Iterative Sampling Methods for Sinkhorn Distributionally Robust Optimization

Jie Wang
School of Artificial Intelligence, School of Data Science
The Chinese University of Hong Kong, Shenzhen
jwang@cuhk.edu.cn

Abstract

Distributionally robust optimization (DRO) has emerged as a powerful paradigm for reliable decision-making under uncertainty. This paper focuses on DRO with ambiguity sets defined via the Sinkhorn discrepancy—an entropy-regularized Wasserstein distance—referred to as Sinkhorn DRO. Existing work primarily addresses Sinkhorn DRO from a dual perspective, leveraging its formulation as a conditional stochastic optimization problem, for which many stochastic gradient methods are applicable. However, the theoretical analyses of such methods often rely on the boundedness of the loss function, and it is indirect to obtain the worst-case distribution associated with Sinkhorn DRO. In contrast, we study Sinkhorn DRO from the primal perspective, by reformulating it as a bilevel program with several infinite-dimensional lower-level subproblems over probability space. This formulation enables us to simultaneously obtain the optimal robust decision and the worst-case distribution, which is valuable in practical settings, such as generating stress-test scenarios or designing robust learning algorithms. We propose both double-loop and single-loop sampling-based algorithms with theoretical guarantees to solve this bilevel program. Finally, we demonstrate the effectiveness of our approach through a numerical study on adversarial classification.

1 Introduction

Distributionally Robust Optimization (DRO) has emerged as a powerful framework for decision-making under uncertainty, aiming to find models that perform well under a set of plausible data distributions, known as the ambiguity set. A prominent approach defines this set using the Wasserstein distance [3, 4, 20, 22]. However, the practical application of Wasserstein DRO is limited by two things. First, the resulting optimization problem is often intractable for general cases, except under some special conditions on the loss function and transportation cost [20, 22, 45, 54, 58], and its worst-case distributions are typically discrete, which can be unrealistic for underlying continuous data and lead to overly conservative decisions.

The Sinkhorn DRO formulation [2, 64, 66, 68, 70] addresses these issues by employing an entropy-regularized Wasserstein distance [15], or called the Sinkhorn discrepancy. This regularization brings two critical benefits. First, this problem is tractable for a broad class of loss functions. Second, it naturally encourages continuous worst-case distributions, often leading to improved generalization. Due to these advantages, Sinkhorn DRO has found success in various applications such as hypothesis testing [69, 73, 80], experimental design [18, 36], machine learning [7, 51, 57, 59], etc.

Despite its promise, the development of efficient algorithms for Sinkhorn DRO remains an active area of research. Existing methods solve the Sinkhorn DRO from its dual perspective, where the dual objective can be expressed as an expectation of the logarithm of another conditional expectation on the exponential of the risk function. It falls into the class of *conditional stochastic*

Preprint.

optimization (CSO) [30], and therefore many stochastic gradient methods, such as multi-level Monte-Carlo gradient methods, can be applied to solve it [30–34]. However, this dual approach has notable limitations. To achieve rate-optimal convergence, existing results require the restrictive assumption that the risk function is bounded [32, 68]. Although recent work has relaxed this assumption [81], it suffers from a sub-optimal convergence rate.

Recently, generative artificial intelligence (GenAI), especially the score-based generative models [1, 40, 42, 60, 61], have drawn much research attention, given their state-of-the-art performance in image and text generation. These models operate by learning the score function from collected samples, and next sampling from the score function to bridge Gaussian noise to plausible data samples via Langevin dynamics. In contrast, Sinkhorn DRO aims to learn a worst-case distribution that maximizes a given risk function, rather than choosing a target distribution as a priori. Based on samples from this worst-case distribution, an optimal robust decision is then trained to minimize the worst-case risk. Inspired by GenAI, this paper investigates the following question:

> *Can we simultaneously obtain the optimal robust decision and corresponding samples from the worst-case distribution in Sinkhorn DRO using a framework similar to that of score-based generative models?*

In this paper, we propose a novel primal perspective and iterative sampling algorithms for solving Sinkhorn DRO. Our main contributions are summarized as follows.

Reformulating Sinkhorn DRO as Infinite-Dimensional Bilevel Program. We consider the soft-constrained Sinkhorn DRO formulation using n samples, where the Sinkhorn discrepancy ball constraint is put in the objective as a penalty term. We show that this type of formulation can be equivalently formulated as a bilevel optimization with n lower-level subproblems (See Problem (5)). Each subproblem seeks the worst-case distribution from the smooth density constructed from the empirical sample point, and the estimated worst-case distribution for Sinkhorn DRO is the average among all worst-case distributions constructed from lower-level problems. This reformulation inspires us to leverage existing methods from the bilevel program to solve Sinkhorn DRO.

Naïve Double-Loop Algorithm. We first develop a double-loop algorithm to solving the bilevel program: at each iteration, we randomly sample a lower-level problem, run an inner loop for a large number of steps to generate the sampling point close to the worst-case distribution of the lower-level problem, and use it to construct a hyper-gradient estimator. We show that under the smoothness assumption of the loss together with a bounded variance condition, this double-loop algorithm finds a ϱ-stationary point of the bilevel program with complexity $\mathcal{O}(\varrho^{-6})$, where the complexity is defined as the number of queries to the gradient of the loss function.

Single-Loop Algorithm. Next, we develop a single-loop algorithm by jointly updating the upper-level decision variable and lower-level sampling points. Specifically, at each iteration, we update several sampling points associated with a randomly-sampled batch of lower-level problems, generate a hypergradient estimator, and employ a momentum update for the upper-level decision variable. For theoretical analysis, we make an extra assumption that the upper-level hypergradient estimator is constructed using the limit of finitely many sampling points described by a mean-field density. Then, we show that the iteration complexity becomes $\mathcal{O}(\varrho^{-4})$ for finding ϱ-stationary point. Compared with the double-loop algorithm, this new algorithm only updates $\mathcal{O}(1)$ lower-level problems by one Langevin dynamics step at each iteration, which is computationally and storage efficient.

Numerical Study. Finally, we examine our algorithms in numerical study of adversarial classification. When taking hypergradient estimator as finitely many sampling points for the single-loop algorithm, we still observe convergence of the proposed algorithm. Also, our single-loop algorithm provides both estimated upper-level decision and estimated sampled points for worst-case distributions of Sinkhorn DRO, which is beneficial for practical high-stake environments that require stress testing.

1.1 Related Literature

During the preparation of this manuscript, we became aware of the concurrent work of Xu and Zhu [79], which also studies sampling-based approaches for general DRO problems and derives convergence rates. Our focus and our analysis differ from theirs in the following ways. First, we presented both double-loop and single-loop algorithms to solve Sinkhorn DRO, whereas they focused only on the double-loop algorithm. Second, there exists a technical flaw in their analysis, since the constant step size makes the last component of the equation above [79, Appendix C4] non-vanishing. Instead, we provided a rigorous analysis with less restrictive assumptions.

Next, we review papers on several related topics.

DRO and GenAI. Recent work seeks to explore the integration of GenAI with DRO [11, 74, 78, 83]. Specifically, Xu et al. [78] and Cheng et al. [11] proposed a flow-based generative model to jointly learn the worst-case distribution and the robust decision for Wasserstein DRO. This framework provides an efficient sampler for the worst-case distribution, which is valuable for stress tests and high-stakes environments [71, 72]. The convergence rate of this framework was later analyzed by [11, 83]. In a different approach, Wen and Yang [74] constructed a new DRO framework with an ambiguity set defined using the diffusion model. While these works integrate generative models into the DRO framework itself, our approach employs Langevin dynamics, a foundational tool for the inference of these models, as a key component for solving the Sinkhorn DRO subproblems.

Bilevel Optimization. Sinkhorn DRO can be formulated as a bilevel optimization with multiple lower-level subproblems, which has been studied in recent literature [25, 28, 29, 33]. Especially, the complexity of the algorithms in [25, 28, 29] grows linearly with the number of lower-level problems n. Hu et al. [33] provided a double-loop algorithm that finds the ϱ-stationary point with complexity $\mathcal{O}(\varrho^{-4})$, which is independent of n and matches the lower bound for general stochastic optimization. However, all these papers assume that the lower-level subproblems are finite-dimensional and strongly convex programs. Our work extends this line to the infinite-dimensional setting, where each lower-level subproblem is an optimization over a probability distribution. Our framework is closely related to Marion et al. [47] and Xiao et al. [77]. These two references only consider bilevel optimization with a single lower-level infinite-dimensional subproblem, which is not applicable to our setting. In addition, Xiao et al. [77] provides a double-loop algorithm, which could be computationally and storage inefficient. We integrate the idea of the single-loop algorithm in [47] and the momentum-based gradient estimator in Hu et al. [29] to design our algorithm in Section 4.

Infinite-Dimensional Optimization. DRO is a special instance of infinite-dimensional optimization. Traditional approaches to these problems can be broadly categorized into three perspectives: discretization, duality, and Frank-Wolfe methods. First, discretization of the decision space provides a straightforward solution by solving a finite-dimensional tractable approximation problem [6, 9, 23, 44, 63]. However, this approach suffers from the curse of dimensionality and becomes infeasible in high-dimensional settings. Second, duality is a powerful approach when the number of constraints is finite, reformulating the primal problem into a finite-dimensional dual via strong duality [3, 22, 55, 56, 75, 82]. A significant drawback is that the resulting dual problem often contains intractable, infinite-dimensional, or nonconvex subproblems. Third, Frank-Wolfe methods provide a flexible framework for optimization over probability spaces by iteratively solving linear minimization subproblems [35, 37, 53, 67]. While each iteration involves a finite-dimensional task, this subproblem is often non-convex and high-dimensional, making it difficult to apply without additional problem structure. A more recent line of work leverages Langevin dynamics for a specific class of problems: entropy-regularized linear optimization over probability distributions [10, 17, 19, 46, 65]. This approach efficiently produces sampled points of the optimal distribution, providing a practical pathway for solving a broad class of infinite-dimensional problems. This insight also lays the foundation of modern diffusion models. In this work, we adapt and extend the Langevin dynamics framework to design efficient Sinkhorn DRO algorithms.

1.2 Notations

For an integer $n \geq 1$, we define $[n] = \{1, 2, \ldots, n\}$. We write $\mathbb{E}_{\mu}[f]$ or $\mathbb{E}_{z\sim\mu}[f(z)]$ to denote the expected value of $f(z)$ with respect to $z \sim \mu$. Denote by $\mathcal{P}(\mathbb{R}^d)$ the set of probability distributions supported on $\mathbb{R}^d$. Throughout the paper, we assume the desired accuracy level $\varrho, \delta > 0$ is sufficiently small. We use the notation $\mathcal{O}(\cdot)$ to denote the order in terms of ϱ or δ, and $\widetilde{\mathcal{O}}(\cdot)$ hides terms having the polylog dependency on ϱ or δ. We provide a list of key mathematical notations and definitions in Appendix A.

1.3 Organizations

In Section 2, we provide a brief review of the existing Sinkhorn DRO results, based on which we reformulate this problem as a bilevel optimization with several infinite-dimensional lower-level problems. In Section 3, we provide a double-loop iterative sampling algorithm for solving the bilevel program, and analyze its convergence rate. In Section 4, we provide a single-loop sampling algorithm for solving the bilevel program, and analyze its convergence rate. In Section 5, we provide the numerical study of the distributionally robust classification problem to validate the effectiveness of our proposed algorithms. In Section 6, we provide several concluding remarks.

2 Setup of Sinkhorn DRO

We consider the Sinkhorn DRO problem, which takes the form of the minimax optimization:

$$\min_{\theta} \max_{\mu:\mathcal{S}_\epsilon(\widehat{\mu},\mu)\leq\rho} \mathbb{E}_{z\sim\mu}[f_\theta(z)], \tag{1}$$

where $\theta \in \mathbb{R}^{d_\theta}$ represents the model parameters, $f_\theta(z)$ is the loss function, and $\mathcal{S}_\epsilon(\widehat{\mu},\cdot)$ is the Sinkhorn Discrepancy (see Definition 2.1) that defines an entropy-regularized ambiguity set around the reference distribution $\widehat{\mu}$. Following reference [68], we provide the definition of Sinkhorn Discrepancy and the strong duality result of Sinkhorn DRO below.

Definition 2.1 (Sinkhorn Discrepancy)**.** For regularization parameter $\epsilon \geq 0$, the Sinkhorn Discrepancy between two distributions μ and ν is defined as

$$\mathcal{S}_\epsilon(\mu,\nu) = \inf_{\gamma\in\Gamma(\mu,\nu)} \left\{ \frac{1}{2}\mathbb{E}_{(x,y)\sim\gamma}[\|x-y\|_2^2] + \epsilon\mathbb{E}_{(x,y)\sim\gamma}\left[\log\frac{\mathrm{d}\gamma(x,y)}{\mathrm{d}\mu(x)\,\mathrm{d}y}\right]\right\},$$

where $\Gamma(\mu,\nu)$ denotes the set of joint distributions with marginal distributions being μ and ν, respectively.

We take the reference distribution $\widehat{\mu}$ as an empirical distribution supported on n available data samples $\{x^{(1)},\ldots,x^{(n)}\}$, i.e., $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n \delta_{x^{(i)}}$. In this context, Problem (1) can be interpreted as an entropy-regularized approximation of the classical Wasserstein DRO problem with a quadratic cost function. While our focus is on this specific setup, the framework can be extended to general cost functions and reference distributions, which we leave for future work. The penalized version of Problem (1) admits a strong duality result [68], leading to a tractable finite-dimensional reformulation that is crucial for our algorithmic development.

Theorem 2.2 (Strong Duality [68])**.** *Consider the penalized counterpart of Problem* (1)*:*

$$\min_{\theta}\left\{\max_{\mu\in\mathcal{P}(\mathbb{R}^d)} \mathbb{E}_{z\sim\mu}[f_\theta(z)] - \lambda\mathcal{S}_\epsilon(\widehat{\mu},\mu)\right\}. \tag{2}$$

Up to an additive constant independent of θ, Problem (2) *is equivalent to:*

$$\min_{\theta}\left\{\frac{\lambda\epsilon}{n}\sum_{i=1}^n\left[\log\mathbb{E}_{z\sim\mathcal{N}(x^{(i)},\epsilon\mathbf{I}_d)}\left[e^{f_\theta(z)/(\lambda\epsilon)}\right]\right]\right\}.$$

Define the density function

$$u_{\theta,i}(z) := \frac{\mathrm{d}\mu_*^{\theta,i}(z)}{\mathrm{d}z} = \alpha_i\cdot\exp\left(\frac{f_\theta(z)-\frac{1}{2}\lambda\|x^{(i)}-z\|_2^2}{\lambda\epsilon}\right), \tag{3}$$

where $\alpha_i = \left(\int\exp\left(\frac{f_\theta(z)-\frac{1}{2}\lambda\|x^{(i)}-z\|_2^2}{\lambda\epsilon}\right)\mathrm{d}z\right)^{-1}, i\in[n]$ *denotes the normalizing constant. For fixed θ, the worst-case distribution of Problem* (2) *has density given by*

$$\frac{\mathrm{d}\mu_*^\theta(z)}{\mathrm{d}z} = \frac{1}{n}\sum_{i=1}^n u_{\theta,i}(z). \tag{4}$$

Remark 2.3 (Soft-Constrained Sinkhorn DRO)**.** The penalized formulation in (2) is a Lagrangian counterpart to the constrained original Problem (1). Here, the radius ρ is replaced by a penalty parameter $\lambda > 0$. In practice, achieving good out-of-sample performance requires tuning hyperparameters: either (ρ, ϵ) in the constrained formulation or the equivalent pair (λ, ϵ) in the penalized one. Our algorithmic development in subsequent sections focuses on solving the penalized formulation (2). The constrained problem can then be addressed via bisection over λ (see Algorithm 2 in [68]).

