot\|_2$  norm and satisfies

$$\|\nabla\Psi(f_1) - \nabla\Psi(f_2)\|_2 \leq (\eta^2 d)^{-1/2} \|f_1 - f_2\|_2.$$

This shows that the predictions of OFTPL are  $(\eta^2 d)^{-1/2}$  stable w.r.t  $\|\cdot\|_2$  norm. We now instantiate Theorem D.1 for this perturbation distribution and for constraint sets which are unit balls centered at origin, and use the above stability bound, together with the fact that  $\mathbb{E}_\sigma [\|\sigma\|_2] \approx \eta\sqrt{d}$ . Suppose  $f$  is smooth w.r.t  $\|\cdot\|_2$  norm and satisfies

$$\|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{x}} f(\mathbf{x}', \mathbf{y}')\|_2 + \|\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{y}} f(\mathbf{x}', \mathbf{y}')\|_2 \leq L\|\mathbf{x} - \mathbf{x}'\|_2 + L\|\mathbf{y} - \mathbf{y}'\|_2.$$

Then Theorem D.1 gives us the following rates of convergence to a NE

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \mathbb{E} \left[ f \left( \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t, \mathbf{y} \right) - f \left( \mathbf{x}, \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t \right) \right] &\leq \frac{2L_1}{m} + \frac{2\eta\sqrt{d}}{T} \\ &+ \frac{20L^2}{\eta\sqrt{d}} \left( \frac{1}{m} \right) + 10L \left( \frac{5L}{\eta\sqrt{d}} \right)^\infty \end{aligned}$$

Choosing  $\eta = 6L/\sqrt{d}$ ,  $m = T$ , we get  $O\left(\frac{L}{T}\right)$  rate of convergence. Although, these rates are dimension independent, we note that our stability bound is only approximate. More accurate analysis is needed to actually claim that Algorithm 3 achieves dimension independent rates in this setting. That being said, for general constraints sets, we believe one can get dimension independent rates by choosing the perturbation distribution appropriately.

## G High Probability Bounds

In this section, we provide high probability bounds for Theorems 4.1, 5.1. Our results rely on the following concentration inequalities.

**Proposition G.1** (Jin et al. [27]). *Let  $X_1, \dots, X_K$  be  $K$  independent mean 0 vector-valued random variables such that  $\|X_i\|_2 \leq B_i$ . Then*

$$\mathbb{P} \left( \left\| \sum_{i=1}^K X_i \right\|_2 \geq t \right) \leq 2 \exp \left( -c \frac{t^2}{\sum_{i=1}^K B_i^2} \right),$$

where  $c > 0$  is a universal constant.

We also need the following concentration inequality for martingales.

**Proposition G.2** (Wainwright [28]). *Let  $X_1, \dots, X_K \in \mathbb{R}$  be a martingale difference sequence, where  $\mathbb{E}[X_i | \mathcal{F}_{i-1}] = 0$ . Assume that  $X_i$  satisfy the following tail condition, for some scalar  $B_i > 0$*

$$\mathbb{P} \left( \left| \frac{X_i}{B_i} \right| \geq z \mid \mathcal{F}_{i-1} \right) \leq 2 \exp(-z^2).$$

Then

$$\mathbb{P} \left( \left| \sum_{i=1}^K X_i \right| \geq z \right) \leq 2 \exp \left( -c \frac{z^2}{\sum_{i=1}^K B_i^2} \right),$$

where  $c > 0$  is a universal constant.

### G.1 Online Convex Learning

In this section, we present a high probability version of Theorem 4.1.

**Theorem G.1.** *Suppose the perturbation distribution  $P_{\text{PRTB}}$  is absolutely continuous w.r.t Lebesgue measure. Let  $D$  be the diameter of  $\mathcal{X}$  w.r.t  $\|\cdot\|$ , which is defined as  $D = \sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}} \|\mathbf{x}_1 - \mathbf{x}_2\|$ . Let  $\eta = \mathbb{E}_\sigma [\|\sigma\|_*]$ , and suppose the predictions of OFTPL are  $C\eta^{-1}$ -stable w.r.t  $\|\cdot\|_*$ , where  $C$  is a constant that depends on the set  $\mathcal{X}$ . Suppose, the sequence of loss functions  $\{f_t\}_{t=1}^T$  are  $G$ -Lipschitz*

w.r.t  $\|\cdot\|$  and satisfy  $\sup_{\mathbf{x} \in \mathcal{X}} \|\nabla f_t(\mathbf{x})\|_* \leq G$ . Moreover, suppose  $\{f_t\}_{t=1}^T$  are Holder smooth and satisfy

$$\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \quad \|\nabla f_t(\mathbf{x}_1) - \nabla f_t(\mathbf{x}_2)\|_* \leq L\|\mathbf{x}_1 - \mathbf{x}_2\|^\alpha,$$

for some constant  $\alpha \in [0, 1]$ . Then the regret of Algorithm 1 satisfies the following with probability at least  $1 - \delta$

