WASSERSTEIN DISTRIBUTIONALLY ROBUST OPTIMIZATION: A THREE-PLAYER GAME FRAMEWORK

Anonymous authors
Paper under double-blind review

ABSTRACT

Wasserstein distributionally robust optimization (DRO) has recently received significant attention in machine learning due to its connection to generalization, robustness and regularization. Existing methods only consider a limited class of loss functions or apply to small values of robustness. In this paper, we present a three-player game framework for solving Wasserstein DRO problem with arbitrary level of robustness, which can handle general loss functions. Specifically, we formulate a min-max game between three players who optimize over probability measures, model parameters and Lagrange multipliers. We also propose new algorithms for finding an equilibrium of the game in convex and non-convex settings which both enjoy provable convergence guarantees. Furthermore, we prove an excess risk bound for the proposed algorithms which shows that the solution returned by the algorithms closely achieves the optimal minimax risk.

1 INTRODUCTION

Distributionally robust optimization (DRO) has become popular in recent years in machine learning due to its ability to improve robustness of learning models. Instead of choosing a model f to minimize an expected loss on instances generated according to P, DRO considers a perturbation to the underlying data distribution within an ambiguity set, whether they are from covariate shifts, changes in the underlying domain or adversarial attacks, and seeks to solve the following saddle-point problem

$$\min_{f\in\mathcal{F}} \sup_{Q\in\mathcal{A}(P)} R_Q(f), \tag{1}$$

where $\mathcal{A}(P)$ is an ambiguity set containing P. The choice of ambiguity set $\mathcal{A}(P)$ influences the richness of the uncertainty set that we wish to consider as well as computability and can be constructed in a variety of ways such as f-divergence ball (Ben-Tal et al., 2013; Namkoong & Duchi, 2017; 2016; Blanchet et al., 2018) or Wasserstein ball (Blanchet et al., 2019b;a; Abadeh et al., 2015; Gao et al., 2017; Volpi et al., 2018).

Despite of the induced robustness, the formulation of DRO problem (1) is intractable for general cases. Prior work on DRO has focused on tractable classes of ambiguity set and loss functions. For example, it has been shown that the DRO problem with Wasserstein distance can be reformulated as a convex optimization problem by proving a duality result (Blanchet & Murthy, 2019; Gao & Kleywegt, 2016; Esfahani & Kuhn, 2018; Zhao & Guan, 2018). But this reformulation is only possible for limited classes of loss functions. To deal with a larger class of losses, Sinha et al. (2018) consider a relaxation of the saddle-point problem (1) and provides a stochastic gradient-type adversarial training procedure to solve it. However, there is no guarantee to find the global optimum in non-convex case even for this relaxed problem, and their method requires additional smoothness condition on loss functions and can only achieve a small amount of robustness. Moreover, the Wasserstein DRO reformulations typically possess complicated structures and can not be solved efficiently by off-the-shelf solvers in large-scale problems. To tackle this difficulty, Li et al. (2019) propose a linearized proximal ADMM algorithm, which can achieve the same accuracy up to hundreds times faster than the standard off-the-shelf solver. However their algorithm can only solve Wasserstein distributionally robust logistic regression problems. There are also some work on DRO with f-divergence ball (Ben-Tal et al., 2013; Namkoong & Duchi, 2017; 2016). But any f-divergence ambiguity set around P can only contain distributions with the same support as P and thus does not allow robustness to unseen data.

In this paper, we present a three-player framework for Wasserstein DRO problem, which applies to general loss functions and arbitrary level of robustness. Instead of reducing the original intractable saddle-point problem (two-player game) to a complicated convex optimization as in the literature, we introduce auxiliary Lagrange multiplier for the Wasserstein constraints in the inner supremum problem and formulate a min-max game between three-players who optimizes over probability distributions, model parameters and Lagrange multiplier. Each of the three players can use different optimization algorithms customized for their problem, as long as the player strategies lead to an equilibrium of the game.

Finding an equilibrium of a game between players remains an active topic of research in machine learning. Much of the focus has been on a two-player game (Agarwal et al., 2018; Donini et al., 2018; Cotter et al., 2019a;b; Kearns et al., 2018). In particular, Agarwal et al. (2018) propose an algorithm for fair classification by computing an equilibrium of a two-player game where one player chooses a mixture of objective functions, and the other player minimizes the loss of the mixture. Recently, Narasimhan et al. (2019) also present a three-player approach for optimizing generalized rate metrics. But their game formulation as well as the algorithms for each player is different from ours. We summarize our contribution as follows:

- We present a three-player game framework for solving Wasserstein DRO problem. Our framework applies to arbitrary level of robustness and general loss functions, whereas prior work focuses on either a limited class of loss functions or small values of robustness.
- We propose new algorithms for solving an equilibrium of the game in convex and non-convex settings which both enjoy provable convergence guarantees. In particular, the player who optimizes over model parameters uses gradient descent algorithm in the convex case and MCMC sampling when the loss function is non-convex.
- We prove an excess risk bound for the proposed algorithms which shows that with high probability the learned model achieves the optimal minimax risk up to the optimization error, which grows linearly with a predefined accuracy parameter μ, and the generalization error, which grows as $1/\sqrt{n}$.

In the next section, we formalize the problem. In Section 3, we present the three-player game framework including the game formulation, proposed algorithms and theoretical guarantees. The experimental results are provided in Section 4. Finally we conclude and discuss future directions. All proofs are deferred to the appendix.

2 PROBLEM SETUP

We consider Wasserstein distributionally robust optimization (DRO) problem. Let $(\mathcal{Z}, d_{\mathcal{Z}})$ be a compact Polish space with metric $d_{\mathcal{Z}}$ and $diam(\mathcal{Z}) := \sup_{z,z' \in \mathcal{Z}} d_{\mathcal{Z}}(z, z')$. Denote with $\mathcal{P}(\mathcal{Z})$ the space of all Borel probability measures on $\mathcal{Z}$ and $\mathcal{P}_1(\mathcal{Z})$ the space of all $P \in \mathcal{P}(\mathcal{Z})$ with finite moment of order 1, i.e., $\mathcal{P}_1(\mathcal{Z}) := \{P \in \mathcal{P}(\mathcal{Z}) : \mathbb{E}_P[d_{\mathcal{Z}}(z, z_0)] < \infty \text{ for } z_0 \in \mathcal{Z}\}$. Then, for any two probability measures $P, Q \in \mathcal{P}_1(\mathcal{Z})$, the Wasserstein distance between P and Q is defined as

$$W(P, Q) := \inf_{M \in \Gamma(P,Q)} (\mathbb{E}_{(z,z') \sim M}[d_{\mathcal{Z}}(z, z')]),$$

where $\Gamma(P, Q)$ denotes the collection of all measures on $\mathcal{Z} \times \mathcal{Z}$ with marginals P and Q on the first and second factors, respectively.

Given a model family Ω and a loss function $l : \Omega \times \mathcal{Z} \to \mathbb{R}_+$, the performance of model ω on instances generated according to P is measured by expected risk denoted by $R_P(\omega) := \mathbb{E}_{z \sim P} l(\omega, z)$. In this work, we do not assume the loss function to be convex in ω. The Wasserstein DRO problem formulation takes the form

$$\min_{\omega \in \Omega} \sup_{Q: W(Q,P) \leq \tilde{\epsilon}} R_Q(\omega), \tag{2}$$

where the parameter $\tilde{\epsilon} > 0$ is the radius of Wasserstein ball and represents the level of robustness.

To deal with general loss function, whether convex or non-convex, we enlarge the space of possible solutions from deterministic models Ω to stochastic models characterized by a distribution over Ω. A stochastic model π first samples a model ω from the distribution π and then use ω to compute

the loss. The resulting expected risk for π is $R_P(\pi) := \int R_P(\omega)\pi(d\omega)$. By replacing $R_Q(\omega)$ with $R_Q(\pi)$, the Wasserstein DRO problem (2) is re-formulated as

$$\min_{\pi\in\Delta} \sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(\pi), \tag{3}$$

where Δ is the set of all distributions over Ω. Note that Problem (3) and (2) are equivalent for convex loss function. In practice, we do not know the true distribution P and only have access to a data set of training samples $\{z_i\}_{i=1}^n$ drawn i.i.d. from P. We therefore replace the true distribution P in problem (3) with empirical distribution $P_n = \frac{1}{n}\sum_{i=1}^n \delta_{z_i}$ and seek to solve the following problem:

$$\min_{\pi\in\Delta} \sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(\pi). \tag{4}$$

3 THREE-PLAYER GAME FRAMEWORK

In this section, we first show how the Wasserstein DRO problem (4) can be formulated as a three-player game. Then we propose algorithms for finding an equilibrium of the expanded game in both convex and non-convex setting. We finally establish excess risk bounds which compare the performance of the proposed algorithms to the optimal solution of (3).

3.1 GAME FORMULATION

To derive the three-player game, we first observe that problem (4) is equivalent to

$$\max_{\pi\in\Delta} \min_{Q:W(Q,P_n)\leq\tilde{\epsilon}} [-R_Q(\pi)]. \tag{5}$$

Then we consider its dual problem (6) obtained by switching max and min and show the equivalence of primal and dual problem in the following lemma.

