Journal of Machine Learning Research 27 (2026) 1-58 Submitted 3/25; Revised 9/26; Published 10/26

Decision-Dependent Stochastic Optimization: The Role of Distribution Dynamics

Zhiyu He ZHIYHE@ETHZ.CH
Automatic Control Laboratory, ETH Zürich
8092 Zürich, Switzerland

Saverio Bolognani BSAVERIO@ETHZ.CH
Automatic Control Laboratory, ETH Zürich
8092 Zürich, Switzerland

Florian Dörfler DORFLER@ETHZ.CH
Automatic Control Laboratory, ETH Zürich
8092 Zürich, Switzerland

Michael Muehlebach MICHAEL.MUEHLEBACH@TUEBINGEN.MPG.DE
Max Planck Institute for Intelligent Systems
72076 Tübingen, Germany

Editor: Sebastian Stich

Abstract

Distribution shifts have long been regarded as troublesome external forces that a decision-maker should either counteract or conform to. An intriguing feedback phenomenon termed *decision dependence* arises when the deployed decision affects the environment and alters the data-generating distribution. In the realm of performative prediction, this is encoded by distribution maps parameterized by decisions due to strategic behaviors. In contrast, we formalize an endogenous distribution shift as a feedback process featuring nonlinear dynamics that couple the evolving distribution with the decision. Stochastic optimization in this dynamic regime provides fertile ground to examine the various roles played by dynamics in the composite problem structure. To this end, we develop an online decision-making algorithm that both adapts to and shapes the dynamic distribution. Throughout the paper, we adopt a distributional perspective and demonstrate how this view facilitates characterizations of distribution dynamics and the optimality and generalization performance of the proposed algorithm. We showcase the theoretical results in an opinion dynamics context, where an opportunistic party maximizes the affinity of a population with evolving, polarized opinions, and in a recommender system scenario, featuring performance optimization with discrete distributions in the probability simplex.

Keywords: Stochastic optimization, distribution shift, dynamics, gradient method, feedback loop.

1. Introduction

Modern decision-making problems in machine learning, operations research, and control often feature intrinsic randomness, streaming data, and large scales. At the heart of such decision-making pipelines is stochastic optimization, which incorporates random objective functions, constraints, and algorithms with random iterative updates (Shapiro et al., 2021). Stochastic optimization often leverages knowledge, estimates, or samples of data distributions to disentangle the complex cou-

©2026 He, Bolognani, Dörfler, Muehlebach.

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v27/25-0514.html.

pling between randomness and decisions, achieve fast processing and adaptation, and navigate vast search spaces to arrive at optimal solutions.

Classical stochastic optimization assumes that the random variables in a problem obey some fixed distributions. In practice, however, distribution shifts are inevitable and can be both *exogenous* and *endogenous*. Exogenous distribution shifts are largely due to changing environmental conditions, e.g., parameter shifts in online estimation or an arbitrary new distribution selected by an adversary. In this regard, online stochastic optimization emphasizes adaptation by sequentially drawing new samples and adjusting decisions (Jiang et al., 2025; Cao et al., 2021).

Endogenous distribution shifts acknowledge the influence of a decision-maker on the data-generating distribution. This influence, i.e., *decision dependence*, contributes to a *closed loop*, whereby the decision and the data distribution interrelate in a repeated decision-making scenario. Various issues cause endogenous shifts and lead to different problem formulations, such as a reinforcement learning agent interacting with its environment, a dominant decision-maker being a price maker in a market, or a content recommender shaping user preferences, among others. In two-stage stochastic programming, models of how first-stage decisions alter distributions of random quantities in the second stage are discussed in Hellemo et al. (2018). Performative prediction (Perdomo et al., 2020; Hardt and Mendler-Dünner, 2025) tackles optimization involving distributions in the form of a general map parameterized by decisions. These parameterized distributions are inspired by strategic behaviors, where individuals intentionally modify features as a response to the deployed predictive model. A predominant assumption made a priori is that distribution shifts are Lipschitz, namely, a bounded change of decisions brings about a bounded change of the resulting distributions.

The aforementioned works largely capture decision dependence through (static) parameterized maps. In contrast, we address decision-making under endogenous distribution shifts represented by a broad class of nonlinear dynamics. This explicit formulation of distribution dynamics is motivated by the interactive feedback loop between a decision-maker and an evolving distribution, exemplified by practical problems in recommender systems (Dean et al., 2025; Lanzetti et al., 2023; Chandrasekaran et al., 2026) and opinion dynamics (Proskurnikov and Tempo, 2017). Each individual random variable (representing a feature, a preference, or intrinsic uncertainty) follows latent dynamics coupling the historical value with the current decision. Through distribution dynamics, the decision affects the individual variable and, thus, the overall distribution. This type of decision dependence features a nonlinear mixture of sequential decisions and non-stationary distributions, which renders the associated stochastic optimization problem challenging and fundamentally different from performative prediction. Our formulation is also closely aligned with the mean-field setting (Laurière et al., 2022). However, major differences exist in terms of the performance measure, the structure of decisions, and the specifications of dynamics and distributions; see Section 1.3 for further discussions.

By exploiting the structure of distribution dynamics, we will provide a fine-grained analysis of distribution shifts, design iterative stochastic algorithms tailored to this dynamic setting, and establish guarantees of optimality and generalization for this decision-dependent stochastic problem. Our design and analysis benefit from a distributional perspective, building connections between stochastic optimization, nonlinear control, and metric probability spaces.

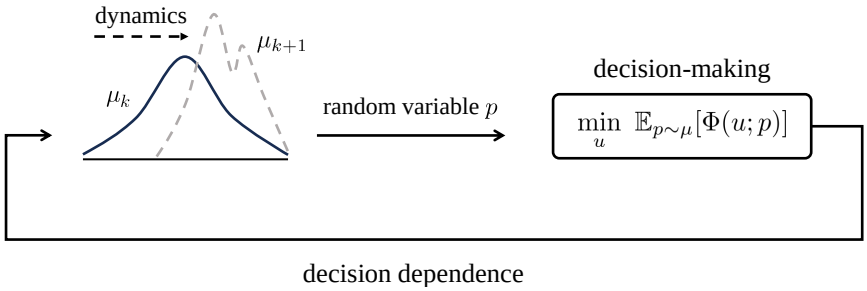


Figure 1: Stochastic optimization with decision dependence features a closed loop, involving endogenous distribution shifts from μ_k to μ_{k+1} due to the decision u and the dynamics. Here Φ is an objective function of the decision u and the random variable p; see problem (6) in Section 2.3 for a formal account.

1.1 Motivations

We present a motivating example of optimal decision-making under endogenous, dynamic distribution shifts in the domain of opinion dynamics.

Example 1 (Ideology tailored to a polarized population)**.** *Consider the interaction between a political party (or a candidate) and a large population. The party is opportunistic, with the aim of gaining power by picking an ideology that aligns with the majority and grants it the most votes.*

Let the adopted ideology of the party and the position (or the preference state) of a random individual in the population at time k be denoted by $q_k \in \mathbb{R}^m$ and $p_k \in \mathbb{R}^m$, respectively. Each coordinate of q_k or p_k indicates a liberal or conservative opinion on a political issue, e.g., taxes, health care, or immigration. The evolution of p_k driven by q_k is

$$p_k = g(p_{k-1}, q_k, p_0), \qquad p_0 \sim \mu_d, \quad k \in \mathbb{N}_+, \tag{1}$$

where $g : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ specifies the opinion dynamics, and μ_d is the distribution of the initial state. For instance, the classical Friedkin–Johnsen model (Proskurnikov and Tempo, 2017) reads $p_k = \Lambda_1 W p_{k-1} + \Lambda_2 q_k + (I - \Lambda_1 - \Lambda_2)p_0$, where $I \in \mathbb{R}^{m\times m}$ is the identity matrix, and $\Lambda_1, \Lambda_2, W \in \mathbb{R}^{m\times m}$ are weight matrices. The nonlinear polarized model (Hązła et al., 2024; Gaitonde et al., 2021) is $p_k \propto \lambda p_{k-1} + (1-\lambda)p_0 + \sigma(p_{k-1}^\top q_k)q_k$, and p_k is always normalized, i.e., $\|p_k\| = 1, \forall k$. At every time k, the party collects samples from the population and opportunistically adjusts the ideology q_k. To maximize the votes at an upcoming election, the steady-state population-wide affinity that the party intends to maximize is $\mathbb{E}_{p_{\rm ss}}[p_{\rm ss}^\top q]$, where $p_{\rm ss}$ satisfies the fixed-point equation $p_{\rm ss} = g(p_{\rm ss}, q, p_0)$ (Yang et al., 2020; Dean and Morgenstern, 2022). Since $p_{\rm ss}$ is rarely available, the party uses the current opinion distribution of p_k, estimated by polls (i.e., sampling), for decision-making.

Due to the dynamics (1) *and the chosen ideology q_k, the distribution followed by p_k changes constantly. This distribution shift in turn affects the decision-making based on p_k and causes a feedback loop; see Fig. 1. The classical paradigm of repeated sampling and retraining helps to adapt to this distribution shift. Nonetheless, this paradigm can suffer from sub-optimality because the dependence of the steady-state position $p_{\rm ss}$ on the decision q is ignored. In Section 5, we will review this motivating example and provide a detailed formulation, analysis, and numerical results.*

Motivated by this example, we explore general stochastic optimization with endogenous distribution shifts arising from the interaction between a decision-maker and a dynamic distribution; see Fig. 1 for an illustration. This setup involves several major challenges. First, the largely unknown distribution dynamics exclude an offline strategy based on the exact (re)formulation of the decision-making problem. Second, while the paradigm of repeated sampling and retraining in performative prediction facilitates adaptation (see Section 1.3), achieving optimality beyond performative stability requires anticipating how the decision affects the distribution and applying proactive adjustments. Finally, the dynamic setup restricts us from sampling from the steady-state distribution corresponding to the decision, which is the *very* distribution we care about while evaluating the overall performance. In this paper, we will address these challenges by developing and analyzing an online stochastic algorithm tailored to this dynamic setting.

1.2 Contributions

We are motivated by applications where a decision-maker drives a constantly evolving distribution and aims to optimize the distribution-level performance. We formulate a decision-dependent stochastic problem featuring endogenous distribution shifts with latent dynamics. For this general and dynamic setting, we adopt a distributional perspective at the intersection of stochastic optimization, nonlinear control, and metric probability spaces and present the following contributions.

- We characterize the distribution shift via the contracting coefficient of distribution dynamics and the change of decisions. To this end, we use as the main metric the Wasserstein distance (Villani, 2009) between the current distribution and its steady-state distribution induced by the decision. This metric plays a similar (albeit not necessarily the same) role as a Lyapunov function in control theory (Khalil, 2002). Although the exact value of such a Wasserstein distance can be elusive, the recursive inequality related to this metric sheds light on the dynamic evolution of the distribution.

- We propose an online stochastic algorithm that leverages samples from the current distribution and takes into account the composite structure of the problem due to dynamics. The iterative update direction consists of two terms. One term focuses on adaptation, and the most recent samples are exploited to adjust decisions. The other term actively shapes future distributions by anticipating the sensitivity of the distribution with respect to the decision, thereby informing optimal decision-making. Our algorithm leverages partial, structural knowledge of the world model governing distribution dynamics. This key information is the so-called sensitivity, an analytical component that can be learned from data under suitable conditions.

- We establish optimality guarantees of the proposed algorithm in a nonconvex setting. In the face of dynamics and the composite structure, the convergence measure (i.e., the expected second moment of gradients) enjoys a favorable $\mathcal{O}(1/\sqrt{T})$ rate, where T is the total number of iterations. This rate is as sharp as that of stochastic gradient descent for static nonconvex problems without decision dependence. Further, we provide high-probability convergence guarantees for a single run of the proposed algorithm. The key insight is to synthesize the coupled evolution of the aforementioned Wasserstein metric and the convergence measure, which correspond to the dynamic distribution and the iterative algorithm, respectively.

- We quantify the finite-sample generalization performance. We explore how the decisions obtained based on an empirical distribution with finite samples will generalize to the original distribution-level problem. We demonstrate that the generalization measure scales polynomially with the number of samples and the number of iterations, as well as poly-logarithmically with the inverse of the failure probability. These results are built on the measure concentration argument and the characterization of the distribution shift, illustrating the benefits of our distributional perspective.
- We illustrate the aforementioned results with practical examples, where an opportunistic party maximizes the affinity of a polarized dynamic population, and a recommender optimizes performance by interacting with a user. In the first setup, the population is modeled as a continuous distribution, whereas in the second example, the user is represented by a discrete distribution evolving in the probability simplex. We demonstrate that respecting the composite problem structure due to decision dependence is crucial for achieving fast convergence and improved optimality.

1.3 Related work

A multitude of works investigate the roles of distribution shifts and decision dependence in machine learning, optimization, and control. We provide a concise review of their setups and focus areas.

In applications, data-generating distributions can change because of various factors, e.g., nonstationary environments, unknown covariate shifts, and adversarial effects. There are two predominant strategies for addressing distribution shifts. One strategy pursues robustness against potential perturbations to distributions. This falls under the umbrella of distributionally robust optimization (Duchi and Namkoong, 2021; Kuhn et al., 2025) and is achieved by optimizing the worst-case cost over a so-called ambiguity set, i.e., the family of distributions that are close to the true distribution under certain metrics. Further, in two-stage stochastic programming, the first-stage decision may change the uncertain distribution in the second stage, thereby producing a decision-dependent ambiguity set (Hellemo et al., 2018). Tractable reformulations are derived in Luo and Mehrotra (2020); Basciftci et al. (2021) to disentangle this dependence and obtain robust solutions.

In the face of constantly evolving distributions, a less conservative and more active strategy is to seek adaptation. Specifically, online stochastic optimization investigates an iterative loop whereby a decision-maker commits to a decision, receives samples from a dynamic distribution as feedback (Jiang et al., 2025; Cao et al., 2021), and then generates a new decision. Such an online framework is versatile and particularly well suited to tackling exogenous distribution shifts, which are outside the influence of a decision-maker. Nonetheless, in terms of optimization subject to endogenous distribution shifts caused by decisions (see also the beginning of Section 1), unique phenomena (e.g., the existence of equilibria and stability issues) arise, requiring tailored methods for this closed-loop setting. We delineate some representative works as follows.

Performative prediction studies optimization under decision-dependent distributions (Perdomo et al., 2020; Hardt and Mendler-Dünner, 2025). A canonical example is strategic classification, where an individual deliberately modifies her features in reaction to the deployed classifier, thereby gaining a more favorable classification result. Such endogenous distribution shifts are usually formalized as static distribution maps parameterized by decisions. This setup leads to a novel equilibrium notion termed performative stability, meaning the decision optimizes the objective given the specific distribution it induces. Through repeated sampling and retraining, various stochastic al-

gorithms converge to performatively stable points (Mendler-Dünner et al., 2020; Drusvyatskiy and Xiao, 2023). A stronger solution concept is performative optimality, requiring that the decision and the induced distribution together lead to an optimal objective value. Convergence to performatively optimal points is achieved with additional structural assumptions on the distribution map, e.g., it belongs to a linear location-scale family (Miller et al., 2021; Jagadeesan et al., 2022; Narang et al., 2023; Yan and Cao, 2023) or an exponential family (Izzo et al., 2021). As a general, elegant, and tractable framework, performative prediction admits numerous extensions, including network scenarios with cooperative (Li et al., 2022) or competing agents (Narang et al., 2023; Piliouras and Yu, 2023), time-varying objectives (Cutler et al., 2023; Wood et al., 2022), saddle point minimax problems (Wood and Dall'Anese, 2023), and coupling constraints (Yan and Cao, 2023).

Our work is closely related to *stateful performative prediction*, which captures historical dependence and considers an evolving distribution that gradually settles at the stationary distribution map. Models of historical dependence include geometrically ergodic Markov chains (Li and Wai, 2022) and geometric decay responses in the form of a linear mixture (Brown et al., 2022; Ray et al., 2022). With Lipschitz distribution shifts assumed in the first place, the algorithms therein converge to performatively stable (albeit not necessarily optimal) points. In contrast, we formalize a decision-dependent distribution shift featuring a broad class of nonlinear dynamics mixing continuous decisions and state distributions. Moreover, we explicitly characterize the distribution shift through the Wasserstein metric (Villani, 2009) and link this shift with dynamics parameters and the change of decisions. We further establish convergence to first-order stationary solutions given a nonconvex objective function involving the decision-dependent structure.

Along the line of *performative decision-making*, some recent works study the role of dynamics in the decision-dependent problem setup. Performative reinforcement learning (Mandal et al., 2023) handles transition probability and reward functions relying on the deployed policy and finds a stable policy given cumulative rewards. Performative control in Cai et al. (2024) addresses linear dynamics with policy-dependent state transition matrices and seeks a performatively stable control solution as a linear combination of states and disturbances. The framework of Conger et al. (2023) represents the dynamics of a strategic population via a gradient flow in the Wasserstein space. The interconnection of this strategic population and a decision-maker results in coupled partial differential equations, which admit asymptotic convergence to optimal solutions for convex (or concave) energy functionals. In contrast, we characterize the distribution shift represented by general nonlinear dynamics. Furthermore, we offer insights into anticipating the sensitivity of the distribution shift, taking into account the composite structure due to dynamics in the algorithmic design, and achieving decision-making in the context of nonconvex objectives. The incorporation of anticipating sensitivity to actively shape distributions is the distinguishing feature of our algorithm compared to the aforementioned performative methods.

Mean-field formulation abstracts the mutual influence in a vast homogeneous population by the interaction between a representative individual and an average density, i.e., the mean field (Caines, 2021). This abstraction facilitates approximately solving an otherwise intractable multi-agent problem, wherein the joint states and actions may grow exponentially with the number of agents. The solutions to mean-field games or control are characterized by two coupled equations, namely, a forward equation for the evolving mean-field distribution and a backward equation associated with the individual value function. Classical approaches rely on the full knowledge of population dynamics and work under restrictive assumptions (Laurière et al., 2022). A recent trend is to apply reinforcement learning to learn models of dynamics (Huang et al., 2024), value functions, or policy

functions (Cui and Koeppl, 2021), thereby obtaining Nash equilibrium policies or socially optimal policies. Different from the mean-field formulation, we examine a simplified case where individuals do not interact with one another. Nevertheless, we provide the following new insights. First, we require less information or learning effort related to distribution dynamics. Rather than constructing well-calibrated models, we only access steady-state sensitivity matrices corresponding to dynamics, which are learnable in appropriate settings; see Section 2.2. Second, we explicitly characterize the distribution shift via dynamics parameters and the change of decisions. Finally, instead of analyzing a cumulative cost and a state-feedback policy, we seek optimal steady-state performance and a general decision vector. For the generic setting with nonconvex objectives and nonlinear dynamics, we quantify the first-order stationarity of the obtained solutions.

Control in probability spaces addresses the formulation where the state of a system is a probability measure instead of a Euclidean vector (Chen et al., 2021). This formulation facilitates the characterization of the evolving uncertainties in a system or the collective behavior of a population, both of which are intrinsically modeled via probability distributions (Terpin et al., 2024; Lanzetti et al., 2023). A typical example is distribution steering, i.e., driving the state distribution from an initial density to a target density in finite time with minimum energy control (Chen et al., 2021). In this regard, tractable control strategies cross-fertilize insights from control theory (e.g., linear state-feedback structures) and optimal transport (e.g., transport map calculations, Villani, 2009). In contrast, we search for an optimal decision vector rather than a state feedback control policy. Further, we do not aim for a specific final distribution. Instead, we hope the decision, together with the distribution induced through the nonlinear dynamics, will lead to an optimal steady-state behavior.

In a broader context, our work aligns with feedback optimization (Simonetto et al., 2020; Hauswirth et al., 2024) that implements optimization iterations as a feedback controller, thereby regulating the steady-state behavior of a dynamical system. Nonetheless, in this paper we are concerned with distribution-level characterizations in a metric space of probability measures, which is drastically different from the system-theoretic analysis of feedback optimization in Euclidean space.

In summary, we capture the distribution shift arising from the dynamic evolution of a distribution driven by a decision-maker. Such dynamics are represented by a broad class of nonlinear equations encompassing continuous states and decisions. We propose and characterize an online stochastic algorithm that respects the composite problem structure due to dynamics, regulates the distribution flow in real time, and yields first-order stationary solutions to the overall nonconvex problem featuring decision dependence.

The remainder of this paper is structured as follows. Section 2 introduces preliminaries of metric spaces of probability measures and formulates the stochastic optimization problem involving dynamic decision-dependent distributions. Section 3 presents the intuitions behind and the design of our online stochastic algorithm. In Section 4, we characterize the distribution dynamics and establish guarantees on the optimality and generalization performance of the proposed algorithm. Section 5 showcases an application in affinity maximization with a polarized, dynamic population following Example 1, as well as another case study in a recommender system context involving discrete distributions. Finally, Section 6 concludes this paper and discusses future directions. All proofs are provided in the appendix.

2. Preliminaries and Problem Formulation

2.1 Metric space of probability measures

We review the background of metric probability spaces and refer readers to Villani (2009) for more details. Let $\mathcal{P}(\mathbb{R}^m)$ be the space of Borel probability distributions on $\mathbb{R}^m$. Let $\mathcal{P}_1(\mathbb{R}^m) \triangleq \left\{\mu \in \mathcal{P}(\mathbb{R}^m) : \int_{\mathbb{R}^m} \|x\| \,\mathrm{d}\mu(x) < \infty\right\}$ be the space of distributions with finite first moments. The Dirac mass at point $x \in \mathbb{R}^m$ is denoted by δ_x, i.e., for any Borel set $A \subseteq \mathbb{R}^m$, $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ otherwise. We use $X \sim \mu$ to indicate that a random variable X is distributed according to μ. The convolution of two distributions $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^m)$ is denoted by $\mu * \nu$. Specifically, if two random variables X and Y are independent and distributed according to μ and ν, respectively, then $X + Y \sim \mu * \nu$. The pushforward of a distribution μ via a Borel map $f : \mathbb{R}^m \to \mathbb{R}^r$ is represented by $f_{\#}\mu$, where $(f_{\#}\mu)[A] \triangleq \mu[f^{-1}(A)]$ for every Borel set $A \subseteq \mathbb{R}^r$. In fact, if $X \sim \mu$, then $f(X) \sim f_{\#}\mu$. The identity map is Id.

Let $\|z\|_P = \sqrt{z^\top P z}$ denote the weighted norm of a vector $z \in \mathbb{R}^m$, where $P \in \mathbb{R}^{m\times m}$ is positive definite. Consider a metric space $(\mathbb{R}^m, c)$ endowed with a norm-induced metric $c : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}_{\geq 0}$. Typical examples of c include the Euclidean distance $c(x, y) = \|x - y\|$, the weighted distance $c(x, y) = \|x - y\|_P$, and other distances defined by composite norms, where $x, y \in \mathbb{R}^m$. The type-1 Wasserstein distance $W_1(\mu, \nu)$ between two distributions $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^m)$ on $(\mathbb{R}^m, c)$ is

$$W_1(\mu, \nu) = \inf_{\gamma\in\Gamma(\mu,\nu)} \int_{\mathbb{R}^m\times\mathbb{R}^m} c(x, y) \,\mathrm{d}\gamma(x, y), \tag{2}$$

where $\Gamma(\mu, \nu)$ is the set of all joint distributions (i.e., couplings) with marginals μ and ν; see Villani (2009, Definition 6.1). Intuitively, the Wasserstein distance is the minimum cost of transporting μ onto ν, where the cost of moving a unit mass from x to y is $c(x, y)$, and the available transport plan is represented by γ. The Wasserstein distance is a flexible and quantitative measure of the discrepancy between distributions, particularly when they have disjoint supports, such as when one distribution is continuous and the other is discrete.

2.2 Distribution dynamics

We generalize the motivating case study in Example 1 and consider the following distribution dynamics with continuous states

$$p_k = f(p_{k-1}, u_k, d), \qquad p_0 \sim \mu_0,\ d \sim \mu_d,\ (p_0, d) \sim \alpha, \quad k \in \mathbb{N}_+. \tag{3}$$

In (3), $p_k \in \mathbb{R}^m$ is a random state at time k distributed according to μ_k, i.e., $p_k \sim \mu_k$. The initial state p_0 follows the distribution $\mu_0 \in \mathcal{P}_1(\mathbb{R}^m)$. Further, $u \in \mathbb{R}^n$ is a decision (or an input) that influences each random state, and $d \in \mathbb{R}^r$ following the distribution $\mu_d \in \mathcal{P}_1(\mathbb{R}^r)$ is an exogenous input that remains constant during iterations. For instance, d can be a bias term (similar to the initial position in Example 1), a disturbance, or a random parameter in a model of f. Each pair (p_0, d) is independently drawn from the joint distribution α, and the first and the second marginals of α are μ_0 and μ_d, respectively. For instance, if p_0 and d as well as μ_0 and μ_d are the same (cf. Example 1), then $\alpha = (\mathrm{Id}, \mathrm{Id})_{\#}\mu_0$; if p_0 and d are independent, then α is the product measure $\mu_0 \times \mu_d$.