The expression for the worst-case distribution μ_*^θ in Theorem 2.2 reveals that Problem (2) is equivalent to a bilevel optimization problem:

$$\begin{aligned} \min_{\theta} \quad F(\theta) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{z\sim\mu_*^{\theta,i}}[f_\theta(z)], && \text{(Upper Level)} \\ \mu_*^{\theta,i} &= \arg\min_{\mu\in\mathcal{P}(\mathbb{R}^d)} \mathcal{D}_{\mathrm{KL}}\Big(\mu \Big\| u_{\theta,i}(\cdot)\Big), \forall i \in [n], \theta, && \text{(Lower Level)} \end{aligned} \tag{5}$$

where the upper-level problem minimizes the worst-case risk, and for fixed θ, each lower-level problem involves approximating the target distribution with density $u_{\theta,i}(\cdot)$. The lower-level problem is not a standard finite-dimensional optimization, but rather a functional problem over a space of probability distributions. Consequently, standard bilevel optimization algorithms for finite-dimensional problems are not directly applicable. To address this, we develop algorithms that leverage the structure of the lower-level problems and combine with sampling techniques inspired by Langevin dynamics. We provide theoretical guarantees for these methods in the following sections.

3 Naïve Double-Loop Iterative Sampling Algorithm

A natural approach for solving the bilevel problem in (5) is stochastic gradient descent (SGD). The hypergradient of the objective in (5) is given by

$$\nabla F(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{z\sim\mu_*^{\theta,i}}[\nabla_\theta f_\theta(z)]. \tag{6}$$

A direct Monte Carlo estimator for $\nabla F(\theta)$ requires i.i.d. samples $z_i \sim \mu_*^{\theta,i}$ for $i \in [n]$. However, exact sampling from $\mu_*^{\theta,i}$ is intractable. The primary challenge is that $\mu_*^{\theta,i}$ is a complicated distribution with density given in (3), and we cannot sample from it exactly, preventing an unbiased gradient estimator. We must therefore rely on biased estimators derived from approximate sampling. Algorithm 1 presents a naïve sampling method by iteratively sampling a point whose distribution is sufficiently close to $\mu_*^{\theta,i}$ and next updating θ using stochastic gradient descent.

Algorithm 1 Naïve Iterative Sampling Algorithm

```
Require: Stepsize parameter $\eta$, initial guess $\theta_0$
1: for $k = 0, 1, 2, \dots, T_{\text{out}} - 1$ do
2:     Sample $i_k$ randomly from $\{1, \dots, n\}$;
3:     Draw a sample $z \sim \widetilde{\mu}$ such that $\mathcal{W}_2(\widetilde{\mu}, \mu_*^{\theta_k, i_k}) \le \delta$;
4:     Update $\theta_{k+1} = \theta_k - \eta\nabla_\theta f_{\theta_k}(z)$.
5: end for
   Output $\widehat{\theta}$ uniformly selected from $\{\theta_1, \dots, \theta_{T_{\text{out}}}\}$.
```

The key step in Algorithm 1 is to sample from $\mu_*^{\theta,i}$, whose density function is given in (3). There are many approaches to finish this task, such as Langevin dynamics and its variants [12, 13, 24, 26], particle-based methods [16, 43, 50], flow-based methods [21, 76], etc. In this work, we provide theoretical guarantees for Algorithm 1 when the sampling in Step 3 is performed using Langevin dynamics, due to its efficiency and well-understood convergence properties.

For fixed $\theta, i \in [n]$, by definition of KL-divergence, the distribution $\mu_*^{\theta,i}$ minimizes the following free energy functional:

$$\mu_*^{\theta,i} = \arg\min_{\mu\in\mathcal{P}} \left\{ \mathbb{E}_\mu \left[\frac{-f_\theta(z)}{\lambda} + \frac{1}{2}\|x^{(i)} - z\|_2^2 \right] - \epsilon\mathcal{H}(\mu) \right\}.$$

A standard method to sample from such a distribution is Langevin dynamics. The corresponding gradient flow is described by the Stochastic Differential Equation (SDE):

$$\mathrm{d}Z_t = -\left(\frac{-\nabla_z f_\theta(Z_t)}{\lambda} + (Z_t - x^{(i)})\right)\mathrm{d}t + \sqrt{2\epsilon}\,\mathrm{d}\mathbf{W}_t,$$

where $\{\mathbf{W}_t\}_t$ is the standard Brownian motion. Algorithm 2 presents the discrete-time implementation of this Langevin dynamics. With a sufficiently small step size and a sufficient number of iterations, the output provides an approximate sample from $\mu_*^{\theta,i}$.

Algorithm 2 Langevin Stochastic Descent

```
Require: Stepsize parameter $\tau$, initial distribution $\mu_0$ that is easy to sample.
Initialize $z_0 \sim \mu_0$
for $t = 0, 1, 2, \ldots, T-1$ do
    Sample $\zeta_t \sim \mathcal{N}(0, \mathbf{I}_d)$
    Update $z_{t+1} \leftarrow z_t - \tau\left(\frac{-\nabla_z f_\theta(z_t)}{\lambda} + (z_t - x^{(i)})\right) + \sqrt{2\tau\epsilon}\zeta_t$
end for
Output $z_T$
```

Remark 3.1 (Comparison with Wasserstein DRO)**.** If we do not add entropy regularization (e.g., set $\epsilon = 0$), Algorithm 2 is the noiseless gradient descent method for unconstrained optimization

$$\max_z \left\{ f_\theta(z) - \frac{\lambda}{2}\|z - x^{(i)}\|_2^2 \right\},$$

and Algorithm 1 reduces to the stochastic gradient descent (SGD) for optimizing the dual objective of the soft-constrained Wasserstein DRO [58]:

$$\min_\theta \left\{ \frac{1}{n}\sum_{i=1}^{n}\left[\max_z\ f_\theta(z) - \frac{\lambda}{2}\|z - x^{(i)}\|_2^2\right] \right\}. \tag{7}$$

This perspective reveals a key computational advantage of Sinkhorn DRO over its unregularized counterpart. The convergence of SGD for the Wasserstein DRO problem (7) typically requires the inner maximization to be strongly concave, which in turn requires a sufficiently large Lagrangian parameter λ. In contrast, as we will show, the entropy regularization ($\epsilon > 0$) in our setting ensures the sampling algorithm converges without requiring strong concavity, leading to convergence guarantees under milder assumptions.

3.1 Convergence Analysis of Algorithm 2

To study the convergence of Algorithm 2, we impose the following two assumptions.

Assumption 3.2. (I) For any (θ, z), the loss function $f_\theta(z)$ is continuously differentiable in z and satisfies for any z, z' that $\|\nabla_z f_\theta(z) - \nabla_z f_\theta(z')\|_2 \le L_{f,2}\|z - z'\|_2$.

(II) For any $\theta, i \in [n]$, the target distribution $\mu_*^{\theta,i}$ satisfies the log-Sobolev inequality (LSI) with constant $\alpha > 0$.

Assumption 3.2(I) is common in the convergence guarantees in optimization. Assumption 3.2(II) is crucial for establishing the global convergence of Algorithm 2, and it holds for a broad class of loss functions. One sufficient condition of Assumption 3.2(II) is the log-concavity of $\mu_*^{\theta,i}$, i.e., λ is sufficiently large such that $z \mapsto f_\theta(z) - \frac{\lambda}{2}\|z - x^{(i)}\|_2^2$ is strongly concave. In the following proposition, we provide some milder sufficient conditions of Assumption 3.2(II). The proof of Proposition 3.3 is provided in Appendix B.

Proposition 3.3 (Sufficient Conditions of LSI)**.** *For fixed θ, the following two conditions ensure $\mu_*^{\theta,i}$ satisfies LSI:*

- *Suppose $f_\theta(z)$ is bounded such that* $\sup f_\theta(z) - \inf f_\theta(z) < B$ *for some constant $B > 0$, then $\mu_*^{\theta,i}$ satisfies LSI with constant* $\alpha = \frac{1}{\epsilon}\exp(-\frac{4B}{\lambda\epsilon})$.

- *Suppose* $\|\nabla_z f_\theta(z)\|_2 \leq M$ *for any* θ *and* z *with some constant* $M > 0$*, then* $\mu_*^{\theta,i}$ *satisfies LSI with constant*

$$\alpha = \frac{1}{2\epsilon} \max\left\{ e^{-4M^2/\lambda^2 \sqrt{2d/\pi}}, \left(4 + (M/\lambda + \sqrt{2})^2 (2 + d + 4M^2/\lambda^2) e^{M^2/(2\lambda^2)}\right)^{-1} \right\}.$$

Next, we present the complexity analysis of Algorithm 2 in the theorem below, following the similar analysis of [65]. Its proof is provided in Appendix B.

Theorem 3.4 (Complexity of Algorithm 2)**.** *Assume Assumption 3.2 holds. Specify the parameters in Algorithm 2 as*

$$\tau = \frac{\alpha\epsilon}{4(1 + L_{f,2}/\lambda)^2} \cdot \min\{1, \delta^2\alpha/(8d)\}, \quad T = \left\lceil \frac{1}{\alpha\tau\epsilon} \log \frac{4D_{\mathrm{KL}}(\mu_0 \| \mu_*^{\theta,i})}{\delta^2\alpha} \right\rceil = \widetilde{\mathcal{O}}(\delta^{-2}).$$

Then, the law of the output of Algorithm 2, denoted as μ_T*, satisfies that* $\mathcal{W}_2(\mu_T, \mu_*^{\theta,i}) \leq \delta$.

Remark 3.5 (Choice of Initial Distribution)**.** We recommend taking the initial distribution μ_0 in Algorithm 2 as $\mathcal{N}(x^{(i)}, \epsilon \mathbf{I}_d)$. In this case, the KL-divergence $D_{\mathrm{KL}}(\mu_0 \| \mu_*^{\theta,i}) = \frac{1}{\lambda\epsilon}\mathbb{E}_{\mu_0}[f_\theta(z)]$. Assume $\|\nabla_z f_\theta(\tilde{z})\|_2 \leq M$ for some $\tilde{z}$, then it can be shown that $D_{\mathrm{KL}}(\mu_0 \| \mu_*^{\theta,i}) = \mathcal{O}(d)$, where $\mathcal{O}(\cdot)$ hides constant related to ϵ, M, and $\|x^{(i)} - \tilde{z}\|_2$.

3.2 Convergence Analysis of Algorithm 1

In this subsection, we provide the convergence analysis of Algorithm 1. In our analysis, we measure computational complexity as the total number of gradient evaluations of $f_\theta(z)$ (with respect to either θ or z). This metric dominates the total computational time from Algorithms 1 and 2. We further impose the following assumptions.

Assumption 3.6. (I) For any θ, the distribution $\mu_*^{\theta,i}, i \in [n]$ satisfies that

$$\mathbb{V}\mathrm{ar}_{(i,z_i)\sim \mathrm{Uniform}([n])\otimes \mu_*^{i,\theta}}\big(\nabla f_\theta(z_i)\big) \leq \sigma^2 \tag{8}$$

for some constant $\sigma^2 > 0$.

(II) For any (θ, z), the loss function $f_\theta(z)$ is continuously differentiable in θ and satisfies for any z, z' that

$$\begin{aligned} \|\nabla f_\theta(z)\|_2 &\leq L_{f,1}, \\ \|\nabla_\theta f_\theta(z) - \nabla_\theta f_\theta(z')\|_2 &\leq L_{f,2}\|z - z'\|_2, \\ \|\nabla_\theta f_\theta(z) - \nabla_{\theta'} f_\theta(z)\|_2 &\leq L_{f,2}\|\theta - \theta'\|_2. \end{aligned}$$

(III) For any (θ, z), the loss function $f_\theta(z)$ is continuously differentiable in z and satisfies for any θ, θ' that $\|\nabla_z f_\theta(z) - \nabla_z f_{\theta'}(z)\|_2 \leq L_{f,2}\|\theta - \theta'\|_2$.

Recall that in Algorithm 1, we constructed the hypergradient estimator of (6) as

$$\widehat{\nabla} F(\theta; z) = \nabla_\theta f_\theta(z), \tag{9}$$

where the random vector z follows the distribution $\widetilde{\mu}$ with $\mathcal{W}_2(\widetilde{\mu}, \mu_*^{\theta,i}) \leq \delta$ and i is a random sample from $[n]$. The following lemma provides bias, variance, and computational complexity analysis regarding our estimator. Its proof is provided in Appendix B.

Lemma 3.7 (Bias, variance, and complexity of hypergradient estimator)**.** *Assume Assumptions 3.2 and 3.6 hold, then the estimator in* (9) *satisfies that*

(I) *(Bias)* $\left\|\mathbb{E}\big[\widehat{\nabla} F(\theta; z)\big] - \nabla F(\theta)\right\|_2 \leq L_{f,2}\delta$.

(II) *(Variance)* $\mathbb{V}\mathrm{ar}(\widehat{\nabla} F(\theta; z)) \leq \mathbf{V} := 2\sigma^2 + 2L_{f,1}^2\sqrt{\alpha}\delta + 2L_{f,2}^2\delta^2$.

(III) *(Complexity) The computational complexity of constructing* (9) *is* $\widetilde{\mathcal{O}}(\delta^{-2})$.

A random vector θ is said to be a ϱ-stationary point if $\mathbb{E}\|\nabla F(\theta)\|_2^2 \leq \varrho^2$. As the bilevel program (5) is nonconvex in general, we focus on finding its stationary point using Algorithm 1. We provide its computational complexity in the theorem below. Its proof is provided in Appendix B.

Theorem 3.8 (Complexity Bounds). *Assume Assumptions 3.2 and 3.6 hold. In order to use Algorithm 1 to obtain a ϱ-stationary point, it suffices to specify parameters in Algorithm 1 as*

$$T_{out} = \mathcal{O}(\varrho^{-4}), \quad \eta = \frac{1}{\sqrt{T_{out}\mathbf{V}}}, \quad \delta = \frac{\varrho}{2L_{f,2}},$$

where $\mathcal{O}(\cdot)$ hides constant depending only on the initial guess, $\mathbf{V}$, and $L_{f,2}$. Consequently, the total computational complexity of Algorithm 1 to find a ϱ-stationary point is $\widetilde{\mathcal{O}}(\varrho^{-6})$.

Remark 3.9 (Comparison with [79]). In this section, we present a double-loop algorithm similar to [79]. We remark that there are some differences in the convergence analysis part. First, the complexity result in [79, Theorem 2] uses a constant stepsize, which leads to a non-vanishing optimization error (bias) in the gradient estimator (see the last component of the equation above [79, Appendix C4]). In contrast, our decaying stepsize schedule ensures this error vanishes asymptotically. Second, our variance assumption (Assumption 3.6(I)) is made on the ideal gradient at the true worst-case distributions $\mu_*^{\theta,i}$. We then explicitly control how the sampling error propagates to the variance of our practical estimator (Lemma 3.7(II)). Their analysis assumes a uniform bound on the variance of the approximate stochastic gradients, which is a stronger condition.

4 Mean-Field Single-Loop Sampling Algorithm

The double-loop algorithm (Algorithm 1) is conceptually simple but can be computationally expensive, as each outer update requires running an inner loop of Langevin dynamics to high accuracy. To improve efficiency, we now propose a single-loop algorithm that interleaves the updates of the upper-level parameter and the lower-level distribution estimators.