Lemma 1. *The strong duality holds, i.e.,*

$$\max_{\pi\in\Delta} \min_{Q:W(Q,P_n)\leq\tilde{\epsilon}} [-R_Q(\pi)] = \min_{Q:W(Q,P_n)\leq\tilde{\epsilon}} \max_{\pi\in\Delta} [-R_Q(\pi)]. \tag{6}$$

From our Lemma 1 and Lemma 36.2 of Rockafellar (1970), the maximum value of problem (5) and the minimum value of (6) are equal and coincide with $-R_{Q^\S}(\pi^\S)$ where $\pi^\S$ and $Q^\S$ denote the solutions of the two problems respectively, i.e., $(Q^\S, \pi^\S)$ is the saddle point. Therefore, by finding the saddle point of problem (6), we can obtain the solution $\pi^\S$ of problem (5). However, because of the Wasserstein ball constraints, it is very hard to directly work on problem (6). A standard approach is to introduce a Lagrange multiplier $\lambda \geq 0$ for the Wasserstein constraint and write the Lagrangian for the problem

$$\tilde{L}(Q, \pi, \lambda) = -R_Q(\pi) + \lambda(W(Q, P_n) - \tilde{\epsilon}).$$

Then one minimize the Lagrangian over probability measures Q, and maximize it over $\pi \in \Delta$ and $\lambda \in \mathbb{R}_+$,

$$\min_{Q} \max_{\pi\in\Delta,\lambda\geq 0} \tilde{L}(Q, \pi, \lambda). \tag{7}$$

We pose this min-max problem as a zero-sum game between three players: a Q-player who minimizes $\tilde{L}$ over Q, a π-player who maximizes $\tilde{L}$ over π and a λ-player who maximizes $\tilde{L}$ over λ. Since $\tilde{L}$ is convex in Q and linear in π and λ, one can solve an equilibrium of this game and obtain a solution for the problem (5). For technical reasons, we impose an additional constraint on λ and aim to solve the constrained version of problem (7)

$$\min_{Q} \max_{\pi\in\Delta, B\geq\lambda\geq 0} L(Q, \pi, \lambda), \tag{8}$$

where B is a predefined parameter and $L(Q, \pi, \lambda) = -R_Q(\pi) + \lambda(W(Q, P_n) - \epsilon)$. Notice that because of the introduction of B we also substitute $\tilde{\epsilon}$ with ϵ and try to optimize L rather than $\tilde{L}$ .We will show how to choose B and ϵ later in our theoretical analysis in Section 3.3. Now all we need to do is to choose the strategy that each player uses to optimize their objective, so that the players converge to an (approximate) equilibrium of this game. We consider two cases: the loss function is (1) convex in ω, and (2) non-convex in ω. For each case, we propose an algorithm for finding a μ-approximate Nash equilibrium of the game, which is a triple $(\hat{Q}, \hat{\pi}, \hat{\lambda})$, where $L(\hat{Q}, \hat{\pi}, \hat{\lambda}) \leq L(Q, \hat{\pi}, \hat{\lambda}) + \mu$ for all Q and $L(\hat{Q}, \hat{\pi}, \hat{\lambda}) \geq L(\hat{Q}, \pi, \lambda) - \mu$ for all $\pi \in \Delta$ and $\lambda \in [0, B]$.

Algorithm 1 Three-player game for convex loss

Input: training sample $\{z_i\}_{i=1}^n$, Wasserstein ball radius ϵ, bound B, accuracy μ_0 and μ_1, learning rate α, β
Initialize $\theta_1 = 0, \omega_1 = 0$
for $t = 1, 2, \cdots$ **do**
$\lambda_t \leftarrow B\dfrac{exp(\theta_t)}{1 + exp(\theta_t)}$
$Q_t \leftarrow BEST_{Q,\mu_1}(\omega_t, \lambda_t)$
$\theta_{t+1} \leftarrow \theta_t + \alpha(W(Q_t, P_n) - \epsilon)$
$\omega_{t+1} \leftarrow \Pi_\Omega(\omega_t + \beta\partial_\omega(-R_{Q_t}(\omega_t)))$
end for
return: $\hat{Q}_t = \frac{1}{t}\sum_{t'=1}^t Q_{t'}$, $\hat{\omega}_t = \frac{1}{t}\sum_{t'=1}^t \omega_{t'}$, $\hat{\lambda}_t = \frac{1}{t}\sum_{t'=1}^t \lambda_{t'}$

Algorithm 2 Three-player game for non-convex loss

Input: same as in Algorithm 1, learning rate α and η, a prior π_1
Initialize $\theta_1 = 0$
for $t = 1, 2, \cdots$ **do**
$\lambda_t \leftarrow B\dfrac{exp(\theta_t)}{1 + exp(\theta_t)}$
Sample $\omega_t \sim \pi_t$
$Q_t \leftarrow BEST_{Q,\mu_1}(\pi_t, \lambda_t)$
$\theta_{t+1} \leftarrow \theta_t + \alpha(W(Q_t, P_n) - \epsilon)$
$\pi_{t+1}(d\omega) = \dfrac{\exp(-\eta R_{Q_t}(\omega))\pi_t(d\omega)}{\int \exp(-\eta R_{Q_t}(\gamma))\pi_t(d\gamma)}$
end for
return: $\hat{Q}_t = \frac{1}{t}\sum_{t'=1}^t Q_{t'}$, $\hat{\pi}_t = \frac{1}{t}\sum_{t'=1}^t \omega_{t'}$, $\hat{\lambda}_t = \frac{1}{t}\sum_{t'=1}^t \lambda_{t'}$

3.2 ALGORITHMS

In this subsection, we propose algorithms for solving problem (8) in both convex and non-convex cases. Then in the following subsection, we show how the solution returned by our algorithms can approximately solved the original problem (4) (or (3)) by choosing appropriate parameters B and ϵ.

3.2.1 CASE OF CONVEX l

We start with the case where $l(\omega, z)$ is convex in ω for any z. We find the approximate equilibrium by using the standard scheme of Freund & Schapire (1996). We proceed iteratively with the Q-player playing best response to the opponents strategies, while running the exponentiated gradient algorithm (Kivinen & Warmuth, 1997) and online gradient descent for λ-player and π-player respectively and terminate as soon as the sub-optimality of the average play falls below the pre-specified accuracy $\mu := \mu_0 + \mu_1$. The resulting algorithm is outlined in Algorithm 1. Here, Π_Ω denotes the l_2 projection onto Ω, and $BEST_{Q,\mu_1}$ represents the μ_1-approximate best response of Q-player, i.e., $L(BEST_{Q,\mu_1}(\omega, \lambda), \omega, \lambda) \leq \min_Q L(Q, \omega, \lambda) + \mu_1$, which is to solve a minimization problem of a probability functional and will be discussed in Section 3.2.3. After the loop terminates, it results in a deterministic model $\hat{\omega}$. Algorithm 1 is guaranteed to find an approximate equilibrium of the game in (8) by the following theorem.

Theorem 1. *Suppose that $l(\omega, z)$ is convex in ω and K-Lipschitz in ω with respect to $||\cdot||_2$ for any z. Let $\rho := diam(\mathcal{Z}) + \epsilon$ and $B_\Omega \geq \max_{\omega\in\Omega} ||\omega||_2$. Then setting $\alpha = \dfrac{\mu_0}{4\rho^2 B}$ and $\beta = \dfrac{\mu_0}{4K^2}$, Algorithm 1 will return a μ-approximate equilibrium in at most $\dfrac{4K^2B_\Omega^2 + 8B^2\rho^2\log(2)}{\mu_0^2}$ iterations.*

Theorem 1 shows that we may need up to $\frac{4K^2B_\Omega^2+8B^2\rho^2\log(2)}{\mu_0^2}$ iterations to achieve a sub-optimality μ. Note that B is the restriction that we place on the Lagrange multiplier λ. In general, large value of B will bring the problem (8) closer to (7) and thus to the primal problem (4), but at the cost of needing more iterations to reach any given suboptimality μ.

3.2.2 CASE OF NON-CONVEX l

In modern machine learning models, like deep neural networks, the loss function is usually non-convex. We now provide a different algorithm for non-convex loss functions where the π-player updates a probability distribution on the set of model parameters Ω using exponentially weighted aggregation procedure (Cesa-Bianchi & Lugosi, 2006; Alquier et al., 2017) instead of online gradient descent. The Q-player and λ-player, however, continue to operate in the same way as in Algorithm 1.

Subroutine 3 Sampling $\omega_t \sim \pi_t$

Start: Draw $\omega_1 \sim \pi_1$
for $t = 1, 2, \cdots$ **do**
  Set $\omega := \omega_t$
  Metropolis-Hastings algorithm: Repeat N times
  for $j = 0, 1, \cdots, N-1$ **do**
    Sample $u \sim \mathcal{U}_{[0,1]}$
    Sample $\omega' \sim \mathcal{N}(\omega, \sigma^2 \mathcal{I})$
    if $u < \min\{1, \exp[\eta(\sum_{t'=1}^{t}(R_{Q_{t'}}(\omega) - R_{Q_{t'}}(\omega')))]\}$ **then**
      $\omega \leftarrow \omega'$
    end if
  end for
  Set $\omega_{t+1} := \omega$
end for

In the proposed algorithm, outlined in Algorithm 2, the π-player maintains a prior distribution on the set of model parameters, which is updated after the encounter of each new task Q_t using the performance of the models. Typically, if the expected risk for model ω is small, then we will assign a large weight to ω. This results in weighting more those models whose accumulative loss is small. We first give a bound for the cumulative regret of the strategy with respect to any model.

Lemma 2. *Assume that $l(\omega, z)$ is K-Lipschitz in ω with respect to $||\cdot||_2$ for any z and uniformly bounded for any $\omega \in \Omega$ and $z \in \mathcal{Z}$, i.e., $0 \leq l(\omega, z) \leq M$. Let $\Omega = \{\omega : ||\omega||_2 \leq \tilde{B}_\Omega, \omega \in \mathbb{R}^d\}$ and π_1 the uniform distribution on Ω. Let Q_1, Q_2, $\cdots$, Q_T, π_1, π_2, $\cdots$, π_T be the iterates generated by Algorithm 2. Then for any $\omega \in \Omega$,*

$$\sum_{t=1}^{T} R_{Q_t}(\pi_t) \leq \sum_{t=1}^{T} R_{Q_t}(\omega) + \tau_T,$$

where $\tau_T := \dfrac{\eta M^2 T}{8} + \dfrac{\mu_0 T}{4} + \dfrac{d}{\eta}\log(\dfrac{8K\tilde{B}_\Omega}{\mu_0})$.