The distribution dynamics (3) feature decision dependence, in that the evolution of the distribution μ_k is driven by the decision u_k. This distribution will in turn determine the optimal decision for an optimal distribution-level behavior. Before we present the formal problem description, we

specify some properties related to the dynamics (3). All these properties serve the purpose of characterization and analysis. Our online algorithm exploits limited information on the model f of the distribution dynamics; it leverages samples and certain learnable *sensitivities* related to (3).

Assumption 1. *The function $f(p,u,d)$ is continuously differentiable, L_f^p-Lipschitz continuous in p with respect to the weighted norm $\|\cdot\|_P$, where $P \in \mathbb{R}^{m\times m}$ is positive definite, L_f^u-Lipschitz continuous in u with respect to $\|\cdot\|$, and L_f^d-Lipschitz continuous in d with respect to $\|\cdot\|$. Here, $L_f^p \in (0,1)$, $L_f^u > 0$, and $L_f^d > 0$. There exists a continuously differentiable steady-state map $h : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^m$ such that $h(u,d) = f(h(u,d),u,d)$. Furthermore, $\nabla_u h(u,d)$ is M_h^d-Lipschitz in d with respect to $\|\cdot\|$.*

Assumption 1 implies that the dynamics (3) are contracting in p with respect to the weighted norm $\|\cdot\|_P$. The existence and conditioning of P (i.e., the so-called contraction metric) are known for incrementally exponentially stable nonlinear dynamics (Tsukamoto et al., 2021). Similar contraction and Lipschitz assumptions appear in Jafarpour et al. (2021, Theorem 3) on perturbed fixed-point equations and Davydov et al. (2025, Lemma 1 and Theorem 2) on parameter-varying dynamics, in the contexts of implicit neural networks and time-varying optimization, respectively. Based on the Lipschitz conditions of f, the Banach contraction theorem ensures that for a fixed input u (i.e., $u_k = u, \forall k \in \mathbb{N}_+$) and a specific exogenous input d, the dynamics (3) admit a unique steady state $p_{\mathrm{ss}} = h(u,d)$ satisfying $p_{\mathrm{ss}} = f(p_{\mathrm{ss}},u,d)$. We can further establish that h is L_h^u-Lipschitz in u and L_h^d-Lipschitz in d with respect to $\|\cdot\|$, where $L_h^u = L_f^u\sqrt{\lambda_{\max}(P)/\lambda_{\min}(P)}/(1-L_f^p)$ and $L_h^d = L_f^d\sqrt{\lambda_{\max}(P)/\lambda_{\min}(P)}/(1-L_f^p)$; see the parametric contraction mapping principle (Dontchev and Rockafellar, 2009, Theorem 1A.4) and also Lemma 13 in Appendix A. The Lipschitz continuity condition on $\nabla_u h(u,d)$ (i.e., the so-called sensitivity matrix) holds when f has bounded Hessians. As we will see in Example 2 below, stable linear dynamics naturally satisfy Assumption 1. Although Assumption 1 enforces contraction and rules out complex behaviors including bifurcation or multiple attractors, we will present extended results that account for non-contracting polarized dynamics in Section 5.1 (albeit retaining a unique steady-state distribution so that the optimization objective is well-defined).

From a distribution-level perspective, all the steady-state samples p_{ss} follow the distribution $\mu_{\mathrm{ss}}(u)$ that depends on u and μ_d

$$\mu_{\mathrm{ss}}(u) = h(u,\cdot)_{\#}\mu_d, \tag{4}$$

where $h(u,\cdot)_{\#}\mu_d$ denotes the pushforward of the distribution μ_d via a Borel map $h(u,\cdot) : \mathbb{R}^r \to \mathbb{R}^m$ parameterized by the decision u. The steady-state distribution $\mu_{\mathrm{ss}}(u)$ is uniquely determined by u and μ_d, a specification similar to those in Brown et al. (2022); Ray et al. (2022); Li and Wai (2022). Let $\nabla_u h(u,d) \in \mathbb{R}^{n\times m}$ be the steady-state sensitivity matrix of $p_{\mathrm{ss}} = h(u,d)$ with respect to the decision u. It follows from the implicit function theorem (Dontchev and Rockafellar, 2009, Theorem 1B.1) that

$$\nabla_u h(u,d) = -\nabla_u f(p_{\mathrm{ss}},u,d)\left[\nabla_p f(p_{\mathrm{ss}},u,d) - I\right]^{-1}, \tag{5}$$

which holds in an open neighborhood of (u,d). The sensitivity $\nabla_u h(u,d)$ quantifies the rate of change of the steady-state sample p_{ss} with respect to the decision u. While (5) may seem daunting at first glance, the sensitivity can be simplified in various scenarios. For instance, when the dynamics (3) are linear, the corresponding sensitivity becomes a constant matrix; see Example 2 below. If

the exogenous input d is additive in (3), then the explicit formula of (5) no longer involves d. More broadly, apart from invoking (5) based on the related knowledge of dynamics and parameters, we can exploit recursive estimation or identification techniques to construct (approximate) sensitivities; see Hauswirth et al. (2024, Sec. 3.3.1) in the context of widely adopted feedback optimization methods. Such a learnable sensitivity is the only model information on (3) used in our online algorithm. All the characterizations in Assumption 1 are for the sake of analysis, and the full, high-fidelity world model (i.e., f) of the distribution dynamics is not required.

In the following example, we review an important special case of (3), where the dynamics function is linear. We will see how Assumption 1 is justified and provide explicit expressions for the distribution μ_k at time k and the steady-state distribution $\mu_{\mathrm{ss}}(u)$.

Example 2 (Linear distribution dynamics)**.** *Suppose that each sample evolves according to a linear dynamical equation $p_k = f(p_{k-1}, u_k, d) = Ap_{k-1} + Bu_k + Ed$, where $A \in \mathbb{R}^{m\times m}$ is non-zero, $B \in \mathbb{R}^{m\times n}$, $E \in \mathbb{R}^{m\times r}$, $p_0 \sim \mu_0, d \sim \mu_d$. Then, Assumption 1 is satisfied if A is Schur stable, i.e., $\rho(A) < 1$. Given a Schur stable matrix A and a positive definite $Q \in \mathbb{R}^{m\times m}$, there exists a unique positive definite matrix $P \in \mathbb{R}^{m\times m}$ satisfying the Lyapunov equation $A^\top PA - P + Q = 0$ (Khalil, 2002). Let $\lambda_{\min}(Q) > 0$ and $\lambda_{\max}(P) > 0$ denote the minimum eigenvalue of Q and the maximum eigenvalue of P, respectively. With the metric $\|\cdot\|_P$, for any $p, \bar{p} \in \mathbb{R}^m$, $u \in \mathbb{R}^n$, and $d \in \mathbb{R}^r$,*

$$\begin{aligned}\|f(p,u,d) - f(\bar{p},u,d)\|_P = \|A(p-\bar{p})\|_P &\overset{(a.1)}{=} \sqrt{(p-\bar{p})^\top (P-Q)(p-\bar{p})} \\ &\overset{(a.2)}{\leq} \underbrace{\sqrt{1 - \frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)}}}_{\in(0,1)} \|p-\bar{p}\|_P,\end{aligned}$$

where (a.1) follows from the aforementioned Lyapunov equation. In (a.2), we use

$$\|x\|_Q \geq \sqrt{\lambda_{\min}(Q)}\|x\| \geq \sqrt{\frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)}}\|x\|_P, \qquad \forall x \in \mathbb{R}^m,$$

see also Lemma 13 in Appendix A. Moreover, since $Q - P = -A^\top PA$ is negative semidefinite, we know $\lambda_{\min}(Q)I \preceq Q \preceq P \preceq \lambda_{\max}(P)I$. If $\lambda_{\min}(Q) = \lambda_{\max}(P)$, then $P = Q$, which implies $A = 0$ and contradicts the setting. Therefore, $\lambda_{\min}(Q)/\lambda_{\max}(P) \in (0,1)$. Further, for any $p \in \mathbb{R}^m$, $u, \bar{u} \in \mathbb{R}^n$, and $d \in \mathbb{R}^r$,

$$\|f(p,u,d) - f(p,\bar{u},d)\| = \|B(u-\bar{u})\| \leq \|B\|\,\|u-\bar{u}\|.$$

When the decision u is fixed (i.e., $u_k = u, \forall k \in \mathbb{N}_+$), the steady-state sample p_{ss} is $p_{\mathrm{ss}} = h(u,d) = (I-A)^{-1}(Bu + Ed)$, and the map h is continuously differentiable. The steady-state sensitivity matrix is $\nabla_u h(u,d) = \left[(I-A)^{-1}B\right]^\top$, which is constant and independent of u and d. In applications, this sensitivity (in engineering lingo called the zero-frequency gain) can be learned from data of decisions and samples (Hauswirth et al., 2024, Sec. 3.3.1). Hence, Assumption 1 is satisfied.

If p_0 and d are independent (i.e., the joint distribution of p_0 and d is the product measure $\mu_0 \times \mu_d$), similar to Aolaritei et al. (2025, Proposition 18), the transient distribution μ_k of (3) *is*

$$\mu_k = \left(A^k x\right)_{\#} \mu_0 * \left(\textstyle\sum_{i=0}^{k-1} A^i B u_{k-i} + \sum_{i=0}^{k-1} A^i E x\right)_{\#} \mu_d,$$

where $(A^k x)_{\#}\mu_0$ denotes the pushforward of μ_0 via the map $\bar{f}(x) = A^k x$, and a similar definition holds for the other term after convolution. The steady-state distribution μ_{ss} for a fixed decision u is

$$\mu_{\mathrm{ss}}(u) = \big((I - A)^{-1}Ex + (I - A)^{-1}Bu\big)_{\#}\,\mu_d.$$

2.3 Problem formulation

We aim to find a decision u that optimizes the steady-state behavior of the dynamic distribution (3):

$$\begin{aligned} \min_{u\in\mathbb{R}^n} \quad & \mathbb{E}_{p\sim\mu_{\mathrm{ss}}(u)}[\Phi(u,p)] \\ \text{s.t.} \quad & \mu_{\mathrm{ss}}(u) = h(u,\cdot)_{\#}\mu_d, \end{aligned} \tag{6}$$

where $\mu_{\mathrm{ss}}(u)$ is the steady-state distribution of (3) induced by u; see also (4). Let the reduced objective function of problem (6) be denoted by

$$\tilde{\Phi}(u) \triangleq \mathbb{E}_{p\sim\mu_{\mathrm{ss}}(u)}[\Phi(u,p)] = \mathbb{E}_{d\sim\mu_d}[\Phi(u,h(u,d))]. \tag{7}$$

The decision-dependent problem (6) formalizes the steady state of the closed loop illustrated by Fig. 1 in Section 1.1. This problem is relevant in many scenarios with a vast population or intrinsic uncertainties, e.g., voting and recommender systems.

Problem (6) involves several major challenges in terms of the nonconvex objective and the unknown dynamics underlying the decision-dependent distribution.

- First, the distribution dynamics (3) induce the steady-state distribution $\mu_{\mathrm{ss}}(u)$ and further bring about a composite structure in problem (6). Even for a convex function Φ, since the steady-state map $h(u,d)$ is nonlinear, the overall objective $\tilde{\Phi}(u)$ can be nonconvex.
- Another challenge originates from the unknown distribution dynamics, which render the structure of decision dependence elusive and preclude an offline numerical scheme based on the exact (re)formulation of problem (6).
- Finally, we cannot directly sample from the steady-state distribution $\mu_{\mathrm{ss}}(u)$ unless we wait sufficiently long, because the distribution μ_k is constantly changing with time k and eventually approaches the steady-state distribution when the decision is held fixed.

To overcome the above challenges, we will propose an online stochastic algorithm in Section 3 that samples from the current distribution μ_k and regulates the distribution shift by anticipating its sensitivity with respect to the decision.

We focus on steady-state performance due to its relevance, generality, and tractability. First, in many problems an optimal steady state matters more than transients, with the latter often not even being modeled; see, for instance, many case studies where feedback optimization or performative prediction is applied. This setup will also allow us to circumvent the need for model knowledge by exploiting distribution sensitivity, which is easier to learn. Second, if we analyze the behavior of a stable and dynamic distribution over a sufficiently long horizon, then the long-run average performance metric (i.e., the cumulative objective values divided by the number of iterations) converges to the steady-state objective under suitable regularity conditions. Finally, in this context we can go beyond the assumptions and policy classes considered in the mean-field literature (Laurière et al., 2022) and establish provable guarantees for continuous decision vectors given a broad class of nonlinear dynamics and nonconvex objective functions. Specifically, we make the following assumptions on the objective.

Assumption 2. *The objective $\tilde{\Phi}(u)$ is well defined (i.e., $\mathbb{E}_{d\sim\mu_d}[|\Phi(u,h(u,d))|]<\infty$), $L_{\tilde{\Phi}}$-smooth (with $L_{\tilde{\Phi}}$-Lipschitz gradients), and bounded below. The function $\Phi(u,p)$ is L_Φ^p-Lipschitz in p. The partial gradients $\nabla_u\Phi(u,p)$ and $\nabla_p\Phi(u,p)$ are M_Φ^u-Lipschitz and M_Φ^p-Lipschitz in p, respectively.*

Assumption 2 requires a well-defined expectation function, which is common in stochastic optimization (see Shapiro et al., 2021, Section 9.2.5) and can be satisfied, e.g., when for each u, $\Phi(u,h(u,d))$ is dominated by an integrable function of d. The smoothness condition is also standard (see Bottou et al., 2018) and holds, e.g., when for every stochastic sample $p_{\text{ss}} = h(u,d)$, the objective $\Phi(u,p_{\text{ss}}) = \Phi(u,h(u,d))$ is smooth and the second moment of the corresponding smoothness constant is bounded. The requirement that the partial gradients are Lipschitz in the random variable p is related to (though not the same as) the joint smoothness property used in performative prediction (Perdomo et al., 2020; Mendler-Dünner et al., 2020). Moreover, the condition that $\tilde{\Phi}(u)$ is bounded below implies that problem (6) admits a finite infimum $\tilde{\Phi}^*\in\mathbb{R}$.

Assumption 3. *There exists a positive random variable $L(d)$ such that $\mathbb{E}_{d\sim\mu_d}[L(d)]<\infty$, and for all $u_1,u_2\in\mathbb{R}^n$, $\big|\Phi(u_1,h(u_1,d))-\Phi(u_2,h(u_2,d))\big|\le L(d)\|u_1-u_2\|$.*

Assumption 3 is similar to the condition in Shapiro et al. (2021, Eq. (9.130)). It is a sufficient condition for the interchangeability of the expectation and gradient operators; see Shapiro et al. (2021, Theorem 9.56) and also (8) in Section 3.1 below. If the function $\Phi(u,h(u,d))$ is Lipschitz continuous in u, uniformly over d, then Assumption 3 holds naturally.

3. Online Stochastic Decision-Making

We present our online stochastic algorithm for solving the decision-dependent problem (6). The main challenges stem from the non-stationary and largely unknown distribution due to the dynamics (3) and the decisions $(u_k)_{k\in\mathbb{N}_+}$. To disentangle the complexity associated with distributions, we leverage samples drawn from the current transient distribution as informative characterizations and feedback. Motivated by the composite structure of the objective, we further enhance the algorithmic update with a term that proactively anticipates and shapes the dynamic distribution based on its sensitivity. Thus, we establish first-order stationarity guarantees for nonconvex problems.

3.1 Intuition behind the stochastic gradient

We provide the intuition for constructing appropriate stochastic gradients to solve problem (6). Ideally, we aim to obtain the gradient of the objective $\tilde{\Phi}$ at u_k, i.e.,

$$\begin{aligned}\nabla\tilde{\Phi}(u_k) =&\nabla\,\mathbb{E}_{p\sim\mu_{\text{ss}}(u_k)}[\Phi(u_k,p)]\\ =&\nabla\,\mathbb{E}_{d\sim\mu_d}[\Phi(u_k,h(u_k,d))]\\ \overset{\text{(a.1)}}{=}&\,\mathbb{E}_{d\sim\mu_d}[\nabla\Phi(u_k,h(u_k,d))]\\ \overset{\text{(a.2)}}{=}&\,\mathbb{E}_{d\sim\mu_d}\left[\nabla_u\Phi(u_k,h(u_k,d))+\nabla_u h(u_k,d)\nabla_p\Phi(u_k,p)|_{p=h(u_k,d)}\right]\\ \overset{\text{(a.3)}}{=}&\,\mathbb{E}_{(p,d)\sim\gamma_{\text{ss}}(u_k)}[\nabla_u\Phi(u_k,p)+\nabla_u h(u_k,d)\nabla_p\Phi(u_k,p)],\end{aligned}\tag{8}$$

where (a.1) leverages interchangeability of the expectation and gradient operators thanks to Assumption 3 (see also Shapiro et al., 2021, Theorem 9.56); (a.2) uses the law of the total derivative;

(a.3) involves expectation with respect to the joint distribution $\gamma_{\text{ss}}(u_k) \triangleq (h(u_k, \cdot), \text{Id})_{\#}\mu_d$ of the steady-state variable $p \sim \mu_{\text{ss}}(u_k) = h(u_k, \cdot)_{\#}\mu_d$ and the exogenous input $d \sim \mu_d$.

However, the exact gradient $\nabla\tilde{\Phi}(u_k)$ can be difficult to calculate for two reasons. First, it involves the expectation with respect to the random variable p, although in practice we can only access finite samples of the distribution. Second, since each sample needs a few iterations to approach its steady state, the steady-state distribution $\mu_{\text{ss}}(u_k)$ of p corresponding to the decision u_k is unavailable at the current time k.

To address these issues, we use samples drawn from the current distribution γ_k to construct a mini-batch stochastic gradient, thereby informing decision-making. The intuition is that γ_k, as reflected by these samples, serves as a reasonable proxy for the steady-state distribution $\gamma_{\text{ss}}(u_k)$, provided that neighboring decisions (i.e., u_{k-1} and u_k) are close and the iteration counter k is large. Consequently, the expected gradient involving γ_k also becomes a close approximation of the exact gradient $\nabla\tilde{\Phi}(u_k)$ involving $\gamma_{\text{ss}}(u_k)$. We will formalize this intuition in Section 4.1.

3.2 Algorithmic design

Guided by the aforementioned intuition, our online stochastic algorithm for solving problem (6) is

$$u_{k+1} = u_k - \eta\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k), \qquad k \in \mathbb{N}, \tag{9}$$

where $\eta > 0$ is a constant step size, k is the iteration counter, and $\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$ is a stochastic gradient based on mini-batches, i.e.,

$$\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k) \triangleq \frac{1}{n_{\text{mb}}}\sum_{i=1}^{n_{\text{mb}}}\Big(\underbrace{\nabla_u\Phi(u_k, p^i_k)}_{①} + \underbrace{\nabla_u h(u_k, d^{i_k})\nabla_p\Phi(u_k, p^i_k)}_{②}\Big). \tag{10}$$

In (10), $n_{\text{mb}} \in \mathbb{N}_+$ is the size of the mini-batch. At every time step k, we draw n_{mb} random samples $\{p^i_k\}_{i=1}^{n_{\text{mb}}}$ afresh from the transient distribution μ_k. The fresh sampling procedure is independent of the past randomness. As such, conditional on the past randomness, the sample pairs (p^i_k, d^{i_k}) are independent and identically distributed, where i_k denotes the index of the exogenous input corresponding to the sample p^i_k. These samples depend on the decisions owing to the dynamics (3), thereby causing a composite structure in the objective $\Phi(u, p)$. Further, $\nabla_u h(u_k, d^{i_k})$ is the steady-state sensitivity of p^i_{ss} with respect to the decision u_k. Such a sensitivity admits various simplifications and can often be learned from data in practice; see the discussion in Section 2.2. For instance, the sensitivity is a constant matrix given linear dynamics and does not explicitly involve d^{i_k} if the exogenous input is additive.

In essence, the mini-batch stochastic gradient (10) is a finite-sample approximation of

$$\widehat{\nabla}^k\tilde{\Phi}(u_k) \triangleq \mathbb{E}_{(p,d)\sim\gamma_k}[\nabla_u\Phi(u_k, p) + \nabla_u h(u_k, d)\nabla_p\Phi(u_k, p)], \tag{11}$$

i.e., the approximate expected gradient at u_k when (p, d) follows the joint distribution γ_k. Specifically, $\gamma_k = (f^{(k)}(x, y), y)_{\#}\alpha(x, y)$ denotes the pushforward of α in (3). Further, $f^{(k)}(p_0, d)$ is the value of p_k given a pair of the initial state p_0 and the exogenous input d sampled from the joint distribution α, a sequence of decisions $(u_t)_{t=1,\ldots,k}$, and the dynamics (3). That is, $f^{(k)} : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}^m$ is a map parameterized by $u_1, \ldots, u_k$. We define the special case $f^{(0)}(p_0, d)$ as p_0. We will show in Section 4 that with suitable algorithmic parameters, $\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$ and $\widehat{\nabla}^k\tilde{\Phi}(u_k)$ are close to $\nabla\tilde{\Phi}(u_k)$, enabling the stochastic algorithm (9) to yield desirable (stationary) solutions.

The mini-batch stochastic gradient $\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$ involves two terms. Term ① in (10) leverages the current sample p^i_k to construct a partial gradient with respect to u. Term ② in (10) anticipates how the decision u_k will influence the sample p (and in turn the objective) and uses this link for achieving optimality. As explained in Section 3.1, these two terms result from the aforementioned composite structure of $\Phi(u,p)$ and the law of the total derivative. While the update rule (9) follows naturally from our presentation, most related online methods do not use an anticipating term such as term ② in (10), as discussed in Section 1.3.

3.3 Properties of the stochastic gradient

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space. Let $\mathcal{F}_k$ be the σ-algebra generated by the random variables $\widehat{\nabla}^0_{\text{mb}}\tilde{\Phi}(u_0), \dots, \widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$. We define the trivial σ-algebra as $\mathcal{F}_{-1} \triangleq \{\varnothing, \Omega\}$. Let $u_0 \in \mathbb{R}^n$ be deterministic. Hence, for any $k \in \mathbb{N}$, the decision u_k is measurable with respect to $\mathcal{F}_{k-1}$. We make the following assumption on the variance of the stochastic gradient constructed from an individual sample. For a stochastic vector $\xi \in \mathbb{R}^n$, we define its variance as $\operatorname{Var}[\xi] \triangleq \mathbb{E}\big[\|\xi - \mathbb{E}[\xi]\|^2\big] = \mathbb{E}[\|\xi\|^2] - \|\mathbb{E}[\xi]\|^2$.

Assumption 4. *For any $k \in \mathbb{N}$ and any $i = 1, \dots, n_{\text{mb}}$, the stochastic gradient satisfies*

$$\operatorname{Var}\left[\nabla_u \Phi(u_k, p^i_k) + \nabla_u h(u_k, d^{i_k}) \nabla_p \Phi(u_k, p^i_k) \middle| \mathcal{F}_{k-1}\right] \leq M + M_V \|\widehat{\nabla}^k \tilde{\Phi}(u_k)\|^2,$$

where $M, M_V \geq 0$ are constants.

Assumption 4 is a standard and relatively weak statement that the variance of the stochastic gradient is restricted (Bottou et al., 2018). It implies that this variance can be nonzero at the point where $\widehat{\nabla}^k \tilde{\Phi}(u_k)$ equals zero and grows at most quadratically in the norm of $\widehat{\nabla}^k \tilde{\Phi}(u_k)$.

As clarified after (10), at time k we draw a mini-batch of random samples from the transient distribution. The fresh sampling process is independent of the σ-algebra $\mathcal{F}_{k-1}$. Hence, when conditioned on $\mathcal{F}_{k-1}$, $\{(p^i_k, d^{i_k})\}_{i=1}^{n_{\text{mb}}}$ are independent and identically distributed according to γ_k. Furthermore, the decision u_k is measurable with respect to $\mathcal{F}_{k-1}$. Building on this observation, we specify some key properties of the mini-batch gradient $\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$ in the lemma below.