For $i \in [n]$, our algorithm initializes the worst-case distribution estimators $\mu_0^{(i)} = \mathcal{N}(x^{(i)}, \epsilon \mathbf{I}_d)$ for the i-th lower-level subproblem. At the beginning of each iteration k, we sample a set of lower-level subproblems with index $i \in I_k$. Next, we update the distribution estimator $\mu_{k+1}^{(i)}$ such that each particle is updated using one-step of Langevin dynamics:

$$z_{k+1}^{(i)} = \begin{cases} z_k^{(i)} - \tau\left(-\dfrac{\nabla_z f_{\theta_k}(z_k^{(i)})}{\lambda} + (z_k^{(i)} - x^{(i)})\right) + \sqrt{2\tau\epsilon}\zeta_k^{(i)}, & \text{if } i \in I_k \\ z_k^{(i)}, & \text{otherwise.} \end{cases} \tag{10}$$
$$\mu_{k+1}^{(i)} = \mathrm{Law}\big(z_{k+1}^{(i)}\big).$$

We then update the gradient estimator of the upper-level objective as

$$v_{k+1} = \frac{1}{|I_k|}\sum_{i\in I_k} \mathbb{E}_{z_k^{(i)}\sim\mu_k^{(i)}}\big[\nabla_\theta f_{\theta_k}(z_k^{(i)})\big]. \tag{11}$$

Compared with the true gradient $\nabla F(\theta_k)$ defined in (6), we replace the average over all indices $i \in [n]$ with the average over sampled indices $i \in [I_k]$, and replace the true worst-case distribution $\mu_*^{\theta_k,i}$ with its estimator $\mu_k^{(i)}$ obtained from the last iteration. Finally, we maintain the moving average estimator r_{k+1} for the gradient estimator v_{k+1} and update the upper-level decision θ_{k+1} using r_{k+1}. The detailed procedure is provided in Algorithm 3.

Remark 4.1 (Mean-Field Update). We call Algorithm 3 a mean-field algorithm because the gradient estimator v_{k+1} in (11) uses the population expectation $\mathbb{E}_{z_k^{(i)}\sim\mu_k^{(i)}}[\cdot]$, corresponding to the limit of an infinite number of particles. In practice, this expectation is approximated by a sample average over a finite set of particles. The technical difficulty of the analysis for the finite-particle case is that this will make the estimator v_{k+1} stochastic provided that I_k is fixed. A possible solution is the propagation of chaos [8, 48, 49, 62, 84], a theory that quantifies the difference between the dynamics of a system of finitely many particles and its limiting behavior described by their mean-field densities. In this work, our convergence analysis (Theorem 4.5) focuses on the idealized mean-field dynamics (population expectations). Analyzing the finite-particle case via propagation of chaos is an important direction for future work.

Algorithm 3 Mean-Field Single-Loop Iterative Sampling Algorithm

```
Require: Stepsize parameters $\eta, \tau$, initial guess $\mu_0^{(i)}, i \in [n]$, moving average estimator $r_0$, moving
    average parameter $\beta_0$, number of iterations $T$, mini-batch size $|I_k|$.
1: for $k = 0, 1, 2, \dots, T-1$ do
2:     Randomly sample indices $I_k \subseteq [n]$
3:     for $i \in [n]$ do
4:         Update $\mu_{k+1}^{(i)}$ according to (10)
5:     end for
6:     Update gradient estimator $v_{k+1}$ according to (11)
7:     Update $r_{k+1} = (1 - \beta_0) r_k + \beta_0 v_{k+1}$
8:     Update $\theta_{k+1} = \theta_k - \tau\eta r_{k+1}$.
9: end for
   Output $\widehat{\theta}$ uniformly selected from $\{\theta_1, \dots, \theta_T\}$.
```

4.1 Convergence Analysis of Algorithm 3

In this subsection, We outline the key steps of the convergence analysis; detailed proofs are deferred to Appendix C. We first build the error bound for the objective at k-th and $(k+1)$-th iterations in the lemma below.

Lemma 4.2 (Descent Lemma). *Assume Assumption 3.6(II) holds and the stepsize parameters satisfy $\eta\tau \le \frac{1}{2L_{f,2}}$. Consider the update in Step 8 of Algorithm 3, it holds that*

$$F(\theta_{k+1}) \le F(\theta_k) + \frac{\tau\eta}{2}\|\nabla F(\theta_k) - r_{k+1}\|_2^2 - \frac{\tau\eta}{2}\|\nabla F(\theta_k)\|_2^2 - \frac{\tau\eta}{4}\|r_{k+1}\|_2^2.$$

The critical step for our analysis is to bound the difference between the true hypergradient $\nabla F(\theta_k)$ and its estimator r_{k+1}. To this end, let us define the auxiliary gradient

$$\nabla F(\theta_k; \mu_k^{(1:n)}) := \frac{1}{n}\sum_{i\in[n]} \mathbb{E}_{z_k^{(i)} \sim \mu_k^{(i)}}[\nabla f_\theta(z_k^{(i)})]. \tag{12}$$

Denote by $\mathcal{G}_k$ the σ-algebra generated by all the randomness up to iteration k. Conditioned on $\mathcal{G}_{k-1}$, it can be shown that v_{k+1} is the unbiased estimator of $\nabla F(\theta_k; \tilde{\mu}_k^{(1:n)})$, as the former uses mini-batch random samples $I_k \subseteq [n]$. In the following, we provide the upper bound of the difference between $\nabla F(\theta_k)$ and r_{k+1} using $\nabla F(\theta_k; \mu_k^{(1:n)})$ and v_{k+1}.

Lemma 4.3 (Gradient Difference Lemma). *Assume Assumption 3.6(II) holds, and the parameter $\beta_0 \in (0, 1]$, then it holds that*

$$\begin{aligned}\mathbb{E}\big[\|\nabla F(\theta_k) - r_{k+1}\|_2^2 \big| \mathcal{G}_{k-1}\big] &\le (1-\beta_0)\|\nabla F(\theta_{k-1}) - r_k\|_2^2 + \frac{4L_{f,2}^2\tau^2\eta^2}{\beta_0}\|r_k\|_2^2 \\ &+ 4\beta_0 \left\|\nabla F(\theta_k) - \nabla F(\theta_k; \mu_k^{(1:n)})\right\|_2^2 + \beta_0^2 \mathbb{E}\left[\left\|\nabla F(\theta_k; \mu_k^{(1:n)}) - v_{k+1}\right\|_2^2 \middle| \mathcal{G}_{k-1}\right].\end{aligned} \tag{13}$$

A key technical challenge arises in bounding the third term of (13). By Pinsker's Inequality [14],

$$\begin{aligned}&\left\|\nabla F(\theta_k) - \nabla F(\theta_k; \mu_k^{(1:n)})\right\|_2^2 \\ =&\left\|\frac{1}{n}\sum_{i\in[n]}\left[\mathbb{E}_{z\sim\mu_k^{(i)}}[\nabla f_{\theta_k}(z)] - \mathbb{E}_{z\sim\mu_*^{\theta_k,i}}[\nabla f_{\theta_k}(z)]\right]\right\|_2^2 \\ \le&\frac{L_{f,1}^2}{n}\sum_{i\in[n]} \mathrm{TV}(\mu_k^{(i)}, \mu_*^{\theta_k,i})^2 \le \frac{L_{f,1}^2}{2n}\sum_{i\in[n]} \mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}, \mu_*^{i,\theta_k}).\end{aligned}$$

This error depends on the KL-divergence between our running estimators $\mu_k^{(i)}$ and the true optimal distributions $\mu*^{i,\theta_k}$. Unlike standard bilevel optimization where lower-level solutions lie in a Euclidean space, here they are probability distributions. Therefore, standard techniques (such as [29, 52]) for Euclidean norm bounds, especially the triangular inequality, do not apply. To overcome this, we adapt and extend the SDE-based techniques from sampling literature [47, 65] to our setup, which contains multiple target distributions, and each distribution is time-varying. Lemma 4.4 provides the desired error bound, relating the cumulative KL error to algorithm parameters and the momentum gradient norm $\|r_{k+1}\|_2^2$.

Lemma 4.4 (KL-Divergence Bound). *Assume Assumptions 3.2 and 3.6 hold, then it holds that*

$$\begin{aligned}&\sum_{k=0}^{T}\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}\|\mu_*^{i,\theta_k})\big]\\ \leq&\frac{4n}{\alpha\tau|I_k|}\mathcal{D}_{\mathrm{KL}}(\mu_0^{(i)}\|\mu_*^{i,\theta_0})+\frac{12\eta L_{f,1}^2Tn}{\lambda\epsilon\alpha|I_k|}+\frac{32\tau\epsilon dL_{G,2}^2T}{\alpha}+\frac{20\eta n}{\lambda\epsilon\alpha|I_k|}\sum_{k=0}^{T-1}\mathbb{E}\big[\|r_{k+1}\|_2^2\big].\end{aligned}$$

Combining the lemmas above, we derive the convergence theorem for Algorithm 3 below.

Theorem 4.5 (Convergence Analysis of Algorithm 3). *Assume Assumptions 3.2 and 3.6 hold, and with the following choices of parameters in Algorithm 3:*

$$\begin{aligned}&\beta_0\leq\frac{\varrho^2|I_k|}{6L_{f,2}^2},\quad\tau\leq\frac{\varrho^2\alpha}{384\epsilon dL_{G,2}^2L_{f,1}^2},\\&\eta\leq\min\left(\frac{\varrho^2\lambda\epsilon\alpha|I_k|}{144L_{f,1}^2L_{f,2}^2n},\frac{\lambda\epsilon\alpha|I_k|}{160L_{f,2}^2n}\right)\\&T\geq\max\left(\frac{12\mathbb{E}[F(\theta_0)-F(\theta_*)]}{\eta\tau\varrho^2},\frac{6\mathbb{E}\|\nabla F(\theta_0)-z_1\|_2^2}{\beta_0\varrho^2},\frac{48L_{f,2}^2}{\alpha\tau|I_k|\varrho^2}\sum_{i=1}^{n}\mathcal{D}_{\mathrm{KL}}(\mu_0^{(i)}\|\mu_*^{i,\theta_0})\right),\end{aligned}$$

Algorithm 3 finds a ϱ-stationary solution of Problem (5). The total computational complexity of Algorithm 1 to obtain a ϱ-stationary point is $\mathcal{O}(\varrho^{-6}\cdot\frac{n}{|I_k|})$, where $\mathcal{O}(\cdot)$ hides constants depending only on $\lambda,\epsilon,\alpha,L_{f,1},L_{f,2}$, and the initial guess.

In Theorem 4.5, the obtained ϱ^{-6} complexity is likely suboptimal compared to the ϱ^{-4} lower bound for general nonconvex stochastic optimization. We hypothesize this gap stems from two sources: (i) the discretization error of the Langevin step, which requires careful control of the KL divergence between iterated distributions, and (ii) the conservative nature of our KL-divergence bound (Lemma 4.4) compared to error bounds typically available for finite-dimensional strongly convex subproblems. Also, in our complexity analysis, we do not specify the mini-batch size $|I_k|$ in each iteration. This offers a trade-off: using $|I_k|=1$ minimizes per-iteration cost, while larger $|I_k|$ reduces the number of iterations T needed, which can be exploited in parallel computing settings

5 Applications

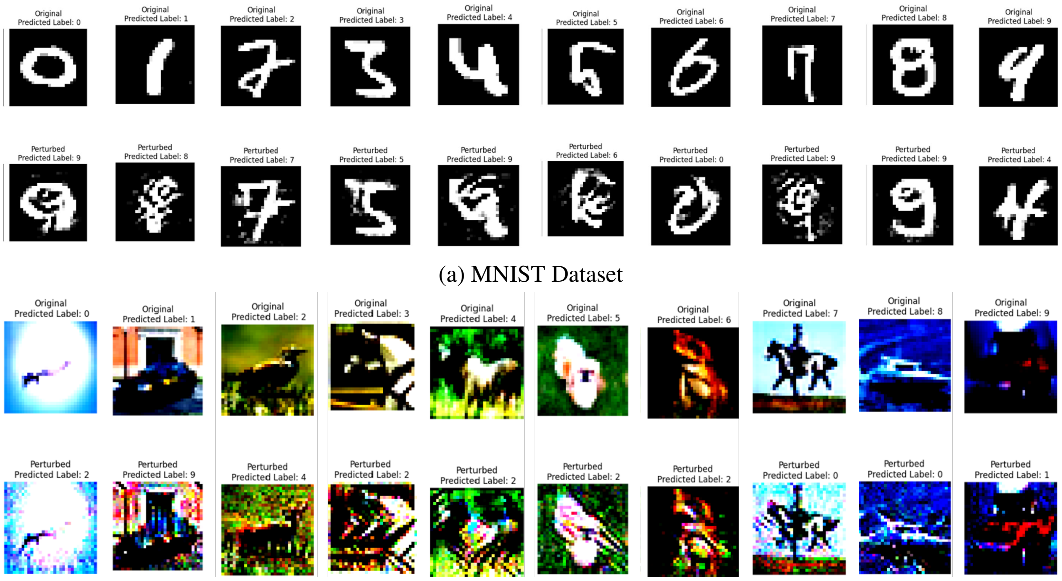
In this section, we validate the performance of our proposed algorithms for the adversarial classification task and follow the similar experiment setup as in [58]. It aims to solve the distributionally robust optimization problem

$$\min_{\theta}\left\{\max_{\mu}\ \mathbb{E}_{(x,y)\sim\mu}[\ell(f_\theta(x),y)]-\lambda\mathcal{S}_\epsilon(\widehat{\mu},\mu)\right\},$$

where $f_\theta(x)$ denotes a neural network classifier consisting of $8*8, 6*6$, and $5*5$ convolutional filer layers with ELU activations followed by a fully connected layer and softmax output, $\ell(\hat{y},y)$ denotes the cross-entropy classification loss, and the Sinkhorn discrepancy peanlty only considers the perturbation for the data feature part instead of the data label part. The reference distribution $\widehat{\mu}$ is constructed using the empirical samples from the MNIST [41] or CIFAR-10 [39] training dataset.

5.1 Visualization of Worst-Case Distributions

In this subsection, we visualize the samples from worst-case distribution by solving the bilevel program (5) using Algorithm 3. We specify the hyper-parameters $\epsilon=0.1$ and $\lambda=20$. The results


Original Predicted Label: 0
Original Predicted Label: 1
Original Predicted Label: 2
Original Predicted Label: 3
Original Predicted Label: 4
Original Predicted Label: 5
Original Predicted Label: 6
Original Predicted Label: 7
Original Predicted Label: 8
Original Predicted Label: 9
Perturbed Predicted Label: 9
Perturbed Predicted Label: 8
Perturbed Predicted Label: 7
Perturbed Predicted Label: 5
Perturbed Predicted Label: 9
Perturbed Predicted Label: 6
Perturbed Predicted Label: 0
Perturbed Predicted Label: 9
Perturbed Predicted Label: 9
Perturbed Predicted Label: 4

(a) MNIST Dataset

Original Predicted Label: 0
Original Predicted Label: 1
Original Predicted Label: 2
Original Predicted Label: 3
Original Predicted Label: 4
Original Predicted Label: 5
Original Predicted Label: 6
Original Predicted Label: 7
Original Predicted Label: 8
Original Predicted Label: 9
Perturbed Predicted Label: 2
Perturbed Predicted Label: 9
Perturbed Predicted Label: 4
Perturbed Predicted Label: 2
Perturbed Predicted Label: 2
Perturbed Predicted Label: 2
Perturbed Predicted Label: 2
Perturbed Predicted Label: 0
Perturbed Predicted Label: 0
Perturbed Predicted Label: 1

(b) CIFAR-10 Dataset

Figure 1: Raw and adversarial samples found by the Sinkhorn DRO. Two subfigures represent numerical experiments for two datasets (MNIST and CIFAR-10). For each subfigure, plots on the top represent the raw samples, and plots on the bottom represent the perturbed samples using Algorithm 3.

are provided in Figure 1, which shows the qualitative changes from the original images to perturbed images for the MNIST or CIFAR-10 datasets. The figures demonstrate that Sinkhorn DRO induces more meaningful contextual changes to the original images to confuse the classifier, which is more aligned with our intuition for the adversarial classification task.