Now we show that the algorithm provably converges to an approximate equilibrium of the game (8) with high probability.

Theorem 2. *Let $\rho := diam(\mathcal{Z}) + \epsilon$. Then, under the assumptions of Lemma 2, by setting $\alpha = \dfrac{\mu_0}{4\rho^2 B}$ and $\eta = \dfrac{2\mu_0}{M^2}$, with probability at least $1 - 2\delta$, Algorithm 2 will return a μ-approximate equilibrium in at most $\dfrac{8M^2\log(1/\delta) + 32B^2\rho^2\log(2) + 4dM^2\log(8K\tilde{B}_\Omega/\mu_0)}{\mu_0^2}$ iterations.*

Remember that in convex setting the algorithm returns a deterministic model $\hat{\omega}_T$. However when the loss function is non-convex, it yields a stochastic model $\hat{\pi}_T$ with a probability mass of $\frac{1}{T}$ on each ω_t. For the case that T is very large, this stochastic model may be undesirable in practice. To select a deterministic model, a heuristic approach is to compute $\hat{Q}$ using Algorithm 2 and then have the π-player play the best response to $\hat{Q}$ which is the maximizer of $-R_{\hat{Q}}(\pi)$ and can always be chosen to put all of the mass on one of the candidate ω.

Remark 1. As will be shown in Section 3.2.3, the step of calculating Q_t is equivalent to a stochastic optimization problem where the objective depends on the random variable $\omega \sim \pi_t$. Stochastic Approximation (SA) methods (Kushner & Yin, 2003) can be applied to this problem, at each iteration using a random sample of ω. Now, the algorithm can be implemented without calculating π_t exactly as in the last step since it only requires being able to sample a model ω from the distribution π_t. This can be achieved by using Markov chain Monte Carlo (MCMC) methods which allow us to sample the distribution without calculating the integral in the denominator (see Subroutine 3).

Remark 2. In Subroutine 3, we use N-steps of Metropolis-Hastings algorithm with a Gaussian proposal distribution (Robert & Casella, 2013) to sample ω_{t+1} from π_{t+1}. To ensure a short burn-in

period, we use previous drawing ω_t as a starting point. Note that we have to compare ω and ω' on the whole distributions that Q-player has yielded so far for computing the acceptance ratio, which might make our algorithm become slow when t grows. In practice, to improve the speed we may truncate past history and use only a fixed number of timesteps.

3.2.3 BEST: THE Q-PLAYER'S BEST RESPONSE

In both algorithms, we let the Q-player play the μ_1-approximate best response to the opponents strategies. Recall that $BEST_{Q,\mu_1}(\omega, \lambda)$ for a given ω and λ is any μ_1-approximate minimizer of a functional $L(Q, \omega, \lambda)$ over all probability measures on the metric space $(\mathcal{Z}, d_{\mathcal{Z}})$. Because the optimization now takes place in the infinite dimensional space of probability measures, standard finite dimensional algorithms like gradient descent are initially unavailable; even the proper notion for the derivate of the probability functional is unclear. Luckily, it can be shown that this probability functional minimization problem is equivalent to solving a transportation map. We first give the definition of the transportation map $T_{\omega,\lambda}(z) : \mathcal{Z} \to \mathcal{Z}$:

$$T_{\omega,\lambda}(z) = \arg\max_{z'} \{l(\omega, z') - \lambda \cdot d_{\mathcal{Z}}(z', z)\}. \tag{9}$$

Then we can prove that $BEST_Q$ can be derived exactly by using the transportation map $T_{\omega,\lambda}$ in the following lemma.

Lemma 3. *Let $Q_{\omega,\lambda} := T_{\omega,\lambda}\#P_n$ be the pushforward of the empirical distribution P_n under the transportation map $T_{\omega,\lambda}$, i.e., $Q_{\omega,\lambda} = \frac{1}{n}\sum_{i=1}^{n} \delta_{T_{\omega,\lambda}(z_i)}$. Then the best response of Q-player for a given ω and λ is attained by*

$$BEST_Q(\omega, \lambda) = \frac{1}{n}\sum_{i=1}^{n} \delta_{T_{\omega,\lambda}(z_i)} \text{ and } W(BEST_Q(\omega, \lambda), P_n) = \frac{1}{n}\sum_{i=1}^{n} d_{\mathcal{Z}}(T_{\omega,\lambda}(z_i), z_i).$$

The above lemma shows that $BEST_Q$ can be obtained directly by solving the transportation map $T_{\omega,\lambda}(z_i)$. Similarly, in the case of non-convex loss functions, computing $BEST_{Q,\mu_1}(\pi, \lambda)$ is equivalent to solving the stochastic transportation map, i.e., $T_{\pi,\lambda}(z) = \arg\max_{z'}\{\mathbb{E}_{\omega\sim\pi}[l(\omega, z') - \lambda \cdot d_{\mathcal{Z}}(z', z)]\}$, which is a stochastic optimization problem and can be tackled with Stochastic Approximation methods by using a random sample ω at each iteration. Moreover, computing the (stochastic) transportation map is a strongly-concave optimization problem for large λ when the loss function is smooth and $d_{\mathcal{Z}}(\cdot, z_i)$ is strongly convex (Sinha et al., 2018). Thus (stochastic) gradient method can solve $T_{(\pi)\omega,\lambda}(z_i)$ efficiently with provable convergence guarantees. In practice, we use (stochastic) gradient ascent steps to solve the map. See Appendix D for the pseudo-code. It should be mentioned that this concave property does not necessarily hold when λ is small and hence gradient ascent in that case can only guarantee to find a local maximum (Reddi et al., 2016; Allen-Zhu & Hazan, 2016). We leave this problem to future work. And in our experimental section, we empirically show that it behaves well on real-world datasets.

Remark 3. We discover that the the algorithm proposed in Sinha et al. (2018) for solving a relaxed DRO problem can be recovered from our framework by having the Q-player play best response, π-player play gradient descent and λ-player play a constant strategy, meaning that he always chooses a fixed λ no matter what the opponent strategies are. Therefore our framework might be viewed as a generalization of their method despite that our motivation and techniques are different.

3.3 RISK BOUNDS

From results in the previous section, Algorithm 1&2 are guaranteed to converge to an approximate equilibrium of problem (8) (or with high probability). Now we show that how such algorithms solve the original problem (3), i.e., generalizing robustness to the perturbations on the true distribution P. Let $(\hat{Q}, \hat{\pi}, \hat{\lambda})$ be the μ-approximate equilibrium returned by our algorithms. Denote $\pi^* = \arg\min_{\pi\in\Delta} \sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(\pi)$ be the solution of the problem (3). The performance of our algorithms is measured in terms of excess risk defined as

$$\sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(\hat{\pi}) - \sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(\pi^*).$$

In this section, we will provide an upper bound for the excess risk of our algorithms. Notice that in the process of deriving Algorithms 1&2 we introduce three different sources of error. First, we

replace the true distribution P with the empirical distribution P_n. Second, we introduce a bound B on the λ and substitute $\tilde{\epsilon}$ with ϵ. Finally, we only run the algorithms for a fixed iterations until it reaches sub-optimality level μ. The first source of error is unavoidable and also called generalization error in statistical learning theory. We can upper bound it by using existing uniform convergence bounds for local worst-case risk (Lee & Raginsky, 2018). The other two sources of error are caused by the optimization algorithm and can be driven arbitrary small via a careful selection of B and ϵ. We first establish that the equilibrium $(\hat{Q}, \hat{\pi}, \hat{\lambda})$ satisfies the following property and then show how to use it to choose B and ϵ.

Lemma 4. *Suppose that the loss function $l(\omega, z)$ is uniformly bounded for any $\omega \in \Omega$ and $z \in \mathcal{Z}$, i.e., $0 \leq l(\omega, z) \leq M$. Then we have*

$$\hat{\lambda} \leq \frac{2\mu + M}{\epsilon} \text{ and } W(\hat{Q}, P_n) - \epsilon \leq \frac{2\mu + M}{B}.$$

Recall that ϵ and B are predefined parameters which are introduced in problem (8). We can always choose a large enough B such that $\frac{2\mu + M}{B} < \tilde{\epsilon}$ and let $\epsilon = \tilde{\epsilon} - \frac{2\mu + M}{B}$. Then by Lemma 4, we have $W(\hat{Q}, P_n) \leq \tilde{\epsilon}$. We now combine the three different sources of error and yield the following excess risk bound.

Theorem 3. *Assume that the loss function $l(\omega, z)$ is Lipschitz with respect to z and uniformly bounded for any $\omega \in \Omega$ and $z \in \mathcal{Z}$, i.e., $0 \leq l(\omega, z) \leq M$. Let $(\hat{Q}, \hat{\pi}, \hat{\lambda})$ be the μ-approximate equilibrium returned by Algorithm 1 or 2 with $B = \frac{(2\mu + M)^2}{\mu\tilde{\epsilon}} + \frac{2\mu + M}{\tilde{\epsilon}}$ and $\epsilon = \frac{2\mu + M}{3\mu + M}\tilde{\epsilon}$. Then, with probability at least $1 - \delta$, $\hat{\pi}$ satisfies*

$$\sup_{Q: W(Q,P) \leq \tilde{\epsilon}} R_Q(\hat{\pi}) - \sup_{Q: W(Q,P) \leq \tilde{\epsilon}} R_Q(\pi^*) \leq 3\mu + \tilde{\mathcal{O}}(1/\sqrt{n}),$$

where $\tilde{\mathcal{O}}$ suppresses polynomial dependence on $\log(1/\delta)$.