Lemma 1. *Given the dynamics* (3)*, $\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$ in* (10) *is an unbiased estimate of $\widehat{\nabla}^k \tilde{\Phi}(u_k)$, i.e.,*

$$\mathbb{E}\left[\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k) \middle| \mathcal{F}_{k-1}\right] = \widehat{\nabla}^k \tilde{\Phi}(u_k), \quad \forall k \in \mathbb{N}. \tag{12}$$

If Assumption 4 holds, the expected second moment of $\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$ satisfies the following bound:

$$\mathbb{E}\left[\|\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)\|^2 \middle| \mathcal{F}_{k-1}\right] \leq \frac{M}{n_{\text{mb}}} + \left(\frac{M_V}{n_{\text{mb}}} + 1\right) \|\widehat{\nabla}^k \tilde{\Phi}(u_k)\|^2, \quad \forall k \in \mathbb{N}. \tag{13}$$

4. Performance Analysis

We analyze the interplay between the distribution dynamics (3) and the proposed online stochastic algorithm (9). In Section 4.1, we characterize the distribution shift driven by the decision-maker through the Wasserstein metric. We then establish the optimality guarantees of the algorithm (9) when applied to the distribution dynamics (3). These guarantees hold in expectation and with high probability and are given in Section 4.2 and Section 4.3, respectively, thereby covering a broad spectrum of the overall performance. Finally, we consider a finite-sample regime and provide generalization certificates in Section 4.4.

4.1 Distribution shifts

The distribution dynamics (3) bring about constant shifts of the joint distribution γ_k. We characterize such distribution shifts through the lens of the Wasserstein distance. Specifically, we focus on the behavior of $W_1(\gamma_k, \gamma_{\text{ss}}(u_k))$, i.e., the Wasserstein distance between the joint distribution γ_k at time k and the joint steady-state distribution $\gamma_{\text{ss}}(u_k)$ induced by the decision u_k.

Throughout this section, we consider the space $(\mathbb{R}^{m+r}, c)$ endowed with the metric

$$c\left((p,d),(p',d')\right) \triangleq \|p-p'\|_P + \|d-d'\|, \tag{14}$$

where $p, p' \in \mathbb{R}^m$, $d, d' \in \mathbb{R}^r$, and $\|\cdot\|_P$ is the weighted norm. Then, $W_1(\gamma_k, \gamma_{\text{ss}}(u_k))$ is defined by

$$W_1(\gamma_k, \gamma_{\text{ss}}(u_k)) = \inf_{\bar{\gamma}\in\Gamma(\gamma_k,\gamma_{\text{ss}}(u_k))} \int_{\mathbb{R}^{m+r}\times\mathbb{R}^{m+r}} c\left((p,d),(p',d')\right)\, \mathrm{d}\bar{\gamma}\left((p,d),(p',d')\right),$$

where (p,d) and (p',d') are distributed according to γ_k and $\gamma_{\text{ss}}(u_k)$, respectively.

The following lemma gives an upper bound on the Wasserstein distance $W_1(\gamma_k, \gamma_{\text{ss}}(u_k))$ and establishes a related recursive inequality. It is built on the properties of the pushforward operation induced by Borel maps; see Aolaritei et al. (2025, Proposition 3). Recall that L_f^p is the Lipschitz constant of the dynamics function f with respect to p under $\|\cdot\|_P$, and that L_h^u is the Lipschitz constant of the steady-state map h of (3) with respect to u under $\|\cdot\|$; see Assumption 1 and the discussion below. Further, $\lambda_{\max}(P) > 0$ is the maximum eigenvalue of the positive definite matrix P, and $f^{(k)}(p_0, d)$ denotes the value of p_k given a pair $(p_0, d) \sim \alpha$ and past decisions; see the paragraph below (11) in Section 3.2.

Lemma 2. *Let Assumption 1 hold. With* (3)*, the joint distributions* $(\gamma_k)_{k\in\mathbb{N}}$ *satisfy*

$$W_1(\gamma_k, \gamma_{\text{ss}}(u_k)) \le \int_{\mathbb{R}^m\times\mathbb{R}^r} \left\| f^{(k)}(p_0,d) - h(u_k,d)\right\|_P \mathrm{d}\alpha(p_0,d) \triangleq V_k, \quad \forall k\in\mathbb{N}, \tag{15a}$$

$$V_k \le L_f^p V_{k-1} + L_f^p L_h^u \sqrt{\lambda_{\max}(P)}\|u_k - u_{k-1}\|, \quad \forall k \in \mathbb{N}_+. \tag{15b}$$

Moreover, for any $u_0 \in \mathbb{R}^n$, $V_0 \triangleq \int_{\mathbb{R}^m\times\mathbb{R}^r} \|f^{(0)}(p_0,d) - h(u_0,d)\|_P \,\mathrm{d}\alpha(p_0,d)$ *is finite.*

Lemma 2 characterizes the behavior of $W_1(\gamma_k, \gamma_{\text{ss}}(u_k))$ through the evolution of the upper bound V_k. Specifically, this upper bound exhibits perturbed contraction (15b), where the contraction coefficient is $L_f^p \in (0,1)$, and the perturbation term is proportional to the difference of consecutive inputs. If the inputs are kept fixed (i.e., $u_k = u, \forall k \in \mathbb{N}$), then V_k and hence $W_1(\gamma_k, \gamma_{\text{ss}}(u))$ converge to 0 as k increases. The implication is that γ_k converges weakly in $\mathcal{P}_1(\mathbb{R}^{m+r})$ to the steady-state distribution $\gamma_{\text{ss}}(u)$; see Villani (2009, Theorem 6.9).

We quantify the cumulative sum of squared Wasserstein distances in the theorem below by building on Lemma 2. As explained in Section 3.1, the difference between γ_k and $\gamma_{\text{ss}}(u_k)$ causes a bias in the gradient at time k. Such a cumulative sum reflects how those biases accumulate when we deploy our online algorithm (9) and will be useful for quantifying the convergence measure of the nonconvex problem; see Theorem 5 in Section 4.2 and Theorem 8 in Section 4.3.

Theorem 3. *Given Assumption 1 and the distribution dynamics* (3)*, the joint distributions* $(\gamma_k)_{k\in\mathbb{N}}$ *satisfy*

$$\sum_{k=0}^{T-1} W_1(\gamma_k, \gamma_{\text{ss}}(u_k))^2 \le \frac{V_0^2}{1-\rho_1} + \frac{\rho_2}{1-\rho_1}\sum_{k=1}^{T}\|u_k - u_{k-1}\|^2, \tag{16}$$

where the coefficients are $\rho_1 = \frac{1+(L_f^p)^2}{2} \in (0,1)$ *and* $\rho_2 = \frac{1+(L_f^p)^2}{1-(L_f^p)^2}(L_f^p L_h^u)^2 \lambda_{\max}(P)$. *Furthermore, the online stochastic algorithm* (9) *applied to* (3) *ensures that*

$$\sum_{k=0}^{T-1} W_1(\gamma_k, \gamma_{\text{ss}}(u_k))^2 \leq \frac{V_0^2}{1-\rho_1} + \frac{\eta^2 \rho_2}{1-\rho_1} \sum_{k=0}^{T-1} \|\widehat{\nabla}_{\text{mb}}^k \tilde{\Phi}(u_k)\|^2. \tag{17}$$

Analogous to (15b), the upper bound on the sum of squared Wasserstein distances depends on the cumulative variation in decisions. After invoking the gradient-based update rule (9), we know that this cumulative variation is proportional to η^2 and the squared norm of the mini-batch gradient $\widehat{\nabla}_{\text{mb}}^k \tilde{\Phi}(u_k)$, where $k = 0, \dots, T-1$. In the following subsections, we will exploit (17) and the connection between $\widehat{\nabla}_{\text{mb}}^k \tilde{\Phi}(u_k)$ and $\nabla \tilde{\Phi}(u_k)$ to establish convergence guarantees.

An important implication of the distribution shift due to the dynamics (3) is that the mini-batch stochastic gradient (10), albeit unbiased with respect to the approximate gradient $\widehat{\nabla}^k \tilde{\Phi}(u_k)$ (see (11)), is a biased estimate of the true gradient $\nabla \tilde{\Phi}(u_k)$. The reason is that we draw samples from the current distribution γ_k instead of the steady-state distribution $\gamma_{\text{ss}}(u_k)$ to construct (10). Let e_k denote the difference between the expected gradients due to the discrepancy between γ_k and $\gamma_{\text{ss}}(u_k)$, that is,

$$\begin{aligned} e_k &\triangleq \widehat{\nabla}^k \tilde{\Phi}(u_k) - \nabla \tilde{\Phi}(u_k) \\ &= \mathbb{E}_{(p,d)\sim\gamma_k}[\nabla_u \Phi(u_k, p) + \nabla_u h(u_k, d) \nabla_p \Phi(u_k, p)] \\ &\quad - \mathbb{E}_{(p,d)\sim\gamma_{\text{ss}}(u_k)}[\nabla_u \Phi(u_k, p) + \nabla_u h(u_k, d) \nabla_p \Phi(u_k, p)]. \end{aligned} \tag{18}$$

The following lemma provides an upper bound on $\|e_k\|$ through $W_1(\gamma_k, \gamma_{\text{ss}}(u_k))$, i.e., the Wasserstein distance between two joint distributions γ_k and $\gamma_{\text{ss}}(u_k)$.

Lemma 4. *If Assumptions 1 to 3 are satisfied, then*

$$\|e_k\| = \|\widehat{\nabla}^k \tilde{\Phi}(u_k) - \nabla \tilde{\Phi}(u_k)\| \leq L W_1(\gamma_k, \gamma_{\text{ss}}(u_k)), \tag{19}$$

where $L = \max\left(L_\Phi^p M_h^d, \frac{M_\Phi^u + L_h^u M_\Phi^p}{\sqrt{\lambda_{\min}(P)}} \right)$.

Lemma 4 indicates that the difference e_k between the approximate gradient $\widehat{\nabla}^k \tilde{\Phi}(u_k)$ involving γ_k and the true gradient $\nabla \tilde{\Phi}(u_k)$ involving $\gamma_{\text{ss}}(u_k)$ admits an upper bound proportional to the Wasserstein distance $W_1(\gamma_k, \gamma_{\text{ss}}(u_k))$. By referring to Lemma 2 on the evolution of $W_1(\gamma_k, \gamma_{\text{ss}}(u_k))$, we know that when neighboring decisions (i.e., u_k and u_{k-1}) are close and the iteration counter is large, $\widehat{\nabla}^k \tilde{\Phi}(u_k)$ is a close approximation of $\nabla \tilde{\Phi}(u_k)$. This observation aligns with the intuition stated at the end of Section 3.1.

4.2 Optimality in expectation

We present the optimality guarantee of our stochastic algorithm (9) when applied to the distribution dynamics (3). The major challenge is that the evolution of the distribution and the iterates of the algorithm are coupled. To disentangle this coupling, we leverage the characterizations of the distribution shift in Theorem 3 and the descent-type iterate (9) and then quantify the overall convergence measure, i.e., the average expected second moment of gradients.

Recall that $\tilde{\Phi}(u) = \mathbb{E}_{p\sim\mu_{\text{ss}}(u)}[\Phi(u,p)]$ is the objective, $\tilde{\Phi}^*$ is the optimal value of problem (6), $\nabla\tilde{\Phi}(u)$ is given by (8), η is the step size, $T \in \mathbb{N}_+$ is the number of iterations, $n_{\text{mb}} \in \mathbb{N}_+$ is the size of the mini-batch, $L_{\tilde{\Phi}}$ is the smoothness constant of $\tilde{\Phi}(u)$, L is the constant specified in Lemma 4, M is the constant in the variance bound in Assumption 4, and ρ_1 and ρ_2 are constants given in Theorem 3. In the following theorem, we provide the first main convergence result of our stochastic algorithm (9).

Theorem 5. *Suppose that Assumptions 1 to 4 hold. Let η be chosen such that*

$$0 < \eta \leq \frac{1}{\sqrt{T}} \cdot \min\left(\frac{1}{4L_{\tilde{\Phi}}\left(\frac{M_V}{n_{\text{mb}}}+1\right)}, \sqrt{\frac{1-\rho_1}{14\rho_2 L^2\left(\frac{M_V}{n_{\text{mb}}}+1\right)}}\right). \tag{20}$$

The stochastic algorithm (9) *applied to the distribution dynamics* (3) *guarantees that*

$$\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] \leq \frac{8(\tilde{\Phi}(u_0)-\tilde{\Phi}^*)}{\eta T} + \frac{4L_{\tilde{\Phi}}M\eta}{n_{\text{mb}}} + \mathcal{O}\left(\frac{1}{T}\right). \tag{21}$$

By choosing η as the upper bound in (20)*, the order of the right-hand side of* (21) *is*

$$\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] = \mathcal{O}\left(\frac{1}{\sqrt{T}}\right). \tag{22}$$

The convergence measure analyzed in Theorem 5 is the average of the expected second moments of $\nabla\tilde{\Phi}(u_k)$. The expectation is taken with respect to the randomness of the proposed algorithm, i.e., the randomness in selecting mini-batch samples. This measure is typical in nonconvex optimization to characterize first-order stationarity of solutions, which is a necessary condition for local optimality (Bottou et al., 2018). The upper bound (21) on this measure involves the number of iterations T, the step size η, and other constants related to the problem and the algorithm. With a step size attaining the upper bound in (20), our stochastic algorithm (9) yields an $\mathcal{O}(1/\sqrt{T})$ rate of convergence, matching that of stochastic gradient descent for static nonconvex problems. It is common in the stochastic optimization literature (Ghadimi and Lan, 2013; Bottou et al., 2018) to let the bound on the step size be a function of the number of iterations, facilitating convergence rate analysis.

The rate (22), nonetheless, is nontrivial given the presence of unknown dynamics and the composite structure of the decision-dependent problem. Online stochastic algorithms reviewed in Section 1.3 that lack an anticipation term such as the one in (10) can incur persistent biases in gradients, causing sub-optimality (namely, the average expected second moment of gradients does not vanish). Although dynamics inhibit us from directly accessing steady-state samples and cause biases in gradients, we demonstrate that the accumulation of these biases does not worsen the overall convergence rate. This is largely due to the contracting distribution dynamics and a relatively slow algorithm (with a bounded step size), à la time-scale separation (Khalil, 2002) as quantified by (20). Finally, we remark that variance reduction techniques offer a promising means of improving the convergence rate (Bottou et al., 2018).

The expected gradient $\nabla\tilde{\Phi}(u_k)$, as part of the convergence measure in (21), involves the steady-state distribution $\gamma_{\text{ss}}(u_k)$. In practice, however, we can only sample from the current distribution γ_k to gain an understanding of the quality of solutions. In the following corollary, we analyze the average expected second moment of $\widehat{\nabla}^k\tilde{\Phi}(u_k)$ involving γ_k; see (11).

Corollary 6. *Let the conditions of Theorem 5 hold. The stochastic algorithm* (9) *acting on the distribution dynamics* (3) *ensures that*

$$\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\widehat{\nabla}^k\tilde{\Phi}(u_k)\|^2\right] \leq \frac{7}{3T}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] + \mathcal{O}\left(\frac{1}{T}\right) = \mathcal{O}\left(\frac{1}{\sqrt{T}}\right). \tag{23}$$

Corollary 6 implies that the average expected second moment of $\widehat{\nabla}^k\tilde{\Phi}(u_k)$ related to the current distribution γ_k is still of the order of $\mathcal{O}(1/\sqrt{T})$. The reason is that the convergence measures in (22) and (23) can be connected via the average expected squared Wasserstein distance, i.e., $\big(\sum_{k=0}^{T-1}\mathbb{E}\left[W_1(\gamma_k,\gamma_{\text{ss}}(u_k))^2\right]\big)/T$, whose order is the same as both measures; see also (58) in Appendix D.1.

Apart from the above optimality guarantees, we are also interested in the convergence of the distribution dynamics (3) while interacting with the decision-maker as per (9). We provide the convergence guarantee in the following theorem.

Theorem 7. *Let the conditions of Theorem 5 hold. Then,*

$$\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[W_1(\gamma_k,\gamma_{\text{ss}}(u_k))\right] \leq \sqrt{\frac{1}{6L^2}\left(\frac{8(\tilde{\Phi}(u_0)-\tilde{\Phi}^*)}{\eta T}+\frac{4L_{\tilde{\Phi}}M\eta}{n_{\text{mb}}}\right)}+\mathcal{O}\left(\frac{1}{\sqrt{T}}\right)=\mathcal{O}\left(\frac{1}{T^{\frac{1}{4}}}\right). \tag{24}$$

In Theorem 7, we characterize the convergence of the distribution dynamics (3) via the average expected Wasserstein distance between the current distribution γ_k and the corresponding steady-state distribution $\gamma_{\text{ss}}(u_k)$. Similar to (21), the upper bound in (24) also depends on the number of iterations, the step size, and other problem- and algorithm-specific constants. As the number of iterations grows, the average expected Wasserstein distance becomes closer to zero, implying that the dynamic distributions approach their steady-state distributions on average. This convergence plays an important role in optimality certificates that couple the distribution dynamics (3) and the algorithm (9); see the discussions below Theorems 3 and 5.

We remark that the Wasserstein metric for the distribution dynamics (3) in Theorem 7 is of the same order as the average expected gradient norm $\sum_{k=0}^{T-1}\mathbb{E}[\|\nabla\tilde{\Phi}(u_k)\|]/T$ rather than the average expected second moment of gradients in Theorem 5. Given the analogy between optimization and sampling, readers may wonder why the convergence rate (24) does not match the $\mathcal{O}(1/\sqrt{T})$ rate of Langevin dynamics for sampling from a target distribution (see, e.g., Chau et al., 2021). Apart from the distinction in problem setups, another major reason for this difference in convergence rates lies in the coupling of the distribution dynamics (3) and the algorithm (9). As characterized by (17), the incurred cumulative Wasserstein distance is related to the second moments of gradients. Since the nonconvex objective of problem (6) limits the achievable order of the average second moment (namely, $\mathcal{O}(1/\sqrt{T})$, see Theorem 5), the order of the average Wasserstein distance between distributions is also restricted.

4.3 Optimality with high probability

The convergence guarantee in Theorem 5 holds in expectation. This result reflects the average performance pertaining to the interplay between the stochastic algorithm (9) and the distribution dynamics (3). However, some cases (e.g., with constraints on computational resources or time) allow only a single trial or a few trials of the stochastic algorithm, rendering the occurrence of an

extreme outcome a key concern, particularly in high-stakes scenarios, such as elections discussed in Example 1. In this regard, another angle for characterizing algorithmic and distributional behaviors is convergence with high probability. Our goal is to quantify the confidence level that a stochastic algorithm yields satisfactory solutions when applied to distribution dynamics after a single trial involving many iterations.

To establish high-probability guarantees, we need an assumption that differs from Assumption 4 and is instead related to the distribution of the gradient noise. Recall from (12) in Lemma 1 that the mini-batch gradient $\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$ is an unbiased estimate of the approximate gradient $\widehat{\nabla}^k\tilde{\Phi}(u_k)$. Let

$$\xi_k = \widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k) - \widehat{\nabla}^k\tilde{\Phi}(u_k) \tag{25}$$

be the noise in $\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$ relative to $\widehat{\nabla}^k\tilde{\Phi}(u_k)$. Hence, $\mathbb{E}\left[\xi_k|\mathcal{F}_{k-1}\right] = 0$, where $\mathcal{F}_{k-1}$ is the σ-algebra generated by $\widehat{\nabla}^0_{\text{mb}}\tilde{\Phi}(u_0), \ldots, \widehat{\nabla}^{k-1}_{\text{mb}}\tilde{\Phi}(u_{k-1})$. Specifically, we assume that ξ_k is sub-Gaussian, i.e., with a tail dominated by the tail of a Gaussian distribution. This specification is formalized by the following assumption.

Assumption 5. *When conditioned on $\mathcal{F}_{k-1}$, the noise ξ_k (25) is sub-Gaussian, i.e., there exists $\sigma_{\text{mb}} > 0$ such that for any $u_k \in \mathbb{R}^n$,*

$$\mathbb{E}\left[\exp\left(\frac{\|\xi_k\|^2}{\sigma^2_{\text{mb}}}\right)\middle|\mathcal{F}_{k-1}\right] \leq \exp(1). \tag{26}$$

Compared to Assumption 4 on the variance, Assumption 5 specifies that the gradient noise ξ_k is light-tailed. Building on Jensen's inequality (i.e., $\forall X \in \mathbb{R}, \exp(\mathbb{E}[X]) \leq \mathbb{E}[\exp(X)]$), Assumption 5 implies that the variance of $\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$ admits a constant upper bound σ^2_{mb}, which is a bit stronger than Assumption 4. In practice, Assumption 5 can be satisfied when the noise in the sample gradient corresponding to each individual is sub-Gaussian. Such light-tailed sub-Gaussian noises are considered in Harvey et al. (2019); Li and Orabona (2020) to analyze the behaviors of stochastic gradient descent for static problems. Further, self-normalized concentration inequalities provide a promising avenue for handling heavy-tailed (e.g., sub-Weibull) noises; see Madden et al. (2024); Li and Liu (2022).

We establish the high-probability guarantee of our stochastic algorithm (9) when applied to the distribution dynamics (3). The involved variables, parameters, and constants are the same as those in Theorem 5. The proof leverages the characterization of the cumulative squared Wasserstein distances in Theorem 3 and some high-probability bounds on terms involving ξ_k.

Theorem 8. *Suppose that Assumptions 1 to 3 and 5 hold. Let η be chosen such that*

$$0 < \eta \leq \frac{1}{\sqrt{T}} \cdot \min\left(\frac{1}{6L_{\tilde{\Phi}}}, \sqrt{\frac{1-\rho_1}{21\rho_2 L^2}}\right). \tag{27}$$

The stochastic algorithm (9) *applied to the distribution dynamics* (3) *guarantees that for any fixed $\tau \in (0,1)$, with probability at least $1-\tau$,*

$$\frac{1}{T}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k)\|^2 \leq \frac{16(\tilde{\Phi}(u_0) - \tilde{\Phi}^*)}{\eta T} + 24\eta L_{\tilde{\Phi}}\sigma^2_{\text{mb}}\left(1 + \ln\frac{2}{\tau}\right) + \mathcal{O}\left(\frac{1}{T}\left(1 + \ln\frac{1}{\tau}\right)\right). \tag{28}$$

By selecting η as the upper bound in (27)*, the order of the right-hand side of* (28) *is*

$$\frac{1}{T}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k)\|^2=\mathcal{O}\bigg(\frac{1}{\sqrt{T}}\left(1+\ln\frac{1}{\tau}\right)\bigg).$$

The upper bound on the step size in (27) is of the same order as that of (20) in Theorem 5, although constant coefficients differ. In the stochastic optimization literature (Ghadimi and Lan, 2013; Bottou et al., 2018), this bound is typically set as a function of the number of iterations to characterize the convergence rate. The convergence measure in Theorem 8 is the average squared gradient norm. In contrast to Theorem 5, however, the measure here does not involve expectation anymore. The upper bound (28) scales inversely with the square root of the number of iterations T. Moreover, it features a logarithmic dependence on the inverse of the failure probability τ. This dependence is favorable particularly when the desired confidence level is high, i.e., when τ is small. For example, $\tau=10^{-3}$ leads to $\ln\frac{1}{\tau}\approx 6.9$, which translates to a moderate increase in magnitude. Hence, under Assumption 5 of sub-Gaussian noises, convergence results similar to those in Theorem 5 also hold with high probability.

Analogous to Theorem 7, we offer a high-probability characterization of the convergence of the distribution dynamics (3) when driven by a decision-maker as per (9).

Theorem 9. *Let the conditions of Theorem 8 hold. For any fixed $\tau\in(0,1)$, with probability at least $1-\tau$,*

$$\frac{1}{T}\sum_{k=0}^{T-1}W_1(\gamma_k,\gamma_{\mathrm{ss}}(u_k))\leq\sqrt{\tfrac{1}{6L^2}\left(\tfrac{16(\tilde{\Phi}(u_0)-\tilde{\Phi}^*)}{\eta T}+24\eta L_{\tilde{\Phi}}\sigma_{\mathrm{mb}}^2\left(1+\ln\tfrac{4}{\tau}\right)\right)}+\mathcal{O}\Big(\tfrac{1}{T^{\frac{1}{2}}}\sqrt{1+\ln\tfrac{1}{\tau}}\Big). \quad (29)$$

By choosing η as the upper bound in (27)*, the order of the right-hand side of* (29) *is*

$$\frac{1}{T}\sum_{k=0}^{T-1}W_1(\gamma_k,\gamma_{\mathrm{ss}}(u_k))=\mathcal{O}\left(\frac{1}{T^{\frac{1}{4}}}\sqrt{1+\ln\frac{1}{\tau}}\right).$$

Theorem 9 ensures that with high probability, the average Wasserstein distance between the current distribution and the associated steady-state distribution approaches zero at a rate of $\mathcal{O}(1/T^{\frac{1}{4}})$ as the number of iterations T increases. Moreover, the upper bound on this average distance exhibits a poly-logarithmic dependence on the inverse of the failure probability τ. Overall, the distributions γ_k and $\gamma_{\mathrm{ss}}(u_k)$ are close on average with high probability.