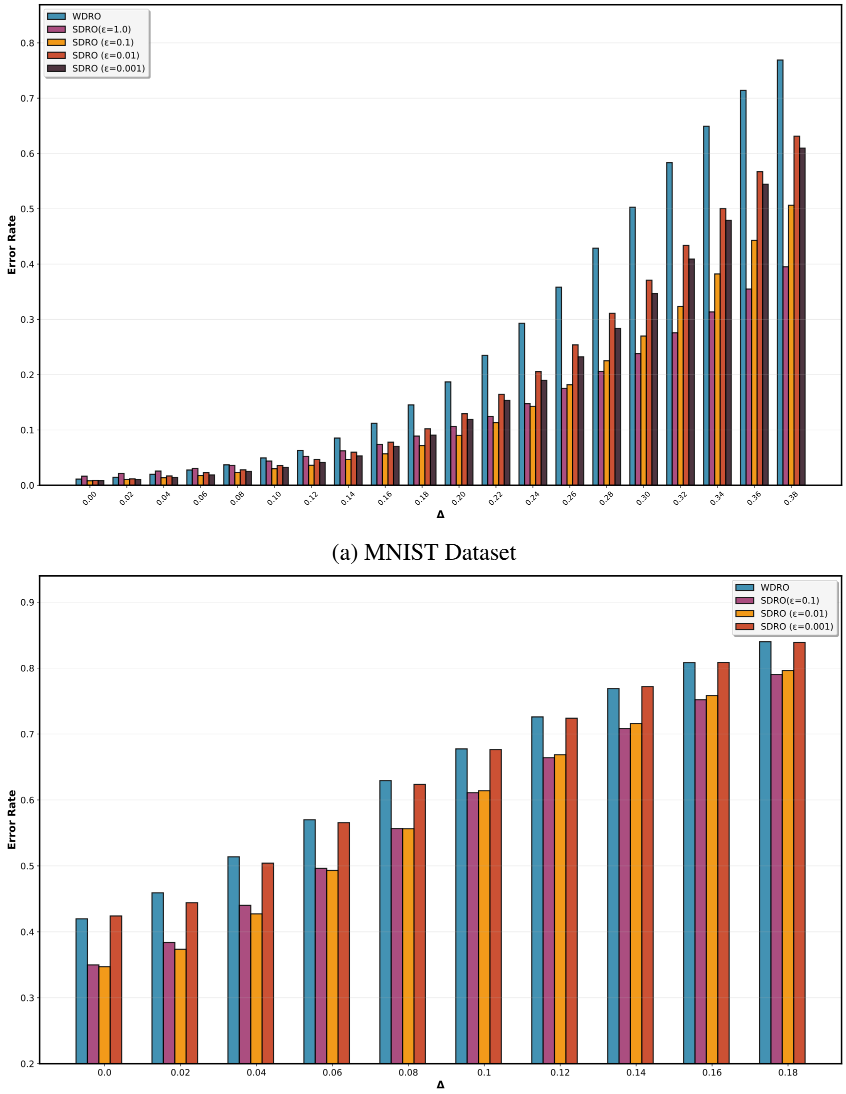
5.2 Ablation Study

Next, we quantitatively validate the effectiveness of Algorithm 3. We fine-tune the regularization for Sinkhorn DRO or Wasserstein DRO such that the ℓ_2-norm between the original image X and the perturbed one is controlled within $C_2 = 0.04 * \mathbb{E}_{\widehat{\mu}}[\|X\|_2]$. To assess the robustness of the proposed models, we apply a Projected Gradient Method (PGM) attack with ℓ_2-norm constraints to the test datasets. The perturbation magnitude Δ is normalized by the average ℓ_2-norm of the test features. We examine the performance using the misclassification rate on perturbed datasets.

Figure 2 reports the classification accuracy of the proposed method and the baseline under various adversarial perturbations. On MNIST, our approach consistently outperforms the baseline across all noise levels. When the perturbation budget is small, $\varepsilon = 0.1$ yields the highest accuracy; as the budget increases, $\varepsilon = 1$ becomes superior. This observation indicates that a larger ε strengthens robustness against stronger attacks. On CIFAR-10, the same trend holds, $\varepsilon = 0.01$ already improves upon the baseline. In our experiments the proposed defense is highly sensitive to the choice of ε. Setting $\varepsilon = 1$ prevents the model from completing the classification task: during training, clean accuracy collapses to 10%, indicating that the training set itself is being perturbed to the point where the network can no longer learn any meaningful image features.

6 Conclusion

In this work, we developed sampling-based algorithms for solving Sinkhorn DRO, by reformulating it as a bilevel program with multiple infinite-dimensional lower-level subproblems over probability spaces. We proposed a double-loop algorithm (Algorithm 1) with a $\widetilde{\mathcal{O}}(\varrho^{-6})$ complexity guarantee and a more efficient single-loop, mean-field algorithm (Algorithm 3) that achieves comparable complexity within an interleaved update framework. Our analysis extends conventional bilevel optimization techniques to handle infinite-dimensional lower-level subproblems by leveraging tools from sampling



(a) MNIST Dataset

(b) CIFAR-10 Dataset

Figure 2: Results of adversarial training using MNIST or CIFAR-10 datasets for Wasserstein DRO or Sinkhorn DRO models. For each plot, the x-axis refers to the level of adversarial perturbation of PGD attack, and the y-axis refers to the misclassification rate on the perturbed testing dataset.

theory (such as properties of LSI and SDE-based analysis) to control the error from approximate sampling.

There are several promising directions to explore. First, our framework could be extended to more general bilevel optimization with infinite-dimensional lower-level variables, such as those arising in meta-learning for generative models (e.g., diffusion or flow-based models). Second, it is an open question to use iterative sampling-based algorithms to achieve the optimal complexity rate. Integrating accelerated sampling techniques and developing sharper theoretical analysis could lead to improved guarantees. Finally, it is interesting to study the Sinkhorn DRO with more general transportation costs and broader classes of loss functions

Acknowledgement

The author would like to thank Jindong Jiang for his help in the numerical study to generate Figure 2. The author is partially sponsored by the Presidential Young Fellowship at The Chinese University of Hong Kong, Shenzhen.

References

[1] Ackley DH, Hinton GE, Sejnowski TJ (1985) A learning algorithm for boltzmann machines. *Cognitive science* 9(1):147–169.

[2] Azizian W, Iutzeler F, Malick J (2023) Regularization for wasserstein distributionally robust optimization. *ESAIM: Control, Optimisation and Calculus of Variations* 29:33.

[3] Blanchet J, Murthy K (2019) Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research* 44(2):565–600.

[4] Blanchet J, Murthy K, Zhang F (2022) Optimal transport-based distributionally robust optimization: Structural properties and iterative schemes. *Mathematics of Operations Research* 47(2):1500–1529.

[5] Bottou L, Curtis FE, Nocedal J (2018) Optimization methods for large-scale machine learning. *SIAM review* 60(2):223–311.

[6] Canuto C, Urban K (2005) Adaptive optimization of convex functionals in banach spaces. *SIAM journal on numerical analysis* 42(5):2043–2075.

[7] Cescon R, Martin A, Ferrari-Trecate G (2025) Data-driven distributionally robust control based on sinkhorn ambiguity sets. *arXiv preprint arXiv:2503.20703* .

[8] Chaintron LP, Diez A (2022) Propagation of chaos: a review of models, methods and applications. i. models and methods. *arXiv preprint arXiv:2203.00446* .

[9] Chen X, Sun H, Xu H (2019) Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems. *Mathematical Programming* 177(1):255–289.

[10] Cheng X, Bartlett P (2018) Convergence of langevin mcmc in kl-divergence. *Algorithmic learning theory*, 186–211 (PMLR).

[11] Cheng X, Xie Y, Zhu L, Zhu Y (2025) Worst-case generation via minimax optimization in wasserstein space. *arXiv preprint arXiv:2512.08176* .

[12] Choi MC (2020) Metropolis–hastings reversiblizations of non-reversible markov chains. *Stochastic Processes and their Applications* 130(2):1041–1073.

[13] Cornish R, Vanetti P, Bouchard-Côté A, Deligiannidis G, Doucet A (2019) Scalable metropolis-hastings for exact bayesian inference with large datasets. *International Conference on Machine Learning*, 1351–1360 (PMLR).

[14] Cover TM (1999) *Elements of information theory* (John Wiley & Sons).

[15] Cuturi M (2013) Sinkhorn distances: Lightspeed computation of optimal transportation distances. *Advances in neural information processing systems*.

[16] Dai B, He N, Dai H, Song L (2016) Provable bayesian inference via particle mirror descent. *Artificial Intelligence and Statistics*, 985–994 (PMLR).

[17] Dalalyan AS (2017) Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 79(3):651–676.

[18] Dapogny C, Iutzeler F, Meda A, Thibert B (2023) Entropy-regularized wasserstein distributionally robust shape and topology optimization. *Structural and Multidisciplinary Optimization* 66(3):42.

[19] Durmus A, Moulines E (2017) Nonasymptotic convergence analysis for the unadjusted langevin algorithm. *The Annals of Applied Probability* .

[20] Esfahani PM, Kuhn D (2018) Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 171(1):115–166.

[21] Fan M, Zhou R, Tian C, Qian X (2024) Path-guided particle-based sampling. *arXiv preprint arXiv:2412.03312* .

[22] Gao R, Kleywegt A (2023) Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research* 48(2):603–655.

[23] Garreis S, Ulbrich M (2019) An inexact trust-region algorithm for constrained problems in hilbert space and its application to the adaptive solution of optimal control problems with pdes. *Preprint, submitted, Technical University of Munich* .

[24] Griffin JE, Walker SG (2013) On adaptive metropolis–hastings methods. *Statistics and Computing* 23(1):123–134.

[25] Guo Z, Hu Q, Zhang L, Yang T (2021) Randomized stochastic variance-reduced methods for multi-task stochastic bilevel optimization. *arXiv preprint arXiv:2105.02266* .

[26] Haario H, Saksman E, Tamminen J (2001) An adaptive metropolis algorithm. *Bernoulli* .

[27] Holley R, Stroock D (1987) Logarithmic Sobolev inequalities and stochastic Ising models. *Journal of Statistical Physics* 46(5):1159–1194, ISSN 1572-9613.

[28] Hu Q, Qiu ZH, Guo Z, Zhang L, Yang T (2023) Blockwise stochastic variance-reduced methods with parallel speedup for multi-block bilevel optimization. *International Conference on Machine Learning*, 13550–13583 (PMLR).

[29] Hu Q, Zhong Y, Yang T (2022) Multi-block min-max bilevel optimization with applications in multi-task deep auc maximization. *Advances in Neural Information Processing Systems* 35:29552–29565.

[30] Hu Y, Chen X, He N (2020) Sample complexity of sample average approximation for conditional stochastic optimization. *SIAM Journal on Optimization* 30(3):2103–2133.

[31] Hu Y, Chen X, He N (2021) On the bias-variance-cost tradeoff of stochastic optimization. *Advances in Neural Information Processing Systems* 34:22119–22131.

[32] Hu Y, Wang J, Chen X, He N (2024) Multi-level monte-carlo gradient methods for stochastic optimization with biased oracles. *arXiv preprint arXiv:2408.11084* .

[33] Hu Y, Wang J, Xie Y, Krause A, Kuhn D (2023) Contextual stochastic bilevel optimization. *Advances in Neural Information Processing Systems* 36.

[34] Hu Y, Zhang S, Chen X, He N (2020) Biased stochastic first-order methods for conditional stochastic optimization and applications in meta learning. *Advances in Neural Information Processing Systems* 33:2759–2770.

[35] Jaggi M (2013) Revisiting frank-wolfe: Projection-free sparse convex optimization. *International conference on machine learning*, 427–435 (PMLR).

[36] Jiang G, Mao T (2025) Sinkhorn distributionally robust conditional quantile prediction with fixed design. *Entropy* 27(6):557.

[37] Kent C, Li J, Blanchet J, Glynn PW (2021) Modified frank wolfe in probability space. *Advances in Neural Information Processing Systems*, volume 34, 14448–14462.

[38] Kim J, Yamamoto K, Oko K, Yang Z, Suzuki T (2023) Symmetric mean-field langevin dynamics for distributional minimax problems. *arXiv preprint arXiv:2312.01127* .

[39] Krizhevsky A (2009) Learning multiple layers of features from tiny images. Technical Report TR-2009, University of Toronto, Toronto, Canada.

[40] Lai CH, Song Y, Kim D, Mitsufuji Y, Ermon S (2025) The principles of diffusion models. *arXiv preprint arXiv:2510.21890* .

[41] LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11):2278–2324.

[42] LeCun Y, Chopra S, Hadsell R, Ranzato M, Huang F, et al. (2006) A tutorial on energy-based learning. *Predicting structured data* 1(0).

[43] Liu Q (2017) Stein variational gradient descent as gradient flow. *Advances in neural information processing systems* 30.

[44] Liu Y, Pichler A, Xu H (2019) Discrete approximation and quantification in distributionally robust optimization. *Mathematics of Operations Research* 44(1):19–37.

[45] Liu Y, Yuan X, Zhang J (2021) Discrete approximation scheme in distributionally robust optimization. *Numer Math Theory Methods Appl* 14(2):285–320.

[46] Liu Z, Liu F, Gao R, Li S (2025) Convergence of mean-field langevin stochastic descent-ascent for distributional minimax optimization. *Forty-second International Conference on Machine Learning*.

[47] Marion P, Korba A, Bartlett P, Blondel M, De Bortoli V, Doucet A, Llinares-López F, Paquette C, Berthet Q (2024) Implicit diffusion: Efficient optimization through stochastic sampling. *arXiv preprint arXiv:2402.05468* .

[48] Mei S, Montanari A, Nguyen PM (2018) A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences* 115(33):E7665–E7671.

[49] Nitanda A (2024) Improved particle approximation error for mean field neural networks. *Advances in Neural Information Processing Systems* 37:113823–113845.

[50] Nitanda A, Suzuki T (2017) Stochastic particle gradient descent for infinite ensembles. *arXiv preprint arXiv:1712.05438* .

[51] Ouasfi A, Jena S, Marchand E, Boukhayma A (2025) Toward robust neural reconstruction from sparse point sets. *Proceedings of the Computer Vision and Pattern Recognition Conference*, 6552–6562.

[52] Qiu ZH, Hu Q, Zhong Y, Zhang L, Yang T (2022) Large-scale stochastic optimization of ndcg surrogates for deep learning with provable convergence. *arXiv preprint arXiv:2202.12183* .

[53] Reddi SJ, Sra S, Póczos B, Smola A (2016) Stochastic frank-wolfe methods for nonconvex optimization. *2016 54th annual Allerton conference on communication, control, and computing (Allerton)*, 1244–1251 (IEEE).