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}} \sum_{t=1}^T f_t(\mathbf{x}_t) - f_t(\mathbf{x}) &\leq \eta D + \sum_{t=1}^T \frac{C}{2\eta} \|\nabla_t - g_t\|_*^2 - \sum_{t=1}^T \frac{\eta}{2C} \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2 \\ &\quad + cGD\sqrt{\frac{T \log 2/\delta}{m}} + cLT \left( \frac{\Psi_1^2 \Psi_2^2 D^2 \log 4T/\delta}{m} \right)^{\frac{1+\alpha}{2}}, \end{aligned}$$

where  $c$  is a universal constant,  $\mathbf{x}_t^\infty = \mathbb{E}[\mathbf{x}_t | g_t, f_{1:t-1}, \mathbf{x}_{1:t-1}]$  and  $\tilde{\mathbf{x}}_{t-1}^\infty = \mathbb{E}[\tilde{\mathbf{x}}_{t-1} | f_{1:t-1}, \mathbf{x}_{1:t-1}]$  and  $\tilde{\mathbf{x}}_{t-1}$  denotes the prediction in the  $t^{\text{th}}$  iteration of Algorithm 1, if guess  $g_t = 0$  was used. Here,  $\Psi_1, \Psi_2$  denote the norm compatibility constants of  $\|\cdot\|$ .

*Proof.* Our proof uses the same notation and similar arguments as in the proof Theorem 4.1. Recall, in Theorem 4.1 we showed that the regret of OFTPL is upper bounded by

$$\begin{aligned} \sum_{t=1}^T f_t(\mathbf{x}_t) - f_t(\mathbf{x}) &\leq \sum_{t=1}^T \langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle + \eta D + \sum_{t=1}^T \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\| \|\nabla_t - g_t\|_* \\ &\quad - \frac{\eta}{2C} \sum_{t=1}^T (\|\tilde{\mathbf{x}}_t^\infty - \mathbf{x}_t^\infty\|^2 + \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2) \\ &\leq \sum_{t=1}^T \langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t \rangle + \eta D + \sum_{t=1}^T \frac{C}{2\eta} \|\nabla_t - g_t\|_*^2 - \sum_{t=1}^T \frac{\eta}{2C} \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2. \end{aligned}$$

From Holder's smoothness assumption, we have

$$\langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla_t - \nabla f_t(\mathbf{x}_t^\infty) \rangle \leq L\|\mathbf{x}_t - \mathbf{x}_t^\infty\|^{1+\alpha}.$$

Substituting this in the previous bound gives us

$$\begin{aligned} \sum_{t=1}^T f_t(\mathbf{x}_t) - f_t(\mathbf{x}) &\leq \underbrace{\sum_{t=1}^T \langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla f_t(\mathbf{x}_t^\infty) \rangle}_{S_1} + \sum_{t=1}^T L \underbrace{\|\mathbf{x}_t - \mathbf{x}_t^\infty\|^{1+\alpha}}_{S_2} + \eta D \\ &\quad + \sum_{t=1}^T \frac{C}{2\eta} \|\nabla_t - g_t\|_*^2 - \sum_{t=1}^T \frac{\eta}{2C} \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2. \end{aligned}$$

We now provide high probability bounds for  $S_1$  and  $S_2$ .

**Bounding  $S_1$ .** Let  $\xi_i = \{g_{i+1}, f_{i+1}, \mathbf{x}_i\}$  and let  $\xi_{0:t}$  denote the union of sets  $\xi_0, \xi_1, \dots, \xi_t$ . Let  $\zeta_t = \langle \mathbf{x}_t - \mathbf{x}_t^\infty, \nabla f_t(\mathbf{x}_t^\infty) \rangle$  with  $\zeta_0 = 0$ . Note that  $\{\zeta_t\}_{t=0}^T$  is a martingale difference sequence w.r.t  $\xi_{0:T}$ . This is because  $\mathbb{E}[\mathbf{x}_t | \xi_{0:t-1}] = \mathbf{x}_t^\infty$  and  $\nabla f_t(\mathbf{x}_t^\infty)$  is a deterministic quantity conditioned on  $\xi_{0:t-1}$ . As a result  $\mathbb{E}[\zeta_t | \xi_{0:t-1}] = 0$ . Moreover, conditioned on  $\xi_{0:t-1}$ ,  $\zeta_t$  is the average of  $m$  independent mean 0 random variables, each of which is bounded by  $GD$ . Using Proposition G.1, we get

$$\mathbb{P}\left(|\zeta_t| \geq s \mid \xi_{0:t-1}\right) \leq 2 \exp\left(-\frac{ms^2}{G^2 D^2}\right).$$

Using Proposition G.2 on the martingale difference sequence  $\{\zeta_t\}_{t=0}^T$ , we get

$$\mathbb{P}\left(\left|\sum_{t=1}^T \zeta_t\right| \geq s\right) \leq 2 \exp\left(-c \frac{ms^2}{G^2 D^2 T}\right),$$

where  $c > 0$  is a universal constant. This shows that with probability at least  $1 - \delta/2$ ,  $S_1$  is upper bounded by  $O\left(\sqrt{\frac{G^2 D^2 T \log \frac{2}{\delta}}{m}}\right)$ .