4.4 Generalization in a finite-sample regime

In this subsection, we examine a finite-sample regime, where access is limited to a specific set of samples rather than allowing unrestricted sampling from the full distribution as in problem (6). When applied in this context, the stochastic algorithm (9) yields decisions that essentially optimize an empirical objective associated with these samples. We are interested in the generalization guarantees on how such decisions perform in the problem involving the overall distribution.

This setup is of interest from an application perspective. Various reasons, e.g., the strategy of seeding trials or the restricted sampling due to privacy concerns, may cause us to only work with specific samples from the distribution. Yet, we hope the decisions obtained from analyzing these

finite samples can generalize to the distribution-level problem. We will show how our distributional perspective allows seamless integration of the measure concentration argument (Fournier and Guillin, 2015) into performance analysis, thus facilitating generalization guarantees.

The distributions of the initial state and the exogenous input are μ_0 and μ_d, respectively. Moreover, let $\{(p_0^1, d^1), \ldots, (p_0^N, d^N)\}$ be independent and identically distributed samples drawn from the joint distribution α, where $N \in \mathbb{N}_+$ is the total number of samples, and the marginals of α are μ_0 and μ_d. The empirical distributions of the initial state and the exogenous input are $\mu_0^N = \frac{1}{N}\sum_{i=1}^N \delta_{p_0^i}$ and $\mu_d^N = \frac{1}{N}\sum_{i=1}^N \delta_{d^i}$, respectively. At every time k, the decision affects the overall distribution μ_k as per (3) and in turn influences the resulting empirical distribution μ_k^N. In a finite-sample regime, the empirical distribution μ_k^N rather than the full distribution μ_k is accessible for sampling. For this setting, the online stochastic algorithm (9) generates decisions $\{u_0^N, \ldots, u_{T-1}^N\}$ for the following finite-sample empirical problem

$$\min_{u\in\mathbb{R}^n} \quad \frac{1}{N}\sum_{i=1}^N \Phi(u, p_{\text{ss}}^i) = \frac{1}{N}\sum_{i=1}^N \Phi(u, h(u, d^i)) \triangleq \tilde{\Phi}^N(u). \tag{30}$$

Our focus is on establishing generalization guarantees on the performance of the decisions $\{u_0^N, \ldots, u_{T-1}^N\}$ in the distribution-level problem (6).

Theorem 5 allows us to characterize the first-order stationarity of the decisions with respect to the empirical objective $\tilde{\Phi}^N$ in (30). In terms of the distribution-level performance, however, we center on stationarity with respect to $\tilde{\Phi}$ in (7), i.e., the reduced objective of problem (6). To this end, we leverage the following decomposition

$$\|\nabla\tilde{\Phi}(u_k^N)\|^2 \le 2\underbrace{\|\nabla\tilde{\Phi}(u_k^N) - \nabla\tilde{\Phi}^N(u_k^N)\|^2}_{①} + 2\underbrace{\|\nabla\tilde{\Phi}^N(u_k^N)\|^2}_{②}. \tag{31}$$

The left-hand side of (31) is the squared norm of the gradient $\nabla\tilde{\Phi}(u_k^N)$, which reflects the quality of decisions for the distribution-level problem (6). Term ① in (31) is the squared norm of the difference between the empirical gradient $\nabla\tilde{\Phi}^N$ and the gradient $\nabla\tilde{\Phi}$, and it represents the generalization error (Li and Liu, 2022). Term ② in (31) is the squared norm of the empirical gradient, constituting the optimization error for the empirical problem (30).

The following lemma gives a high-probability bound on the generalization error averaged over iterations. It is built on the measure concentration result derived in Wasserstein distances; see Fournier and Guillin (2015, Theorem 2). Recall that r is the size of the exogenous input d.

Lemma 10. *Let Assumptions 1 to 3 hold and $r \ge 3$. Suppose that there exist $\theta > 1$ and $\kappa > 0$ such that $\mathcal{E}_{\theta,\kappa}(\mu_d) \triangleq \int_{\mathbb{R}^r} e^{\kappa\|x\|^\theta}\,\mathrm{d}\mu_d(x)$ is finite. For any fixed $\tau \in (0,1)$, with probability at least $1-\frac{\tau}{2}$, the decisions $\{u_0^N, \ldots, u_{T-1}^N\}$ satisfy*

$$\frac{1}{T}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k^N) - \nabla\tilde{\Phi}^N(u_k^N)\|^2 \le$$
$$\begin{cases} L^2\left(L_h^d\sqrt{\lambda_{\max}(P)}+1\right)^2\left(\frac{1}{c_2 N}\ln\left(\frac{2c_1}{\tau}\right)\right)^{\frac{2}{r}}, & \text{if } N \ge \frac{1}{c_2}\ln\left(\frac{2c_1}{\tau}\right), \\ L^2\left(L_h^d\sqrt{\lambda_{\max}(P)}+1\right)^2\left(\frac{1}{c_2 N}\ln\left(\frac{2c_1}{\tau}\right)\right)^{\frac{2}{\theta}}, & \text{if } 1 \le N < \frac{1}{c_2}\ln\left(\frac{2c_1}{\tau}\right), \end{cases}$$

where c_1 and c_2 are positive constants that only depend on r, θ, κ, and $\mathcal{E}_{\theta,\kappa}(\mu_d)$.

The finite moment condition (namely, on $\mathcal{E}_{\theta,\kappa}(\mu_d)$) in Lemma 10 specifies the degree to which the distribution μ_d is light-tailed. This condition is common in the literature (Fournier and Guillin, 2015; Kuhn et al., 2025). Lemma 10 encompasses both the low-data and high-data regimes.

We now characterize the generalization performance of our stochastic algorithm (9). We focus on the decisions $\{u_0^N, \ldots, u_{T-1}^N\}$ obtained by the algorithm (9) in solving the finite-sample problem (30), with the mini-batch size n_{mb} satisfying $1 \leq n_{\text{mb}} \leq N$. The quality of these decisions is evaluated by the average squared gradient norm of the objective $\tilde{\Phi}$ involving the distribution μ_d and associated with problem (6). Note that r is the size of d, $\theta > 1$ is a constant present in $\mathcal{E}_{\theta,\kappa}(\mu_d)$, and c_1 and c_2 are constants specified in Lemma 10.

Theorem 11. *Let Assumptions 1 to 3 and 5 and the conditions of Lemma 10 hold, and assume $\tilde{\Phi}^N(u)$ is $L_{\tilde{\Phi}}^N$-smooth. Consider the use of the stochastic algorithm* (9) *to solve the finite-sample problem* (30)*, with the step size satisfying $0 < \eta \leq \frac{1}{\sqrt{T}} \cdot \min\left(\frac{1}{6L_{\tilde{\Phi}}^N}, \sqrt{\frac{1-\rho_1}{21\rho_2 L^2}}\right)$. Let $\tau \in (0,1)$ be fixed. When $N \geq \frac{1}{c_2}\ln\left(\frac{2c_1}{\tau}\right)$, with probability at least $1-\tau$,*

$$\frac{1}{T}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k^N)\|^2 \leq 2L^2\left(L_h^d\sqrt{\lambda_{\max}(P)}+1\right)^2\left(\frac{1}{c_2 N}\ln\left(\frac{2c_1}{\tau}\right)\right)^{\frac{2}{r}} + \frac{32(\tilde{\Phi}^N(u_0^N)-\tilde{\Phi}^{N,*})}{\eta T} + 48\eta L_{\tilde{\Phi}}^N\sigma_{\text{mb}}^2\left(1+\ln\frac{4}{\tau}\right) + \mathcal{O}\left(\frac{1}{T}\left(1+\ln\frac{1}{\tau}\right)\right).$$

When the step size η attains its upper bound, the overall order is

$$\frac{1}{T}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k^N)\|^2 = \mathcal{O}\left(\frac{1}{N^{\frac{2}{r}}}\left(\ln\frac{1}{\tau}\right)^{\frac{2}{r}}\right) + \mathcal{O}\left(\frac{1}{\sqrt{T}}\left(1+\ln\frac{1}{\tau}\right)\right).$$

When $1 \leq N < \frac{1}{c_2}\ln\left(\frac{2c_1}{\tau}\right)$, with probability at least $1-\tau$,

$$\frac{1}{T}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k^N)\|^2 \leq 2L^2\left(L_h^d\sqrt{\lambda_{\max}(P)}+1\right)^2\left(\frac{1}{c_2 N}\ln\left(\frac{2c_1}{\tau}\right)\right)^{\frac{2}{\theta}} + \underbrace{\frac{32(\tilde{\Phi}^N(u_0^N)-\tilde{\Phi}^{N,*})}{\eta T} + 48\eta L_{\tilde{\Phi}}^N\sigma_{\text{mb}}^2\left(1+\ln\frac{4}{\tau}\right)}_{\sim\mathcal{O}\left((1+\ln(1/\tau))/\sqrt{T}\right)} + \mathcal{O}\left(\frac{1}{T}\left(1+\ln\frac{1}{\tau}\right)\right).$$

Theorem 11 quantifies how the performance measure (i.e., the average squared gradient norm) scales polynomially with the number of samples N and the number of iterations T, as well as polylogarithmically with the inverse of the specified failure probability τ. As N and T increase, this performance measure approaches zero. It implies that the decisions $\{u_0^N, \ldots, u_{T-1}^N\}$ arising from the finite-sample problem (30) generalize to the distribution-level problem (6).

5. Case Studies

We present two case studies to demonstrate how decision dependence in stochastic optimization can be effectively addressed with our algorithm. The first case study delineates Example 1 in opinion

dynamics and involves continuous population distributions. The second example originates from a recommender system context and focuses on discrete distributions over the probability simplex. All experiments are conducted on an HTCondor cluster using CPU-only nodes, with 6 GB of memory allocated per job. Our code is publicly available[1].

5.1 Affinity maximization in a polarized population

We revisit Example 1 and consider the interaction of an opportunistic party and a polarized dynamic population. While the case study is streamlined, a relevant real-world example is the U.S. presidential election (Yang et al., 2020). The ideological position of a party causes a shift in the position of each individual in the population. This party focuses on picking an ideology to gain the most votes. A tractable proxy for this goal is to maximize the steady-state population-wide affinity.

The shift of the individual position is described by the following polarized dynamics model adapted from Hązła et al. (2024); Gaitonde et al. (2021)

$$\widetilde{p}_{k+1} = \lambda p_k + (1-\lambda)p_0 + \sigma \cdot (p_k^\top q)q, \qquad p_{k+1} = \frac{\widetilde{p}_{k+1}}{\|\widetilde{p}_{k+1}\|}, \qquad p_0 \sim \mu_0,\ \|p_0\| = 1,\ k \in \mathbb{N}, \tag{32}$$

where $p_k \in \mathbb{R}^m$ is the position of an individual at time k, $q \in \mathbb{R}^m$ is the ideological position of the party, $\lambda \in (0,1)$ is a coefficient for blending the current and initial positions, $\sigma > 0$ regulates the position evolution, and μ_0 is the distribution of initial positions. In (32), the term $(1-\lambda)p_0$ captures the persistent influence of the initial position p_0, which is also present in the classical linear Friedkin–Johnsen model (Proskurnikov and Tempo, 2017). The closer λ is to 1, the less an individual sticks to her initial position. Thanks to the normalization step, the position p_k is of unit norm. The coefficients λ and σ are the same for each individual, and therefore the overall population is homogeneous.

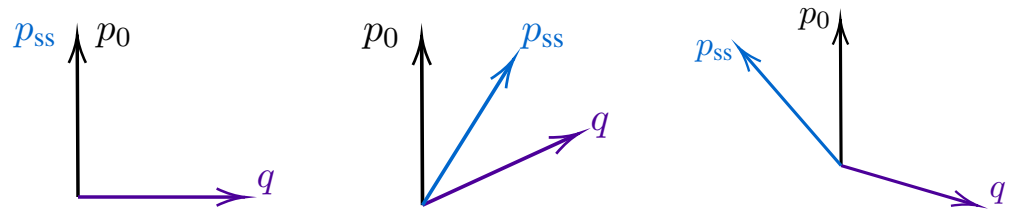
The polarized dynamics (32) feature biased assimilation (Dean and Morgenstern, 2022). If a specific individual prefers an ideology q (i.e., $p_k^\top q > 0$), then her position will move closer to q. Conversely, if the ideology is disliked (i.e., $p_k^\top q < 0$), then her position will instead move away from q, or more specifically, closer to $-q$. As illustrated in Fig. 2, the dynamics (32) admit a steady-state (i.e., limiting) position p_{ss}, and the location of p_{ss} relative to q depends on the angle between p_0 and q and the coefficients λ and σ. Hence, the dynamics (32) are not necessarily globally contracting. The affinity of an individual towards an ideology q at time k is given by the inner product $p_k^\top q$.

Given p_0 and a fixed decision q, the fixed-point equation corresponding to the dynamics (32) is

$$\widetilde{p}_{\text{fix}} = \lambda p_{\text{fix}} + (1-\lambda)p_0 + \sigma(p_{\text{fix}}^\top q)q, \qquad p_{\text{fix}} = \frac{\widetilde{p}_{\text{fix}}}{\|\widetilde{p}_{\text{fix}}\|}, \qquad p_{\text{fix}} \in \mathbb{R}^m, \tag{33}$$

which may admit multiple solutions depending on λ, σ, and $p_0^\top q$. While algebraically valid, some of these solutions neither arise as limits of the trajectory nor reflect the behaviors in an opinion dynamics context (Dean and Morgenstern, 2022). Our focus is on the *unique limit* $p_{\text{ss}} = h_{\text{ss}}(q, p_0)$ that is approached by the trajectory starting from p_0, evolving by (32), and driven by q. This limit is the final interaction outcome between an opportunistic party and an individual and affects affinity maximization. We formalize the properties of the polarized dynamics (32) in the following theorem.

1. https://github.com/zyhe/distribution-dynamics-opt



(a) Not influenced ($p_0^\top q = 0$) (b) Acute angle remains ($p_0^\top q > 0$) (c) Obtuse angle remains ($p_0^\top q < 0$)

Figure 2: The polarized dynamics (32) admit a unique limiting state depending on the sign of $p_0^\top q$.

Theorem 12. *For a fixed ideology q, the dynamics* (32) *starting at p_0 converge to a unique limit p_{ss}; this limit p_{ss} is a solution of* (33)*. As $\lambda \in (0,1)$ and $\sigma > 0$ increase, p_{ss} becomes closer to q (meaning $p_{ss}^\top q \geq p_0^\top q$) when $p_0^\top q > 0$, and to $-q$ (meaning $p_{ss}^\top q \leq p_0^\top q$) when $p_0^\top q < 0$, indicating a stronger steering ability. For (q, p_0) such that $p_0^\top q \neq 0$ and the relevant Jacobian is non-singular, the steady-state sensitivity $\nabla_q h_{ss}(q, p_0)$ in an open neighborhood of (q, p_0) is*

$$\nabla_q h_{ss}(q,p_0) = -\sigma \left((p_{ss}^\top q) I + p_{ss} q^\top \right) (I - p_{ss} p_{ss}^\top) \left[(\lambda I + \sigma q q^\top)(I - p_{ss} p_{ss}^\top) - \|\widetilde{p}_{ss}\| I \right]^{-1}, \quad (34)$$

where $\widetilde{p}_{ss} = \lambda p_{ss} + (1-\lambda) p_0 + \sigma (p_{ss}^\top q) q$.

More broadly, nonlinear opinion dynamics models are capable of capturing bifurcation or multiple polarized attractors through saturation bounds (Bizyaeva et al., 2023), hyperbolic tangent functions (Leonard et al., 2021), or weighted medians (Mei et al., 2024). In contrast, we adopt the geometric polarized model (32) to maintain tractability in performance optimization, while still permitting non-trivial polarized behaviors. When strong nonlinearities invalidate the unique steady-state condition and render the objective set-valued, the current guarantees no longer apply. However, similar *local* guarantees may be established by restricting the analysis to the regions of attraction associated with various equilibria. Another important direction is incorporating principled selection rules across multiple steady states.

To align with the majority and gain the most votes, the party aims to find an ideology q that optimizes the steady-state population-wide affinity:

$$\begin{aligned} \max_{q \in \mathbb{R}^m} \quad & \mathbb{E}_{p \sim \mu_{ss}(q)}[p^\top q] \\ \text{s.t.} \quad & \mu_{ss}(q) = h_{ss}(q, \cdot)_{\#} \mu_0, \\ & \|q\| \leq 1, \end{aligned} \quad (35)$$

where $\mu_{ss}(q)$ is the steady-state (i.e., limiting) distribution induced by q, and the randomness is due to the persistent influence of the stochastic initial position acting as an exogenous input d in (3). In problem (35), the norm constraint on q represents restrictions (e.g., of resources or social norms) and ensures that the optimal solution to (35) is bounded. Since the polarized dynamics (32) are not globally contracting, the results herein show behaviors beyond the presented theoretical setting.

As established in Dean and Morgenstern (2022, Proposition 1), when $\lambda = 1$ (i.e., the initial position disappears from (32)), for a constant ideology q, $p_{ss} = q/\|q\|$ if $p_0^\top q > 0$, and $p_{ss} = -q/\|q\|$ if $p_0^\top q < 0$. In this case, the goal of individual affinity maximization is trivial: the key is to

identify the hemisphere of p_0 and then stick to some q that ensures $p_0^\top q > 0$ (Dean and Morgenstern, 2022). Then, the maximum achievable affinity equals 1. However, for the general case that involves $\lambda \in [0, 1)$ and unknown p_0 and extends from the individual to the population, affinity maximization is not straightforward anymore, and a systematic means of searching for q is required.

Problem (35) entails the underlying distribution dynamics (32). In practice, the difficulty of accessing an accurate dynamics model and the need for real-time decision-making can render the steady-state distribution $\mu_{\mathrm{ss}}(q)$ elusive, thereby inhibiting offline numerical pipelines that rely on the formula or samples of $\mu_{\mathrm{ss}}(q)$. One online strategy to solve problem (35) is to apply a stochastic algorithm in the style of performative prediction (Hardt and Mendler-Dünner, 2025); see the discussions in Section 1.3. This algorithm exploits repeated sampling and retraining but is unaware of the composite structure wherein the steady-state distribution of positions depends on the ideology. The update rule of such an online vanilla algorithm is

$$q_{k+1} = \mathrm{Proj}_{\|q\| \le 1} \left(q_k + \frac{\eta}{n_{\mathrm{mb}}} \sum_{i=1}^{n_{\mathrm{mb}}} p_k^i \right), \tag{36}$$

where $\mathrm{Proj}_{\|q\| \le 1}(\cdot)$ denotes the projection onto the norm ball $\{q \in \mathbb{R}^m \,|\, \|q\| \le 1\}$, and $\{p_k^i\}$ are collected after applying q_k to the dynamics (32). This form of projected gradient ascent is due to maximization in (35). Furthermore, we implement the following online stateful algorithm building on the delayed repeated risk minimization in Brown et al. (2022)

$$q_{k+1} = \begin{cases} \mathrm{Proj}_{\|q\| \le 1} \left(q_k + \frac{\eta}{n_{\mathrm{mb}}} \sum_{i=1}^{n_{\mathrm{mb}}} p_k^i \right), & \text{if } k \equiv 0 \pmod{\bar{k}}, \\ q_k, & \text{otherwise}, \end{cases} \tag{37}$$

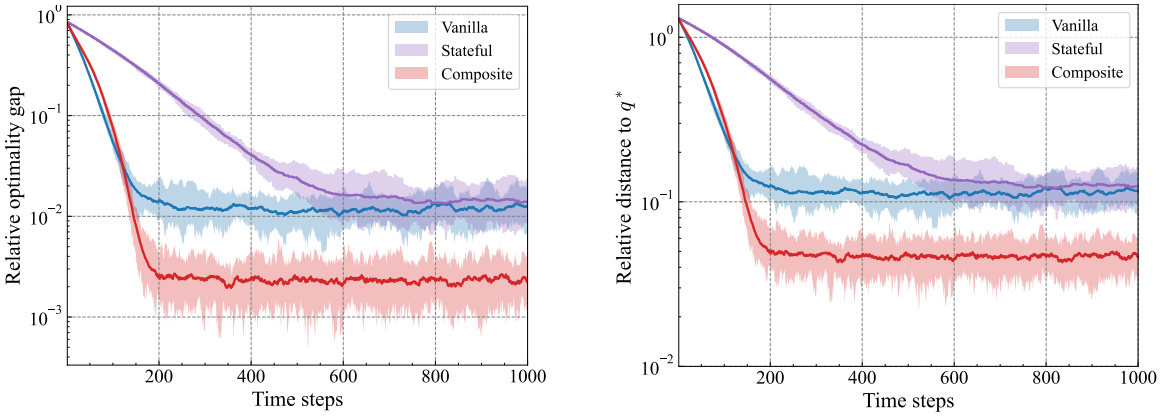
where $\bar{k} \in \mathbb{N}_+$ is the epoch length. The intuition behind (37) is to let the polarized dynamics (32) evolve for $\bar{k}$ steps before periodically updating the decision q. The stateful algorithm (37) with $\bar{k} = 1$ recovers the vanilla algorithm (36). Further, the above stateful algorithm forms the backbone of the state-dependent stochastic approximation algorithm for controlled Markov chains (Li and Wai, 2022) and motivates the epoch-based first-order algorithm in Ray et al. (2022) that exploits the linear location-scale family structure. For comparison, we leverage the proposed online stochastic algorithm (9) interacting with the dynamics (32) with the following adjustments. First, we add a projection step within iterations to satisfy the norm constraint $\|q\| \le 1$. Second, we use real-time samples to construct in an online manner the sensitivity $\widehat{H}(q_k, p_k)$ of the dynamics (32), namely,

$$\widehat{H}(q_k, p_k) = -\sigma \left(p_k^\top q_k I + p_k q_k^\top \right) (I - p_k p_k^\top) \left[(\lambda I + \sigma q_k q_k^\top)(I - p_k p_k^\top) - \|\widetilde{p}_k\| I \right]^{-1}, \tag{38}$$

where $\widetilde{p}_k = \lambda p_{k-1} + (1 - \lambda) p_0 + \sigma \cdot (p_{k-1}^\top q_k) q_k$. The difference between (38) and (34) lies in the replacement of p_{ss} with the current position p_k. In Appendix G.2, we characterize the effect of this replacement on the gradient difference (similar to Lemma 4) and show that the related theoretical guarantees hold analogously. Overall, our online stochastic algorithm (9) tailored to (35) reads

$$q_{k+1} = \mathrm{Proj}_{\|q\| \le 1} \left(q_k + \frac{\eta}{n_{\mathrm{mb}}} \sum_{i=1}^{n_{\mathrm{mb}}} \left(p_k^i + \widehat{H}(q_k, p_k^i) q_k \right) \right). \tag{39}$$

Here the ascent-based update corresponds to maximization in (35).



(a) Relative optimality gap

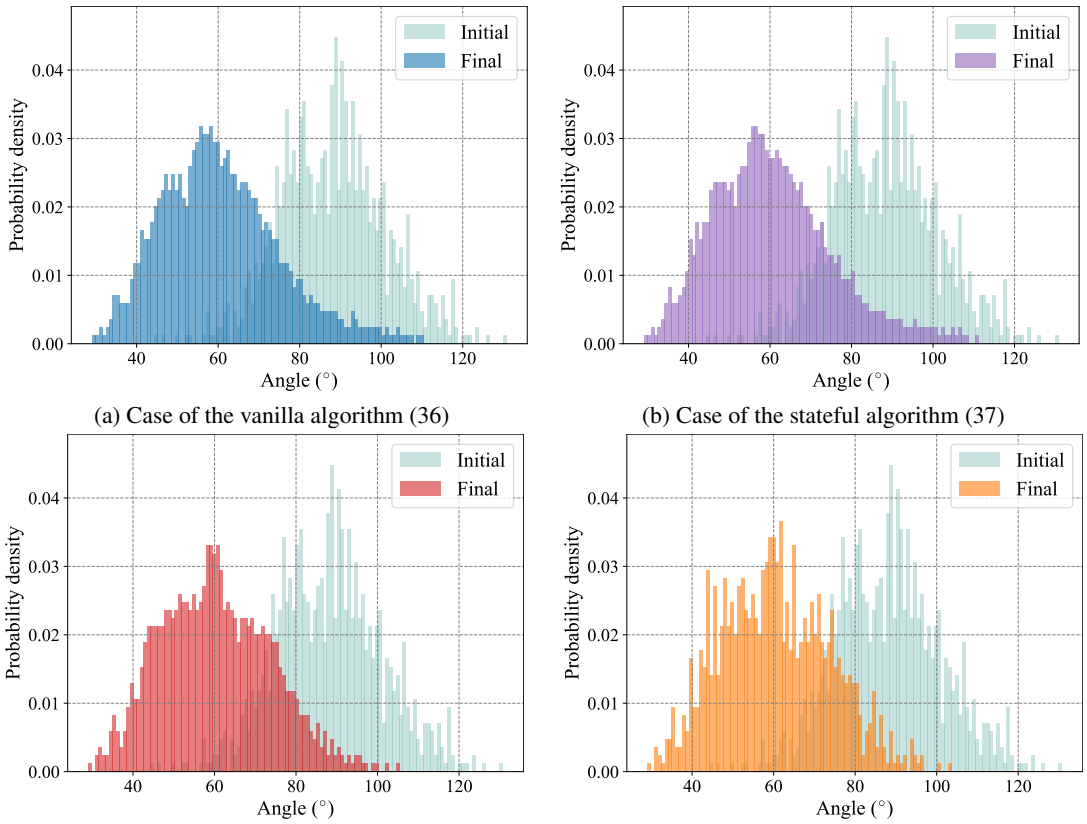
(b) Relative distance to the benchmark solution q^*

Figure 3: This figure illustrates the convergence behaviors of our online stochastic algorithm (39) (termed "composite"), the vanilla algorithm (36) (termed "vanilla"), and the stateful algorithm (37) (termed "stateful") for the normal-mixing scenario. The solid lines represent the average values of convergence measures, whereas the shaded areas indicate the ranges of variation across independent trials.