This theorem shows that the solution returned by our algorithms achieves the optimal worst-case loss on the true distribution up to the optimization error, which grows linearly with μ, and the generalization error, which grows as $1/\sqrt{n}$. For Algorithm 1, we can set $\mu \propto 1/\sqrt{n}$ to guarantee that the optimization error does not dominate the generalization error. Then, by Theorem 3 and Theorem 1, we know that Algorithm 1 with $B \propto \sqrt{n}, \epsilon \approx \tilde{\epsilon}, \alpha \propto \rho^{-2} n^{-1}$ and $\beta \propto 1/\sqrt{n}$ will terminate in at most $\mathcal{O}(n^2\rho^2)$ iterations and return $\hat{\omega}$, which with probability at least $1 - \delta$ satisfies $\sup_{Q: W(Q,P) \leq \tilde{\epsilon}} R_Q(\hat{\omega}) - \sup_{Q: W(Q,P) \leq \tilde{\epsilon}} R_Q(\pi^*) \leq \tilde{\mathcal{O}}(1/\sqrt{n})$. A similar analysis also applies to Algorithm 2.

4 EXPERIMENTS

We evaluate our algorithms on real-world datasets in convex and non-convex settings with the simple binary Logistic Regression (LR) and the widely-used convolutional neural networks (CNNs) respectively. We train on MNIST (LeCun et al., 1998) and test on MNIST-M (Ganin & Lempitsky, 2015), SVHN (Netzer et al., 2011), SYN (Ganin & Lempitsky, 2015) and USPS (Denker et al., 1989) datasets, using test accuracy as a metric for evaluating distributional robustness.

For convex setting, we consider the number-pairwise classification, and take the raw pixel values in $[0, 1]$ as the features accommodated with a LR classifier. For each class pair, we randomly sample 60% images for training and determine parameters by cross-validation with grid search (e.g., $\epsilon = 0.5$ and $B = 3000$). Besides the classic LR, we also compare our method with a distributionally robust variant DRLR (Li et al., 2019). For non-convex setting, we use a ConvNet (Volpi et al., 2018) for multi-class classification, which is composed of two convolutional layers with two fully connected layers, and set the hyperparameter $N = 50$ for MCMC sampling (results with LeNet (LeCun et al., 1998) in Appendix E). We use 10000 digit images for training. The results are compared to that of Empirical Risk Minimization (ERM), the iterative method (ITP) (Volpi et al., 2018) and the stochastic gradient descent procedure (WRM) (Sinha et al., 2018), the latter two of which are aimed to solve a relaxed Wasserstein DRO problem. We use $T_Q = 15$ iterations for solving the transportation map.

Table 1: Average classification accuracy on test datasets with CNNs.

DATASETS	ERM	ITP	WRM	TPG-DRO
USPS	78.9%±1.7%	78.4%±1.3%	80.0%±1.3%	78.9%±1.9%
SVHN	28.3%±3.2%	34.6%±3.3%	35.4%±1.4%	35.6%±3.1%
MNIST-M	54.8%±2.1%	59.8%±1.8%	58.6%±1.3%	58.7%±2.0%
SYN	40.6%±2.2%	45.0%±1.4%	42.6%±0.8%	45.1%±1.4%

For the transportation cost of the underlying space, we use the Euclidean norm cost for the feature vectors and define the overall as $d_{\mathcal{Z}}((x,y),(x',y')) = ||x-x'||_2^2 + \infty \mathbb{1}_{y \neq y'}$.

The results are averaged over 10 independent sampling, and reported in Figure 1 and Table 1. We observe that our method TPG-DRO (LR) achieves higher or at least the same accuracy on almost all test datasets compared to the models trained with LR and DRLR. In non-convex setting, our method TPG-DRO outperforms ERM on all datasets and improves the performances of ITP on USPS, SVHN and SYN datasets. On MNIST-M, the performance of our method is slightly inferior to ITP. Compared to WRM, TPG-DRO also achieves higher accuracy on all datasets except for USPS. The running time of our algorithms as well as the baselines is provided in Appendix E.

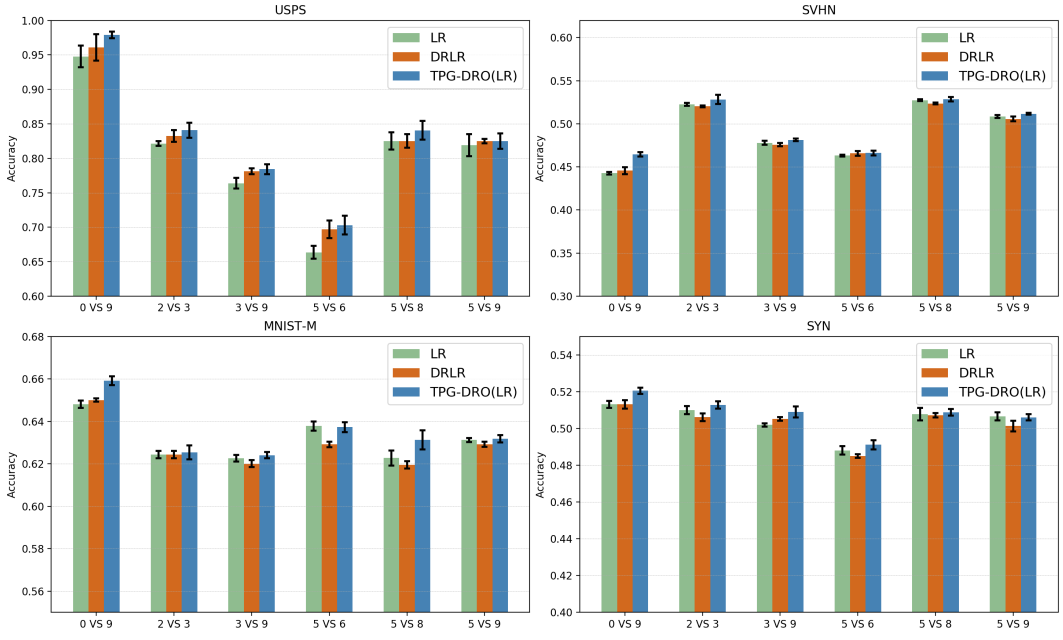

Figure 1: Average classification accuracy on test datasets with LR.

5 Conclusion

We introduce a three-player game framework for solving Wasserstein DRO problem with arbitrary level of robustness, which applies to convex and non-convex loss functions. One advantage of the framework is that it makes new algorithms possible by designing different strategies for each player.

There remain many avenues for future investigation. Our algorithm for non-convex loss functions uses MCMC sampling method instead of gradient-based algorithms. A major benefit of adapting this strategy is that MCMC is (in theory) able to fully explore the parameter space. Thus it may find a better optima for a non-convex function. As shown in Theorem 2, it returns an approximate equilibrium with high probability. However compared to gradient-based algorithms MCMC sampling

is computationally expensive and requires more tuning such as choosing a proper proposal distribution. In the future, one interesting problem is to develop more efficient and theoretically grounded (gradient-based) algorithm for the π-player in non-convex setting.

REFERENCES

Soroosh Shafieezadeh Abadeh, Peyman Mohajerin Mohajerin Esfahani, and Daniel Kuhn. Distributionally robust logistic regression. In *Advances in Neural Information Processing Systems*, pp. 1576–1584, 2015.

Alekh Agarwal, Alina Beygelzimer, Miroslav Dudik, John Langford, and Hanna Wallach. A reductions approach to fair classification. In *International Conference on Machine Learning*, pp. 60–69, 2018.

Zeyuan Allen-Zhu and Elad Hazan. Variance reduction for faster non-convex optimization. In *International Conference on Machine Learning*, pp. 699–707, 2016.

P Alquier, TT Mai, and M Pontil. Regret bounds for lifelong learning. In *AISTATS*, volume 54, pp. 261–269. PMLR (Proceedings of Machine Learning Research), 2017.

Aharon Ben-Tal, Dick Den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

Jose Blanchet and Karthyek Murthy. Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600, 2019.

Jose Blanchet, Karthyek Murthy, and Fan Zhang. Optimal transport based distributionally robust optimization: Structural properties and iterative schemes. *arXiv preprint arXiv:1810.02403*, 2018.

Jose Blanchet, Peter W Glynn, Jun Yan, and Zhengqing Zhou. Multivariate distributionally robust convex regression under absolute error loss. In *Advances in Neural Information Processing Systems*, pp. 11794–11803. 2019a.

Jose Blanchet, Yang Kang, and Karthyek Murthy. Robust wasserstein profile inference and applications to machine learning. *Journal of Applied Probability*, 56(3):830–857, 2019b.

Olivier Catoni. A pac-bayesian approach to adaptive classification. *preprint*, 840, 2003.

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

Andrew Cotter, Heinrich Jiang, and Karthik Sridharan. Two-player games for efficient non-convex constrained optimization. In *Algorithmic Learning Theory*, pp. 300–332, 2019a.

Andrew Cotter, Heinrich Jiang, Serena Wang, Taman Narayan, Seungil You, Karthik Sridharan, and Maya R Gupta. Optimization with non-differentiable constraints with applications to fairness, recall, churn, and other goals. *Journal of Machine Learning Research*, 20(172):1–59, 2019b.

JS Denker, WR Gardner, HP Graf, D Henderson, RE Howard, W Hubbard, LD Jackel, HS Baird, and I Guyon. Advances in neural information processing systems 1. chapter neural network recognizer for hand-written zip code digits. 1989.