[54] Shafieezadeh Abadeh S, Mohajerin Esfahani PM, Kuhn D (2015) Distributionally robust logistic regression. *Advances in Neural Information Processing Systems*, volume 28.

[55] Shapiro A (2001) On duality theory of conic linear problems. *Nonconvex Optimization and its Applications* 57:135–155.

[56] Shapiro A, Dentcheva D, Ruszczynski A (2021) *Lectures on stochastic programming: modeling and theory* (SIAM).

[57] Shen Y, Xu P, Zavlanos MM (2023) Wasserstein distributionally robust policy evaluation and learning for contextual bandits. *arXiv preprint arXiv:2309.08748* .

[58] Sinha A, Namkoong H, Duchi J (2018) Certifiable distributional robustness with principled adversarial training. *International Conference on Learning Representations*.

[59] Song J, He N, Ding L, Zhao C (2024) Provably convergent policy optimization via metric-aware trust region methods. *Transactions on Machine Learning Research* .

[60] Song Y, Ermon S (2019) Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems* 32.

[61] Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B (2021) Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*.

[62] Suzuki T, Nitanda A, Wu D (2023) Uniform-in-time propagation of chaos for the mean-field gradient langevin dynamics. *The Eleventh International Conference on Learning Representations*.

[63] Ulbrich S, Ziems JC (2017) Adaptive multilevel trust-region methods for time-dependent pde-constrained optimization. *Portugaliae Mathematica* 74(1):37–67.

[64] van Maarschalkerwaart F, Mukherjee S, Landman MS, Brune C, Carioni M (2025) Perturbation-aware distributionally robust optimization for inverse problems. *International Conference on Scale Space and Variational Methods in Computer Vision*, 109–122 (Springer).

[65] Vempala S, Wibisono A (2019) Rapid convergence of the unadjusted langevin algorithm: Isoperimetry suffices. *Advances in neural information processing systems* 32.

[66] Vincent F, Azizian W, Iutzeler F, Malick J (2024) skwdro: a library for wasserstein distributionally robust machine learning. *arXiv preprint arXiv:2410.21231* .

[67] Wang C, Wang Y, Schapire R, et al. (2015) Functional frank-wolfe boosting for general loss functions. *arXiv preprint arXiv:1510.02558* .

[68] Wang J, Gao R, Xie Y (2021) Sinkhorn distributionally robust optimization. *arXiv preprint arXiv:2109.11926* .

[69] Wang J, Gao R, Xie Y (2024) Non-convex robust hypothesis testing using Sinkhorn uncertainty sets. *arXiv preprint arXiv:2403.14822* .

[70] Wang J, Gao R, Xie Y (2024) Regularization for adversarial robust learning. *arXiv preprint arXiv:2408.09672* .

[71] Wang J, Gao R, Zha H (2024) Reliable off-policy evaluation for reinforcement learning. *Operations Research* 72(2):699–716.

[72] Wang J, Moore R, Xie Y, Kamaleswaran R (2022) Improving sepsis prediction model generalization with optimal transport. *Machine Learning for Health*, 474–488 (PMLR).

[73] Wang J, Xie Y (2022) A data-driven approach to robust hypothesis testing using sinkhorn uncertainty sets. *arXiv preprint arXiv:2202.04258* .

[74] Wen J, Yang J (2025) Distributionally robust optimization via diffusion ambiguity modeling. *arXiv preprint arXiv:2510.22757* .

[75] Wiesemann W, Kuhn D, Sim M (2014) Distributionally robust convex optimization. *Operations Research* 62(6):1358–1376.

[76] Wu D, Xie Y (2024) Annealing flow generative models towards sampling high-dimensional and multi-modal distributions. *arXiv preprint arXiv:2409.20547* .

[77] Xiao Q, Yuan H, Saif A, Liu G, Kompella R, Wang M, Chen T (2025) A first-order generative bilevel optimization framework for diffusion models. *arXiv preprint arXiv:2502.08808* .

[78] Xu C, Lee J, Cheng X, Xie Y (2024) Flow-based distributionally robust optimization. *IEEE Journal on Selected Areas in Information Theory* .

[79] Xu Z, Zhu JJ (2025) Gradient flow sampler-based distributionally robust optimization. *arXiv preprint arXiv:2510.25956* .

[80] Yang SB, Li Z (2023) Distributionally robust chance-constrained optimization with sinkhorn ambiguity set. *AIChE Journal* 69(10):e18177.

[81] Yang Y, Zhou Y, Lu Z (2025) Nested stochastic algorithm for generalized sinkhorn distance-regularized distributionally robust optimization. *arXiv preprint arXiv:2503.22923* .

[82] Zhen J, Kuhn D, Wiesemann W (2025) A unified theory of robust and distributionally robust optimization via the primal-worst-equals-dual-best principle. *Operations Research* 73(2):862–878.

[83] Zhu L, Xie Y (2024) Distributionally robust optimization via iterative algorithms in continuous probability spaces. *arXiv preprint arXiv:2412.20556* .

[84] Zhu Y, Zhang Y, Wang Z, Yang Z, Chen X (2024) A mean-field analysis of neural stochastic gradient descent-ascent for functional minimax optimization. *arXiv preprint arXiv:2404.12312* .

A Notations and Definitions

We provide a table consisting of related mathematical notations and a list of useful definitions below.

Table 1: Common Mathematical Notations

Notation	Meaning
θ	Decision
z	Random vector
d	Dimension of random vector z
ϵ	Entropic regularization parameter of Sinkhorn DRO
λ	Soft-constrained penalty parameter of Sinkhorn DRO
$x^{(i)}$	i-th collected sample
μ_*^{θ}	Worst-case distribution for fixed decision θ
$\mu_*^{\theta,i}$	Worst-case distribution for i-th observation and fixed decision θ
$L_{f,1}, L_{f,2}$	Lipschitz constants for $f_\theta(z)$ and $\nabla f_\theta(z)$, respectively
α	Log-Sobolev inequality constant for the worst-case distribution
σ^2	Variance of gradient estimator $\nabla f_\theta(z)$ for $z \sim \mu_*^{\theta,i}$ and $i \sim \mathrm{Uniform}([n])$
δ	Accuracy level of Algorithm 2
ϱ	Error tolerance of bilevel optimization Problem (5)
$\mu_k^{(i)}$	Estimated worst-case distribution for i-th observation at k-th iteration of Algorithm 3
$z_k^{(i)}$	Random sample following distribution $\mu_k^{(i)}$
v_{k+1}	Constructed gradient estimator at k-th iteration of Algorithm 3
r_{k+1}	Constructed momentum gradient estimator at k-th iteration of Algorithm 3
β_0	Momentum parameter
τ	Stepsize parameter for one-step update of Langevin dynamics
η	Stepsize parameter for Algorithm 1 or scaled stepsize parameter for Algorithm 3
T_{out}	Number of iterations for Algorithm 1
T	Number of iterations for Algorithm 2 or 3
I_k	Random sampled set of indices from $[n]$ at k-th iteration

KL-divergence. For two probability distributions μ and ν, define the KL-divergence

$$\mathcal{D}_{\mathrm{KL}}(\mu\|\nu) = \int \log\left(\frac{\mathrm{d}\mu(z)}{\mathrm{d}\nu(z)}\right)\,\mathrm{d}\mu(z).$$

Suppose $v(\cdot)$ is the density function of ν, we also write $\mathcal{D}_{\mathrm{KL}}(\mu\|v(\cdot))$ to represent $\mathcal{D}_{\mathrm{KL}}(\mu\|\nu)$ for notational simplicity.

Wasserstein Distance. For two probability distributions μ and ν and $p \in [1,\infty)$, define the p-Wasserstein distance

$$\mathcal{W}_p(\mu,\nu) = \inf_{\gamma\in\Gamma(\mu,\nu)} \left\{ \left(\mathbb{E}_{(x,y)\sim\gamma}[\|x-y\|_2^p]\right)^{1/p} \right\},$$

where $\Gamma(\mu,\nu)$ denotes the set of joint distributions with marginal distributions being μ and ν, respectively. Especially, when $p=1$, the Wasserstein distance has the strong dual reformulation:

$$\mathcal{W}_1(\mu,\nu) = \sup_{f:\ \|f\|_{\mathrm{Lip}}\le 1} \Big\{ \mathbb{E}_{z\sim\mu}[f(z)] - \mathbb{E}_{z\sim\nu}[f(z)] \Big\}.$$

Total variation distance. For two probability distributions μ and ν, define the total variation distance

$$\mathrm{TV}(\mu,\nu) = \sup_{f:\ \|f\|_\infty\le 1} \Big\{ \mathbb{E}_{z\sim\mu}[f(z)] - \mathbb{E}_{z\sim\nu}[f(z)] \Big\}.$$

Fisher divergence. For two probability distributions μ and ν, define the Fisher divergence

$$\mathrm{FD}(\mu\|\nu) = \int \left\| \nabla_z \log\left(\frac{\mu[z]}{\nu[z]}\right) \right\|^2 \mu[z]\,\mathrm{d}z,$$

where for a distribution μ, $\mu[z]$ denote its density function at z.

Entropy. For a probability distribution μ, define its (differential) entropy

$$\mathcal{H}(\mu) = -\int \log(\mu[z])\mu[z]\,\mathrm{d}z.$$

B Proofs of Technical Results in Section 3

Proof of Proposition 3.3. The first part of this proposition follows from [27]. The second part follows a similar proof idea as in [38]. □

Proof of Theorem 3.4. It can be shown that $V_i(\theta, z) := -\frac{1}{\lambda\epsilon} f_\theta(z) + \frac{1}{2\epsilon}\|z - x^{(i)}\|_2^2$ is $L_{V,2}$-smooth with

$$L_{V,2} := \frac{1}{\epsilon}\left(1 + \frac{L_{f,2}}{\lambda}\right).$$

Also, $\mu_*^{i,\theta}[z] := \frac{\mathrm{d}\mu_*^{i,\theta}}{\mathrm{d}z}(z) \propto \exp(-V_i(\theta, z))$ satisfies LSI with constant α. Based on [65, Lemma 3], it holds that

$$D_{\mathrm{KL}}(\mu_T \| \mu_*^{\theta,i}) \le e^{-\alpha\tau\epsilon T} D_{\mathrm{KL}}(\mu_0 \| \mu_*^{\theta,i}) + \frac{8\tau\epsilon d L_{V,2}^2}{\alpha}.$$

Since $\tau \le \frac{\alpha\epsilon}{4(1+L_{f,2}/\lambda)^2} \cdot \frac{\delta^2\alpha}{8d}$, we derive

$$\frac{8\tau\epsilon d L_{V,2}^2}{\alpha} \le \frac{\alpha\delta^2}{4}.$$

Since $T \ge \frac{1}{\alpha\tau\epsilon}\log\frac{4D_{\mathrm{KL}}(\mu_0\|\mu_*^{\theta,i})}{\delta^2\alpha}$, we derive

$$e^{-\alpha\tau\epsilon T} D_{\mathrm{KL}}(\mu_0 \| \mu_*^{\theta,i}) \le \frac{\alpha\delta^2}{4}.$$

Combining those terms together, we obtain that

$$D_{\mathrm{KL}}(\mu_T \| \mu_*^{\theta,i}) \le \frac{\alpha\delta^2}{2}.$$

As $\mu_*^{\theta,i}$ satisfies Talagrand's inequality with constant α, we derive that

$$\mathcal{W}_2(\mu_T, \mu_*^{\theta,i}) \le \sqrt{\frac{2D_{\mathrm{KL}}(\mu_T\|\mu_*^{\theta,i})}{\alpha}} \le \delta.$$

The proof is completed. □

Proof of Lemma 3.7.

- For the bias part, it holds that

$$\begin{aligned}
&\left\|\mathbb{E}\big[\widehat{\nabla} F(\theta; z)\big] - \nabla F(\theta)\right\|_2 \\
\le&\mathbb{E}_{i\sim\mathrm{Uniform}([n])}\left\|\mathbb{E}_{\tilde{\mu}^{i,\theta}}\big[\nabla_\theta f_\theta(z)\big] - \mathbb{E}_{\mu_*^{i,\theta}}\big[\nabla_\theta f_\theta(z)\big]\right\|_2 \\
\le&\mathbb{E}_{i\sim\mathrm{Uniform}([n])}\big[\mathcal{W}_1(\tilde{\mu}^{i,\theta}, \mu_*^{i,\theta}) \cdot L_{f,2}\big] \\
\le&\mathbb{E}_{i\sim\mathrm{Uniform}([n])}\big[\mathcal{W}_2(\tilde{\mu}^{i,\theta}, \mu_*^{i,\theta}) \cdot L_{f,2}\big] \le \delta \cdot L_{f,2},
\end{aligned}$$

  where in the first inequality, we denote by $\tilde{\mu}^{i,\theta}$ the probability distribution such that $\mathcal{W}_2(\tilde{\mu}^{i,\theta}, \mu_*^{i,\theta}) \le \delta$.

- For the variance part, it holds that

$$\mathbb{V}\mathrm{ar}(\widehat{\nabla} F(\theta; z)) = \frac{1}{n}\sum_{i=1}^n A_i,$$

  where

$$A_i = \mathbb{E}_{z\sim\tilde{\mu}^{\theta,i}}\left\|\nabla_\theta f_\theta(z) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{z\sim\tilde{\mu}^{\theta,i}}[\nabla_\theta f_\theta(z)]\right\|_2^2.$$

Based on the triangular inequality,

$$\begin{aligned} A_i \leq& \mathbb{E}_{z\sim\mu_*^{\theta,i}} \left\| \nabla_\theta f_\theta(z) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{z\sim\tilde{\mu}^{\theta,i}}[\nabla_\theta f_\theta(z)] \right\|_2^2 \\ &+ \left| \mathbb{E}_{z\sim\mu_*^{\theta,i}-\tilde{\mu}^{\theta,i}} \left\| \nabla_\theta f_\theta(z) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{z\sim\tilde{\mu}^{\theta,i}}[\nabla_\theta f_\theta(z)] \right\|_2^2 \right| \end{aligned} \tag{14}$$

Note that in the expression above, we write $\mathbb{E}_{\mu-\nu}[f(z)]$ to denote the difference between $\mathbb{E}_\mu[f(z)]$ and $\mathbb{E}_\nu[f(z)]$ for generic probability distributions μ, ν and measurable function $f(z)$. For the first component on the right-hand side of (14), it can be further bounded as

$$\begin{aligned} &2\mathbb{E}_{z\sim\mu_*^{\theta,i}} \left\| \nabla_\theta f_\theta(z) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{z\sim\mu_*^{\theta,i}}[\nabla_\theta f_\theta(z)] \right\|_2^2 + 2\left\| \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{z\sim\tilde{\mu}^{\theta,i}-\mu_*^{\theta,i}}[\nabla_\theta f_\theta(z)] \right\|_2^2 \\ \leq& 2\mathbb{E}_{z\sim\mu_*^{\theta,i}} \left\| \nabla_\theta f_\theta(z) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{z\sim\mu_*^{\theta,i}}[\nabla_\theta f_\theta(z)] \right\|_2^2 + 2L_{f,2}^2\delta^2. \end{aligned}$$

For the second component on the right-hand side of (14), from the proof of Theorem 3.4, it can be shown that

$$\mathrm{TV}(\mu_*^{\theta,i}, \tilde{\mu}^{\theta,i}) \leq \sqrt{\frac{1}{2}\mathcal{D}_{\mathrm{KL}}(\tilde{\mu}^{\theta,i}\|\mu_*^{\theta,i})} \leq \frac{\delta\sqrt{\alpha}}{2},$$

and

$$\left\| \nabla_\theta f_\theta(z) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{z\sim\tilde{\mu}^{\theta,i}}[\nabla_\theta f_\theta(z)] \right\|_2^2 \leq 4L_{f,1}^2,$$

and therefore

$$\left| \mathbb{E}_{z\sim\mu_*^{\theta,i}-\tilde{\mu}^{\theta,i}} \left\| \nabla_\theta f_\theta(z) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{z\sim\tilde{\mu}^{\theta,i}}[\nabla_\theta f_\theta(z)] \right\|_2^2 \right| \leq 2\sqrt{\alpha}\delta L_{f,1}^2.$$

Combining these bounds togehter, we arrive that

$$\begin{aligned} &\mathbb{V}\mathrm{ar}(\widehat{\nabla} F(\theta; z)) \\ \leq& \frac{1}{n}\sum_{i=1}^n \left(2\mathbb{E}_{z\sim\mu_*^{\theta,i}} \left\| \nabla_\theta f_\theta(z) - \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{z\sim\mu_*^{\theta,i}}[\nabla_\theta f_\theta(z)] \right\|_2^2 + 2L_{f,2}^2\delta^2 \right) + 2\sqrt{\alpha}\delta L_{f,1}^2 \\ \leq& 2\sigma^2 + 2L_{f,2}^2\delta^2 + 2\sqrt{\alpha}\delta L_{f,1}^2. \end{aligned}$$

- The last part of this lemma follows from Theorem 3.4.