5.1.1 SCENARIO 1: NORMAL MIXING

We consider a normal-mixing scenario, where the coefficients in the distribution dynamics (32) are $\lambda = 0.4$ and $\sigma = 0.5$. We set p_k and q to be 20-dimensional vectors, i.e., $m = 20$. We generate a population of 1000 individuals and sample their initial positions from a unit hemisphere in the 20-dimensional space, i.e., the set of unit vectors that form an angle not more than 90° with a randomly generated reference vector. For the stochastic algorithms (36), (37), and (39), we tune the step sizes over the grid spanning $[5 \times 10^{-3}, 0.5]$. The selected step sizes that balance empirical convergence rates and accuracy are 0.125, 0.175, and 0.035, respectively. The mini-batch size adopted in these algorithms is $n_{\text{mb}} = 50$. We set the epoch length $\bar{k} = 5$ for the stateful algorithm (37). As a benchmark, we use the optimizer IPOPT (Wächter and Biegler, 2006) to calculate a (locally) optimal solution q^* and the corresponding optimal value related to problem (35). This optimizer starts from a random initial guess and accesses all the parameters and data across the population. In practice, however, such full access can be prohibitive, and our online algorithm (39) is more desirable. We run 20 independent trials of the stochastic algorithms (36), (37), and (39). Given the hardware setting at the beginning of Section 5, the total wall-clock runtime for all the algorithms is 6 minutes and 40 seconds.

Fig. 3 illustrates the convergence behaviors of different algorithms while interacting with the polarized dynamics (32) and solving problem (35). In Fig. 3, each time step corresponds to one iteration of the dynamics (32). The vanilla algorithm (36) and our composite algorithm (39) perform updates at every iteration, whereas the stateful algorithm (37) adjusts only once every $\bar{k} = 5$ time steps. Overall, our algorithm (39) achieves a lower optimality gap and attains a smaller distance to the benchmark solution compared with the vanilla (36) and the stateful (37) algorithms. The key reason stems from the algorithmic structure. The vanilla algorithm (36) is oblivious to the composite characteristic of problem (35) and solely focuses on adaptation. While the stateful algorithm (37)



(a) Case of the vanilla algorithm (36)

(b) Case of the stateful algorithm (37)

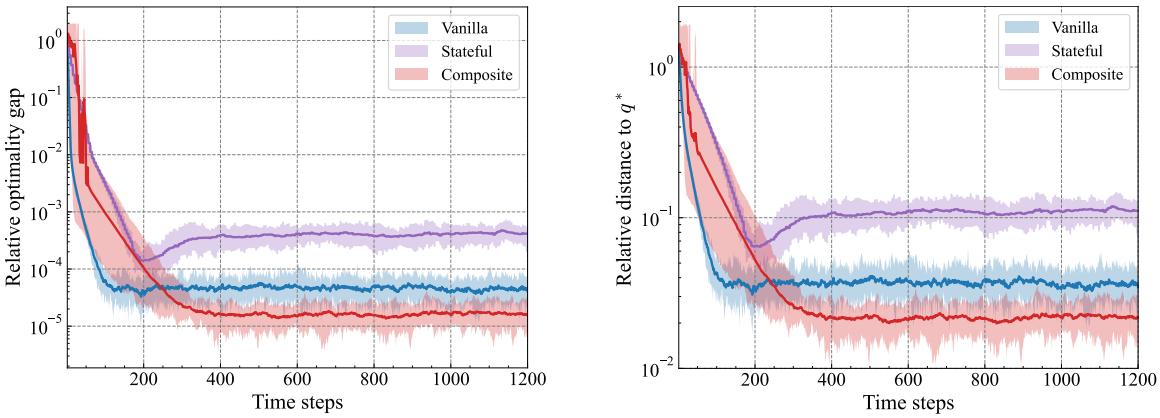
(c) Case of our composite algorithm (39)

(d) Case with the benchmark solution q^* via IPOPT

Figure 4: This figure illustrates the histograms of the angles between the position and the decision across the population in the normal-mixing scenario. Initially, the angles are largely concentrated in $[80°, 100°]$, implying the average population-wide affinity is close to zero. In the end (i.e., when the population reaches the steady state), these angles mostly fall into $[40°, 80°]$, which indicates that the average affinity becomes positive.

employs periodic updates and incurs lower computational cost, it does not address such a composite characteristic. In contrast, our algorithm (39) takes into account the composite problem structure and actively regulates the shift of distribution dynamics (32), thereby yielding smaller optimality gaps. We remark that the non-vanishing errors in Fig. 3 are due to the variance of the stochastic gradients constructed during iterations. Further experiments confirm that a larger mini-batch size n_{mb} and a smaller (or diminishing) step size η contribute to a lower asymptotic value of the convergence measure. Extensive experiments where the algorithms and IPOPT start from various random initial points yield consistent comparison results.

The above convergence behaviors are also linked with the histograms of the angles between the position and the decision across the population; see Fig. 4. Since the objective function of (35) is an inner product of the position p and the decision q, the angle between p and q is a useful indicator of the sign and the value of the inner product $p^\top q$. As shown in Fig. 4, the initial angles are largely distributed in $[80°, 100°]$, rendering the average population-wide affinity close to zero. The benchmark solution returned by IPOPT and the solutions obtained by the algorithms (36), (37),



(a) Relative optimality gap

(b) Relative distance to the benchmark solution q^*

Figure 5: This figure illustrates the convergence behaviors of our online stochastic algorithm (39) (termed "composite"), the vanilla algorithm (36) (termed "vanilla"), and the stateful algorithm (37) (termed "stateful") for the strong-polarization scenario. The solid lines represent the average values of convergence measures, whereas the shaded areas indicate the ranges of variation across independent trials.

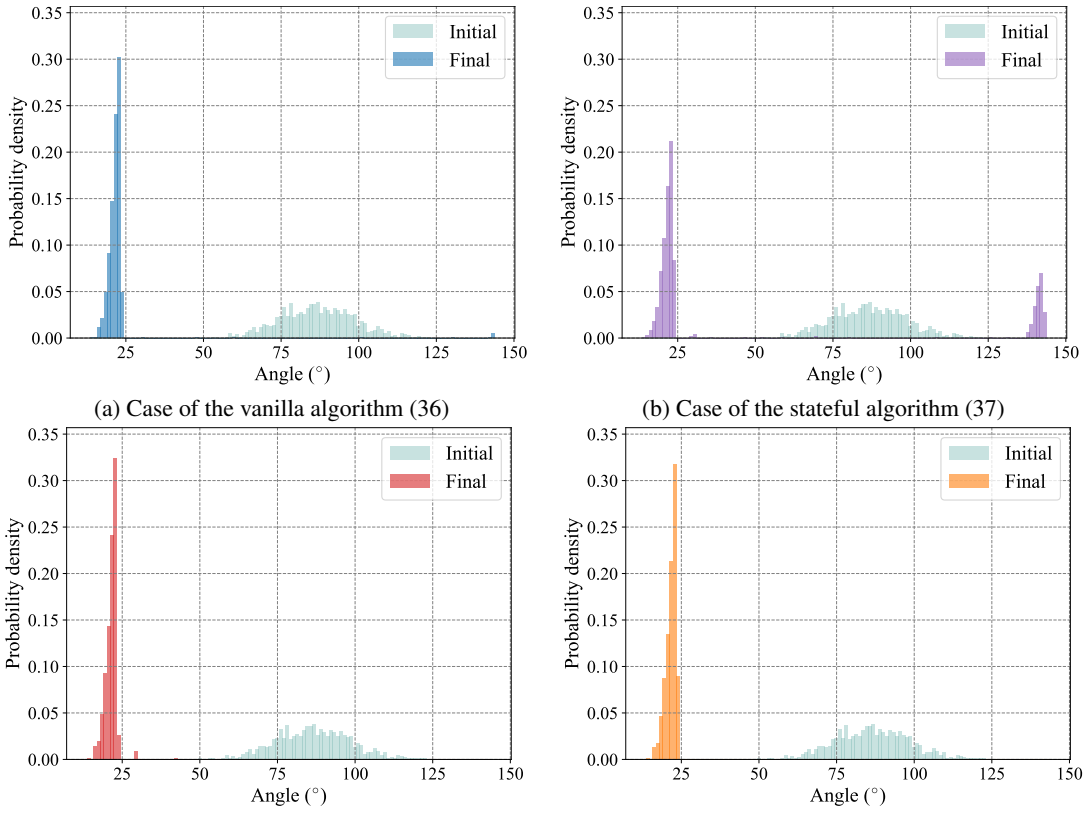
and (39) all result in a final configuration wherein most angles fall into $[40^\circ, 80^\circ]$. Hence, the final average affinity is positive. Since the vanilla (36) and the stateful (37) algorithms focus less on steering the distribution, in Figs. 4a and 4b, there are relatively more final angles in $[100^\circ, 110^\circ]$, which accounts for a smaller affinity and a larger optimality gap as shown in Fig. 3. In contrast, the solution found by our algorithm (39) induces population-wide behavior nearly identical to the IPOPT benchmark.

5.1.2 SCENARIO 2: STRONG POLARIZATION

We further examine a strong-polarization regime and demonstrate performance comparisons. In this setting, the coefficients in (32) are set to $\lambda = 0.9$ and $\sigma = 0.2$, indicating that the initial position only exerts a weak influence. The problem dimension and the population size are identical to those in the normal-mixing scenario. The angular range of the unit spherical cap used for sampling initial positions is 110°. We tune the step sizes of the algorithms (36), (37), and (39) over the grid spanning $[5 \times 10^{-3}, 0.5]$ and settle on 0.375, 0.425, and 0.325, respectively, to balance empirical convergence rates and accuracy. The mini-batch size $n_{\text{mb}} = 50$ and the epoch length $\bar{k} = 5$ are the same. Under the hardware configuration described at the beginning of Section 5, the wall-clock runtime covering all the algorithms is 8 minutes and 10 seconds.

In the scenario with $\lambda = 0.9$, the dynamics (32) converge to the steady state more slowly than in the normal-mixing case. Although theoretical intuition suggests that small step sizes may be necessary for convergence, the experimental results indicate the opposite, in that moderate step sizes remain effective. This phenomenon may arise from the polarization property of (32). Since the influence of the initial position is limited, the relative steering ability of the decision-maker becomes stronger, thereby facilitating the co-evolution of fast polarization and algorithmic updates.

Fig. 5 demonstrates the convergence behaviors of different algorithms in this strong-polarization scenario. Each unit increment on the horizontal axis of Fig. 5 (i.e., time step) represents one iteration



(a) Case of the vanilla algorithm (36)

(b) Case of the stateful algorithm (37)

(c) Case of our composite algorithm (39)

(d) Case with the benchmark solution q^* via IPOPT

Figure 6: This figure illustrates the histograms of the angles between the position and the decision across the population in the strong-polarization scenario. The initial angles mostly fall in $[75°, 100°]$. However, due to the polarization property of (32) and the stronger influence of the decision compared to the initial position, the final angles largely lie in $[15°, 25°]$. In Fig. 6b, the stateful algorithm (37) induces polarization, as shown by the two peaks in the distribution of final angles.

of the polarized dynamics (32). Thanks to exploiting the composite problem structure, our algorithm (39) achieves a lower optimality gap and a smaller distance to the benchmark solution. Such performance improvements are less pronounced than in the normal-mixing case, as reflected in the angle histograms of different algorithms in Fig. 6. The initial angles largely lie within $[75°, 100°]$. However, in sharp contrast to Fig. 4, most final angles fall in $[15°, 25°]$. The stateful algorithm (37) causes polarization, as shown by the two peaks in the final distribution in Fig. 6b. The vanilla algorithm (36) and our algorithm (39) yield final configurations similar to the optimal benchmark, though Fig. 6c exhibits a relatively higher concentration in the region below $22°$ than Fig. 6a.

5.2 Performance optimization given discrete choice distributions

Section 5.1 addresses affinity maximization involving continuous population distributions. In contrast, we now consider decision-making given discrete distributions lying in the probability simplex $\Delta^{m-1} = \{p \in \mathbb{R}^m \mid 1^\top p = 1, 0 \leq p(i) \leq 1, i = 1, \ldots, m\}$, where $m \in \mathbb{N}_+$, and $p(i)$ denotes

the i-th element of p. This setting is motivated by a recommender system scenario, where a user selects a specific item (e.g., product, movie, or plan) from a set of candidates, and the choice model is characterized by a discrete distribution in Δ^{m-1}. The recommender aims to optimize certain performance measures (e.g., gain and diversity) by interacting with this user and adjusting its decision. This setting is also related to the multi-armed bandit problem, in that the user action described by a choice distribution $p \in \Delta^{m-1}$ generates performance feedback determined by the decision-maker.

We consider the following distribution dynamics

$$p_{k+1} = \lambda_1 p_k + \lambda_2 \operatorname{softmax}(-\epsilon q) + (1 - \lambda_1 - \lambda_2) p_0, \qquad k \in \mathbb{N}, \tag{40}$$

where $p_k \in \Delta^{m-1}$ denotes the choice distribution at time k, $q \in \mathbb{R}^m$ is a decision vector whose i-th element represents, e.g., the price or loss of item i in a set, $\lambda_1, \lambda_2 \in (0,1)$ are combination coefficients that satisfy $\lambda_1 + \lambda_2 \leq 1$, and the i-th element of $\operatorname{softmax}(-\epsilon q)$ is given by $\exp(-\epsilon q(i))/\sum_{j=1}^m \exp(-\epsilon q(j))$, where $\epsilon > 0$ is a coefficient, and $i = 1, \dots, m$. The term $\operatorname{softmax}(-\epsilon q)$ gives a distribution consisting of probabilities proportional to the exponential of $-\epsilon q$. One interpretation of this softmax operation is that an item i with a low price $q(i)$ is selected with a high probability. As in (32), the last term of (40) captures the persistent influence of the initial choice distribution. The steady-state distribution of (40) given a fixed decision q is

$$p_{\text{ss}} = \frac{\lambda_2}{1-\lambda_1} \operatorname{softmax}(-\epsilon q) + \frac{1-\lambda_1-\lambda_2}{1-\lambda_1} p_0 \triangleq h(q, p_0). \tag{41}$$

The recommender aims to maximize its expected gain while preserving diversity, subject to certain budget requirements. This is formalized by the following problem

$$\begin{aligned} \max_{q \in \mathbb{R}^m} \quad & p^\top q - \rho \sum_{i=1}^m p(i) \log p(i) \\ \text{s.t.} \quad & p = h(q, p_0), \\ & 1^\top q = b, \\ & 0 \leq q(i) \leq \overline{q}, \qquad i = 1, \dots, m. \end{aligned} \tag{42}$$

In problem (42), the first term $p^\top q$ of the objective function represents the expected value given the steady-state discrete distribution $p = h(q, p_0)$; see also (41). This term indicates the expected price paid by the user, or conversely, the expected gain of the recommender. The second term of the objective is an entropy regularization term with the coefficient $\rho > 0$, and its purpose is to promote diversity by favoring high-entropy distributions p. The remaining constraints of problem (42) specify budget requirements of the decision q, where b and $\overline{q} > 0$ are constants.

Problem (42) involves the underlying distribution dynamics (40). An offline numerical solver hinges on an explicit and accurate model of (40), and therefore it can be restrictive in practice. To solve problem (42) in an online fashion, the vanilla algorithm, which is inspired by performative prediction (Hardt and Mendler-Dünner, 2025) and is oblivious to the composite problem structure, reads as follows

$$q_{k+1} = \operatorname{Proj}_{\mathcal{C}}(q_k + \eta p_k), \tag{43}$$

where $\operatorname{Proj}_{\mathcal{C}}(\cdot)$ denotes the projection to the feasible set $\mathcal{C} = \{q \in \mathbb{R}^m | 1^\top q = b, 0 \leq q(i) \leq \overline{q}\}$, $q(i)$ is the i-th element of q, $\eta > 0$ is the step size, and p_k is obtained after q_k is applied to

the dynamics (40) with previous state p_{k-1}. Similar to (37), we implement the following stateful algorithm with periodic updates based on Brown et al. (2022)

$$q_{k+1} = \begin{cases} \mathrm{Proj}_{\mathcal{C}}(q_k + \eta p_k), & \text{if } k \equiv 0 \pmod{\bar{k}}, \\ q_k, & \text{otherwise}, \end{cases} \tag{44}$$

where $\bar{k} \in \mathbb{N}_+$ is the epoch length. As discussed in Section 5.1, subsequent stateful algorithms (Ray et al., 2022; Li and Wai, 2022) adopt the same periodic updating structure and apply to various evolving distribution dynamics, including linear mixtures or geometrically ergodic Markov chains. Another choice is the derivative-free algorithm in Zhang et al. (2022) that relies on the evaluations of the objective $\Phi(q,p) \triangleq p^\top q - \rho \sum_{i=1}^m p(i) \log p(i)$:

$$q_{k+1} = \mathrm{Proj}_{\mathcal{C}} \left(q_k + \eta \frac{m}{\delta} (\Phi(q_k + \delta v_k, p_k) - \Phi(q_{k-1} + \delta v_{k-1}, p_{k-1})) v_k \right), \tag{45}$$

where $\eta > 0$ is the step size, m is the dimension of q_k, $\delta > 0$ is the smoothing parameter, v_k, v_{k-1} are independent random vectors uniformly sampled from the unit sphere in $\mathbb{R}^m$, $q_k + \delta v_k$ is the deployed decision that generates p_k, and $\Phi(q_k + \delta v_k, p_k)$ is the objective value evaluated at time k. Algorithm (45) exploits bandit feedback of the objective, and a similar version of (45) is adopted in Miller et al. (2021); Ray et al. (2022). In contrast, our algorithm taking into account the composite structure due to decision dependence is

$$q_{k+1} = \mathrm{Proj}_{\mathcal{C}} \Big(q_k + \eta \big(p_k + H_k (q_k - \rho (1 + \log p_k)) \big) \Big). \tag{46}$$

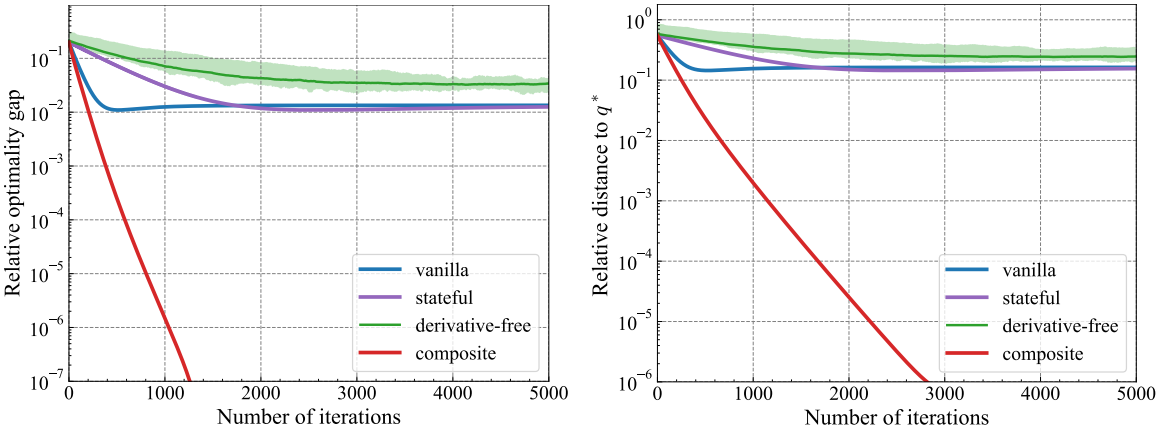
In (46), $\eta > 0$ is the step size, and the distribution sensitivity H_k (independent of p_k and p_0) is

$$H_k = \nabla_q h(q_k, p_0) = -\frac{\epsilon \lambda_2}{1 - \lambda_1} \cdot \frac{(1^\top z_k) \operatorname{diag}(z_k) - z_k z_k^\top}{(1^\top z_k)^2}, \tag{47}$$

where the vector $z_k \in \mathbb{R}^m$ is given by $\exp(-\epsilon q_k)$, and $\operatorname{diag}(z_k)$ denotes a diagonal matrix with z_k as its diagonal elements. The ascent-based updates in (43), (44), (45), and (46) are due to maximization in (42). Since the recommender directly interacts with a user endowed with the choice distribution p_k, the update (46) differs slightly from (9), although both are grounded in the same principle.

In experiments, the choice distribution p and the decision q are 100-dimensional vectors, i.e., $m = 100$. For the distribution dynamics (40), we set $\lambda_1 = 0.2, \lambda_2 = 0.5$, and $\epsilon = 0.5$. In problem (42), we set the regularization coefficient $\rho = 0.1$, the budget $b = 250$, and the upper bound $\overline{q} = 5$. For the algorithms (43), (44), and (46), we tune the step sizes over the grid spanning $[0.1, 5]$ and select the values 1.5, 1.5, and 1.25, respectively, to balance empirical convergence rates and accuracy. For the stateful algorithm (44), we set the epoch length $\bar{k} = 5$. For the algorithm (45), we search over η and δ on the grid spanning $[5 \times 10^{-2}, 1] \times [0.5, 5]$, settle on $\eta = 0.15, \delta = 1.5$, and run 20 independent trials. We use IPOPT (Wächter and Biegler, 2006) to calculate a (locally) optimal solution of problem (42) in an offline manner as a benchmark. However, as discussed earlier, numerical solvers can be restrictive in practice when the dynamics model (40) is inaccurate and real-time decision-making is necessary. Given the hardware setting stated at the beginning of Section 5, the wall-clock runtime to evaluate all the algorithms is 38 seconds.

Fig. 7 shows the convergence results of online algorithms (43), (44), (45), and (46) while interacting with the dynamics (40) to solve problem (42). Here, each time step corresponds to one iteration of (40). The vanilla algorithm (43) is online in nature, and therefore it adapts to the changing



(a) Relative optimality gap

(b) Relative distance to the benchmark solution q^*

Figure 7: This figure compares the convergence behaviors of the vanilla algorithm (43) (labeled "vanilla"), the stateful algorithm (44) (labeled "stateful"), the derivative-free algorithm (45) (labeled "derivative-free"), and our online stochastic algorithm (46) (labeled "composite") when interacting with the distribution dynamics (40) to solve problem (42).

distribution (40), eventually reaching a fixed-point solution. However, this algorithm suffers from significant sub-optimality, because it neglects the composite problem structure where the steady-state distribution is a function of the decision. The stateful algorithm (44) reduces computational costs thanks to its periodic updates. Nonetheless, this algorithm requires more interactions with the dynamics (40) than the vanilla algorithm (43) and still incurs sub-optimality. The derivative-free algorithm (45) converges slowly due to the stochasticity of gradient estimates. In contrast, our algorithm (46) respects the composite structure by leveraging distribution sensitivities and regulating the shift of the choice distribution. Consequently, it attains significantly higher solution accuracy. We remark that compared to Fig. 3 in Section 5.1, the curves of algorithms (43) and (46) in Fig. 7 do not exhibit stochastic oscillations, because we no longer resort to samples of a population distribution for mini-batch gradients.