Michele Donini, Luca Oneto, Shai Ben-David, John S Shawe-Taylor, and Massimiliano Pontil. Empirical risk minimization under fairness constraints. In *Advances in Neural Information Processing Systems*, pp. 2791–2801, 2018.

Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166, 2018.

Yoav Freund and Robert E Schapire. Game theory, on-line prediction and boosting. In *COLT*, volume 96, pp. 325–332. Citeseer, 1996.

Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning*, pp. 1180–1189, 2015.

Rui Gao and Anton J Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *arXiv preprint arXiv:1604.02199*, 2016.

Rui Gao, Xi Chen, and Anton J Kleywegt. Wasserstein distributional robustness and regularization in statistical learning. *arXiv preprint arXiv:1712.06050*, 2017.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, pp. 13–30, 1963.

Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. In *International Conference on Machine Learning*, pp. 2569–2577, 2018.

Jyrki Kivinen and Manfred K Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *information and computation*, 132(1):1–63, 1997.

Harold J. Kushner and G. George Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2003.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Jaeho Lee and Maxim Raginsky. Minimax statistical learning with wasserstein distances. In *Advances in Neural Information Processing Systems*, pp. 2687–2696, 2018.

Jiajin Li, Sen Huang, and Anthony Man-Cho So. A first-order algorithmic framework for distributionally robust logistic regression. In *Advances in Neural Information Processing Systems*, pp. 3939–3949. 2019.

Hongseok Namkoong and John C Duchi. Stochastic gradient methods for distributionally robust optimization with f-divergences. In *Advances in Neural Information Processing Systems*, pp. 2208–2216, 2016.

Hongseok Namkoong and John C Duchi. Variance-based regularization with convex objectives. In *Advances in Neural Information Processing Systems*, pp. 2971–2980, 2017.

Harikrishna Narasimhan, Andrew Cotter, and Maya Gupta. Optimizing generalized rate metrics with three players. In *Advances in Neural Information Processing Systems*, pp. 10746–10757, 2019.

Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng. Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011*, 2011.

J v Neumann. Zur theorie der gesellschaftsspiele. *Mathematische annalen*, 100(1):295–320, 1928.

Sashank J. Reddi, Ahmed Hefny, Suvrit Sra, Barnabas Poczos, and Alex Smola. Stochastic variance reduction for nonconvex optimization. In *International Conference on Machine Learning*, pp. 314–323, 2016.

Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

R Tyrrell Rockafellar. *Convex analysis*, volume 28. Princeton university press, 1970.

Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.

Aman Sinha, Hongseok Namkoong, and John Duchi. Certifiable distributional robustness with principled adversarial training. In *International Conference on Learning Representations*, 2018. URL `https://openreview.net/forum?id=Hk6kPgZA-`.

Maurice Sion et al. On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176, 1958.

Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. In *Advances in Neural Information Processing Systems*, pp. 5334–5344, 2018.

Chaoyue Zhao and Yongpei Guan. Data-driven risk-averse stochastic optimization with wasserstein metric. *Operations Research Letters*, 46(2):262–267, 2018.

A PROOFS IN SECTION 3.1

A.1 PROOF OF LEMMA 1

Proof. We first prove that the Wasserstein ball is compact and convex. Since $\mathcal{Z}$ is compact by our assumption, by Theorem 6.18 of Villani (2008), the Wasserstein space $\mathcal{P}_1(\mathcal{Z})$ is also compact. Notice that the Wasserstein ball is a closed subset of $\mathcal{P}_1(\mathcal{Z})$ since Wasserstein distance defines a distance on $\mathcal{P}_1(\mathcal{Z})$, and a closed subset of a compact space is compact. So the Wasserstein ball is compact. The convexity of the Wasserstein ball is straightforward since the Wasserstein distance is convex. Then by the minimax theorem (Neumann, 1928; Sion et al., 1958), the strong duality holds because $-R_Q(\pi)$ is linear in π and Q and the domains of Q and π are compact and convex. □

B PROOFS IN SECTION 3.2

B.1 PROOF OF THEOREM 1

Proof. We first derive the regret bounds for the updates of λ-player and π-player using standard convergence analysis for online convex optimization. Let $\Lambda := \{\lambda : 0 \leq \lambda \leq B\}$ and $\Lambda' := \{\lambda' \in \mathbb{R}^2_+ : ||\lambda'||_1 = B\}$. We associate every $\lambda \in \Lambda$ with the $\lambda' \in \Lambda'$ which is equal to λ on the first dimension and puts the remaining mass on the second dimension. Consider a run of Algorithm 1. For each λ_t, let $\lambda'_t \in \Lambda'$ be the associated element of Λ'. Let $r_t := W(Q_t, P_n) - \epsilon$ and let $r'_t \in \mathbb{R}^2$ be equal to r_t on the first coordinate and 0 on the second coordinate. By definition of Wasserstein distance, it is easy to see that $||r'_t||_\infty = |r_t| \leq W(Q_t, P_n) + \epsilon \leq diam(\mathcal{Z}) + \epsilon = \rho$. Moreover, for any λ and the associated λ', we have for all t

$$\lambda r_t = (\lambda')^T r'_t,$$

and in particular,

$$\lambda_t r_t = (\lambda'_t)^T r'_t.$$

If we interpret r'_t as the reward vector for the λ-player, then the choices of λ'_t correspond to those of the exponentiated gradient algorithm. By Corollary 2.14 and 2.16 of Shalev-Shwartz et al. (2012), for any $\lambda' \in \Lambda'$, we have

$$\sum_{t=1}^{T} (\lambda')^T r'_t \leq \sum_{t=1}^{T} (\lambda'_t)^T r'_t + \xi_T,$$

where $\xi_T := \dfrac{B \log(2)}{\alpha} + \alpha \rho^2 BT$. Therefore, we also have for any $\lambda \in \Lambda$,

$$\sum_{t=1}^{T} \lambda r_t \leq \sum_{t=1}^{T} \lambda_t r_t + \xi_T.$$

Similarly, for the π-player, the choices of $\omega_t \in \Omega$ correspond to those of the online gradient descent with a sequence of loss $R_{Q_1}(w), R_{Q_2}(w), \cdots, R_{Q_T}(w)$. Since $l(\omega, z)$ is convex in ω for any z by our assumption, $R_{Q_t}(\omega)$ is also convex in ω for any t. Then, by Corollary 2.17 and 2.13 of Shalev-Shwartz et al. (2012), for any $\omega \in \Omega$, we have

$$\sum_{t=1}^{T} -R_{Q_t}(\omega) \leq \sum_{t=1}^{T} -R_{Q_t}(\omega_t) + \zeta_T,$$

where $\zeta_T = \frac{B_\Omega^2}{2\beta} + \beta T K^2$. Now we use these regret bounds to bound the suboptimality of $L(\hat{Q}_T, \hat{\omega}_T, \hat{\lambda}_T)$. First, for any (π, λ),

$$\begin{aligned}
&L(\hat{Q}_T, \pi, \lambda) = -R_{\hat{Q}_T}(\pi) + \lambda(W(\hat{Q}_T, P_n) - \epsilon). \\
&\leq \frac{1}{T}\sum_{t=1}^T(-R_{Q_t}(\pi) + \lambda(W(Q_t, P_n) - \epsilon)) \\
&\leq \frac{1}{T}\sum_{t=1}^T(-R_{Q_t}(\omega_t) + \lambda_t(W(Q_t, P_n) - \epsilon)) + \frac{\xi_T}{T} + \frac{\zeta_T}{T} \\
&= \frac{1}{T}\sum_{t=1}^T L(Q_t, \omega_t, \lambda_t) + \frac{\xi_T}{T} + \frac{\zeta_T}{T} \\
&\leq \frac{1}{T}\sum_{t=1}^T L(\hat{Q}_T, \omega_t, \lambda_t) + \mu_1 + \frac{\xi_T}{T} + \frac{\zeta_T}{T} \\
&\leq L(\hat{Q}_T, \hat{\omega}_T, \hat{\lambda}_T) + \mu_1 + \frac{\xi_T}{T} + \frac{\zeta_T}{T}
\end{aligned},$$

where the first inequality uses the linearity of $-R_Q(\pi)$ in Q and the convexity of mapping $u \to W(u, v)$, the second inequality follows from the regret bounds, the third inequality is by the choice of Q_t, and the last inequality uses convexity of $R_{\hat{Q}_T}(\omega)$ in ω.

Also, for any Q,

$$\begin{aligned}
&L(Q, \hat{\omega}_T, \hat{\lambda}_T) \geq \frac{1}{T}\sum_{t=1}^T L(Q, \omega_t, \lambda_t)) \\
&\geq \frac{1}{T}\sum_{t=1}^T L(Q_t, \omega_t, \lambda_t) - \mu_1 \\
&= \frac{1}{T}\sum_{t=1}^T(-R_{Q_t}(\omega_t) + \lambda_t(W(Q_t, P_n) - \epsilon)) - \mu_1 \\
&\geq \frac{1}{T}\sum_{t=1}^T(-R_{Q_t}(\hat{\omega}_T) + \hat{\lambda}_T(W(Q_t, P_n) - \epsilon)) - \frac{\xi_T}{T} - \frac{\zeta_T}{T} - \mu_1 \\
&\geq -R_{\hat{Q}_T}(\hat{\omega}_T) + \hat{\lambda}_T(W(\hat{Q}_T, P_n) - \epsilon) - \frac{\xi_T}{T} - \frac{\zeta_T}{T} - \mu_1 \\
&= L(\hat{Q}_T, \hat{\omega}_T, \hat{\lambda}_T) - \frac{\xi_T}{T} - \frac{\zeta_T}{T} - \mu_1
\end{aligned},$$

where the first inequality uses the convexity of $R_Q(\omega)$, the second inequality follows from the definition of Q_t, the third inequality uses the regret bounds, and the last inequality is due to convexity of Wasserstein distance.