□

Proof of Theorem 3.8. Denote by $\widehat{F}(\theta)$ the objective corresponding to the gradient $\mathbb{E}\widehat{\nabla} F(\theta)$. We have the following error decomposition:

$$\begin{aligned} &\mathbb{E}\|\nabla F(\widehat{\theta})\|_2^2 \leq 2\mathbb{E}\|\nabla \widehat{F}(\widehat{\theta})\|_2^2 + 2\mathbb{E}\|\nabla \widehat{F}(\widehat{\theta}) - \nabla F(\widehat{\theta})\|_2^2 \\ \leq& 2\mathbb{E}\|\nabla \widehat{F}(\widehat{\theta})\|_2^2 + 2L_{f,2}^2\delta^2. \end{aligned}$$

As $\widehat{F}(\widehat{\theta})$ is $L_{f,2}$-smooth in θ, according to the convergence analysis in [5], the output in Algorithm 1 with constant stepsize η satisfies

$$\mathbb{E}\|\nabla \widehat{F}(\widehat{\theta})\|_2^2 \leq \frac{2\big(\widehat{F}(\theta_0) - \min_\theta \widehat{F}(\theta)\big)}{\eta T_{\text{out}}} + L_{f,2}\eta\mathbf{V}.$$

We specify the stepsize $\eta = 1/\sqrt{T_{\text{out}}\mathbf{V}}$, then

$$\mathbb{E}\|\nabla \widehat{F}(\widehat{\theta})\|_2^2 \leq \frac{\sqrt{\mathbf{V}}\big(2\big(\widehat{F}(\theta_0) - \min_\theta \widehat{F}(\theta)\big) + L_{f,2}\big)}{\sqrt{T_{\text{out}}}}$$

In order to ensure $\mathbb{E}\|\nabla F(\widehat{\theta})\|_2^2 \leq \varrho^2$, we take

$$2L_{f,2}^2\delta^2 \leq \frac{1}{2}\varrho^2, \quad \frac{\sqrt{\mathbf{V}}\big(2\big(\widehat{F}(\theta_0) - \min_\theta \widehat{F}(\theta)\big) + L_{f,2}\big)}{\sqrt{T_{\text{out}}}} \leq \frac{1}{4}\varrho^2.$$

Thus, we take $\delta = \frac{\varrho}{2L_{f,2}}$ and

$$T_{\text{out}} \geq 16\mathbf{V}\big(2\big(\widehat{F}(\theta_0) - \min_\theta \widehat{F}(\theta)\big) + L_{f,2}\big)^2\varrho^{-4}.$$

The proof is completed. □

C Proofs of Technical Results in Section 4.1

Proof of Lemma 4.2. It is easy to verify that $F(\theta)$ is $L_{f,2}$-smooth in θ. It follows that

$$\begin{aligned}
F(\theta_{k+1}) &\leq F(\theta_k) + \nabla F(\theta_k)^\top(\theta_{k+1} - \theta_k) + \frac{L_{f,2}}{2}\|\theta_{k+1} - \theta_k\|_2^2 \\
&= F(\theta_k) - \eta\tau\nabla F(\theta_k)^\top r_{k+1} + \frac{L_{f,2}\tau^2\eta^2}{2}\|r_{k+1}\|_2^2 \\
&= F(\theta_k) + \frac{\eta\tau}{2}\|\nabla F(\theta_k) - r_{k+1}\|_2^2 - \frac{\tau\eta}{2}\|\nabla F(\theta_k)\|_2^2 + \left(\frac{L_{f,2}\eta^2\tau^2}{2} - \frac{\eta\tau}{2}\right)\|r_{k+1}\|_2^2,
\end{aligned}$$

where the first equality is by the relation that $\theta_{k+1} = \theta_k - \eta\tau r_{k+1}$.

Since $\eta\tau \leq \frac{1}{2L_{f,2}}$, it holds that $\frac{L_{f,2}\eta^2\tau^2}{2} - \frac{\eta\tau}{2} \leq -\frac{\eta\tau}{4}$. Therefore, the desired result holds. □

Proof of Lemma 4.3. By the relation $r_{k+1} = (1-\beta_0)r_k + \beta_0 v_{k+1}$, it follows that

$$\begin{aligned}
&\mathbb{E}\big[\|\nabla F(\theta_k) - r_{k+1}\|_2^2 \big| \mathcal{G}_{k-1}\big] \\
=&\mathbb{E}\big[\|\nabla F(\theta_k) - (1-\beta_0)r_k - \beta_0 v_{k+1}\|_2^2 \big| \mathcal{G}_{k-1}\big] \\
=&\mathbb{E}\big[\big\|(1-\beta_0)(\nabla F(\theta_{k-1}) - r_k) + (1-\beta_0)(\nabla F(\theta_k) - \nabla F(\theta_{k-1})) \\
&\quad + \beta_0(\nabla F(\theta_k) - \nabla F(\theta_k; \mu_k^{(1:n)})) + \beta_0(\nabla F(\theta_k; \mu_k^{(1:n)}) - v_{k+1})\big\|_2^2 \big| \mathcal{G}_{k-1}\big] \\
=&\big\|(1-\beta_0)(\nabla F(\theta_{k-1}) - r_k) + (1-\beta_0)(\nabla F(\theta_k) - \nabla F(\theta_{k-1})) \\
&\quad + \beta_0(\nabla F(\theta_k) - \nabla F(\theta_k; \mu_k^{(1:n)}))\big\|_2^2 + \beta_0^2\mathbb{E}\big[\big\|\nabla F(\theta_k; \mu_k^{(1:n)}) - v_{k+1}\big\|_2^2 \big| \mathcal{G}_{k-1}\big],
\end{aligned} \tag{15}$$

where the last equality is because, conditioned on $\mathcal{G}_{k-1}$, the term

$$\mathbf{A}_1 := (1-\beta_0)(\nabla F(\theta_{k-1}) - r_k) + (1-\beta_0)(\nabla F(\theta_k) - \nabla F(\theta_{k-1})) + \beta_0(\nabla F(\theta_k) - \nabla F(\theta_k; \mu_k^{(1:n)}))$$

is deterministic, and $\mathbb{E}[\nabla F(\theta_k; \mu_k^{(1:n)}) - v_{k+1} \mid \mathcal{G}_{k-1}] = 0$.

Furthermore, due to the relation $\|a+b\|_2^2 \leq (1+\beta)\|a\|_2^2 + (1+\frac{1}{\beta})\|b\|_2^2$ for any $\beta > 0$, it holds that

$$\begin{aligned}
\|\mathbf{A}_1\|_2^2 &\leq (1+\beta_0)(1-\beta_0)^2\|\nabla F(\theta_{k-1}) - r_k\|_2^2 \\
&\qquad + (1 + \frac{1}{\beta_0})\|(1-\beta_0)(\nabla F(\theta_k) - \nabla F(\theta_{k-1})) + \beta_0(\nabla F(\theta_k) - \nabla F(\theta_k; \mu_k^{(1:n)}))\|_2^2 \\
&\leq (1+\beta_0)(1-\beta_0)^2\|\nabla F(\theta_{k-1}) - r_k\|_2^2 \\
&\qquad + 2(1 + \frac{1}{\beta_0})\Big[(1-\beta_0)^2\|\nabla F(\theta_k) - \nabla F(\theta_{k-1})\|_2^2 + \beta_0^2\|\nabla F(\theta_k) - \nabla F(\theta_k; \mu_k^{(1:n)})\|_2^2\Big] \\
&\leq (1-\beta_0)\|\nabla F(\theta_{k-1}) - r_k\|_2^2 + \frac{4}{\beta_0}\|\nabla F(\theta_k) - \nabla F(\theta_{k-1})\|_2^2 \\
&\qquad + 4\beta_0\|\nabla F(\theta_k) - \nabla F(\theta_k; \mu_k^{(1:n)})\|_2^2
\end{aligned} \tag{16}$$

where the second inequality uses the relation $\|a+b\|_2^2 \leq 2\|a\|_2^2 + 2\|b\|_2^2$, and the last inequality uses the assumption that $\beta_0 \in (0, 1]$.

By the Lipschitz smoothness assumption of $F(\theta)$,

$$\|\nabla F(\theta_k) - \nabla F(\theta_{k-1})\|_2^2 \leq L_{f,2}^2\|\theta_k - \theta_{k-1}\|_2^2 = L_{f,1}^2\tau^2\eta^2\|r_k\|_2^2. \tag{17}$$

Combining relations (15), (16), and (17), we obtain the desired result. □

Proof of Lemma 4.4. Throughout the proof, we use the stepsize parameters

$$\tau \leq \min\left(1, \frac{1}{L_{G,2}}, \sqrt{\frac{\lambda}{\epsilon L_{f,2}^2}}, \frac{1}{\alpha}, \frac{\alpha}{4L_{G,2}^2}\right), \qquad \eta \leq 1. \tag{18}$$

Let us define the auxiliary distribution $\tilde{\mu}_{k+1}^{(i)}$ as the law of $\tilde{z}_{k+1}^{(i)}$, where

$$\tilde{z}_{k+1}^{(i)} = z_k^{(i)} - \tau \left(-\frac{\nabla_z f_{\theta_k}(z_k^{(i)})}{\lambda} + (z_k^{(i)} - x^{(i)}) \right) + \sqrt{2\tau\epsilon}\zeta_k^{(i)}, \quad \mathrm{Law}(z_k^{(i)}) = \mu_k^{(i)}.$$

Therefore, it holds that

$$\begin{aligned} &\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\mu_{k+1}^{(i)} \| \mu_*^{i,\theta_{k+1}}) \big| \mathcal{G}_k\big] \\ =&\left(1 - \frac{|I_k|}{n}\right) \mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)} \| \mu_*^{i,\theta_{k+1}}) \big| \mathcal{G}_k\big] + \frac{|I_k|}{n} \mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\tilde{\mu}_{k+1}^{(i)} \| \mu_*^{i,\theta_{k+1}}) \big| \mathcal{G}_k\big], \end{aligned} \tag{19}$$

where $\mathbb{E}[\cdot \mid \mathcal{G}_k]$ refers to the expected value with respect to the randomness conditioned on $\mathcal{G}_k$, i.e., taking the expected value over the randomness of the sampling indices I_k. We upper bound each of the component on the right-hand-side in the following.

Bounding $\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)} \| \mu_*^{i,\theta_{k+1}}) \big| \mathcal{G}_k\big]$. Define θ_{k+1} as the output at time τ of the dynamics

$$\vartheta_t = \theta_k - t\eta r_{k+1}, \quad t \in [0, \tau]. \tag{20}$$

For a given probability measure μ, we write $\mu[z]$ for the density function $\frac{\mathrm{d}\mu(z)}{\mathrm{d}z}$. For fixed I_k, by the chain rule, it holds that

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)} \| \mu_*^{i,\vartheta_t}) = -\int \frac{\mu_k^{(i)}[z]}{\mu_*^{i,\vartheta}[z]} \frac{\mathrm{d}}{\mathrm{d}t} \left(\mu_*^{i,\vartheta_t}[z]\right) \mathrm{d}z. \tag{21}$$

Define the function $V_i(\theta, z) := -\frac{1}{\lambda\epsilon} f_\theta(z) + \frac{1}{2\epsilon}\|z - x^{(i)}\|_2^2$ and scalar $Z_{i,\theta} := \int V_i(\theta, z)\, \mathrm{d}z$. Then the density $\mu_*^{i,\vartheta_t}[z] = \exp(-V_i(\vartheta_t, z))/Z_{i,\vartheta_t}$. By the chain rule, it holds that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\left(\mu_*^{i,\vartheta_t}[z]\right) = \left\langle \frac{\mathrm{d}}{\mathrm{d}\vartheta_t}\left(\mu_*^{i,\vartheta_t}[z]\right),\ \frac{\mathrm{d}\vartheta_t}{\mathrm{d}t} \right\rangle \\ =&\left\langle -\nabla_1 V_i(\vartheta_t, z) \cdot \mu_*^{i,\vartheta_t}[z] + \mathbb{E}_{z\sim\mu_*^{i,\vartheta_t}}\left[\nabla_1 V_i(\vartheta_t, z)\right] \cdot \mu_*^{i,\vartheta_t}[z],\ -\eta r_{k+1} \right\rangle \\ =&\frac{\eta}{\lambda\epsilon} \mu_*^{i,\vartheta_t}[z] \cdot \langle \nabla_\theta f_{\vartheta_t}(z) + \mathbb{E}_{z\sim\mu_*^{i,\vartheta_t}}\left[f_{\vartheta_t}(z)\right],\ r_{k+1} \rangle. \end{aligned} \tag{22}$$

Substituting (22) into (21), we obtain that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)} \| \mu_*^{i,\vartheta_t}) = -\frac{\eta}{\lambda\epsilon} \int \langle \nabla_\theta f_{\vartheta_t}(z) + \mathbb{E}_{z\sim\mu_*^{i,\vartheta_t}}\left[f_{\vartheta_t}(z)\right],\ r_{k+1} \rangle \cdot \mu_k^{(i)}[z]\, \mathrm{d}z \\ \le&\frac{\eta}{\lambda\epsilon}(L_{f,1}^2 + \|r_{k+1}\|_2^2). \end{aligned}$$

By the fundamental theorem of calculus, it holds that

$$\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)} \| \mu_*^{i,\vartheta_t}) \le \mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)} \| \mu_*^{i,\theta_k}) + \frac{\eta t(L_{f,1}^2 + \|r_{k+1}\|_2^2)}{\lambda\epsilon}, \quad \forall t \in [0, \tau].$$

Therefore, we obtain that

$$\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)} \| \mu_*^{i,\theta_{k+1}}) \big| \mathcal{G}_k\big] \le \mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)} \| \mu_*^{i,\theta_k}) + \frac{\eta\tau L_{f,1}^2}{\lambda\epsilon} + \frac{\eta\tau}{\lambda\epsilon} \mathbb{E}\big[\|r_{k+1}\|_2^2 \big| \mathcal{G}_k\big]. \tag{23}$$