The aforementioned difference in solution quality also causes the discrepancy in the evolution of the Wasserstein distances $W_1(p_k, p_{\mathrm{ss}}(q^*))$ between the dynamic choice distribution p_k and the steady-state distribution $p_{\mathrm{ss}}(q^*)$ induced by the benchmark decision q^* via IPOPT. To evaluate $W_1(p_k, p_{\mathrm{ss}}(q^*))$, we regard both vectors as distributions on the common support $\{1, 2, \ldots, m\}$ endowed with the ground metric $c(i, j) = |i - j|$, where $i, j = 1, \ldots, m$. Fig. 8 illustrates such a difference in $W_1(p_k, p_{\mathrm{ss}}(q^*))$. Overall, these distances converge to limiting values, thanks to the stable distribution dynamics (40) and the convergence of the algorithms to fixed decisions. We note that the vanilla algorithm (43), the stateful algorithm (44), and the derivative-free algorithm (45) bring about biased final choice distributions, whereas our algorithm (46) drives the distribution to closely approach $p_{\mathrm{ss}}(q^*)$ corresponding to the benchmark decision. As pointed out earlier, this major difference is due to our algorithmic structure, which explicitly addresses the composite problem characteristic resulting from decision dependence.

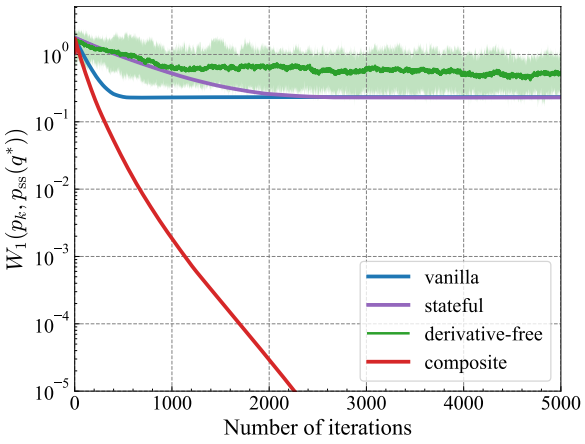


Figure 8: This figure demonstrates the evolution of the Wasserstein distances $W_1(p_k, p_{\mathrm{ss}}(q^*))$ between the choice distribution p_k and the steady-state distribution $p_{\mathrm{ss}}(q^*)$ induced by q^*. The legends "vanilla", "stateful", "derivative-free", and "composite" correspond to algorithms (43), (44), (45), and (46), respectively.

6. Conclusion

We formulated a decision-dependent stochastic optimization problem, where the dependence originates from the closed-loop interaction between a decision-maker and a dynamically evolving distribution. We presented an online stochastic algorithm that leverages samples from the dynamic distribution, takes into account the composite structure of the problem by anticipating the sensitivity of the distribution with respect to the decision, and shapes the overall distribution to inform optimal decision-making. We established the optimality guarantees of the proposed algorithm both in expectation and with high probability. Furthermore, we quantified the generalization performance in a finite-sample regime with an empirical distribution.

Future directions include, but are not limited to, i) addressing constraints related to the decision and the distribution (represented by, e.g., bounds on risk measures), ii) integrating sensitivity-learning schemes and quantifying the impact of sensitivity modeling or learning errors on performance guarantees, iii) developing model-free methods (through, e.g., derivative-free optimization) that bypass the need for sensitivity matrices of distribution dynamics, and iv) extending case studies to address general non-contracting opinion dynamics beyond the combination structure in Section 5.1. We hope this work sparks further advances at the intersection of machine learning, stochastic optimization, and nonlinear control.

Acknowledgments

We thank the Max Planck ETH Center for Learning Systems, the Swiss National Science Foundation via NCCR Automation (51NF40_225155), and the German Research Foundation for their support. We acknowledge N. Lanzetti for useful insights and comments, particularly on Lemma 2, and thank G. De Pasquale, P. Grontas, L. Aolaritei, S. Li, X. He, and S. Hall for helpful discussions.

Appendix A. Useful Lemmas

First, we present a lemma that establishes the upper and lower bounds on the weighted norm.

Lemma 13. *Let $x \in \mathbb{R}^m$, and let $P \in \mathbb{R}^{m\times m}$ be positive definite. Then,*

$$\sqrt{\lambda_{\min}(P)}\|x\| \leq \|x\|_P \leq \sqrt{\lambda_{\max}(P)}\|x\|. \tag{48}$$

Proof For a positive definite matrix P, both $\lambda_{\max}(P)I - P$ and $P - \lambda_{\min}(P)I$ are positive semidefinite. Hence, for any $x \in \mathbb{R}^m$, $\lambda_{\max}(P)x^\top x - x^\top Px \geq 0$ and $x^\top Px - \lambda_{\min}(P)x^\top x \geq 0$. By rearranging terms and taking the square root, we know that (48) holds. ■

Some implications of Lemma 13 are as follows. If a function $\Psi : \mathbb{R}^m \to \mathbb{R}$ is L_Ψ-Lipschitz with respect to $\|\cdot\|_P$, then it is $L_\Psi\sqrt{\lambda_{\max}(P)}$-Lipschitz with respect to $\|\cdot\|$. Conversely, if Ψ is L'_Ψ-Lipschitz with respect to $\|\cdot\|$, then it is $L'_\Psi/\sqrt{\lambda_{\min}(P)}$-Lipschitz with respect to $\|\cdot\|_P$.

The following lemma provides upper bounds related to the square of a nonnegative sequence.

Lemma 14. *Suppose that a nonnegative sequence $(a_k)_{k\in\mathbb{N}}$ satisfies $a_{k+1} \leq ca_k + b_k, \forall k \in \mathbb{N}$, where $c \in (0,1), a_k, b_k \in \mathbb{R}$. Then,*

$$a_{k+1}^2 \leq \frac{1+c^2}{2}a_k^2 + \frac{1+c^2}{1-c^2}b_k^2, \tag{49a}$$

$$\sum_{k=0}^{T-1} a_k^2 \leq \frac{2}{1-c^2}\left(a_0^2 + \frac{1+c^2}{1-c^2}\sum_{k=0}^{T-1} b_k^2\right). \tag{49b}$$

Proof We establish a recursive inequality of $(a_k^2)_{k\in\mathbb{N}}$ as follows

$$a_{k+1}^2 \leq c^2a_k^2 + 2ca_kb_k + b_k^2 \leq c^2a_k^2 + \frac{1-c^2}{2}a_k^2 + \frac{2c^2}{1-c^2}b_k^2 + b_k^2,$$

where the last inequality holds because $2ab \leq ta^2 + \frac{b^2}{t}, \forall a, b \in \mathbb{R}, t > 0$. Hence, (49a) is proved. We sum both sides of (49a) over $k = 0, \ldots, T-1$ and obtain

$$\sum_{k=1}^{T} a_k^2 \leq \frac{1+c^2}{2}\sum_{k=0}^{T-1} a_k^2 + \frac{1+c^2}{1-c^2}\sum_{k=0}^{T-1} b_k^2.$$

It follows that

$$\frac{1-c^2}{2}\sum_{k=0}^{T-1} a_k^2 \leq a_0^2 - a_T^2 + \frac{1+c^2}{1-c^2}\sum_{k=0}^{T-1} b_k^2 \leq a_0^2 + \frac{1+c^2}{1-c^2}\sum_{k=0}^{T-1} b_k^2.$$

We multiply both sides of the above inequality by $2/(1-c^2)$ and arrive at (49b). ■

We present a useful inequality relating the arithmetic and quadratic means. It can be proved by the Cauchy-Schwarz inequality or Jensen's inequality.

Lemma 15. *Consider $K \in \mathbb{N}_+$ real numbers $X_1, \ldots, X_K \in \mathbb{R}$. Then, $\frac{1}{K}\sum_{i=1}^K X_i \leq \sqrt{\frac{\sum_{i=1}^K X_i^2}{K}}$.*

The following lemma uses the Wasserstein distance between distributions to give an upper bound on the difference between the expectations of a Lipschitz function under these distributions. It is a corollary of the Kantorovich–Rubinstein theorem (Villani, 2009, Particular case 5.16).

Lemma 16 (Perdomo et al. (2020, Lemma D.4)). *Consider a metric space $(\mathbb{R}^m, c)$. Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^m)$, and let $\Psi : \mathbb{R}^m \to \mathbb{R}^n$ be L_Ψ-Lipschitz, i.e., $\forall x, y \in \mathbb{R}^m, \|\Psi(x) - \Psi(y)\| \leq L_\Psi c(x, y)$. With $W_1(\mu, \nu)$ defined by* (2)*, we have*

$$\|\mathbb{E}_{p\sim\mu}[\Psi(p)] - \mathbb{E}_{p\sim\nu}[\Psi(p)]\| \leq L_\Psi W_1(\mu, \nu). \tag{50}$$

Next, we present Hölder's inequality for probability measures. It enables us to bound the expectation of the product of random variables.

Lemma 17 (Hölder's inequality). *Consider $K \in \mathbb{N}_+$ random variables $X_1, \ldots, X_K \in \mathbb{R}$. Let constants $p_1, \ldots, p_K > 0$ satisfy $\sum_{i=1}^K 1/p_i = 1$. Then, $\mathbb{E}\left[\left|\prod_{i=1}^K X_i\right|\right] \leq \prod_{i=1}^K \left(\mathbb{E}\left[|X_i|^{p_i}\right]\right)^{\frac{1}{p_i}}$.*

For a set of random variables endowed with tail bounds, we can quantify the probability that their sums do not exceed a certain threshold, as discussed in the following lemma.

Lemma 18. *Suppose that $K \in \mathbb{N}_+$ random variables $X_i \in \mathbb{R}$ satisfy $\Pr(X_i \leq a_i) \geq 1 - \tau_i$, where $i \in \{1, \ldots, K\}$, with constants $a_i \in \mathbb{R}$ and $\tau_i \in (0, 1)$ such that $\sum_{i=1}^K \tau_i \in (0, 1)$. Then, we obtain $\Pr\left(\sum_{i=1}^K X_i \leq \sum_{i=1}^K a_i\right) \geq 1 - \sum_{i=1}^K \tau_i$.*

Proof Let A_i denote the event that $X_i \leq a_i$. Its complementary event A_i^c is $X_i > a_i$. The condition of the lemma implies that $\forall i \in \{1, \ldots, K\}, \Pr(A_i^c) \leq \tau_i$. Then,

$$\Pr\left(\bigcup_{i=1}^K A_i^c\right) \overset{\text{(a.1)}}{\leq} \sum_{i=1}^K \Pr(A_i^c) \leq \sum_{i=1}^K \tau_i,$$

where (a.1) follows from Boole's inequality. Consequently,

$$\Pr\left(\sum_{i=1}^K X_i \leq \sum_{i=1}^K a_i\right) \overset{\text{(a.1)}}{\geq} \Pr\left(\bigcap_{i=1}^K A_i\right) = 1 - \Pr\left(\bigcup_{i=1}^K A_i^c\right) \geq 1 - \sum_{i=1}^K \tau_i,$$

where (a.1) holds because the event $X_i \leq a_i, \forall i \in \{1, \ldots, K\}$ implies that $\sum_{i=1}^K X_i \leq \sum_{i=1}^K a_i$, but not vice versa. ■

Appendix B. Proof of Lemma 1

Proof The sampling procedure to obtain mini-batches is independent of the σ-algebra $\mathcal{F}_{k-1}$. Hence, conditional on $\mathcal{F}_{k-1}$, the sample pairs $\{(p_k^i, d^{i_k})\}_{i=1}^{n_{\text{mb}}}$ are independent and identically distributed according to γ_k. Moreover, the decision u_k is measurable with respect to $\mathcal{F}_{k-1}$. Therefore, the mini-batch stochastic gradient $\widehat{\nabla}_{\text{mb}}^k \tilde{\Phi}(u_k)$ given in (10) satisfies

$$\mathbb{E}\left[\widehat{\nabla}_{\text{mb}}^k \tilde{\Phi}(u_k)\middle|\mathcal{F}_{k-1}\right] = \frac{1}{n_{\text{mb}}} \sum_{i=1}^{n_{\text{mb}}} \mathbb{E}_{(p_k^i, d^{i_k})\sim\gamma_k}\left[\nabla_u \Phi(u_k, p_k^i) + \nabla_u h(u_k, d^{i_k}) \nabla_p \Phi(u_k, p_k^i)\middle|\mathcal{F}_{k-1}\right]$$

$$= \widehat{\nabla}^k \tilde{\Phi}(u_k), \tag{51}$$

where the expectation is taken over the randomness of sampling, and the last equality is due to (11). Furthermore, if Assumption 4 is satisfied,

$$\begin{aligned}\operatorname{Var}\left[\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)\middle|\mathcal{F}_{k-1}\right] &\overset{\text{(a.1)}}{=} \frac{1}{n_{\text{mb}}^2}\sum_{i=1}^{n_{\text{mb}}}\operatorname{Var}\left[\nabla_u\Phi(u_k,p_k^i)+\nabla_u h(u_k,d^{i_k})\nabla_p\Phi(u_k,p_k^i)\middle|\mathcal{F}_{k-1}\right]\\ &\overset{\text{(a.2)}}{\leq} \frac{M}{n_{\text{mb}}}+\frac{M_V}{n_{\text{mb}}}\|\widehat{\nabla}^k\tilde{\Phi}(u_k)\|^2,\end{aligned} \tag{52}$$

where (a.1) holds because the sample pairs $\{p_k^i, d^{i_k}\}$ are independent when conditioned on $\mathcal{F}_{k-1}$, and for any random variable $\xi \in \mathbb{R}^n$ and real number $a \in \mathbb{R}$, $\operatorname{Var}[a\xi] = \mathbb{E}\big[\|a\xi\|^2\big] - \|\mathbb{E}[a\xi]\|^2 = a^2\operatorname{Var}[\xi]$; (a.2) follows from Assumption 4. Therefore,

$$\begin{aligned}\mathbb{E}\left[\|\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)\|^2\middle|\mathcal{F}_{k-1}\right] &= \operatorname{Var}\left[\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)\middle|\mathcal{F}_{k-1}\right] + \left\|\,\mathbb{E}[\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)\middle|\mathcal{F}_{k-1}]\right\|^2\\ &\leq \frac{M}{n_{\text{mb}}}+\left(\frac{M_V}{n_{\text{mb}}}+1\right)\|\widehat{\nabla}^k\tilde{\Phi}(u_k)\|^2,\end{aligned}$$

where the last inequality uses (51) and (52). Consequently, Lemma 1 is proved. ■

Appendix C. Proofs for Section 4.1

C.1 Proof of Lemma 2

Proof The joint distributions γ_k and $\gamma_{\text{ss}}(u_k)$ lie in the space $\mathcal{P}_1(\mathbb{R}^{m+r})$, endowed with the Wasserstein distance W_1 induced by the metric c in (14). Hence,

$$\begin{aligned}W_1(\gamma_k,\gamma_{\text{ss}}(u_k)) &\overset{\text{(a.1)}}{=} \inf_{\beta\in\Gamma(\alpha(p_0,d),\mu_d(d'))}\int_{\mathbb{R}^m\times\mathbb{R}^r\times\mathbb{R}^r} c\left((f^{(k)}(p_0,d),d),(h(u_k,d'),d')\right)\,\mathrm{d}\beta(p_0,d,d')\\ &\overset{\text{(a.2)}}{\leq} \int_{\mathbb{R}^m\times\mathbb{R}^r\times\mathbb{R}^r} c\left((f^{(k)}(p_0,d),d'),(h(u_k,d),d')\right)\,\mathrm{d}\left((\pi_1,\pi_2,\pi_2)_{\#}\tilde{\beta}\right)(p_0,d,d')\\ &\overset{\text{(a.3)}}{=} \int_{\mathbb{R}^m\times\mathbb{R}^r\times\mathbb{R}^r} c\left((f^{(k)}(p_0,d),d),(h(u_k,d),d)\right)\,\mathrm{d}\tilde{\beta}(p_0,d,d')\\ &\overset{\text{(a.4)}}{=} \int_{\mathbb{R}^m\times\mathbb{R}^r} c\left((f^{(k)}(p_0,d),d),(h(u_k,d),d)\right)\,\mathrm{d}\alpha(p_0,d)\\ &\overset{\text{(a.5)}}{=} \int_{\mathbb{R}^m\times\mathbb{R}^r}\left\|f^{(k)}(p_0,d)-h(u_k,d)\right\|_P\,\mathrm{d}\alpha(p_0,d).\end{aligned} \tag{53}$$

In (53), (a.1) holds because $f^{(k)}$ and h are continuous and hence Borel maps, allowing us to use the property of the pushforward operation; see Aolaritei et al. (2025, Proposition 3). In (a.2), $\tilde{\beta}$ is a specific coupling in $\Gamma(\alpha(p_0,d),\mu_d(d'))$, and π_1 and π_2 denote the projections on the first and second variables, respectively, i.e., $\pi_1 : (p_0,d,d') \mapsto p_0$ and $\pi_2 : (p_0,d,d') \mapsto d$. Consequently, the joint distribution of the first two variables of $\big((\pi_1,\pi_2,\pi_2)_{\#}\tilde{\beta}\big)(p_0,d,d')$, namely $(\pi_1,\pi_2)_{\#}\tilde{\beta}$, is α, and the marginal distribution of the last variable is μ_d. Since $\big((\pi_1,\pi_2,\pi_2)_{\#}\tilde{\beta}\big)(p_0,d)$ is also a

coupling in $\Gamma(\alpha(p_0, d), \mu_d(d'))$, (a.2) is true. Further, (a.3) invokes the change-of-variables formula for pushforward measures. Moreover, (a.4) follows from the fact that the integrand is independent of d', allowing the integration to be performed with respect to the joint distribution α of p_0 and d. Finally, (a.5) leverages the metric c in (14), where the distance related to the second component is zero for the same d. Therefore, (15a) is proved.

We proceed to establish a recursive inequality for V_k. For any $k \in \mathbb{N}_+$,

$$\begin{aligned}
V_k &= \int_{\mathbb{R}^m \times \mathbb{R}^r} \left\| f^{(k)}(p_0, d) - h(u_k, d) \right\|_P \mathrm{d}\alpha(p_0, d) \\
&\overset{\text{(a.1)}}{=} \int_{\mathbb{R}^m \times \mathbb{R}^r} \left\| f\big(f^{(k-1)}(p_0, d), u_k, d\big) - f(h(u_k, d), u_k, d) \right\|_P \mathrm{d}\alpha(p_0, d) \\
&\overset{\text{(a.2)}}{\leq} L_f^p \int_{\mathbb{R}^m \times \mathbb{R}^r} \left\| f^{(k-1)}(p_0, d) - h(u_k, d) \right\|_P \mathrm{d}\alpha(p_0, d) \\
&\overset{\text{(a.3)}}{\leq} L_f^p \underbrace{\int_{\mathbb{R}^m \times \mathbb{R}^r} \left\| f^{(k-1)}(p_0, d) - h(u_{k-1}, d) \right\|_P \mathrm{d}\alpha(p_0, d)}_{①=V_{k-1}} \\
&\quad + L_f^p \underbrace{\int_{\mathbb{R}^m \times \mathbb{R}^r} \|h(u_{k-1}, d) - h(u_k, d)\|_P \mathrm{d}\alpha(p_0, d)}_{②},
\end{aligned} \tag{54}$$

where (a.1) is due to the distribution dynamics (3) and the steady state $h(u_k, d)$ satisfying the fixed-point equation $f(h(u_k, d), u_k, d) = h(u_k, d)$; (a.2) leverages the property that f is L_f^p-Lipschitz in p, see Assumption 1; (a.3) uses the triangle inequality. In (54), term ① corresponds to V_{k-1}, and term ② can be bounded from above by

$$\begin{aligned}
② &\overset{\text{(a.1)}}{\leq} \sqrt{\lambda_{\max}(P)} \int_{\mathbb{R}^m \times \mathbb{R}^r} \|h(u_{k-1}, d) - h(u_k, d)\| \mathrm{d}\alpha(p_0, d) \\
&\overset{\text{(a.2)}}{\leq} L_h^u \sqrt{\lambda_{\max}(P)} \int_{\mathbb{R}^m \times \mathbb{R}^r} \|u_{k-1} - u_k\| \mathrm{d}\alpha(p_0, d) \\
&\overset{\text{(a.3)}}{\leq} L_h^u \sqrt{\lambda_{\max}(P)} \|u_k - u_{k-1}\|,
\end{aligned}$$

where (a.1) uses Lemma 13 in Appendix A; (a.2) follows from the fact that h is L_h^u-Lipschitz in u; see the discussion below Assumption 1; (a.3) is due to $\int 1 \, \mathrm{d}\alpha(p_0, d) = 1$. We incorporate this upper bound into (54) and prove (15b).

Finally, we demonstrate that the initial upper bound V_0 on $W_1(\gamma_0, \gamma_{\text{ss}}(u_0))$ is finite. Note that

$$\begin{aligned}
V_0 &\overset{\text{(a.1)}}{=} \int_{\mathbb{R}^m \times \mathbb{R}^r} \|p_0 - h(u_0, d)\|_P \mathrm{d}\alpha(p_0, d) \\
&\overset{\text{(a.2)}}{\leq} \sqrt{\lambda_{\max}(P)} \int_{\mathbb{R}^m \times \mathbb{R}^r} \|p_0 - h(u_0, d)\| \mathrm{d}\alpha(p_0, d) \\
&\overset{\text{(a.3)}}{\leq} \sqrt{\lambda_{\max}(P)} \bigg(\int_{\mathbb{R}^m \times \mathbb{R}^r} \|p_0\| \mathrm{d}\alpha(p_0, d) + \int_{\mathbb{R}^m \times \mathbb{R}^r} \|h(u_0, d)\| \mathrm{d}\alpha(p_0, d) \bigg)
\end{aligned}$$

$$\overset{\text{(a.4)}}{\leq} \sqrt{\lambda_{\max}(P)}\Big(\underbrace{\int_{\mathbb{R}^m} \|p_0\| \,\mathrm{d}\mu_0(p_0)}_{①} + \underbrace{\int_{\mathbb{R}^r} \|h(u_0,d)\| \,\mathrm{d}\mu_d(d)}_{②}\Big) \tag{55}$$

where (a.1) follows from the definition $f^{(0)}(p_0,d)=p_0$; (a.2) uses Lemma 13 in Appendix A; (a.3) is due to the triangle inequality; (a.4) holds because the integrands only depend on the corresponding random variables and not on their joint distribution. In (55), term ① is finite thanks to $\mu_0 \in \mathcal{P}_1(\mathbb{R}^m)$, implying that μ_0 admits a finite first moment. We now show that term ② in (55) is finite. In fact,

$$\begin{aligned}\int_{\mathbb{R}^r} \|h(u_0,d)\| \,\mathrm{d}\mu_d(d) &\overset{\text{(a.1)}}{\leq} \int_{\mathbb{R}^r} \|h(u_0,d)-h(u_0,0)\| \,\mathrm{d}\mu_d(d) + \int_{\mathbb{R}^r} \|h(u_0,0)\| \,\mathrm{d}\mu_d(d)\\ &\overset{\text{(a.2)}}{\leq} L_h^d \int_{\mathbb{R}^r} \|d\| \,\mathrm{d}\mu_d(d) + \|h(u_0,0)\| < \infty,\end{aligned}$$

where (a.1) invokes the triangle inequality, and (a.2) holds because $h(u_0,d)$ is L_h^d-Lipschitz in d and $\int_{\mathbb{R}^r} 1 \,\mathrm{d}\mu_d(d)=1$. We know that $\int_{\mathbb{R}^r} \|d\| \,\mathrm{d}\mu_d(d)$ is finite due to $\mu_d \in \mathcal{P}_1(\mathbb{R}^r)$. Since both terms ① and ② in (55) are finite, the initial bound V_0 is also finite. ■

C.2 Proof of Theorem 3

Proof For any $k \in \mathbb{N}$, the Wasserstein distance $W_1(\gamma_k, \gamma_{\text{ss}}(u_k))$ is always nonnegative, because the metric c given in (14) will not be negative. Therefore, the corresponding upper bound V_k is also nonnegative. Hence, we start from (15b) in Lemma 2, apply (49b) in Lemma 14, and obtain

$$\sum_{k=0}^{T-1} V_k^2 \leq \frac{V_0^2}{1-\rho_1} + \frac{\rho_2}{1-\rho_1} \sum_{k=1}^{T} \|u_k - u_{k-1}\|^2,$$

where ρ_1 and ρ_2 are given in Theorem 3. The upper bound (15a) and the non-negativity of the Wasserstein distance imply that $0 \leq W_1(\gamma_k, \gamma_{\text{ss}}(u_k)) \leq V_k$, which leads to

$$\sum_{k=0}^{T-1} W_1(\gamma_k, \gamma_{\text{ss}}(u_k))^2 \leq \sum_{k=0}^{T-1} V_k^2. \tag{56}$$

Therefore, (16) holds. The update rule (9) indicates that $u_k - u_{k-1} = -\eta \widehat{\nabla}_{\text{mb}}^{k-1} \tilde{\Phi}(u_{k-1})$. We incorporate this expression into (16) and arrive at (17). ■

C.3 Proof of Lemma 4

Proof Let $u_k \in \mathbb{R}^n$ be given. Consider the following function $\Psi : \mathbb{R}^m \times \mathbb{R}^r \to \mathbb{R}^n$ representing the full gradient inside the expectations in (8) and (11)

$$\Psi(p,d) \triangleq \nabla_u \Phi(u_k,p) + \nabla_u h(u_k,d) \nabla_p \Phi(u_k,p).$$

We now demonstrate that $\Psi(p,d)$ is Lipschitz continuous in (p,d) with respect to the metric c defined in (14). For any $p_1, p_2 \in \mathbb{R}^m$ and $d_1, d_2 \in \mathbb{R}^r$,

$$\begin{aligned}
\|\Psi(p_1,d_1)-\Psi(p_2,d_2)\| &\overset{\text{(a.1)}}{\leq} \|\Psi(p_1,d_1)-\Psi(p_1,d_2)\| + \|\Psi(p_1,d_2)-\Psi(p_2,d_2)\| \\
&\overset{\text{(a.2)}}{\leq} \|\nabla_u h(u_k,d_1)-\nabla_u h(u_k,d_2)\|\,\|\nabla_p \Phi(u_k,p_1)\| \\
&\quad + \|\nabla_u \Phi(u_k,p_1)-\nabla_u \Phi(u_k,p_2)\| + \|\nabla_u h(u_k,d_2)\|\,\|\nabla_p \Phi(u_k,p_1)-\nabla_p \Phi(u_k,p_2)\| \\
&\overset{\text{(a.3)}}{\leq} L_\Phi^p M_h^d \|d_1-d_2\| + \left(M_\Phi^u + L_h^u M_\Phi^p\right)\|p_1-p_2\| \\
&\overset{\text{(a.4)}}{\leq} L(\|p_1-p_2\|_P + \|d_1-d_2\|),
\end{aligned}$$

where (a.1) follows from the triangle inequality; (a.2) uses the fact that norms are sub-multiplicative and the triangle inequality; (a.3) invokes the Lipschitz continuity of $\nabla_u h(u_k,d)$, $\nabla_u \Phi(u_k,p)$, $\nabla_p \Phi(u_k,p)$, $\Phi(u_k,p)$, and $h(u_k,d)$ thanks to Assumptions 1 and 2, indicating $\|\nabla_p \Phi(u_k,p_1)\| \leq L_\Phi^p$ and $\|\nabla_u h(u_k,d_2)\| \leq L_h^u$; (a.4) uses (48) in Lemma 13 to relate the Euclidean norm to the weighted norm $\|\cdot\|_P$. Afterward, we leverage Lemma 16 in Appendix A to obtain (19). ■

Appendix D. Proofs for Section 4.2

D.1 Proof of Theorem 5

The overarching idea is to analyze the coupled evolution of the distribution dynamics (3) and the algorithm (9). To this end, we first quantify the cumulative Wasserstein metric related to the distribution (3). Afterward, we focus on the iterative update of the algorithm (9) and synthesize the overall convergence measure, i.e., the average expected second moment of gradients.