Combining the above results immediately implies that for any $T \geq 1$, the suboptimality μ_T of $L(\hat{Q}_T, \hat{\omega}_T, \hat{\lambda}_T)$ satisfies

$$\mu_T \leq \frac{\xi_T}{T} + \frac{\zeta_T}{T} + \mu_1 = \frac{B\log(2)}{T\alpha} + \alpha\rho^2 B + \frac{B_\Omega^2}{2T\beta} + \beta K^2 + \mu_1.$$

By plugging $\alpha = \frac{\mu_0}{4\rho^2 B}$ and $\beta = \frac{\mu_0}{4K^2}$, if $T \geq \frac{4K^2 B_\Omega^2 + 8B^2\rho^2\log(2)}{\mu_0^2}$, we can verify that

$$\mu_T \leq \frac{4\rho^2 B^2 \log(2)}{\mu_0 \frac{4K^2 B_\Omega^2 + 8B^2\rho^2\log(2)}{\mu_0^2}} + \frac{2K^2 B_\Omega^2}{\mu_0 \frac{4K^2 B_\Omega^2 + 8B^2\rho^2\log(2)}{\mu_0^2}} + \frac{\mu_0}{2} + \mu_1 = \mu.$$

□

B.2 Proof of Lemma 2

Proof. First, note that

$$\pi_t(d\omega) = \frac{\exp(-\eta\sum_{t'=1}^{t-1} R_{Q_{t'}}(\omega))\pi_1(d\omega)}{\int \exp(-\eta\sum_{t'=1}^{t-1} R_{Q_{t'}}(\gamma))\pi_1(d\gamma)}.$$

Denote $H_t = \int \exp(-\eta\sum_{t'=1}^{t-1} R_{Q_{t'}}(\gamma))\pi_1(d\gamma)$. Since $R_{Q_t}(\omega) \in [0, M]$ by assumption, using Hoeffdings inequality (Hoeffding, 1963), for any t, we have

$$\log \mathbb{E}_{\pi_t}(\exp(-\eta R_{Q_t}(\omega))) \leq -\eta R_{Q_t}(\pi_t) + \frac{\eta^2 M^2}{8},$$

which can be rewritten as

$$\mathbb{E}_{\pi_t}(\exp(-\eta R_{Q_t}(\omega)))\exp(-\frac{\eta^2 M^2}{8}) \leq \exp(-\eta R_{Q_t}(\pi_t)).$$

For $\mathbb{E}_{\pi_t}(\exp(-\eta R_{Q_t}(\omega)))$, it can be calculated as

$$\begin{aligned}
&\mathbb{E}_{\pi_t}(\exp(-\eta R_{Q_t}(\omega))) = \textstyle\int \exp(-\eta R_{Q_t}(\omega))\pi_t(d\omega) \\
&= \int \exp(-\eta R_{Q_t}(\omega)) \frac{\exp(-\eta \sum_{t'=1}^{t-1} R_{Q_{t'}}(\omega))\pi_1(d\omega)}{H_t} \\
&= \frac{\int \exp(-\eta \sum_{t'=1}^{t} R_{Q_{t'}}(\omega))\pi_1(d\omega)}{H_t} \\
&= \frac{H_{t+1}}{H_t}
\end{aligned}.$$

Multiplying over $t = 1, 2, \cdots, T$, we get

$$\exp(-\frac{\eta^2 M^2 T}{8}) H_{T+1} \leq \exp(-\eta \sum_{t=1}^{T} R_{Q_t}(\pi_t)).$$

Taking $\log$ on both sides and rearranging the terms, we obtain

$$\begin{aligned}
&\textstyle\sum_{t=1}^{T} R_{Q_t}(\pi_t) \leq \frac{\eta M^2 T}{8} - \frac{\log(H_{T+1})}{\eta} \\
&= \frac{\eta M^2 T}{8} - \frac{\log(\int \exp(-\eta \sum_{t=1}^{T} R_{Q_t}(\gamma))\pi_1(d\gamma))}{\eta} \\
&= \frac{\eta M^2 T}{8} - \frac{\log(\mathbb{E}_{\pi_1}[\exp(-\eta \sum_{t=1}^{T} R_{Q_t}(\gamma))])}{\eta} \\
&= \frac{\eta M^2 T}{8} + \inf_{\pi} \{\mathbb{E}_{\pi} \textstyle\sum_{t=1}^{T} R_{Q_t}(\gamma) + \frac{KL(\pi||\pi_1)}{\eta}\} \\
&= \frac{\eta M^2 T}{8} + \inf_{\pi} \{ \textstyle\sum_{t=1}^{T} R_{Q_t}(\pi) + \frac{KL(\pi||\pi_1)}{\eta}\}
\end{aligned},$$

where the third equality follows from the duality formula for Kullback-Leibler divergence (Catoni, 2003). Denote $\omega^* = \arg\min \sum_{t=1}^{T} R_{Q_t}(\omega)$. Consider a parametric distribution family $p_c(d\omega) \propto \mathbb{1}\{||\omega - \omega^*||_2 \leq c\}\pi_1(d\omega)$ where $\mathbb{1}\{\cdot\}$ is the indicator function. Note that when c is small, p_c highly concentrates on ω^*. Now we have

$$KL(p_c||\pi_1) = -\log(\pi_1(||\omega - \omega^*||_2 \leq c)),$$

and

$$\pi_1(||\omega - \omega^*||_2 \leq c) \geq \frac{\pi^{d/2}(c/2)^d}{\Gamma(d/2+1)} \Big/ \frac{\pi^{d/2}(\tilde{B}_\Omega)^d}{\Gamma(d/2+1)} = (\frac{c}{2\tilde{B}_\Omega})^d,$$

where the inequality is due to the fact that since π_1 is uniformly distributed on the $\tilde{B}_\Omega$ d-ball, the probability to be calculated is greater than the ratio between the volume of d-ball with radius of $c/2$ and the volume of d-ball with radius $\tilde{B}_\Omega$. So we get

$$KL(p_c||\pi_1) \leq d\log(\frac{2\tilde{B}_\Omega}{c}).$$

Furthermore, by Lipschitz assumption,

$$\sum_{t=1}^{T} R_{Q_t}(p_c) \leq \sum_{t=1}^{T} R_{Q_t}(\omega^*) + TKc.$$

Therefore,

$$\begin{aligned}
&\textstyle\sum_{t=1}^{T} R_{Q_t}(\pi_t) \leq \frac{\eta M^2 T}{8} + \inf_{\pi} \{ \textstyle\sum_{t=1}^{T} R_{Q_t}(\pi) + \frac{KL(\pi||\pi_1)}{\eta}\} \\
&\leq \frac{\eta M^2 T}{8} + \inf_{c} \{ \textstyle\sum_{t=1}^{T} R_{Q_t}(p_c) + \frac{KL(p_c||\pi_1)}{\eta}\} \\
&\leq \textstyle\sum_{t=1}^{T} R_{Q_t}(\omega^*) + \frac{\eta M^2 T}{8} + \inf_{c} \{TKc + \frac{d}{\eta}\log(\frac{2\tilde{B}_\Omega}{c})\} \\
&\leq \textstyle\sum_{t=1}^{T} R_{Q_t}(\omega^*) + \frac{\eta M^2 T}{8} + \frac{\mu_0 T}{4} + \frac{d}{\eta}\log(\frac{8K\tilde{B}_\Omega}{\mu_0})
\end{aligned},$$

where the last equality follows by choosing $c = \frac{\mu_0}{4K}$. We complete the proof. □

B.3 PROOF OF THEOREM 2

Proof. The proof follows the same logic as that of Theorem 1. First, since the strategy for λ-player remains the same, the regret bound obtained in Theorem 1 still holds, that is, for any $\lambda \in \Lambda$,

$$\sum_{t=1}^{T} \lambda r_t \leq \sum_{t=1}^{T} \lambda_t r_t + \xi_T.$$

Then the suboptimality of $L(\hat{Q}_T, \hat{\pi}_T, \hat{\lambda}_T)$ is bounded as follows. For any (π, λ),

$$\begin{aligned}
&L(\hat{Q}_T, \pi, \lambda) = -R_{\hat{Q}_T}(\pi) + \lambda(W(\hat{Q}_T, P_n) - \epsilon). \\
&\leq \frac{1}{T}\textstyle\sum_{t=1}^{T}(-R_{Q_t}(\pi) + \lambda(W(Q_t, P_n) - \epsilon)) \\
&\leq \frac{1}{T}\textstyle\sum_{t=1}^{T}(-R_{Q_t}(\pi_t) + \lambda_t(W(Q_t, P_n) - \epsilon)) + \frac{\xi_T}{T} + \frac{\tau_T}{T} \\
&= \frac{1}{T}\textstyle\sum_{t=1}^{T} L(Q_t, \pi_t, \lambda_t) + \frac{\xi_T}{T} + \frac{\tau_T}{T} \\
&\leq \frac{1}{T}\textstyle\sum_{t=1}^{T} L(\hat{Q}_T, \pi_t, \lambda_t) + \mu_1 + \frac{\xi_T}{T} + \frac{\tau_T}{T} \\
&\leq \frac{1}{T}\textstyle\sum_{t=1}^{T} L(\hat{Q}_T, \omega_t, \lambda_t) + \mu_1 + \frac{\xi_T}{T} + \frac{\tau_T}{T} + M\sqrt{\frac{\log(1/\delta)}{2T}} \; (with\ probability\ at\ least\ 1-\delta) \\
&= L(\hat{Q}_T, \hat{\pi}_T, \hat{\lambda}_T) + \mu_1 + \frac{\xi_T}{T} + \frac{\tau_T}{T} + M\sqrt{\frac{\log(1/\delta)}{2T}}
\end{aligned},$$

where the first inequality uses the linearity of $-R_Q(\pi)$ in Q and the convexity of mapping $u \to W(u, v)$, the second inequality follows from the regret bound for λ-player and Lemma 2, the third inequality is by the choice of Q_t, and the last inequality uses Chernoff bound.