**Bounding  $S_2$ .** Conditioned on  $\{g_t, f_{1:t-1}, \mathbf{x}_{1:t-1}\}$ ,  $\mathbf{x}_t - \mathbf{x}_t^\infty$  is the average of  $m$  independent mean 0 random variables which are bounded by  $D$  in  $\|\cdot\|$  norm. From our definition of norm compatibility constant  $\Psi_2$ , this implies the random variables are bounded by  $\Psi_2 D$  in  $\|\cdot\|_2$ . Using Proposition G.1, we get

$$\mathbb{P}\left(\|\mathbf{x}_t - \mathbf{x}_t^\infty\|_2 \geq \Psi_2 D \sqrt{\frac{c \log 4T/\delta}{m}} \mid g_t, f_{1:t-1}, \mathbf{x}_{1:t-1}\right) \leq \frac{\delta}{2T}.$$

Since the above bound holds for any set of  $\{g_t, f_{1:t}, \mathbf{x}_{1:t-1}\}$ , the same tail bound also holds without the conditioning. This shows that

$$\mathbb{P}\left(\|\mathbf{x}_t - \mathbf{x}_t^\infty\|^{1+\alpha} \geq \left(\frac{c\Psi_1^2\Psi_2^2 D^2 \log 4T/\delta}{m}\right)^{\frac{1+\alpha}{2}}\right) \leq \frac{\delta}{2T},$$

where we converted back to  $\|\cdot\|$  by introducing the norm compatibility constant  $\Psi_1$ .

**Bounding the regret.** Plugging the above high probability bounds for  $S_1, S_2$  in the previous regret bound and using union bound, we get the following regret bound which holds with probability at least  $1 - \delta$

$$\begin{aligned} \sum_{t=1}^T f_t(\mathbf{x}_t) - f_t(\mathbf{x}) &\leq cGD\sqrt{\frac{T \log 2/\delta}{m}} + cLT \left(\frac{\Psi_1^2\Psi_2^2 D^2 \log 4T/\delta}{m}\right)^{\frac{1+\alpha}{2}} + \eta D \\ &\quad + \sum_{t=1}^T \frac{C}{2\eta} \|\nabla_t - g_t\|_*^2 - \sum_{t=1}^T \frac{\eta}{2C} \|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|^2, \end{aligned}$$

where  $c > 0$  is a universal constant. □

## G.2 Convex-Concave Games

In this section, we present a high probability version of Theorem 5.1.

**Theorem G.2.** Consider the minimax game in Equation (1). Suppose both the domains  $\mathcal{X}, \mathcal{Y}$  are compact subsets of  $\mathbb{R}^d$ , with diameter  $D = \max\{\sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}} \|\mathbf{x}_1 - \mathbf{x}_2\|_2, \sup_{\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{Y}} \|\mathbf{y}_1 - \mathbf{y}_2\|_2\}$ . Suppose  $f$  is convex in  $\mathbf{x}$ , concave in  $\mathbf{y}$  and is Lipschitz w.r.t  $\|\cdot\|_2$  and satisfies

$$\max\left\{\sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})\|_2, \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \|\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})\|_2\right\} \leq G.$$

Moreover, suppose  $f$  is smooth w.r.t  $\|\cdot\|_2$

$$\|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{x}} f(\mathbf{x}', \mathbf{y}')\|_2 + \|\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{y}} f(\mathbf{x}', \mathbf{y}')\|_2 \leq L\|\mathbf{x} - \mathbf{x}'\|_2 + L\|\mathbf{y} - \mathbf{y}'\|_2.$$

Suppose Algorithm 3 is used to solve the minimax game. Suppose the perturbation distributions used by both the players are the same and equal to the uniform distribution over  $\{\mathbf{x} : \|\mathbf{x}\|_2 \leq (1 + d^{-1})\eta\}$ . Suppose the guesses used by  $\mathbf{x}, \mathbf{y}$  players in the  $t^{\text{th}}$  iteration are  $\nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1}), \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1})$ , where  $\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1}$  denote the predictions of  $\mathbf{x}, \mathbf{y}$  players in the  $t^{\text{th}}$  iteration, if guess  $g_t = 0$  was used. If Algorithm 3 is run with  $\eta = 6dD(L+1)$ ,  $m = T$ , then the iterates  $\{(\mathbf{x}_t, \mathbf{y}_t)\}_{t=1}^T$  satisfy the following bound with probability at least  $1 - \delta$

$$\sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \left[ f\left(\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t, \mathbf{y}\right) - f\left(\mathbf{x}, \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t\right) \right] = O\left(\frac{GD\sqrt{\log \frac{8}{\delta}}}{T} + \frac{D^2(L+1)(d + \log \frac{16T}{\delta})}{T}\right).$$