We start with the following lemma that characterizes the cumulative sum of the squared Wasserstein distances between the distribution at each time and the corresponding steady-state distribution. It is built on Lemma 2 and further uses the condition (20) on the step size in Theorem 5, thereby exposing the true gradient $\nabla\tilde{\Phi}(u_k)$ in the upper bound on this cumulative sum.

Lemma 19. *Under the conditions of Theorem 5, we have*

$$\sum_{k=0}^{T-1} \mathbb{E}\left[W_1(\gamma_k,\gamma_{\text{ss}}(u_k))^2\right] \leq \frac{1}{6L^2}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] + \frac{7V_0^2}{6(1-\rho_1)} + \frac{M}{12L^2(M_V+n_{\text{mb}})}. \quad (57)$$

Proof We first prove the following intermediate inequality

$$\sum_{k=0}^{T-1} \mathbb{E}\left[W_1(\gamma_k,\gamma_{\text{ss}}(u_k))^2\right] \leq \frac{V_0^2}{1-\rho_1} + \frac{\eta^2\rho_2 MT}{(1-\rho_1)n_{\text{mb}}} + \frac{\eta^2\rho_2}{1-\rho_1}\cdot\left(\frac{M_V}{n_{\text{mb}}}+1\right)\sum_{k=0}^{T-1}\mathbb{E}\left[\|\widehat{\nabla}^k\tilde{\Phi}(u_k)\|^2\right], \quad (58)$$

where the coefficients ρ_1 and ρ_2 are specified in Theorem 3. We know from the update rule (9) that $u_k - u_{k-1} = -\eta\widehat{\nabla}_{\text{mb}}^{k-1}\tilde{\Phi}(u_{k-1})$. We plug this expression into (15b), invoke (49a) in Lemma 14, Appendix A (note that $(V_k)_{k\in\mathbb{N}}$ is a nonnegative sequence; see Appendix C.2), and obtain

$$V_k^2 \leq \rho_1 V_{k-1}^2 + \eta^2\rho_2\|\widehat{\nabla}_{\text{mb}}^{k-1}\tilde{\Phi}(u_{k-1})\|^2. \quad (59)$$

Recall that $\mathcal{F}_k$ is the σ-algebra generated by $\widehat{\nabla}^0_{\text{mb}}\tilde{\Phi}(u_0), \ldots, \widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$. It follows that

$$\begin{aligned}\mathbb{E}\left[V_k^2 \middle| \mathcal{F}_{k-2}\right] &\leq \rho_1 \, \mathbb{E}\left[V_{k-1}^2 \middle| \mathcal{F}_{k-2}\right] + \eta^2\rho_2 \mathbb{E}\left[\|\widehat{\nabla}^{k-1}_{\text{mb}}\tilde{\Phi}(u_{k-1})\|^2 \middle| \mathcal{F}_{k-2}\right] \\ &\overset{\text{(a.1)}}{\leq} \rho_1 \, \mathbb{E}\left[V_{k-1}^2 \middle| \mathcal{F}_{k-2}\right] + \frac{\eta^2\rho_2 M}{n_{\text{mb}}} + \eta^2\rho_2\left(\frac{M_V}{n_{\text{mb}}}+1\right)\|\widehat{\nabla}^{k-1}\tilde{\Phi}(u_{k-1})\|^2,\end{aligned}$$

where (a.1) is due to (13) in Lemma 1. We further take the total expectation of both sides of the above inequality, use the tower rule, and arrive at

$$\mathbb{E}\left[V_k^2\right] \leq \rho_1 \, \mathbb{E}\left[V_{k-1}^2\right] + \frac{\eta^2\rho_2 M}{n_{\text{mb}}} + \eta^2\rho_2\left(\frac{M_V}{n_{\text{mb}}}+1\right)\mathbb{E}\left[\|\widehat{\nabla}^{k-1}\tilde{\Phi}(u_{k-1})\|^2\right]. \tag{60}$$

Similar to the proof of (49b) in Lemma 14, Appendix A, we sum both sides of (60) over $k = 1, \ldots, T$, use the upper bound $V_0^2 - \mathbb{E}\left[V_T^2\right] \leq V_0^2$, and invoke (56) to obtain the inequality (58).

We now focus on the last term of (58). Recall from (8) and (11) in Section 3.1 that $\nabla\tilde{\Phi}(u_k)$ and $\widehat{\nabla}^k\tilde{\Phi}(u_k)$ are the true gradient and the approximate gradient at u_k, respectively. Moreover, e_k is the difference of $\widehat{\nabla}^k\tilde{\Phi}(u_k)$ and $\nabla\tilde{\Phi}(u_k)$; see (18). Therefore,

$$\begin{aligned}\mathbb{E}\left[\|\widehat{\nabla}^k\tilde{\Phi}(u_k)\|^2\right] &\overset{\text{(a.1)}}{\leq} 2\,\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] + 2\,\mathbb{E}\left[\|e_k\|^2\right] \\ &\overset{\text{(a.2)}}{\leq} 2\,\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] + 2L^2\,\mathbb{E}\left[W_1(\gamma_k, \gamma_{\text{ss}}(u_k))^2\right],\end{aligned} \tag{61}$$

where (a.1) uses the inequality $\|a+b\|^2 \leq 2\|a\|^2 + 2\|b\|^2, \forall a, b \in \mathbb{R}^n$, and (a.2) applies (19) in Lemma 4. We plug this upper bound into (58), rearrange terms, and obtain

$$\begin{aligned}&\left(1 - \frac{2\eta^2\rho_2 L^2}{1-\rho_1}\left(\frac{M_V}{n_{\text{mb}}}+1\right)\right)\sum_{k=0}^{T-1}\mathbb{E}\left[W_1(\gamma_k, \gamma_{\text{ss}}(u_k))^2\right] \\ &\quad\leq \frac{2\eta^2\rho_2}{1-\rho_1}\left(\frac{M_V}{n_{\text{mb}}}+1\right)\sum_{k=0}^{T-1}\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] + \frac{V_0^2}{1-\rho_1} + \frac{\eta^2\rho_2 MT}{(1-\rho_1)n_{\text{mb}}}.\end{aligned} \tag{62}$$

The parametric condition (20) of Theorem 5 leads to $\frac{2\eta^2\rho_2 L^2}{1-\rho_1}\left(\frac{M_V}{n_{\text{mb}}}+1\right) \leq \frac{1}{7T} \leq \frac{1}{7}$. It follows that

$$\begin{aligned}&\frac{1}{1-\frac{2\eta^2\rho_2 L^2}{1-\rho_1}\left(\frac{M_V}{n_{\text{mb}}}+1\right)}\cdot\frac{2\eta^2\rho_2}{1-\rho_1}\left(\frac{M_V}{n_{\text{mb}}}+1\right) \leq \frac{1}{1-\frac{1}{7}}\cdot\frac{1}{7L^2} = \frac{1}{6L^2}, \\ &\frac{1}{1-\frac{2\eta^2\rho_2 L^2}{1-\rho_1}\left(\frac{M_V}{n_{\text{mb}}}+1\right)}\cdot\frac{\eta^2\rho_2 MT}{(1-\rho_1)n_{\text{mb}}} \leq \frac{1}{1-\frac{1}{7}}\cdot\frac{M}{14L^2(M_V+n_{\text{mb}})} = \frac{M}{12L^2(M_V+n_{\text{mb}})}.\end{aligned}$$

We divide both sides of (62) by $1 - \frac{2\eta^2\rho_2 L^2}{1-\rho_1}\left(\frac{M_V}{n_{\text{mb}}}+1\right)$, use the above upper bounds on coefficients, and obtain (57). ■

With Lemma 19 in hand, we are ready to prove Theorem 5.

Proof Since $\tilde{\Phi}(u)$ is $L_{\tilde{\Phi}}$-smooth, we have

$$\begin{aligned}\tilde{\Phi}(u_{k+1}) \leq & \tilde{\Phi}(u_k) + \nabla\tilde{\Phi}(u_k)^\top(u_{k+1}-u_k) + \frac{L_{\tilde{\Phi}}}{2}\|u_{k+1}-u_k\|^2 \\ \overset{\text{(a.1)}}{=} & \tilde{\Phi}(u_k) - \eta\nabla\tilde{\Phi}(u_k)^\top\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k) + \frac{L_{\tilde{\Phi}}\eta^2}{2}\|\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)\|^2, \end{aligned} \tag{63}$$

where (a.1) follows from the update $u_{k+1}-u_k = -\eta\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$; see (9). Hence,

$$\begin{aligned}\mathbb{E}\left[\tilde{\Phi}(u_{k+1})\middle|\mathcal{F}_{k-1}\right] \overset{\text{(a.1)}}{\leq} & \tilde{\Phi}(u_k) - \eta\nabla\tilde{\Phi}(u_k)^\top\widehat{\nabla}^k\tilde{\Phi}(u_k) + \frac{L_{\tilde{\Phi}}\eta^2}{2}\,\mathbb{E}\left[\|\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)\|^2\middle|\mathcal{F}_{k-1}\right] \\ \overset{\text{(a.2)}}{\leq} & \tilde{\Phi}(u_k) - \eta\|\nabla\tilde{\Phi}(u_k)\|^2 - \eta\nabla\tilde{\Phi}(u_k)^\top e_k \\ & + \frac{L_{\tilde{\Phi}}\eta^2}{2}\left(\frac{M}{n_{\text{mb}}} + \left(\frac{M_V}{n_{\text{mb}}}+1\right)\|\widehat{\nabla}^k\tilde{\Phi}(u_k)\|^2\right) \\ \overset{\text{(a.3)}}{\leq} & \tilde{\Phi}(u_k) - \left(\frac{\eta}{2} - L_{\tilde{\Phi}}\eta^2\left(\frac{M_V}{n_{\text{mb}}}+1\right)\right)\|\nabla\tilde{\Phi}(u_k)\|^2 \\ & + \left(\frac{\eta}{2} + L_{\tilde{\Phi}}\eta^2\left(\frac{M_V}{n_{\text{mb}}}+1\right)\right)\|e_k\|^2 + \frac{L_{\tilde{\Phi}}M\eta^2}{2n_{\text{mb}}}, \end{aligned} \tag{64}$$

where (a.1) uses $\mathbb{E}[\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)|\mathcal{F}_{k-1}] = \widehat{\nabla}^k\tilde{\Phi}(u_k)$, see (12); (a.2) uses the equality $\widehat{\nabla}^k\tilde{\Phi}(u_k) = \nabla\tilde{\Phi}(u_k) + e_k$ (see (18)) and the upper bound (13); (a.3) follows from the inequalities

$$-\nabla\tilde{\Phi}(u_k)^\top e_k \leq \frac{1}{2}\|\nabla\tilde{\Phi}(u_k)\|^2 + \frac{1}{2}\|e_k\|^2, \qquad \|\widehat{\nabla}^k\tilde{\Phi}(u_k)\|^2 \leq 2\|\nabla\tilde{\Phi}(u_k)\|^2 + 2\|e_k\|^2.$$

We take expectation of both sides of (64), use the tower rule, and obtain

$$\begin{aligned}\mathbb{E}\left[\tilde{\Phi}(u_{k+1})\right] \leq & \mathbb{E}\left[\tilde{\Phi}(u_k)\right] - \left(\frac{\eta}{2} - L_{\tilde{\Phi}}\eta^2\left(\frac{M_V}{n_{\text{mb}}}+1\right)\right)\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] \\ & + \left(\frac{\eta}{2} + L_{\tilde{\Phi}}\eta^2\left(\frac{M_V}{n_{\text{mb}}}+1\right)\right)\mathbb{E}\left[\|e_k\|^2\right] + \frac{L_{\tilde{\Phi}}M\eta^2}{2n_{\text{mb}}}. \end{aligned}$$

We sum up both sides of the above inequality for $k = 0, \ldots, T-1$, reorganize terms, and obtain

$$\begin{aligned}&\left(\frac{\eta}{2} - L_{\tilde{\Phi}}\eta^2\left(\frac{M_V}{n_{\text{mb}}}+1\right)\right)\sum_{k=0}^{T-1}\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] \\ &\quad\leq \tilde{\Phi}(u_0) - \mathbb{E}\left[\tilde{\Phi}(u_T)\right] + \left(\frac{\eta}{2} + L_{\tilde{\Phi}}\eta^2\left(\frac{M_V}{n_{\text{mb}}}+1\right)\right)\sum_{k=0}^{T-1}\mathbb{E}\left[\|e_k\|^2\right] + \frac{L_{\tilde{\Phi}}M\eta^2T}{2n_{\text{mb}}}. \end{aligned}$$

The parametric condition (20) of Theorem 5 ensures that $\frac{\eta}{2} - L_{\tilde{\Phi}}\eta^2\left(\frac{M_V}{n_{\text{mb}}}+1\right) \geq \frac{\eta}{4}$ and $\frac{\eta}{2} + L_{\tilde{\Phi}}\eta^2\left(\frac{M_V}{n_{\text{mb}}}+1\right) \leq \frac{3}{4}\eta$. We leverage (19) and obtain

$$\sum_{k=0}^{T-1}\mathbb{E}\left[\|e_k\|^2\right] \leq L^2\sum_{k=0}^{T-1}\mathbb{E}\left[W_1(\gamma_k, \gamma_{\text{ss}}(u_k))^2\right].$$

Furthermore, we invoke (57) in Lemma 19. Hence, we establish the following upper bound

$$\frac{\eta}{8}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] \leq \tilde{\Phi}(u_0) - \mathbb{E}\left[\tilde{\Phi}(u_T)\right] + \frac{7\eta L^2 V_0^2}{8(1-\rho_1)} + \frac{\eta M}{16(M_V+n_{\text{mb}})} + \frac{L_{\tilde{\Phi}} M \eta^2 T}{2n_{\text{mb}}}.$$

We multiply both sides of the above inequality by $8/(\eta T)$ and arrive at

$$\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] \leq \underbrace{\frac{8(\tilde{\Phi}(u_0)-\tilde{\Phi}^*)}{\eta T}}_{\sim\mathcal{O}(1/\sqrt{T})} + \underbrace{\frac{4L_{\tilde{\Phi}}M\eta}{n_{\text{mb}}}}_{\sim\mathcal{O}(1/\sqrt{T})} + \underbrace{\frac{7L^2V_0^2}{(1-\rho_1)T}}_{\sim\mathcal{O}(1/T)} + \underbrace{\frac{M}{2T(M_V+n_{\text{mb}})}}_{\sim\mathcal{O}(1/T)},$$

where we additionally use the fact that $\mathbb{E}[\tilde{\Phi}(u_T)] \geq \tilde{\Phi}^*$, because $\tilde{\Phi}^*$ is the optimal value of problem (6). Therefore, (21) and (22) are proved. ■

D.2 Proof of Corollary 6

Proof We know from (61) in Appendix D.1 that

$$\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\widehat{\nabla}^k\tilde{\Phi}(u_k)\|^2\right] \leq \frac{2}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] + \frac{2L^2}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[W_1(\gamma_k,\gamma_{\text{ss}}(u_k))^2\right].$$

Furthermore, we leverage Lemma 19 in Appendix D.1 to obtain

$$\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\widehat{\nabla}^k\tilde{\Phi}(u_k)\|^2\right] \leq \frac{7}{3T}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] + \underbrace{\frac{1}{T}\left(\frac{7L^2V_0^2}{3(1-\rho_1)} + \frac{M}{6(M_V+n_{\text{mb}})}\right)}_{\sim\mathcal{O}(1/T)}.$$

By invoking (21) and (22), we prove (23). ■

D.3 Proof of Theorem 7

Proof We first derive the following upper bound

$$\begin{aligned}\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[W_1(\gamma_k,\gamma_{\text{ss}}(u_k))^2\right] &\overset{\text{(a.1)}}{\leq} \frac{1}{6L^2T}\sum_{k=0}^{T-1}\mathbb{E}\left[\|\nabla\tilde{\Phi}(u_k)\|^2\right] + \underbrace{\frac{1}{T}\left(\frac{7V_0^2}{6(1-\rho_1)} + \frac{M}{12L^2(M_V+n_{\text{mb}})}\right)}_{\sim\mathcal{O}(1/T)} \\ &\overset{\text{(a.2)}}{\leq} \underbrace{\frac{1}{6L^2}\left(\frac{8(\tilde{\Phi}(u_0)-\tilde{\Phi}^*)}{\eta T} + \frac{4L_{\tilde{\Phi}}M\eta}{n_{\text{mb}}}\right)}_{\sim\mathcal{O}(1/\sqrt{T})} + \mathcal{O}\left(\frac{1}{T}\right),\end{aligned}$$

where (a.1) follows from Lemma 19 in Appendix D.1, and (a.2) invokes the upper bound (21) in Theorem 5. Furthermore, Lemma 15 in Appendix A implies that

$$\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[W_1(\gamma_k,\gamma_{\text{ss}}(u_k))\right] \leq \sqrt{\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[W_1(\gamma_k,\gamma_{\text{ss}}(u_k))^2\right]}.$$

We combine the above inequalities and prove Theorem 7. ■

Appendix E. Proofs for Section 4.3

E.1 Proof of Theorem 8

The proof resembles what we have presented in Appendix D.1. The key idea is to quantify the coupled evolution of optimization iterations and dynamic distributions. However, to establish convergence with high probability rather than in expectation, we will derive and exploit some tail bounds related to stochastic errors of gradients.

Throughout this subsection, we rely on the following decomposition

$$\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k) = \widehat{\nabla}^k\tilde{\Phi}(u_k) + \xi_k = \nabla\tilde{\Phi}(u_k) + e_k + \xi_k, \tag{65}$$

where e_k and ξ_k are given in (18) and (25), respectively. Specifically, e_k denotes the difference of expected gradients (i.e., $\widehat{\nabla}^k\tilde{\Phi}(u_k)$ and $\nabla\tilde{\Phi}(u_k)$) arising from the discrepancy between γ_k and $\gamma_{\text{ss}}(u_k)$, and ξ_k indicates the stochastic noise in the mini-batch gradient $\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$ relative to its expectation $\widehat{\nabla}^k\tilde{\Phi}(u_k)$.

Along with the iteration (9), the stochastic noise ξ_k (25) exerts a cumulative influence on the accuracy of solutions. In the following lemma, we provide high-probability bounds on some cumulative quantities that involve ξ_k and are relevant to the overall convergence measure.

Lemma 20. *Let Assumption 5 hold. For any $\tau_1, \tau_2 \in (0,1)$ and $\lambda > 0$, the iteration* (9) *ensures*

$$\sum_{k=0}^{T-1} -\nabla\tilde{\Phi}(u_k)^\top\xi_k \le \frac{3}{4}\lambda\sigma^2_{\text{mb}}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k)\|^2 + \frac{1}{\lambda}\ln\frac{1}{\tau_1}, \tag{66a}$$

$$\sum_{k=0}^{T-1}\|\xi_k\|^2 \le T\sigma^2_{\text{mb}}\left(1+\ln\frac{1}{\tau_2}\right), \tag{66b}$$

with probabilities at least $1-\tau_1$ and $1-\tau_2$, respectively.

Proof Recall that $\mathcal{F}_k$ is the σ-algebra generated by $\widehat{\nabla}^0_{\text{mb}}\tilde{\Phi}(u_0), \ldots, \widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$. Hence, as a function of $\widehat{\nabla}^0_{\text{mb}}\tilde{\Phi}(u_0), \ldots, \widehat{\nabla}^{k-1}_{\text{mb}}\tilde{\Phi}(u_{k-1})$ determined by the update (9), u_k is measurable with respect to $\mathcal{F}_{k-1}$. Moreover, when conditioned on $\mathcal{F}_{k-1}$, $\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k)$ is an unbiased estimate of $\widehat{\nabla}^k\tilde{\Phi}(u_k)$; see (12). Therefore,

$$\mathbb{E}\left[-\nabla\tilde{\Phi}(u_k)^\top\xi_k\middle|\mathcal{F}_{k-1}\right] = -\nabla\tilde{\Phi}(u_k)^\top\,\mathbb{E}\left[\widehat{\nabla}^k_{\text{mb}}\tilde{\Phi}(u_k) - \widehat{\nabla}^k\tilde{\Phi}(u_k)\middle|\mathcal{F}_{k-1}\right] = 0.$$

Hence, $\big(-\nabla\tilde{\Phi}(u_k)^\top\xi_k\big)_{k\in\mathbb{N}}$ is a martingale difference sequence. Whenever $\nabla\tilde{\Phi}(u_k) = 0$, we define $\frac{(-\nabla\tilde{\Phi}(u_k)^\top\xi_k)^2}{(\sigma_{\text{mb}}\|\nabla\tilde{\Phi}(u_k)\|)^2} \triangleq 0$. Furthermore, $\sigma_{\text{mb}}\|\nabla\tilde{\Phi}(u_k)\|$ is measurable with respect to $\mathcal{F}_{k-1}$, and

$$\mathbb{E}\left[\exp\left(\frac{(-\nabla\tilde{\Phi}(u_k)^\top\xi_k)^2}{(\sigma_{\text{mb}}\|\nabla\tilde{\Phi}(u_k)\|)^2}\right)\middle|\mathcal{F}_{k-1}\right] \overset{\text{(a.1)}}{\le} \mathbb{E}\left[\exp\left(\frac{\|\nabla\tilde{\Phi}(u_k)\|^2\|\xi_k\|^2}{(\sigma_{\text{mb}}\|\nabla\tilde{\Phi}(u_k)\|)^2}\right)\middle|\mathcal{F}_{k-1}\right]$$

$$= \mathbb{E}\left[\exp\left(\frac{\|\xi_k\|^2}{\sigma_{\text{mb}}^2}\right)\middle|\mathcal{F}_{k-1}\right] \overset{\text{(a.2)}}{\leq} \exp(1),$$

where (a.1) uses the Cauchy-Schwarz inequality, and (a.2) is due to (26). Then, we apply the martingale concentration inequality (see Li and Orabona, 2020, Lemma 1) and obtain (66a).