Also, for any Q,

$$\begin{aligned}
&L(Q, \hat{\pi}_T, \hat{\lambda}_T) = \frac{1}{T}\textstyle\sum_{t=1}^{T} L(Q, \omega_t, \lambda_t)) \\
&\geq \frac{1}{T}\textstyle\sum_{t=1}^{T} L(Q, \pi_t, \lambda_t)) - M\sqrt{\frac{\log(1/\delta)}{2T}} \; (with\ probability\ at\ least\ 1-\delta) \\
&\geq \frac{1}{T}\textstyle\sum_{t=1}^{T} L(Q_t, \pi_t, \lambda_t) - \mu_1 - M\sqrt{\frac{\log(1/\delta)}{2T}} \\
&= \frac{1}{T}\textstyle\sum_{t=1}^{T}(-R_{Q_t}(\pi_t) + \lambda_t(W(Q_t, P_n) - \epsilon)) - \mu_1 - M\sqrt{\frac{\log(1/\delta)}{2T}} \\
&\geq \frac{1}{T}\textstyle\sum_{t=1}^{T}(-R_{Q_t}(\hat{\pi}_T) + \hat{\lambda}_T(W(Q_t, P_n) - \epsilon)) - \frac{\xi_T}{T} - \frac{\tau_T}{T} - \mu_1 - M\sqrt{\frac{\log(1/\delta)}{2T}} \\
&\geq -R_{\hat{Q}_T}(\hat{\pi}_T) + \hat{\lambda}_T(W(\hat{Q}_T, P_n) - \epsilon) - \frac{\xi_T}{T} - \frac{\tau_T}{T} - \mu_1 - M\sqrt{\frac{\log(1/\delta)}{2T}} \\
&= L(\hat{Q}_T, \hat{\pi}_T, \hat{\lambda}_T) - \frac{\xi_T}{T} - \frac{\tau_T}{T} - \mu_1 - M\sqrt{\frac{\log(1/\delta)}{2T}}
\end{aligned},$$

where the first inequality uses Chernoff bound, the second inequality follows from the definition of Q_t, the third inequality uses the regret bound and Lemma 2, and the last inequality is due to convexity of Wasserstein distance.

The above results immediately imply that with probability at least $1 - 2\delta$,

$$\begin{aligned}
&\mu_T \leq \frac{\xi_T}{T} + \frac{\tau_T}{T} + \mu_1 + M\sqrt{\frac{\log(1/\delta)}{2T}} \\
&= \frac{B\log(2)}{T\alpha} + \alpha\rho^2 B + \frac{\eta M^2}{8} + \frac{\mu_0}{4} + \frac{d}{\eta T}\log(\frac{8K\tilde{B}_\Omega}{\mu_0}) + M\sqrt{\frac{\log(1/\delta)}{2T}} + \mu_1
\end{aligned}.$$

By plugging $\alpha = \frac{\mu_0}{4\rho^2 B}$ and $\eta = \frac{2\mu_0}{M^2}$, if $T \geq \frac{8M^2\log(1/\delta) + 32B^2\rho^2\log(2) + 4dM^2\log(8K\tilde{B}_\Omega/\mu_0)}{\mu_0^2}$, we can verify that $\mu_T \leq \mu$. □

B.4 PROOF OF LEMMA 3

The proof of Lemma 3 uses the following strong duality result by Blanchet & Murthy (2019):

Proposition 1. *Let $f : \mathcal{Z} \to \mathbb{R}$ be upper semicontinuous. Denote $\chi_{f,\vartheta}(z) := \sup_{z' \in \mathcal{Z}}\{f(z') - \vartheta \cdot d_{\mathcal{Z}}(z, z')\}$. Then for any $P \in \mathcal{P}_1(\mathcal{Z})$,*

$$\sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(f) = \min_{\vartheta\geq 0}\{\vartheta\tilde{\epsilon} + \mathbb{E}_P[\chi_{f,\vartheta}(z)]\},$$

and for any $\vartheta \geq 0$,

$$\sup_Q\{R_Q(f) - \vartheta W(Q, P)\} = \mathbb{E}_P[\chi_{f,\vartheta}(z)].$$

Proof. First by definition, $BEST_Q(\omega, \lambda) = \arg\min_Q L(Q, \omega, \lambda) = \arg\min_Q[-R_Q(\omega) + \lambda(W(Q, P_n) - \epsilon)] = \arg\max_Q[R_Q(\omega) - \lambda W(Q, P_n)]$, i.e., $BEST_Q$ is the maximizer of $R_Q(\omega) - \lambda W(Q, P_n)$. Then for $Q_{\omega,\lambda}$, we have

$$\begin{array}{l} R_{Q_{\omega,\lambda}}(\omega) - \lambda W(Q_{\omega,\lambda}, P_n) \leq \sup_Q\{R_Q(\omega) - \lambda W(Q, P_n)\} = \mathbb{E}_{P_n}[\chi_{\omega,\lambda}(z)] \\ = \mathbb{E}_{P_n}[l(\omega, T_{\pi,\lambda}(z))] - \lambda\mathbb{E}_{P_n}[d_{\mathcal{Z}}(T_{\omega,\lambda}(z), z)] \\ = R_{Q_{\omega,\lambda}}(\omega) - \lambda\mathbb{E}_{P_n}[d_{\mathcal{Z}}(T_{\omega,\lambda}(z), z)] \leq R_{Q_{\omega,\lambda}}(\omega) - \lambda W(Q_{\omega,\lambda}, P_n) \end{array},$$

where the first equality follows from Proposition 1 and the last inequality uses the definition of Wasserstein distance. Therefore we obtain $R_{Q_{\omega,\lambda}}(\omega) - \lambda W(Q_{\omega,\lambda}, P_n) = \sup_Q\{R_Q(\omega) - \lambda W(Q, P_n)\}$ and $W(Q_{\omega,\lambda}, P_n) = \mathbb{E}_{P_n}[d_{\mathcal{Z}}(T_{\omega,\lambda}(z), z)]$, which means that the best response $BEST_Q$ can be attained by $Q_{\omega,\lambda}$. We complete the proof. □

C PROOFS IN SECTION 3.3

C.1 PROOF OF LEMMA 4

Proof. Define $\lambda = 0$ if $W(\hat{Q}, P_n) \leq \epsilon$ otherwise $\lambda = B$. Then we have

$$\begin{array}{l} \mu - R_{P_n}(\hat{\pi}) + \hat{\lambda}(W(P_n, P_n) - \epsilon) \\ = \mu + L(P_n, \hat{\pi}, \hat{\lambda}) \geq L(\hat{Q}, \hat{\pi}, \hat{\lambda}) \\ \geq L(\hat{Q}, \hat{\pi}, \lambda) - \mu \\ = -R_{\hat{Q}}(\hat{\pi}) + \lambda(W(\hat{Q}, P_n) - \epsilon) - \mu \\ = -R_{\hat{Q}}(\hat{\pi}) + B(W(\hat{Q}, P_n) - \epsilon)_+ - \mu \end{array},$$

where the fist and second inequalities follows from the definition of μ-approximate equilibrium and the last equality uses the definition of λ and notation $x_+ = \max\{x, 0\}$. Arranging the terms on both sides and using $R_{\hat{Q}}(\hat{\pi}) - R_{P_n}(\hat{\pi}) \leq M$ by assumption gives

$$2\mu + M - \hat{\lambda}\epsilon \geq B(W(\hat{Q}, P_n) - \epsilon)_+ \geq 0.$$

Thus, we obtain $\hat{\lambda} \leq \dfrac{2\mu + M}{\epsilon}$, proving the first part of the lemma. The second part of the lemma follows immediately by $(W(\hat{Q}, P_n) - \epsilon)_+ \geq W(\hat{Q}, P_n) - \epsilon$ and $2\mu + M \geq 2\mu + M - \hat{\lambda}\epsilon$. □

C.2 PROOF OF THEOREM 3

We begin by establishing an auxiliary lemma:

Lemma C.1. *Under the assumptions of Theorem 3, denote $\mathcal{F} = \{l(\omega, z) : \omega \in \Omega\}$ and $\mathcal{F}_T = \{\frac{1}{T}\sum_{i=1}^T l(\omega_i, z) : \omega_i \in \Omega\}$ for any given T, then with probability at least $1 - \delta$, for all $g \in \mathcal{F}_T$,*

$$\sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(g) - \sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(g) \leq \frac{48\sqrt{T}\mathfrak{C}(\mathcal{F})}{\sqrt{n}} + M\sqrt{\frac{\log(\frac{1}{\delta})}{2n}} + \frac{48}{\sqrt{n}}C \cdot diam(\mathcal{Z}) \ ,$$

where $\mathfrak{C}(\mathcal{F}) = \int_0^\infty \sqrt{log\mathcal{N}(\mathcal{F}, ||\cdot||_\infty, u/2)}du$ and $\mathcal{N}(\mathcal{F}, ||\cdot||_\infty, u/2)$ denotes the covering number of $\mathcal{F}$.