*Proof.* We use the same notation and proof technique as Theorems D.1, 5.1. From Theorem 4.1 we know that the predictions of OFTPL are  $dD\eta^{-1}$  stable w.r.t  $\|\cdot\|_2$ , for the particular perturbation distribution we consider here. We use this stability bound in our proof. From Theorem G.1, we have

the following regret bound for both the players, which holds with probability at least  $1 - \delta/2$

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}_t) - f(\mathbf{x}, \mathbf{y}_t) \right] &\leq cGD \sqrt{\frac{T \log 8/\delta}{m}} + cLT \left( \frac{D^2 \log 16T/\delta}{m} \right) + \eta D \\ &\quad + \frac{dD}{2\eta} \sum_{t=1}^T [\|\nabla_{\mathbf{x}} f(\mathbf{x}_t, \mathbf{y}_t) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1})\|_2^2] \\ &\quad - \frac{\eta}{2dD} \sum_{t=1}^T [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|_2^2]. \end{aligned}$$

$$\begin{aligned} \sup_{\mathbf{y} \in \mathcal{Y}} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}) - f(\mathbf{x}_t, \mathbf{y}_t) \right] &\leq cGD \sqrt{\frac{T \log 8/\delta}{m}} + cLT \left( \frac{D^2 \log 16T/\delta}{m} \right) + \eta D \\ &\quad + \frac{dD}{2\eta} \sum_{t=1}^T [\|\nabla_{\mathbf{y}} f(\mathbf{x}_t, \mathbf{y}_t) - \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1})\|_2^2] \\ &\quad - \frac{\eta}{2dD} \sum_{t=1}^T [\|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|_2^2]. \end{aligned}$$

First, consider the regret of the  $\mathbf{x}$  player. From the proof of Theorem D.1, we have

$$\begin{aligned} \|\nabla_{\mathbf{x}} f(\mathbf{x}_t, \mathbf{y}_t) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}, \tilde{\mathbf{y}}_{t-1})\|_2^2 &\leq 5L^2 \|\mathbf{x}_t - \mathbf{x}_t^\infty\|_2^2 + 5L^2 \|\tilde{\mathbf{x}}_{t-1} - \tilde{\mathbf{x}}_{t-1}^\infty\|_2^2 \\ &\quad + 5L^2 \|\mathbf{y}_t - \mathbf{y}_t^\infty\|_2^2 + 5L^2 \|\tilde{\mathbf{y}}_{t-1} - \tilde{\mathbf{y}}_{t-1}^\infty\|_2^2 \\ &\quad + 5 \|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_2^2. \end{aligned}$$

Moreover, from the proof of Theorem G.1, we know that  $\|\mathbf{x}_t - \mathbf{x}_t^\infty\|_2^2$  satisfies the following tail bound

$$\mathbb{P} \left( \|\mathbf{x}_t - \mathbf{x}_t^\infty\|_2^2 \geq \frac{cD^2 \log 16T/\delta}{m} \right) \leq \frac{\delta}{8T}.$$

Similar bounds hold for the quantities appearing in the regret bound of  $\mathbf{y}$  player. Plugging this in the previous regret bounds, we get the following which hold with probability at least  $1 - \delta$

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}_t) - f(\mathbf{x}, \mathbf{y}_t) \right] &\leq cGD \sqrt{\frac{T \log 8/\delta}{m}} + \left( L + \frac{10dDL^2}{\eta} \right) \left( \frac{cD^2 \log 16T/\delta}{m} \right) T \\ &\quad + \eta D + \frac{5dD}{2\eta} \sum_{t=1}^T [\|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_2^2] \\ &\quad - \frac{\eta}{2dD} \sum_{t=1}^T [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|_2^2]. \end{aligned}$$

$$\begin{aligned} \sup_{\mathbf{y} \in \mathcal{Y}} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}) - f(\mathbf{x}_t, \mathbf{y}_t) \right] &\leq cGD \sqrt{\frac{T \log 8/\delta}{m}} + \left( L + \frac{10dDL^2}{\eta} \right) \left( \frac{cD^2 \log 16T/\delta}{m} \right) T \\ &\quad + \eta D + \frac{5dD}{2\eta} \sum_{t=1}^T [\|\nabla_{\mathbf{y}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_2^2] \\ &\quad - \frac{\eta}{2dD} \sum_{t=1}^T [\|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|_2^2]. \end{aligned}$$

Summing these two regret bounds, we get

$$\begin{aligned}
\sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}) - f(\mathbf{x}, \mathbf{y}_t) \right] &\leq 2cGD \sqrt{\frac{T \log 8/\delta}{m}} + \left( L + \frac{10dDL^2}{\eta} \right) \left( \frac{2cD^2 \log 16T/\delta}{m} \right) T + 2\eta D \\
&+ \frac{10dD}{2\eta} \sum_{t=1}^T [\|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_2^2] \\
&+ \frac{10dD}{2\eta} \sum_{t=1}^T [\|\nabla_{\mathbf{y}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_2^2] \\
&- \frac{\eta}{2dD} \sum_{t=1}^T [\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|_2^2 + \|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|_2^2].
\end{aligned}$$