Bounding $\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\tilde{\mu}_{k+1}^{(i)} \| \mu_*^{i,\theta_{k+1}}) \big| \mathcal{G}_k\big]$. Define

$$G_i(\theta, z) = -\frac{f_\theta(z)}{\lambda} + \frac{1}{2}\|z - x^{(i)}\|_2^2. \tag{24}$$

It can be shown that $G_i(z, \theta)$ is $L_{G,2}$-smooth in z with

$$L_{G,2} := 1 + \frac{1}{\lambda} L_{f,2}. \tag{25}$$

Let us define $\rho_0 := \mu_k^{(i)}$ with $z_0 \sim \rho_0$, and

$$\mathrm{d}z_t = -\nabla_2 G_i(\theta_k, z_0)\,\mathrm{d}t + \sqrt{2\epsilon}\,\mathrm{d}\mathbf{B}_t, \quad \mathrm{Law}(z_t) = \rho_t.$$

Then, $\tilde{\mu}_{k+1}^{(i)}$ has the same distribution of the output of the SDE above at time τ. Recall we define θ_{k+1} as the output at time τ of (20). Following the same calculation procedure as in [47, Appendix B], we find the time derivative

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}_{\mathrm{KL}}(\rho_t \| \mu_*^{i,\vartheta_t}) = -\mathrm{FD}(\rho_t \| \mu_*^{i,\vartheta_t}) \\
&\quad + \int \left\langle \nabla \log\left(\frac{\rho_t[z_t]}{\mu_*^{i,\vartheta_t}[z_t]}\right),\ \nabla_2 G_i(\theta_k, z_t) - \mathbb{E}_{z_0 \sim \rho_{0|t}}[\nabla_2 G_i(\theta_k, z_0) \mid z_t] \right\rangle \rho_t[z_t]\,\mathrm{d}z_t \\
&\quad + \int \left\langle \nabla \log\left(\frac{\rho_t[z_t]}{\mu_*^{i,\vartheta_t}[z_t]}\right),\ \nabla_2 G_i(\vartheta_t, z_t) - \nabla_2 G_i(\theta_k, z_t) \right\rangle \rho_t[z_t]\,\mathrm{d}z_t \\
&\quad + \frac{\eta}{\lambda\epsilon}\int \Big[\langle \nabla f_{\vartheta_t}(z_t), r_{k+1}\rangle - \mathbb{E}_{\mu_*^{i,\vartheta_t}}[\langle \nabla f_{\vartheta_t}(z_t), r_{k+1}\rangle]\Big]\,\rho_t[z_t]\,\mathrm{d}z_t.
\end{aligned} \tag{26}$$

Denote by the second, third and fourth components of the right-hand-side above as $\mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4$, respectively. We provide upper bound on those expressions below.

- By the law of total expectation, it holds that

$$\begin{aligned}
\mathbf{A}_2 &= \int \left\langle \nabla \log\left(\frac{\rho_t[z_t]}{\mu_*^{i,\vartheta_t}[z_t]}\right),\ \nabla_2 G_i(\theta_k, z_t) - \nabla_2 G_i(\theta_k, z_0) \right\rangle \rho_{0,t}[z_0, z_t]\,\mathrm{d}(z_0, z_t) \\
&\leq \int \Bigg(\|\nabla_2 G_i(\theta_k, z_t) - \nabla_2 G_i(\theta_k, z_0)\|_2^2 \\
&\qquad\qquad + \frac{1}{4}\left\|\nabla \log\left(\frac{\rho_t[z_t]}{\mu_*^{i,\vartheta_t}[z_t]}\right)\right\|_2^2 \Bigg) \rho_{0,t}[z_0, z_t]\,\mathrm{d}(z_0, z_t) \\
&\leq L_{G,2}^2 \mathbb{E}_{(z_0,z_t)\sim\rho_{0,t}}\|z_t - z_0\|_2^2 + \frac{1}{4}\mathrm{FD}(\rho_t \| \mu_*^{i,\vartheta_t}),
\end{aligned}$$

  where the first inequality is based on the relation $\langle a, b\rangle \leq \|a\|_2^2 + \frac{1}{4}\|b\|_2^2$, and the second inequality is because $G_i(\theta, z)$ is $L_{G,2}$-smooth in z. Based on Lemma D.1, it further holds that

$$\mathbf{A}_2 \leq \frac{4t^2 L_{G,2}^4}{\alpha}\mathcal{D}_{\mathrm{KL}}(\rho_0 \| \mu_*^{i,\theta_k}) + 2t^2 d L_{G,2}^3 \epsilon + 2t\epsilon d L_{G,2}^2 + \frac{1}{4}\mathrm{FD}(\rho_t \| \mu_*^{i,\vartheta_t}). \tag{27}$$

- Using the relation $\langle a, b\rangle \leq \|a\|_2^2 + \frac{1}{4}\|b\|_2^2$, again, it holds that

$$\begin{aligned}
\mathbf{A}_3 &\leq \int \left(\|\nabla_2 G_i(\vartheta_t, z_t) - \nabla_2 G_i(\theta_k, z_t)\|_2^2 + \frac{1}{4}\left\|\nabla \log\left(\frac{\rho_t[z_t]}{\mu_*^{i,\vartheta_t}[z_t]}\right)\right\|_2^2 \right) \rho_t[z_t]\,\mathrm{d}z_t \\
&\leq \frac{L_{f,2}^2}{\lambda^2}\|\vartheta_t - \theta_k\|_2^2 + \frac{1}{4}\mathrm{FD}(\rho_t \| \mu_*^{i,\vartheta_t}) \\
&= \frac{L_{f,2}^2 t^2 \eta^2}{\lambda^2}\|r_{k+1}\|_2^2 + \frac{1}{4}\mathrm{FD}(\rho_t \| \mu_*^{i,\vartheta_t}),
\end{aligned} \tag{28}$$

  where the second inequality is because

$$\|\nabla_2 G_i(\theta, z) - \nabla_2 G_i(\theta', z)\|_2 = \frac{1}{\lambda}\|\nabla_z f_\theta(z) - \nabla_z f_{\theta'}(z)\|_2 \leq L_{f,2}/\lambda.$$

- Using the relation $\langle a, b\rangle \leq \frac{1}{2}\|a\|_2^2 + \frac{1}{2}\|b\|_2^2$, it holds that

$$\mathbf{A}_4 \leq \frac{\eta(L_{f,1}^2 + \|r_{k+1}\|_2^2)}{\lambda\epsilon}. \tag{29}$$

As μ_*^{i,ϑ_t} satisfies α-LSI,

$$\mathcal{D}_{\mathrm{KL}}(\rho_t \| \mu_*^{i,\vartheta_t}) \le \frac{1}{2\alpha}\mathrm{FD}(\rho_t \| \mu_*^{i,\vartheta_t}). \tag{30}$$

Sustituting (27), (28), (29), and (30) into (26), we obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}_{\mathrm{KL}}(\rho_t \| \mu_*^{i,\vartheta_t}) \\
\le& -\alpha\mathcal{D}_{\mathrm{KL}}(\rho_t \| \mu_*^{i,\vartheta_t}) + \frac{4L_{G,2}^4 t^2}{\alpha}\mathcal{D}_{\mathrm{KL}}(\rho_0 \| \mu_*^{i,\theta_k}) \\
&\qquad + 2dL_{G,2}^3\epsilon t^2 + 2\epsilon d L_{G,2}^2 t + \frac{L_{f,2}^2\eta^2 t^2}{\lambda^2}\|r_{k+1}\|^2 + \frac{\eta(L_{f,1}^2 + \|r_{k+1}\|_2^2)}{\lambda\epsilon} \\
\le& -\alpha\mathcal{D}_{\mathrm{KL}}(\rho_t \| \mu_*^{i,\vartheta_t}) + \frac{4L_{G,2}^4 \tau^2}{\alpha}\mathcal{D}_{\mathrm{KL}}(\rho_0 \| \mu_*^{i,\theta_k}) \\
&\qquad + 2dL_{G,2}^3\epsilon \tau^2 + 2\epsilon d L_{G,2}^2 \tau + \frac{L_{f,2}^2\eta^2 \tau^2}{\lambda^2}\|r_{k+1}\|^2 + \frac{\eta(L_{f,1}^2 + \|r_{k+1}\|_2^2)}{\lambda\epsilon}.
\end{aligned}$$

Since $\tau \le \frac{1}{L_{G,2}}, \eta \le 1$, and $\frac{L_{f,2}^2\tau^2}{\lambda} \le \frac{1}{\epsilon}$, we further obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{D}_{\mathrm{KL}}(\rho_t \| \mu_*^{i,\vartheta_t}) \\
\le& -\alpha\mathcal{D}_{\mathrm{KL}}(\rho_t \| \mu_*^{i,\vartheta_t}) + \frac{4L_{G,2}^4 \tau^2}{\alpha}\mathcal{D}_{\mathrm{KL}}(\rho_0 \| \mu_*^{i,\theta_k}) \\
&\qquad + 4\epsilon d L_{G,2}^2 \tau + \frac{L_{f,2}^2\eta^2 \tau^2}{\lambda^2}\|r_{k+1}\|^2 + \frac{\eta(L_{f,1}^2 + \|r_{k+1}\|_2^2)}{\lambda\epsilon} \\
\le& -\alpha\mathcal{D}_{\mathrm{KL}}(\rho_t \| \mu_*^{i,\vartheta_t}) + \frac{4L_{G,2}^4 \tau^2}{\alpha}\mathcal{D}_{\mathrm{KL}}(\rho_0 \| \mu_*^{i,\theta_k}) + 4\epsilon d L_{G,2}^2 \tau + \frac{\eta(L_{f,1}^2 + 2\|r_{k+1}\|_2^2)}{\lambda\epsilon}.
\end{aligned}$$

Define constants

$$\begin{aligned}
\mathbf{C}_1 &= \frac{4L_{G,2}^4\tau^2}{\alpha}\mathcal{D}_{\mathrm{KL}}(\rho_0 \| \mu_*^{i,\theta_k}), \\
\mathbf{C}_2 &= 4\epsilon d L_{G,2}^2\tau + \frac{\eta(L_{f,1}^2 + 2\|r_{k+1}\|_2^2)}{\lambda\epsilon}.
\end{aligned}$$

By Grönwall's inequality in Lemma D.3, we find

$$\begin{aligned}
&\mathcal{D}_{\mathrm{KL}}(\rho_t \| \mu_*^{i,\vartheta_t}) \\
\le& \frac{(\mathbf{C}_1 + \mathbf{C}_2)e^{-\alpha t}(e^{\alpha t} - 1)}{\alpha} + \mathcal{D}_{\mathrm{KL}}(\rho_0 \| \mu_*^{i,\theta_k})e^{-\alpha t} \\
\le& 2(\mathbf{C}_1 + \mathbf{C}_2)\tau e^{-\alpha\tau} + \mathcal{D}_{\mathrm{KL}}(\rho_0 \| \mu_*^{i,\theta_k})e^{-\alpha t} \\
=& 2\mathbf{C}_2\tau e^{-\alpha\tau} + \left(1 + \frac{8\tau^3 L_{G,2}^4}{\alpha}\right)\mathcal{D}_{\mathrm{KL}}(\rho_0 \| \mu_*^{i,\theta_k})e^{-\alpha t} \\
\le& 2\mathbf{C}_2\tau e^{-\alpha\tau} + \left(1 + \frac{\alpha\tau}{2}\right)\mathcal{D}_{\mathrm{KL}}(\rho_0 \| \mu_*^{i,\theta_k})e^{-\alpha\tau},
\end{aligned}$$

where the second inequality is because $\alpha t \le \alpha\tau \le 1$ and $e^{\alpha t} \le 1 + 2\alpha t$ and the last inequality is because $\frac{8\tau^3 L_{G,2}^4}{\alpha} \le \frac{\alpha\tau}{2}$. By taking the time $t = \tau$, we finally obtain that

$$\mathcal{D}_{\mathrm{KL}}(\tilde{\mu}_{k+1}^{(i)} \| \mu_*^{i,\theta_{k+1}}) \le 2\mathbf{C}_2\tau e^{-\alpha\tau} + \left(1 + \frac{\alpha\tau}{2}\right)\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)} \| \mu_*^{i,\theta_k})e^{-\alpha\tau}.$$

Therefore,

$$\begin{aligned}
\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\tilde{\mu}_{k+1}^{(i)}\|\mu_*^{i,\theta_{k+1}})\big|\mathcal{G}_k\big] &\leq \left(1+\frac{\alpha\tau}{2}\right)e^{-\alpha\tau}\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}\|\mu_*^{i,\theta_k}) + 2\tau e^{-\alpha\tau}\left(4\epsilon d L_{G,2}^2\tau + \frac{\eta L_{f,1}^2}{\lambda\epsilon}\right) \\
&\qquad + \frac{4\eta\tau}{\lambda\epsilon}\cdot e^{-\alpha\tau}\cdot\mathbb{E}\big[\|r_{k+1}\|_2^2\big|\mathcal{G}_k\big] \\
&\leq \left(1-\frac{\alpha\tau}{4}\right)\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}\|\mu_*^{i,\theta_k}) + 2\tau e^{-\alpha\tau}\left(4\epsilon d L_{G,2}^2\tau + \frac{\eta L_{f,1}^2}{\lambda\epsilon}\right) \\
&\qquad + \frac{4\eta\tau}{\lambda\epsilon}\cdot e^{-\alpha\tau}\cdot\mathbb{E}\big[\|r_{k+1}\|_2^2\big|\mathcal{G}_k\big],
\end{aligned} \tag{31}$$

where the second inequality is due to $\frac{\alpha\tau}{2}\leq\frac{1}{2}$ and $\left(1+\frac{\alpha\tau}{2}\right)e^{-\alpha\tau}\leq e^{-\alpha\tau/2}\leq 1-\frac{\alpha\tau}{4}$.