Proof. First, by Theorem 2 of Lee & Raginsky (2018), we have with probability at least $1-\delta$, for all $g \in \mathcal{F}_T$,

$$\sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(g) - \sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(g) \leq \frac{48\mathfrak{C}(\mathcal{F}_T)}{\sqrt{n}} + M\sqrt{\frac{\log(\frac{1}{\delta})}{2n}} + \frac{48}{\sqrt{n}}C \cdot diam(\mathcal{Z}) \ .$$

Then, for any u, suppose $\mathcal{F}'$ is a $u/2$-covering set of $\mathcal{F}$. Define $\mathcal{F}'_T = \{\frac{1}{T}\sum_{i=1}^T f_i : f_i \in \mathcal{F}'\}$. It is easy to verify that $\mathcal{F}'_T$ is also a $u/2$-covering set of $\mathcal{F}_T$. Therefore we have $\mathcal{N}(\mathcal{F}_T, ||\cdot||_\infty, u/2) \leq \mathcal{N}(\mathcal{F}, ||\cdot||_\infty, u/2)^T$. The lemma now follows by $\mathfrak{C}(\mathcal{F}_T) \leq \sqrt{T}\mathfrak{C}(\mathcal{F})$. □

Now we are ready to prove Theorem 3.

Proof. We start by decomposing the excess risk.

$$\begin{aligned}
&\sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(\hat{\pi}) - \sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(\pi^*) \\
&= \sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(\hat{\pi}) - \sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(\hat{\pi}) + \sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(\hat{\pi}) - \sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(\pi^*) + \\
&\sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(\pi^*) - \sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(\pi^*)
\end{aligned} \ .$$

For the first difference term, by Lemma C.1, we have with probability at least $1-\delta/2$,

$$\sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(\hat{\pi}) - \sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(\hat{\pi}) \leq \mathcal{O}(1/\sqrt{n}) + M\sqrt{\frac{\log(\frac{2}{\delta})}{2n}} \ .$$

We next bound the second difference term. Denote $\bar{Q} = \arg\min_{Q:W(Q,P_n)\leq\tilde{\epsilon}} -R_Q(\hat{\pi})$. Since $\tilde{\epsilon} = \epsilon + \dfrac{2\mu+M}{B}$ in our setting, by Lemma 4, we have $W(\hat{Q}, P_n) \leq \tilde{\epsilon}$. Then,

$$\begin{aligned}
&\sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(\hat{\pi}) - \sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(\pi^*) \\
&= \min_{Q:W(Q,P_n)\leq\tilde{\epsilon}} -R_Q(\pi^*) - \min_{Q:W(Q,P_n)\leq\tilde{\epsilon}} -R_Q(\hat{\pi}) \\
&\leq -R_{\hat{Q}}(\pi^*) - \min_{Q:W(Q,P_n)\leq\tilde{\epsilon}} -R_Q(\hat{\pi}) \\
&= L(\hat{Q}, \pi^*, 0) - \min_{Q:W(Q,P_n)\leq\tilde{\epsilon}} -R_Q(\hat{\pi}) \\
&\leq L(\hat{Q}, \hat{\pi}, \hat{\lambda}) - \min_{Q:W(Q,P_n)\leq\tilde{\epsilon}} -R_Q(\hat{\pi}) + \mu \\
&\leq L(\bar{Q}, \hat{\pi}, \hat{\lambda}) + R_{\bar{Q}}(\hat{\pi}) + 2\mu \\
&= -R_{\bar{Q}}(\hat{\pi}) + \hat{\lambda}(W(\bar{Q}, P_n) - \epsilon) + R_{\bar{Q}}(\hat{\pi}) + 2\mu \\
&= \hat{\lambda}(W(\bar{Q}, P_n) - \epsilon) + 2\mu \\
&\leq \hat{\lambda}(\tilde{\epsilon} - \epsilon) + 2\mu \leq \frac{(2\mu+M)^2}{B\epsilon} + 2\mu = 3\mu
\end{aligned} \ ,$$

where the first inequality follows from $W(\hat{Q}, P_n) \leq \tilde{\epsilon}$, the second and third inequalities use the definition of μ-approximate equilibrium, the fourth inequality is due to the fact that $W(\bar{Q}, P_n) \leq \tilde{\epsilon}$ by the definition of $\bar{Q}$ and the last inequality uses Lemma 4.

Finally, by Proposition 1, the last difference term can be rewritten as

$$\begin{aligned}
&\sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(\pi^*) - \sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(\pi^*) \\
&= \min_{\vartheta\geq 0}\{\vartheta\tilde{\epsilon} + \int_{\mathcal{Z}} \chi_{\vartheta, l_{\pi^*}}(z) P_n(dz)\} - \min_{\vartheta\geq 0}\{\vartheta\tilde{\epsilon} + \int_{\mathcal{Z}} \chi_{\vartheta, l_{\pi^*}}(z) P(dz)\}
\end{aligned} \ .$$

Denote $\bar{\vartheta} = \arg\min_{\vartheta\geq 0}\{\vartheta\tilde{\epsilon} + \int_{\mathcal{Z}} \chi_{\vartheta, l_{\pi^*}}(z) P(dz)\}$. Then, we have

$$\sup_{Q:W(Q,P_n)\leq\tilde{\epsilon}} R_Q(\pi^*) - \sup_{Q:W(Q,P)\leq\tilde{\epsilon}} R_Q(\pi^*) \leq \int_{\mathcal{Z}} \chi_{\bar{\vartheta}, l_{\pi^*}}(z)(P_n - P)(dz) \ .$$

Algorithm 4 Pseudo-code of the three players

Sample $\{z_i\}_{i=1}^m \sim P_n$
Q-player update:
for $i = 1, 2, \cdots, m$ **do**
  $z_i^k \leftarrow z_i$
  Update z_i^k by ascending its stochastic descents
  if non-convex **then**
    Sample ω from π
  end if
  $z_i^k \leftarrow z_i^k + \kappa \partial_z [l(\omega, z_i^k) - \lambda d_{\mathcal{Z}}(z_i^k, z_i)]$
end for
λ-player update:
Update θ by
$\theta \leftarrow \theta + \alpha(\frac{1}{m}\sum_{i=1}^m d_{\mathcal{Z}}(z_i^k, z_i) - \epsilon)$
π-player update:
Case 1: Convex loss
Update ω by
$\omega \leftarrow \Pi_\Omega(w - \beta(\frac{1}{m}\sum_{i=1}^m \partial_\omega l(\omega, z_i^k)))$
Case 2: Non-convex loss
Metropolis-Hastings algorithm: Repeat N times
for $j = 0, 1, \cdots, N-1$ **do**
  Sample $u \sim \mathcal{U}_{[0,1]}$
  Sample $\omega' \sim \mathcal{N}(\omega, \sigma^2 \mathcal{I})$
  if $u < \min\{1, \exp[\eta(\sum_{t'=1}^t \frac{1}{m}\sum_{i=1}^m (l(\omega, z_{it'}^k) - l(\omega', z_{it'}^k)))]\}$ **then**
    $\omega \leftarrow \omega'$
  end if
end for

Since $l(\omega, z)$ takes values in $[0, M]$ by assumption, the same holds for l_{π^*} and $\chi_{\bar{\vartheta}, l_{\pi^*}}$. From Hoeffdings inequality, it follows that

$$\sup_{Q: W(Q, P_n) \leq \tilde{\epsilon}} R_Q(\pi^*) - \sup_{Q: W(Q,P) \leq \tilde{\epsilon}} R_Q(\pi^*) \leq M\sqrt{\frac{\log(\frac{2}{\delta})}{2n}}$$

holds with probability at least $1 - \delta/2$.

Collecting all these terms and applying the union bound, we have

$$\sup_{Q: W(Q,P) \leq \tilde{\epsilon}} R_Q(\hat{\pi}) - \sup_{Q: W(Q,P) \leq \tilde{\epsilon}} R_Q(\pi^*) \leq \mathcal{O}(1/\sqrt{n}) + 2M\sqrt{\frac{\log(\frac{2}{\delta})}{2n}} + 3\mu$$

with probability at least $1 - \delta$. □

D PSEUDO-CODE

The pseudo-code of each player is summarized in Algorithm 4.

E ADDITIONAL EXPERIMENTAL RESULTS

Here, we provide the time consumption of our algorithms and the baselines used in the experimental section. All the experiments are conducted on a single NVIDIA TITAN Xp GPU. We compute the overall training time on MNIST for 20 epochs. Table 2 summarizes the running time for different algorithms. We observe that in convex case the time cost of our algorithm is close to LR and DRLR,

Table 2: The running time for different algorithms on MNIST.

	Methods	**Training Duration**
Non-convex (hours)	ERM	0.1
	ITP	0.1
	WRM	0.6
	TPG-DRO	1
Convex (mins)	LR	3
	DRLR	3
	TPG-DRO (LR)	6

Table 3: Average classification accuracy on test datasets with LeNet.

DATASETS	**ERM**	**ITP**	**WRM**	**TPG-DRO**
USPS	76.2%$\pm$2.1%	76.0%$\pm$1.4%	81.0%$\pm$1.0%	81.6%$\pm$2.3%
SVHN	24.1%$\pm$1.2%	24.4%$\pm$1.5%	32.2%$\pm$2.1%	33.4%$\pm$2.5%
MNIST-M	48.0%$\pm$1.9%	51.4%$\pm$0.2%	58.6%$\pm$0.7%	57.2%$\pm$1.0%
SYN	34.7%$\pm$1.2%	37.0%$\pm$0.6%	44.7%$\pm$0.8%	45.1% $\pm$1.0%

but in non-convex case our algorithm is a little bit slower than other baselines mainly because of the sampling step. We believe that this time gap in non-convex setting can be greatly reduced by using more advanced MCMC sampling method. Table 3 reports the results using LeNet. As in the body, the model trained with our algorithm achieves higher or at least comparable accuracy on all test datasets compared to the models trained with ERM, ITP and WRM.