From Holder's smoothness assumption on  $f$ , we have

$$\begin{aligned}
\|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_2^2 &\leq 2\|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_2^2 \\
&+ 2\|\nabla_{\mathbf{x}} f(\mathbf{x}_t^\infty, \tilde{\mathbf{y}}_{t-1}^\infty) - \nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_2^2 \\
&\leq 2L^2\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|_2^2 + 2L^2\|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|_2^2,
\end{aligned}$$

Using a similar argument, we get

$$\|\nabla_{\mathbf{y}} f(\mathbf{x}_t^\infty, \mathbf{y}_t^\infty) - \nabla_{\mathbf{y}} f(\tilde{\mathbf{x}}_{t-1}^\infty, \tilde{\mathbf{y}}_{t-1}^\infty)\|_2^2 \leq 2L^2\|\mathbf{x}_t^\infty - \tilde{\mathbf{x}}_{t-1}^\infty\|_2^2 + 2L^2\|\mathbf{y}_t^\infty - \tilde{\mathbf{y}}_{t-1}^\infty\|_2^2.$$

Plugging this in the previous bound, and setting  $\eta = 6dD(L+1)$ ,  $m = T$ , we get the following bound which holds with probability at least  $1 - \delta$

$$\sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \left[ \sum_{t=1}^T f(\mathbf{x}_t, \mathbf{y}) - f(\mathbf{x}, \mathbf{y}_t) \right] \leq O \left( GD \sqrt{\log \frac{8}{\delta}} + D^2(L+1) \left( d + \log \frac{16T}{\delta} \right) \right).$$

□

### G.3 Nonconvex-Nonconcave Games

In this section, we present a high probability version of Theorem 5.2.

**Theorem G.3.** *Consider the minimax game in Equation (1). Suppose the domains  $\mathcal{X}, \mathcal{Y}$  are compact subsets of  $\mathbb{R}^d$  with diameter  $D = \max\{\sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}} \|\mathbf{x}_1 - \mathbf{x}_2\|_1, \sup_{\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{Y}} \|\mathbf{y}_1 - \mathbf{y}_2\|_1\}$ . Suppose  $f$  is Lipschitz w.r.t  $\|\cdot\|_1$  and satisfies*

$$\max \left\{ \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})\|_\infty, \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \|\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})\|_\infty \right\} \leq G.$$

Moreover, suppose  $f$  satisfies the following smoothness property

$$\|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{x}} f(\mathbf{x}', \mathbf{y}')\|_\infty + \|\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) - \nabla_{\mathbf{y}} f(\mathbf{x}', \mathbf{y}')\|_\infty \leq L\|\mathbf{x} - \mathbf{x}'\|_1 + L\|\mathbf{y} - \mathbf{y}'\|_1.$$

Suppose both  $\mathbf{x}$  and  $\mathbf{y}$  players use Algorithm 4 to solve the game with linear perturbation functions  $\sigma(\mathbf{z}) = \langle \bar{\sigma}, \mathbf{z} \rangle$ , where  $\bar{\sigma} \in \mathbb{R}^d$  is such that each of its entries is sampled independently from  $\text{Exp}(\eta)$ . Suppose the guesses used by  $\mathbf{x}$  and  $\mathbf{y}$  players in the  $t^{\text{th}}$  iteration are  $f(\cdot, \hat{Q}_{t-1}), f(\hat{P}_{t-1}, \cdot)$ , where  $\hat{P}_{t-1}, \hat{Q}_{t-1}$  denote the predictions of  $\mathbf{x}, \mathbf{y}$  players in the  $t^{\text{th}}$  iteration, if guess  $g_t = 0$  was used. If Algorithm 4 is run with  $\eta = 10d^2D(L+1)$ ,  $m = T$ , then the iterates  $\{(P_t, Q_t)\}_{t=1}^T$  satisfy the following with probability at least  $1 - \delta$

$$\begin{aligned}
\sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \sum_{t=1}^T f(P_t, \mathbf{y}) - f(\mathbf{x}, Q_t) &= O \left( \frac{d^2 D^2 (L+1) \log d}{T} + \frac{GD}{T} \sqrt{\log \frac{8}{\delta}} \right) \\
&+ O \left( \min \left\{ D^2 L, \frac{d^2 G^2 \log T + dG^2 \log \frac{8}{\delta}}{LT} \right\} \right).
\end{aligned}$$

*Proof.* We use the same notation used in the proofs of Theorems 4.2, E.1. Let  $\mathcal{F}, \mathcal{F}'$  be the set of Lipschitz functions over  $\mathcal{X}, \mathcal{Y}$ , and  $\|g_1\|_{\mathcal{F}}, \|g_2\|_{\mathcal{F}'}$  be the Lipschitz constants of functions  $g_1 : \mathcal{X} \rightarrow \mathbb{R}, g_2 : \mathcal{Y} \rightarrow \mathbb{R}$  w.r.t  $\|\cdot\|_1$ . Recall, in Corollary 4.2 we showed that for our choice of perturbation distribution,  $\mathbb{E}_\sigma [\|\sigma\|_{\mathcal{F}}] = \eta \log d$  and OFTPL is  $O(d^2 D \eta^{-1})$  stable. We use this in our proof.