Bounding $\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\mu_{k+1}^{(i)}\|\mu_*^{i,\theta_{k+1}})\big|\mathcal{G}_k\big]$. Substituting (23) and (31) into (19), we obtain that

$$\begin{aligned}
&\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\mu_{k+1}^{(i)}\|\mu_*^{i,\theta_{k+1}})\big|\mathcal{G}_k\big] \\
=&\left(1-\frac{|I_k|}{n}\right)\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}\|\mu_*^{i,\theta_{k+1}})\big|\mathcal{G}_k\big] + \frac{|I_k|}{n}\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\tilde{\mu}_{k+1}^{(i)}\|\mu_*^{i,\theta_{k+1}})\big|\mathcal{G}_k\big] \\
\leq&\left(1-\frac{|I_k|}{n}\right)\left(\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}\|\mu_*^{i,\theta_k}) + \frac{\eta\tau L_{f,1}^2}{\lambda\epsilon} + \frac{\eta\tau}{\lambda\epsilon}\mathbb{E}\big[\|r_{k+1}\|_2^2\big|\mathcal{G}_k\big]\right) \\
&\quad + \frac{|I_k|}{n}\Bigg(\left(1-\frac{\alpha\tau}{4}\right)\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}\|\mu_*^{i,\theta_k}) + 2\tau e^{-\alpha\tau}\left(4\epsilon d L_{G,2}^2\tau + \frac{\eta L_{f,1}^2}{\lambda\epsilon}\right) \\
&\qquad\qquad + \frac{4\eta\tau}{\lambda\epsilon}\cdot e^{-\alpha\tau}\cdot\mathbb{E}\big[\|r_{k+1}\|_2^2\big|\mathcal{G}_k\big]\Bigg) \\
\leq&\left(1-\frac{\alpha\tau|I_k|}{4n}\right)\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}\|\mu_*^{i,\theta_k}) + \frac{3\eta\tau L_{f,1}^2}{\lambda\epsilon} + \frac{8\tau^2\epsilon d L_{G,2}^2|I_k|}{n} + \frac{5\eta\tau}{\lambda\epsilon}\mathbb{E}\big[\|r_{k+1}\|_2^2\big|\mathcal{G}_k\big].
\end{aligned}$$

Taking the full expectation, we obtain that

$$\begin{aligned}
&\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\mu_{k+1}^{(i)}\|\mu_*^{i,\theta_{k+1}})\big] \\
\leq&\left(1-\frac{\alpha\tau|I_k|}{4n}\right)\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}\|\mu_*^{i,\theta_k})\big] + \frac{3\eta\tau L_{f,1}^2}{\lambda\epsilon} + \frac{8\tau^2\epsilon d L_{G,2}^2|I_k|}{n} + \frac{5\eta\tau}{\lambda\epsilon}\mathbb{E}\big[\|r_{k+1}\|_2^2\big].
\end{aligned}$$

Taking summation over $k=0,\ldots,T$, we have that

$$\begin{aligned}
&\sum_{k=0}^{T}\mathbb{E}\big[\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}\|\mu_*^{i,\theta_k})\big] \\
\leq&\frac{4n}{\alpha\tau|I_k|}\mathcal{D}_{\mathrm{KL}}(\mu_0^{(i)}\|\mu_*^{i,\theta_0}) + \frac{12\eta L_{f,1}^2 T n}{\lambda\epsilon\alpha|I_k|} + \frac{32\tau\epsilon d L_{G,2}^2 T}{\alpha} + \frac{20\eta n}{\lambda\epsilon\alpha|I_k|}\sum_{k=0}^{T-1}\mathbb{E}\big[\|r_{k+1}\|_2^2\big].
\end{aligned}$$

The proof is complete. □

Before showing the proof of Theorem 4.5, we provide the following error bounds:

- Based on the expression in (12), it holds that

$$\begin{aligned}
&\left\|\nabla F(\theta_k) - \nabla F(\theta_k;\mu_k^{(1:n)})\right\|_2 \leq \frac{1}{n}\sum_{i\in[n]}\left\|\mathbb{E}_{\mu_k^{(i)}-\mu_*^{i,\theta_k}}[\nabla f_{\theta_k}(z)]\right\|_2 \\
\leq&\frac{1}{n}\sum_{i\in[n]}L_{f,1}\cdot\mathrm{TV}(\mu_k^{(i)},\mu_*^{i,\theta_k}) \\
\leq&\frac{L_{f,1}}{n}\sum_{i\in[n]}\sqrt{\frac{1}{2}\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)},\mu_*^{i,\theta_k})},
\end{aligned}$$

which further implies

$$\left\|\nabla F(\theta_k) - \nabla F(\theta_k; \mu_k^{(1:n)})\right\|_2^2 \leq \frac{L_{f,1}^2}{2n} \sum_{i\in[n]} \mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}, \mu_*^{i,\theta_k}).$$

- Since $\mathbb{V}\mathrm{ar}(v_{k+1} \mid \mathcal{G}_{k-1}) \leq \mathbb{E}[\|v_{k+1}\|_2^2 \mid \mathcal{G}_{k-1}] \leq L_{f,1}^2$, it holds that

$$\mathbb{E}\left[\left\|\nabla F(\theta_k; \mu_k^{(1:n)}) - v_{k+1}\right\|_2^2 \middle| \mathcal{G}_{k-1}\right] \leq \frac{L_{f,1}^2}{|I_k|}.$$

- Substituting the error bounds above into (13), we obtain

$$\begin{aligned}\mathbb{E}\big[\|\nabla F(\theta_k) - r_{k+1}\|_2^2 \big| \mathcal{G}_{k-1}\big] \leq (1-\beta_0)\|\nabla F(\theta_{k-1}) - r_k\|_2^2 + \frac{4L_{f,2}^2\tau^2\eta^2}{\beta_0}\|r_k\|_2^2 \\ + \frac{2\beta_0 L_{f,1}^2}{n} \sum_{i\in[n]} \mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}, \mu_*^{i,\theta_k}) + \frac{\beta_0^2 L_{f,1}^2}{|I_k|}.\end{aligned}$$

Consequently, it holds that

$$\begin{aligned}\sum_{k=0}^{T} \mathbb{E}\big[\|\nabla F(\theta_k) - r_{k+1}\|_2^2\big] \leq \frac{1}{\beta_0}\mathbb{E}\big[\|\nabla F(\theta_0) - r_1\|_2^2\big] \\ + \frac{4L_{f,2}^2\tau^2\eta^2}{\beta_0^2}\sum_{k=1}^{T}\|r_k\|_2^2 + \frac{2L_{f,1}^2}{n}\sum_{i\in[n]}\sum_{k=1}^{T}\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}, \mu_*^{i,\theta_k}) + \frac{\beta_0 L_{f,1}^2 T}{|I_k|}.\end{aligned} \tag{32}$$

Proof of Theorem 4.5. The average over gradient norms can be bounded as

$$\begin{aligned}&\frac{1}{T+1}\sum_{k=0}^{T}\mathbb{E}\big[\|\nabla F(\theta_k)\|_2^2\big] \\ \leq&\frac{2\mathbb{E}[F(\theta_0) - F(\theta_*)]}{\eta\tau T} + \frac{1}{T}\sum_{k=0}^{T}\mathbb{E}\big[\|\nabla F(\theta_k) - r_{k+1}\|_2^2\big] - \frac{1}{2T}\sum_{k=0}^{T}\mathbb{E}[\|r_{k+1}\|_2^2] \\ \leq&\frac{2\mathbb{E}[F(\theta_0) - F(\theta_*)]}{\eta\tau T} + \frac{\mathbb{E}\big[\|\nabla F(\theta_0) - r_1\|_2^2\big]}{\beta_0 T} + \frac{4L_{f,2}^2\tau^2\eta^2}{\beta_0^2 T}\sum_{k=0}^{T}\mathbb{E}\big[\|r_k\|_2^2\big] \\ &\quad + \frac{2L_{f,1}^2}{nT}\sum_{i\in[n]}\sum_{k=1}^{T}\mathbb{E}\left[\mathcal{D}_{\mathrm{KL}}(\mu_k^{(i)}, \mu_*^{i,\theta_k})\right] + \frac{\beta_0 L_{f,2}^2}{|I_k|} - \frac{1}{2T}\sum_{k=0}^{T}\mathbb{E}[\|r_{k+1}\|_2^2],\end{aligned}$$

wjere the first inequality is based on Lemma 4.2 and we denote θ_* the optimal solution of $\min_\theta \ F(\theta)$, and the second inequality is by (32).

Based on Lemma 4.4, we further obtain that

$$\begin{aligned}
&\frac{1}{T+1}\sum_{k=0}^{T}\mathbb{E}\big[\|\nabla F(\theta_k)\|_2^2\big] \\
\leq&\frac{2\mathbb{E}[F(\theta_0)-F(\theta_*)]}{\eta\tau T}+\frac{\mathbb{E}\big[\|\nabla F(\theta_0)-r_1\|_2^2\big]}{\beta_0 T}+\frac{4L_{f,2}^2\tau^2\eta^2}{\beta_0^2 T}\sum_{k=0}^{T}\mathbb{E}\big[\|r_k\|_2^2\big]+2L_{f,2}^2\Bigg[\\
&\quad\frac{4}{\alpha\tau|I_k|T}\sum_{i=1}^{n}\mathcal{D}_{\mathrm{KL}}(\mu_0^{(i)}\|\mu_*^{i,\theta_0})+\frac{12\eta L_{f,1}^2 n}{\lambda\epsilon\alpha|I_k|}+\frac{32\tau\epsilon dL_{G,2}^2}{\alpha}+\frac{20\eta n}{\lambda\epsilon\alpha|I_k|T}\sum_{t=0}^{T-1}\mathbb{E}\big[\|r_{k+1}\|_2^2\big]\Bigg] \\
&+\frac{\beta_0 L_{f,2}^2}{|I_k|}-\frac{1}{2T}\sum_{k=0}^{T}\mathbb{E}[\|r_{k+1}\|_2^2] \\
=&\frac{1}{T}\left\{\frac{2\mathbb{E}[F(\theta_0)-F(\theta_*)]}{\eta\tau}+\frac{\mathbb{E}\|\nabla F(\theta_0)-z_1\|_2^2}{\beta_0}+\frac{8L_{f,2}^2}{\alpha\tau|I_k|}\sum_{i=1}^{n}\mathcal{D}_{\mathrm{KL}}(\mu_0^{(i)}\|\mu_*^{i,\theta_0})\right\} \\
&+\frac{\sum_{k=0}^{T}\mathbb{E}\big[\|r_k\|_2^2\big]}{T}\left[\frac{40\eta L_{f,2}^2 n}{\lambda\epsilon\alpha|I_k|}+\frac{4L_{f,2}^2\tau^2\eta^2}{\beta_0^2}-\frac{1}{2}\right] \\
&+\frac{\beta_0 L_{f,2}^2}{|I_k|}+\frac{64\tau\epsilon dL_{G,2}^2L_{f,2}^2}{\alpha}+\frac{24\eta L_{f,1}^2L_{f,2}^2 n}{\lambda\epsilon\alpha|I_k|}.
\end{aligned}$$

By setting

$$\beta_0\leq\frac{\varrho^2|I_k|}{6L_{f,2}^2},\quad \tau\leq\frac{\varrho^2\alpha}{384\epsilon dL_{G,2}^2L_{f,2}^2},\quad \eta\leq\frac{\varrho^2\lambda\epsilon\alpha|I_k|}{144L_{f,1}^2L_{f,2}^2 n}, \tag{33}$$

it holds that

$$\frac{\beta_0 L_{f,2}^2}{|I_k|}+\frac{64\tau\epsilon dL_{G,2}^2L_{f,2}^2}{\alpha}+\frac{24\eta L_{f,1}^2L_{f,2}^2 n}{\lambda\epsilon\alpha|I_k|}\leq\frac{\varrho^2}{6}+\frac{\varrho^2}{6}+\frac{\varrho^2}{6}\leq\frac{\varrho^2}{2}.$$

By setting the number of iterations

$$T\geq\max\left(\frac{12\mathbb{E}[F(\theta_0)-F(\theta_*)]}{\eta\tau\varrho^2},\frac{6\mathbb{E}\|\nabla F(\theta_0)-z_1\|_2^2}{\beta_0\varrho^2},\frac{48L_{f,2}^2}{\alpha\tau|I_k|\varrho^2}\sum_{i=1}^{n}\mathcal{D}_{\mathrm{KL}}(\mu_0^{(i)}\|\mu_*^{i,\theta_0})\right), \tag{34}$$

it holds that

$$\begin{aligned}
&\frac{1}{T}\left\{\frac{2\mathbb{E}[F(\theta_0)-F(\theta_*)]}{\eta\tau}+\frac{\mathbb{E}\|\nabla F(\theta_0)-z_1\|_2^2}{\beta_0}+\frac{8L_{f,2}^2}{\alpha\tau|I_k|}\sum_{i=1}^{n}\mathcal{D}_{\mathrm{KL}}(\mu_0^{(i)}\|\mu_*^{i,\theta_0})\right\} \\
\leq&\frac{\varrho^2}{6}+\frac{\varrho^2}{6}+\frac{\varrho^2}{6}\leq\frac{\varrho^2}{2}.
\end{aligned}$$

By setting

$$\eta\leq\frac{\lambda\epsilon\alpha|I_k|}{160nL_{f,2}^2},\quad \eta\tau\leq\frac{\beta_0}{4L_{f,2}}, \tag{35}$$

where the latter condition is automatically satisfied for sufficiently small ϱ, it holds that

$$\frac{\sum_{k=0}^{T}\mathbb{E}\big[\|r_k\|_2^2\big]}{T}\left[\frac{40\eta L_{f,2}^2 n}{\lambda\epsilon\alpha|I_k|}+\frac{4L_{f,2}^2\tau^2\eta^2}{\beta_0^2}-\frac{1}{2}\right]\leq 0.$$

In summary, with proper choices of parameters as in (33), (34), and (35), it holds that

$$\frac{1}{T+1}\sum_{k=0}^{T}\mathbb{E}\|\nabla F(\theta_k)\|_2^2\leq\varrho^2,$$

and the output of Algorithm 3 finds ϱ-stationary point in T iterations. □

D Additional Results

Lemma D.1. *Let us define $\rho_0 := \mu_k^{(i)}$ with $x_0 \sim \rho_0$, and*

$$\mathrm{d}z_t = -\nabla_2 G_i(\theta_k, z_0)\,\mathrm{d}t + \sqrt{2\epsilon}\,\mathrm{d}\mathbf{B}_t, \quad \mathrm{Law}(z_t) = \rho_t.$$

Then

$$\mathbb{E}_{\rho_{0,t}}\|z_t - z_0\|_2^2 \le \frac{4t^2 L_{G,2}^2}{\alpha}\mathcal{D}_{\mathrm{KL}}(\rho_0\|\mu_*^{i,\theta_k}) + 2t^2 d L_{G,2}\epsilon + 2t\epsilon d.$$

Proof of Lemma D.1. It is noteworthy that z_t has the same distribution as

$$z_0 - t\nabla_2 G_i(\theta_k, z_0) + \sqrt{2t\epsilon}\zeta_0,$$

where $\zeta_0 \sim \mathcal{N}(0, \mathbf{I}_d)$ is an independent random vector. Therefore,

$$\begin{aligned}
&\mathbb{E}_{\rho_{0,t}}\|z_t - z_0\|_2^2 = \mathbb{E}_{\rho_{0,t}}\|t\nabla_2 G_i(\theta_k, z_0) - \sqrt{2t\epsilon}\zeta_0\|_2^2\\
\le& t^2\mathbb{E}_{\rho_0}\|\nabla_2 G_i(\theta_k, z_0)\|_2^2 + 2t\epsilon d\\
\le& t^2\left[\frac{4L_{G,2}^2}{\alpha}\mathcal{D}_{\mathrm{KL}}(\rho_0\|\mu_*^{i,\theta_k}) + 2dL_{G,2}\epsilon\right] + 2t\epsilon d\\
\le& \frac{4t^2 L_{G,2}^2}{\alpha}\mathcal{D}_{\mathrm{KL}}(\rho_0\|\mu_*^{i,\theta_k}) + 2t^2 d L_{G,2}\epsilon + 2t\epsilon d,
\end{aligned}$$

where the second inequality is by Lemma D.2 and $L_{G,2} = 1 + L_{f,2}/\lambda$. □

Lemma D.2 (Lemma 12 in [65])**.** *Suppose ν satisfies α-LSI and $\nu[z] \propto e^{-f(z)}$, with $f(z)$ being L-smooth in $z \in \mathbb{R}^d$. For any distribution ρ, it holds that*

$$\mathbb{E}_{z\sim\rho}[\|\nabla f(z)\|_2^2] \le \frac{4L^2}{\alpha}\mathcal{D}_{\mathrm{KL}}(\rho\|\nu) + 2dL.$$

Lemma D.3 (Grönwall's inequality)**.** *For any given function $u : [0, T) \to \mathbb{R}$ with $T \in (0, \infty]$, of class $\mathcal{C}^1$ satisfying the differential inequality*

$$u' \le au$$

for some $a \in \mathbb{R}$, it also satisfies the pointwise estimate

$$u(t) \le e^{at}u(0), \quad \forall t \in [0, T).$$