From Theorem 4.2, we know that the regret of  $\mathbf{x}, \mathbf{y}$  players satisfy

$$\begin{aligned} \sum_{t=1}^T f(P_t, Q_t) - f(\mathbf{x}, Q_t) &\leq \eta D \log d + \underbrace{\sum_{t=1}^T \langle P_t - P_t^\infty, f(\cdot, Q_t) \rangle}_{S_1} \\ &\quad + \underbrace{\sum_{t=1}^T \frac{cd^2 D}{2\eta} \|f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1})\|_{\mathcal{F}}^2}_{S_2} \\ &\quad - \sum_{t=1}^T \frac{\eta}{2cd^2 D} \gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_{t-1}^\infty)^2 \\ \sum_{t=1}^T f(P_t, \mathbf{y}) - f(P_t, Q_t) &\leq \eta D \log d + \sum_{t=1}^T \langle Q_t - Q_t^\infty, f(P_t, \cdot) \rangle \\ &\quad + \sum_{t=1}^T \frac{cd^2 D}{2\eta} \|f(P_t, \cdot) - f(\tilde{P}_{t-1}, \cdot)\|_{\mathcal{F}'}^2 \\ &\quad - \sum_{t=1}^T \frac{\eta}{2cd^2 D} \gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty)^2, \end{aligned}$$

where  $c > 0$  is a positive constant. We now provide high probability bounds for  $S_1, S_2$ .

**Bounding  $S_1$ .** Let  $\xi_i = \{\tilde{P}_i, \tilde{Q}_i, P_i, Q_{i+1}\}$  with  $\xi_0 = \{Q_1\}$  and let  $\xi_{0:t}$  denote the union of sets  $\xi_0, \dots, \xi_t$ . Let  $\zeta_t = \langle P_t - P_t^\infty, f(\cdot, Q_t) \rangle$  with  $\zeta_0 = 0$ . Note that  $\{\zeta_t\}_{t=0}^T$  is a martingale difference sequence w.r.t  $\xi_{0:T}$ . This is because  $\mathbb{E}[P_t | \xi_{0:t-1}] = P_t^\infty$  and  $f(\cdot, Q_t)$  is a deterministic quantity conditioned on  $\xi_{0:t-1}$ . As a result  $\mathbb{E}[\zeta_t | \xi_{0:t-1}] = 0$ . Moreover, conditioned on  $\xi_{0:t-1}$ ,  $\zeta_t$  is the average of  $m$  independent mean 0 random variables, each of which is bounded by  $2GD$ . Using Proposition G.1, we get

$$\mathbb{P}\left(|\zeta_t| \geq s \mid \xi_{0:t-1}\right) \leq 2 \exp\left(-\frac{ms^2}{4G^2 D^2}\right).$$

Using Proposition G.2 on the martingale difference sequence  $\{\zeta_t\}_{t=0}^T$ , we get

$$\mathbb{P}\left(\left|\sum_{t=1}^T \zeta_t\right| \geq s\right) \leq 2 \exp\left(-c \frac{ms^2}{G^2 D^2 T}\right),$$

where  $c > 0$  is a universal constant. This shows that with probability at least  $1 - \delta/8$ ,  $S_1$  is upper bounded by  $O\left(\sqrt{\frac{G^2 D^2 T \log \frac{8}{\delta}}{m}}\right)$ .

**Bounding  $S_2$ .** We upper bound  $S_2$  as

$$\begin{aligned} \|f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1})\|_{\mathcal{F}}^2 &\leq 3\|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}}^2 \\ &\quad + 3\|f(\cdot, Q_t^\infty) - f(\cdot, \tilde{Q}_{t-1}^\infty)\|_{\mathcal{F}}^2 \\ &\quad + 3\|f(\cdot, \tilde{Q}_{t-1}^\infty) - f(\cdot, \tilde{Q}_{t-1})\|_{\mathcal{F}}^2. \end{aligned}$$

We first provide a high probability bound for  $\|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}}^2$ . A trivial bound for this quantity is  $L^2 D^2$ , which can be obtained as follows

$$\begin{aligned} \|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}} &= \sup_{\mathbf{x} \in \mathcal{X}} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_\infty \\ &= \|\mathbb{E}_{\mathbf{y}_1 \sim Q_t, \mathbf{y}_2 \sim Q_t^\infty} [\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}_1) - \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}_2)]\|_\infty \\ &\stackrel{(a)}{\leq} LD, \end{aligned}$$

where (a) follows from the smoothness assumption on  $f$  and the fact that the diameter of  $\mathcal{X}$  is  $D$ . A better bound for this quantity can be obtained as follows. From proof of Theorem E.1, we have

$$\|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}}^2 \leq 2 \sup_{\mathbf{x} \in \mathcal{N}_\epsilon} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_\infty^2 + 8L^2\epsilon^2.$$

where  $\mathcal{N}_\epsilon$  be the  $\epsilon$ -net of  $\mathcal{X}$  w.r.t  $\|\cdot\|$ . Recall, in the proof of Theorem E.1, we showed the following high probability bound for the RHS quantity

$$\mathbb{P}\left(\sup_{\mathbf{x} \in \mathcal{N}_\epsilon} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_2^2 > \frac{4dG^2}{m}(d + 2\sqrt{ds} + 2s)\right) \leq e^{-s+d\log(1+2D/\epsilon)}.$$

Choosing  $\epsilon = Dm^{-1/2}$ ,  $s = \log \frac{8}{\delta} + d\log(1 + 2m^{1/2})$ , we get the following bound for  $\sup_{\mathbf{x} \in \mathcal{N}_\epsilon} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_2^2$  which holds with probability at least  $1 - \delta/8$

$$\sup_{\mathbf{x} \in \mathcal{N}_\epsilon} \|\nabla_{\mathbf{x}} f(\mathbf{x}, Q_t) - \nabla_{\mathbf{x}} f(\mathbf{x}, Q_t^\infty)\|_2^2 \leq \frac{20dG^2}{m} \left(\log \frac{8}{\delta} + d\log(1 + 2m^{1/2})\right).$$

Together with our trivial bound of  $D^2L^2$ , this gives us the following bound for  $\|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}}^2$ , which holds with probability at least  $1 - \delta/8$

$$\|f(\cdot, Q_t) - f(\cdot, Q_t^\infty)\|_{\mathcal{F}}^2 \leq \min\left(\frac{20dG^2}{m} \left(\log \frac{8}{\delta} + d\log(1 + 2m^{1/2})\right), D^2L^2\right) + \frac{8D^2L^2}{m}.$$

Next, we bound  $\|f(\cdot, Q_t^\infty) - f(\cdot, \tilde{Q}_{t-1}^\infty)\|_{\mathcal{F}}^2$ . From our smoothness assumption on  $f$ , we have

$$\|f(\cdot, Q_t^\infty) - f(\cdot, \tilde{Q}_{t-1}^\infty)\|_{\mathcal{F}} \leq L\gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty).$$

Combining the previous two results, we get the following upper bound for  $S_2$  which holds with probability at least  $1 - \delta/8$

$$\begin{aligned} \|f(\cdot, Q_t) - f(\cdot, \tilde{Q}_{t-1}^\infty)\|_{\mathcal{F}}^2 &\leq 3L^2\gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty)^2 + \frac{48D^2L^2}{m} \\ &\quad + \min\left(\frac{120dG^2}{m} \left(\log \frac{8}{\delta} + d\log(1 + 2m^{1/2})\right), 6D^2L^2\right). \end{aligned}$$

**Regret bound.** Substituting the above bounds for  $S_1, S_2$  in the regret bound for  $\mathbf{x}$  player gives us the following bound, which holds with probability at least  $1 - \delta/2$

$$\begin{aligned} \sum_{t=1}^T f(P_t, Q_t) - f(\mathbf{x}, Q_t) &\leq \eta D \log d + O\left(GD\sqrt{\frac{T \log \frac{8}{\delta}}{m}} + \frac{d^2 D^3 L^2 T}{\eta m}\right) \\ &\quad + O\left(\min\left(\frac{d^3 DG^2 T}{\eta m} \left(\log \frac{8}{\delta} + d\log(2m)\right), \frac{d^2 D^3 L^2 T}{\eta}\right)\right) \\ &\quad + \sum_{t=1}^T \frac{3cd^2 DL^2}{2\eta} \gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty)^2 - \sum_{t=1}^T \frac{\eta}{2cd^2 D} \gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_{t-1}^\infty)^2 \end{aligned}$$

Using a similar analysis, we get the following regret bound for the  $\mathbf{y}$  player

$$\begin{aligned} \sum_{t=1}^T f(P_t, Q_t) - f(\mathbf{x}, Q_t) &\leq \eta D \log d + O\left(GD\sqrt{\frac{T \log \frac{8}{\delta}}{m}} + \frac{d^2 D^3 L^2 T}{\eta m}\right) \\ &\quad + O\left(\min\left(\frac{d^3 DG^2 T}{\eta m} \left(\log \frac{8}{\delta} + d\log(2m)\right), \frac{d^2 D^3 L^2 T}{\eta}\right)\right) \\ &\quad + \sum_{t=1}^T \frac{3cd^2 DL^2}{2\eta} \gamma_{\mathcal{F}}(P_t^\infty, \tilde{P}_{t-1}^\infty)^2 - \sum_{t=1}^T \frac{\eta}{2cd^2 D} \gamma_{\mathcal{F}'}(Q_t^\infty, \tilde{Q}_{t-1}^\infty)^2 \end{aligned}$$

Choosing,  $\eta = 10d^2 D(L+1)$ ,  $m = T$ , and adding the above two regret bounds, we get

$$\begin{aligned} \sup_{\mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}} \sum_{t=1}^T f(P_t, \mathbf{y}) - f(\mathbf{x}, Q_t) &= O\left(d^2 D^2 (L+1) \log d + GD\sqrt{\log \frac{8}{\delta}}\right) \\ &\quad + O\left(\min\left\{D^2 LT, \frac{d^2 G^2 \log T}{L} + \frac{dG^2 \log \frac{8}{\delta}}{L}\right\}\right). \end{aligned}$$

□

## H Background on Convex Analysis

**Fenchel Conjugate.** The Fenchel conjugate of a function  $f$  is defined as

$$f^*(x^*) = \sup_x \langle x, x^* \rangle - f(x).$$

We now state some useful properties of Fenchel conjugates. These properties can be found in Rockafellar [22].

**Theorem H.1.** *Let  $f$  be a proper convex function. The conjugate function  $f^*$  is then a closed and proper convex function. Moreover, if  $f$  is lower semi-continuous then  $f^{**} = f$ .*

**Theorem H.2.** *For any proper convex function  $f$  and any vector  $x$ , the following conditions on a vector  $x^*$  are equivalent to each other*

- $x^* \in \partial f(x)$
- $\langle z, x^* \rangle - f(z)$  achieves its supremum in  $z$  at  $z = x$
- $f(x) + f^*(x^*) = \langle x, x^* \rangle$

If  $(clf)(x) = f(x)$ , the following condition can be added to the list

- $x \in \partial f^*(x^*)$

**Theorem H.3.** *If  $f$  is a closed proper convex function,  $\partial f^*$  is the inverse of  $\partial f$  in the sense of multivalued mappings, i.e.,  $x \in \partial f^*(x^*)$  iff  $x^* \in \partial f(x)$ .*

**Theorem H.4.** *Let  $f$  be a closed proper convex function. Let  $\partial f$  be the subdifferential mapping. The effective domain of  $\partial f$ , which is the set  $\text{dom}(\partial f) = \{x | \partial f \neq \emptyset\}$ , satisfies*

$$\text{ri}(\text{dom}(f)) \subseteq \text{dom}(\partial f) \subseteq \text{dom}(f).$$

The range of  $\partial f$  is defined as  $\text{range} \partial f = \cup \{\partial f(x) | x \in \mathbb{R}^d\}$ . The range of  $\partial f$  is the effective domain of  $\partial f^*$ , so

$$\text{ri}(\text{dom}(f^*)) \subseteq \text{range} \partial f \subseteq \text{dom}(f^*).$$

**Strong Convexity and Smoothness.** We now define strong convexity and strong smoothness and show that these two properties are duals of each other.

**Definition H.1** (Strong Convexity). A function  $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{\infty\}$  is  $\beta$ -strongly convex w.r.t a norm  $\|\cdot\|$  if for all  $x, y \in \text{ri}(\text{dom}(f))$  and  $\alpha \in (0, 1)$  we have

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \frac{1}{2}\beta\alpha(1 - \alpha)\|x - y\|^2.$$

This definition of strong convexity is equivalent to the following condition on  $f$  [see Lemma 13 of 24]

$$f(y) \geq f(x) + \langle g, y - x \rangle + \frac{1}{2}\beta\|y - x\|^2, \quad \text{for any } x, y \in \text{ri}(\text{dom}(f)), g \in \partial f(x)$$

**Definition H.2** (Strong Smoothness). A function  $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{\infty\}$  is  $\beta$ -strongly smooth w.r.t a norm  $\|\cdot\|$  if  $f$  is everywhere differentiable and if for all  $x, y$  we have

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2}\beta\|y - x\|^2.$$

**Theorem H.5** (Kakade et al. [29]). *Assume that  $f$  is a proper closed and convex function. Suppose  $f$  is  $\beta$ -strongly smooth w.r.t a norm  $\|\cdot\|$ . Then its conjugate  $f^*$  satisfies the following for all  $a, x$  with  $u = \nabla f(x)$*

$$f^*(a + u) \geq f^*(u) + \langle x, a \rangle + \frac{1}{2\beta}\|a\|_*^2.$$

**Theorem H.6** (Kakade et al. [29]). *Assume that  $f$  is a closed and convex function. Then  $f$  is  $\beta$ -strongly convex w.r.t a norm  $\|\cdot\|$  iff  $f^*$  is  $\frac{1}{\beta}$ -strongly smooth w.r.t the dual norm  $\|\cdot\|_*$ .*