We proceed to prove (66b). We know from (26) and the tower rule that

$$\mathbb{E}\left[\exp\left(\frac{\|\xi_k\|^2}{\sigma_{\text{mb}}^2}\right)\right] = \mathbb{E}\left[\mathbb{E}\left[\exp\left(\frac{\|\xi_k\|^2}{\sigma_{\text{mb}}^2}\right)\middle|\mathcal{F}_{k-1}\right]\right] \leq \exp(1), \quad \forall k \in \mathbb{N}. \tag{67}$$

Let $\bar{\sigma}_{\text{mb}} \triangleq \sqrt{T}\sigma_{\text{mb}} > 0$. We note that

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\frac{\sum_{k=0}^{T-1}\|\xi_k\|^2}{\bar{\sigma}_{\text{mb}}^2}\right)\right] &= \mathbb{E}\left[\prod_{k=0}^{T-1}\exp\left(\frac{\|\xi_k\|^2}{\bar{\sigma}_{\text{mb}}^2}\right)\right] \\
&\overset{\text{(a.1)}}{\leq} \prod_{k=0}^{T-1}\mathbb{E}\left[\exp\left(\frac{\|\xi_k\|^2}{\bar{\sigma}_{\text{mb}}^2}\right)^T\right]^{\frac{1}{T}} = \prod_{k=0}^{T-1}\mathbb{E}\left[\exp\left(\frac{T\|\xi_k\|^2}{T\sigma_{\text{mb}}^2}\right)\right]^{\frac{1}{T}} \\
&\overset{\text{(a.2)}}{\leq} \prod_{k=0}^{T-1}[\exp(1)]^{\frac{1}{T}} = \exp(1),
\end{aligned} \tag{68}$$

where (a.1) follows from Hölder's inequality, see Lemma 17 in Appendix A, and (a.2) uses (67). Therefore, for any constant $a > 0$,

$$\begin{aligned}
\Pr\left(\sum_{k=0}^{T-1}\|\xi_k\|^2 > a\right) &= \Pr\left(\exp\left(\frac{\sum_{k=0}^{T-1}\|\xi_k\|^2}{\bar{\sigma}_{\text{mb}}^2}\right) > \exp\left(\frac{a}{\bar{\sigma}_{\text{mb}}^2}\right)\right) \\
&\overset{\text{(a.1)}}{\leq} \frac{\mathbb{E}\left[\exp\left(\sum_{k=0}^{T-1}\|\xi_k\|^2/\bar{\sigma}_{\text{mb}}^2\right)\right]}{\exp(a/\bar{\sigma}_{\text{mb}}^2)} \overset{\text{(a.2)}}{\leq} \frac{\exp(1)}{\exp(a/\bar{\sigma}_{\text{mb}}^2)},
\end{aligned}$$

where (a.1) holds because of Markov's inequality, and (a.2) uses (68). For any fixed $\tau_2 \in (0,1)$, we set $a = T\sigma_{\text{mb}}^2\big(1 + \ln\frac{1}{\tau_2}\big)$. It follows that (66b) holds. ■

Similar to Lemma 19, the following lemma provides an upper bound on the cumulative squared Wasserstein distances. This cumulative sum quantifies the evolution of the distribution driven by the decision-maker. Due to the coupling between the distribution dynamics and the decision, the upper bound involves quantities related to gradients, which originate from optimization iterations.

Lemma 21. *Under the conditions of Theorem 8, we have*

$$\sum_{k=0}^{T-1} W_1(\gamma_k, \gamma_{\text{ss}}(u_k))^2 \leq \frac{1}{6L^2}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k)\|^2 + \frac{7V_0^2}{6(1-\rho_1)} + \frac{1}{6L^2T}\sum_{k=0}^{T-1}\|\xi_k\|^2. \tag{69}$$

Proof We rely on (17) in Theorem 3, Section 4.1 and proceed by deriving an upper bound on the cumulative squared norm of $\widehat{\nabla}_{\text{mb}}^k\tilde{\Phi}(u_k)$. Based on (65), we know

$$\|\widehat{\nabla}_{\text{mb}}^k\tilde{\Phi}(u_k)\|^2 = \|\nabla\tilde{\Phi}(u_k) + e_k + \xi_k\|^2$$

$$\begin{aligned}
&\overset{\text{(a.1)}}{\leq} 3\|\nabla\tilde{\Phi}(u_k)\|^2 + 3\|e_k\|^2 + 3\|\xi_k\|^2 \\
&\overset{\text{(a.2)}}{\leq} 3\|\nabla\tilde{\Phi}(u_k)\|^2 + 3\|\xi_k\|^2 + 3L^2 W_1(\gamma_k, \gamma_{\text{ss}}(u_k))^2,
\end{aligned} \tag{70}$$

where (a.1) uses the inequality $\forall a, b, c \in \mathbb{R}^n, \|a+b+c\|^2 \leq 3(\|a\|^2 + \|b\|^2 + \|c\|^2)$, which is a vector-form extension of Lemma 15 in Appendix A, and (a.2) follows from (19). We plug (70) into the right-hand side of (17), rearrange terms, and obtain

$$\left(1 - \frac{3\eta^2\rho_2 L^2}{1-\rho_1}\right) \sum_{k=0}^{T-1} W_1(\gamma_k, \gamma_{\text{ss}}(u_k))^2 \leq \frac{V_0^2}{1-\rho_1} + \frac{3\eta^2\rho_2}{1-\rho_1} \sum_{k=0}^{T-1} \|\nabla\tilde{\Phi}(u_k)\|^2 + \frac{3\eta^2\rho_2}{1-\rho_1} \sum_{k=0}^{T-1} \|\xi_k\|^2. \tag{71}$$

The parametric condition (27) of Theorem 8 ensures that $\frac{3\eta^2\rho_2 L^2}{1-\rho_1} \leq \frac{1}{7T} \leq \frac{1}{7}$, which leads to

$$\frac{1}{1 - \frac{3\eta^2\rho_2 L^2}{1-\rho_1}} \cdot \frac{3\eta^2\rho_2}{1-\rho_1} \leq \frac{1}{1-\frac{1}{7}} \cdot \frac{1}{7L^2T} = \frac{1}{6L^2T} \leq \frac{1}{6L^2}.$$

We divide both sides of (71) by $1 - \frac{3\eta^2\rho_2 L^2}{1-\rho_1}$, use the above bounds on coefficients, and obtain (69). ■

We give the complete proof of Theorem 8. The key idea is to synthesize the convergence measure (i.e., the average squared gradient norm) from optimization iterations and the characterization of distribution dynamics in Lemma 21. We also leverage the high-probability bounds in Lemma 20.

Proof We incorporate (65) into (63) and have

$$\begin{aligned}
\tilde{\Phi}(u_{k+1}) \leq& \tilde{\Phi}(u_k) - \eta\nabla\tilde{\Phi}(u_k)^\top(\nabla\tilde{\Phi}(u_k) + e_k + \xi_k) + \frac{L_{\tilde{\Phi}}\eta^2}{2}\|\nabla\tilde{\Phi}(u_k) + e_k + \xi_k\|^2 \\
\overset{\text{(a.1)}}{\leq}& \tilde{\Phi}(u_k) - \left(\frac{\eta}{2} - \frac{3}{2}L_{\tilde{\Phi}}\eta^2\right)\|\nabla\tilde{\Phi}(u_k)\|^2 - \eta\nabla\tilde{\Phi}(u_k)^\top\xi_k + \frac{3}{2}L_{\tilde{\Phi}}\eta^2\|\xi_k\|^2 \\
&+ \left(\frac{\eta}{2} + \frac{3}{2}L_{\tilde{\Phi}}\eta^2\right) L^2 W_1(\gamma_k, \gamma_{\text{ss}}(u_k))^2,
\end{aligned}$$

where (a.1) follows from the inequality $-\nabla\tilde{\Phi}(u_k)^\top e_k \leq \frac{1}{2}\|\nabla\tilde{\Phi}(u_k)\|^2 + \frac{1}{2}\|e_k\|^2$, (70), and (19). We sum up both sides of the above inequality for $k = 0, \ldots, T-1$, reorganize terms, and obtain

$$\begin{aligned}
\left(\frac{\eta}{2} - \frac{3}{2}L_{\tilde{\Phi}}\eta^2\right)\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k)\|^2 \leq& \tilde{\Phi}(u_0) - \tilde{\Phi}(u_T) - \eta\sum_{k=0}^{T-1}\nabla\tilde{\Phi}(u_k)^\top\xi_k + \frac{3}{2}L_{\tilde{\Phi}}\eta^2\sum_{k=0}^{T-1}\|\xi_k\|^2 \\
&+ \left(\frac{\eta}{2} + \frac{3}{2}L_{\tilde{\Phi}}\eta^2\right) L^2 \sum_{k=0}^{T-1} W_1(\gamma_k, \gamma_{\text{ss}}(u_k))^2.
\end{aligned}$$

The parametric condition (27) of Theorem 8 guarantees that $\frac{\eta}{2} - \frac{3}{2}L_{\tilde{\Phi}}\eta^2 \geq \frac{\eta}{4}$ and $\frac{\eta}{2} + \frac{3}{2}L_{\tilde{\Phi}}\eta^2 \leq \frac{3}{4}\eta$. Furthermore, we use the fact that $\tilde{\Phi}(u_T)$ is not less than the optimal value $\tilde{\Phi}^*$ of problem (6),

incorporate the upper bound (69), rearrange terms, and arrive at

$$\frac{\eta}{8}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k)\|^2 \leq \tilde{\Phi}(u_0)-\tilde{\Phi}^* \underbrace{-\eta\sum_{k=0}^{T-1}\nabla\tilde{\Phi}(u_k)^\top\xi_k}_{①} + \underbrace{\left(\frac{3}{2}L_{\tilde{\Phi}}\eta^2+\frac{\eta}{8T}\right)\sum_{k=0}^{T-1}\|\xi_k\|^2}_{②} + \frac{7}{8}\eta L^2\frac{V_0^2}{1-\rho_1}. \tag{72}$$

For terms ① and ② in (72), we exploit the high-probability bounds in Lemma 20. Specifically, we select $\tau_1=\tau_2=\tau/2$ and apply both Lemmas 18 and 20. Then, with probability at least $1-\tau$,

$$① + ② \leq \frac{3}{4}\eta\lambda\sigma_{\text{mb}}^2\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k)\|^2 + \frac{\eta}{\lambda}\ln\frac{2}{\tau} + \left(\frac{3}{2}L_{\tilde{\Phi}}\eta^2+\frac{\eta}{8T}\right)T\sigma_{\text{mb}}^2\left(1+\ln\frac{2}{\tau}\right). \tag{73}$$

Since λ can be any positive constant, we select $\lambda=\frac{1}{12\sigma_{\text{mb}}^2}$. We then plug (73) into (72) and know that with probability at least $1-\tau$,

$$\frac{\eta}{16}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k)\|^2 \leq \tilde{\Phi}(u_0)-\tilde{\Phi}^*+\frac{7}{8}\eta L^2\frac{V_0^2}{1-\rho_1}+12\eta\sigma_{\text{mb}}^2\ln\frac{2}{\tau}+\left(\frac{3}{2}L_{\tilde{\Phi}}\eta^2+\frac{\eta}{8T}\right)T\sigma_{\text{mb}}^2\left(1+\ln\frac{2}{\tau}\right).$$

We multiply both sides of the above inequality by $16/(\eta T)$ and arrive at

$$\begin{aligned}\frac{1}{T}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k)\|^2 \leq{}& \underbrace{\frac{16(\tilde{\Phi}(u_0)-\tilde{\Phi}^*)}{\eta T}}_{\sim\mathcal{O}\left(1/\sqrt{T}\right)} + \underbrace{\left(24\eta L_{\tilde{\Phi}}+\frac{2}{T}\right)\sigma_{\text{mb}}^2}_{\sim\mathcal{O}\left(1/\sqrt{T}\right)} + \underbrace{24\eta L_{\tilde{\Phi}}\sigma_{\text{mb}}^2\ln\frac{2}{\tau}}_{\sim\mathcal{O}\left(\ln(1/\tau)/\sqrt{T}\right)} \\ &+ \underbrace{\frac{194}{T}\sigma_{\text{mb}}^2\ln\frac{2}{\tau}}_{\sim\mathcal{O}(\ln(1/\tau)/T)} + \underbrace{\frac{14L^2V_0^2}{(1-\rho_1)T}}_{\sim\mathcal{O}(1/T)},\end{aligned}$$

which holds with probability at least $1-\tau$. Hence, Theorem 8 is proved. ■

E.2 Proof of Theorem 9

Proof Starting from Lemma 21 in Appendix E.1, we know that

$$\frac{1}{T}\sum_{k=0}^{T-1}W_1(\gamma_k,\gamma_{\text{ss}}(u_k))^2 \leq \frac{1}{6L^2}\cdot\frac{1}{T}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k)\|^2 + \frac{1}{6L^2T^2}\sum_{k=0}^{T-1}\|\xi_k\|^2 + \mathcal{O}\left(\frac{1}{T}\right). \tag{74}$$

Theorem 8 in Section 4.3 indicates that with probability at least $1-\frac{\tau}{2}$,

$$\frac{1}{T}\sum_{k=0}^{T-1}\|\nabla\tilde{\Phi}(u_k)\|^2 \leq \frac{16(\tilde{\Phi}(u_0)-\tilde{\Phi}^*)}{\eta T} + 24\eta L_{\tilde{\Phi}}\sigma_{\text{mb}}^2\left(1+\ln\frac{4}{\tau}\right) + \mathcal{O}\left(\frac{1}{T}\left(1+\ln\frac{1}{\tau}\right)\right). \tag{75}$$

Moreover, Lemma 20 in Appendix E.1 implies that with probability at least $1-\frac{\tau}{2}$,

$$\sum_{k=0}^{T-1}\|\xi_k\|^2 \le T\sigma_{\text{mb}}^2\left(1+\ln\frac{2}{\tau}\right). \tag{76}$$

By incorporating the high-probability bounds (75) and (76) into (74) and using Lemma 18 in Appendix A, we know that with probability at least $1-\tau$,

$$\frac{1}{T}\sum_{k=0}^{T-1}W_1(\gamma_k,\gamma_{\text{ss}}(u_k))^2 \le \underbrace{\frac{1}{6L^2}\left(\frac{16(\tilde{\Phi}(u_0)-\tilde{\Phi}^*)}{\eta T}+24\eta L_{\tilde{\Phi}}\sigma_{\text{mb}}^2\left(1+\ln\frac{4}{\tau}\right)\right)}_{\sim\mathcal{O}\left((1+\ln(1/\tau))/\sqrt{T}\right)}+\mathcal{O}\left(\frac{1}{T}\left(1+\ln\frac{1}{\tau}\right)\right).$$

To link with the non-squared Wasserstein distance, we apply Lemma 15 in Appendix A and obtain

$$\frac{1}{T}\sum_{k=0}^{T-1}W_1(\gamma_k,\gamma_{\text{ss}}(u_k)) \le \sqrt{\frac{1}{T}\sum_{k=0}^{T-1}W_1(\gamma_k,\gamma_{\text{ss}}(u_k))^2}.$$

Consequently, Theorem 9 is proved. ■

Appendix F. Proofs for Section 4.4

F.1 Proof of Lemma 10

We rely on the concentration of the empirical measure evaluated by Wasserstein distances (Fournier and Guillin, 2015). Such a concentration is concerned with the closeness between μ_d and μ_d^N with high probability. Since $\mu_d, \mu_d^N \in \mathcal{P}(\mathbb{R}^r)$, we define $W_1(\mu_d,\mu_d^N)$ as

$$W_1(\mu_d,\mu_d^N) = \inf_{\bar{\beta}\in\Gamma(\mu_d,\mu_d^N)}\int_{\mathbb{R}^r\times\mathbb{R}^r}\|d-d'\|\,\mathrm{d}\bar{\beta}(d,d'). \tag{77}$$

The measure concentration argument is delineated in the following lemma.

Lemma 22 (from Fournier and Guillin (2015, Theorem 2))**.** *Let the conditions of Lemma 10 hold. Then, for any $N \ge 1$,*

$$\Pr\left(W_1(\mu_d,\mu_d^N)\le\epsilon\right) \ge \begin{cases}1-c_1e^{-c_2N\epsilon^r} & \text{if } \epsilon\in(0,1],\\ 1-c_1e^{-c_2N\epsilon^\theta} & \text{if } \epsilon\in(1,\infty),\end{cases}$$

where c_1 and c_2 are positive constants that only depend on r,θ,κ, and $\mathcal{E}_{\theta,\kappa}(\mu_d)$.

The proof of Lemma 10 is as follows.

Proof The empirical objective $\tilde{\Phi}^N(u)$ can be equivalently written as $\tilde{\Phi}^N(u)=\mathbb{E}_{p\sim\mu_{\text{ss}}^N(u)}[\Phi(u,p)]$. Accordingly, the gradient $\nabla\tilde{\Phi}^N(u)$ involves the empirical distribution $\gamma_{\text{ss}}^N(u)$. For any $u\in\mathbb{R}^n$, in terms of the difference between the gradients given the distribution $\gamma_{\text{ss}}(u)$ and the empirical distribution $\gamma_{\text{ss}}^N(u)\triangleq(h(u,\cdot),\mathrm{Id})_{\#}\mu_d^N$, we have

$$\left\|\nabla\tilde{\Phi}(u)-\nabla\tilde{\Phi}^N(u)\right\| \overset{\text{(a.1)}}{\le} LW_1(\gamma_{\text{ss}}(u),\gamma_{\text{ss}}^N(u)) \overset{\text{(a.2)}}{\le} L\left(L_h^d\sqrt{\lambda_{\max}(P)}+1\right)W_1(\mu_d,\mu_d^N), \tag{78}$$

where (a.1) is similar to (19) in Lemma 4, Section 4.1. We derive (a.2) in (78) as follows. Let $\overline{\beta}^*$ be an optimal coupling of (μ_d, μ_d^N) in the sense of (77), i.e., $\int \|d - d'\| \, \mathrm{d}\overline{\beta}^*(d, d') = W_1(\mu_d, \mu_d^N)$. Such an optimal coupling exists; see Villani (2009, Theorem 4.1) for a formal proof. Then,

$$\begin{aligned}
W_1(\gamma_{\mathrm{ss}}(u), \gamma_{\mathrm{ss}}^N(u)) &= W_1\big((h(u,\cdot), \mathrm{Id})_{\#}\mu_d, (h(u,\cdot), \mathrm{Id})_{\#}\mu_d^N\big) \\
&\overset{\text{(a.1)}}{=} \inf_{\overline{\beta}(d,d') \in \Gamma(\mu_d, \mu_d^N)} \int_{\mathbb{R}^r \times \mathbb{R}^r} c\big((h(u,d),d), (h(u,d'),d')\big) \, \mathrm{d}\overline{\beta}(d,d') \\
&\overset{\text{(a.2)}}{\leq} \int_{\mathbb{R}^r \times \mathbb{R}^r} c\big((h(u,d),d), (h(u,d'),d')\big) \, \mathrm{d}\overline{\beta}^*(d,d') \\
&\overset{\text{(a.3)}}{\leq} \int_{\mathbb{R}^r \times \mathbb{R}^r} \big(\|h(u,d) - h(u,d')\|_P + \|d - d'\|\big) \, \mathrm{d}\overline{\beta}^*(d,d') \\
&\overset{\text{(a.4)}}{\leq} \Big(L_h^d \sqrt{\lambda_{\max}(P)} + 1\Big) \int_{\mathbb{R}^r \times \mathbb{R}^r} \|d - d'\| \, \mathrm{d}\overline{\beta}^*(d,d') \\
&\overset{\text{(a.5)}}{=} \Big(L_h^d \sqrt{\lambda_{\max}(P)} + 1\Big) W_1(\mu_d, \mu_d^N),
\end{aligned}$$

where (a.1) follows from the property of the pushforward operation (Aolaritei et al., 2025, Proposition 3), and the metric c is given by (14); (a.2) holds because $\overline{\beta}^*$ is a specific coupling of (μ_d, μ_d^N), i.e., $\overline{\beta}^* \in \Gamma(\mu_d, \mu_d^N)$; (a.3) uses the expression (14); (a.4) leverages the property that $h(u,d)$ is L_h^d-Lipschitz in d and (48) in Lemma 13, Appendix A; (a.5) is true because $\overline{\beta}^*$ is an optimal coupling of (μ_d, μ_d^N). Since (78) holds for any $u \in \mathbb{R}^n$, we plug in $\{u_0^N, \dots, u_{T-1}^N\}$ and know that

$$\begin{aligned}
\frac{1}{T}\sum_{k=0}^{T-1} \|\nabla\tilde{\Phi}(u_k^N) - \nabla\tilde{\Phi}^N(u_k^N)\|^2 &\leq \frac{1}{T}\sum_{k=0}^{T-1} L^2 \Big(L_h^d \sqrt{\lambda_{\max}(P)} + 1\Big)^2 W_1(\mu_d, \mu_d^N)^2 \\
&= L^2 \Big(L_h^d \sqrt{\lambda_{\max}(P)} + 1\Big)^2 W_1(\mu_d, \mu_d^N)^2. \qquad (79)
\end{aligned}$$

Furthermore, Lemma 22 ensures that with probability at least $1 - \frac{\tau}{2}$,

$$W_1(\mu_d, \mu_d^N) \leq \begin{cases} \Big(\frac{1}{c_2 N} \ln\big(\frac{2c_1}{\tau}\big)\Big)^{\frac{1}{r}}, & \text{if } N \geq \frac{1}{c_2} \ln\big(\frac{2c_1}{\tau}\big), \\ \Big(\frac{1}{c_2 N} \ln\big(\frac{2c_1}{\tau}\big)\Big)^{\frac{1}{\theta}}, & \text{if } 1 \leq N < \frac{1}{c_2} \ln\big(\frac{2c_1}{\tau}\big). \end{cases} \qquad (80)$$

We combine (80) with (79) and prove Lemma 10. ■

F.2 Proof of Theorem 11

Proof The key idea is to start from the decomposition (31) and then leverage the high-probability bounds on the generalization error (see Lemma 10) and the optimization error (see Theorem 8).

We exploit (31) in Section 4.4 to obtain

$$\frac{1}{T}\sum_{k=0}^{T-1} \|\nabla\tilde{\Phi}(u_k^N)\|^2 \leq \underbrace{\frac{2}{T}\sum_{k=0}^{T-1} \|\nabla\tilde{\Phi}(u_k^N) - \nabla\tilde{\Phi}^N(u_k^N)\|^2}_{①} + \underbrace{\frac{2}{T}\sum_{k=0}^{T-1} \|\nabla\tilde{\Phi}^N(u_k^N)\|^2}_{②}. \qquad (81)$$

Lemma 10 provides the following upper bounds on term ① in (81):

$$① \leq \begin{cases} 2L^2\left(L_h^d\sqrt{\lambda_{\max}(P)}+1\right)^2\left(\frac{1}{c_2 N}\ln\left(\frac{2c_1}{\tau}\right)\right)^{\frac{2}{r}}, & \text{if } N \geq \frac{1}{c_2}\ln\left(\frac{2c_1}{\tau}\right), \\ 2L^2\left(L_h^d\sqrt{\lambda_{\max}(P)}+1\right)^2\left(\frac{1}{c_2 N}\ln\left(\frac{2c_1}{\tau}\right)\right)^{\frac{2}{\theta}}, & \text{if } 1 \leq N < \frac{1}{c_2}\ln\left(\frac{2c_1}{\tau}\right), \end{cases} \tag{82}$$

which hold with probability at least $1-\tau/2$. For term ② in (81), we invoke Theorem 8 and know that with probability at least $1-\tau/2$,

$$② \leq \underbrace{\frac{32(\tilde{\Phi}^N(u_0^N)-\tilde{\Phi}^{N,*})}{\eta T}+48\eta L_{\tilde{\Phi}}^N\sigma_{\text{mb}}^2\left(1+\ln\frac{4}{\tau}\right)}_{\sim\mathcal{O}\left((1+\ln(1/\tau))/\sqrt{T}\right)}+\mathcal{O}\left(\frac{1}{T}\left(1+\ln\frac{1}{\tau}\right)\right). \tag{83}$$

Supported by Lemma 18, we combine the upper bounds (82) and (83) and prove Theorem 11. ■

Appendix G. Proof and Additional Results for Section 5

G.1 Proof of Theorem 12

We first present a lemma concerning the angles between vectors involving strict conic combinations.

Lemma 23. *Consider any nonzero vectors $v_1, v_2 \in \mathbb{R}^m$ that form an acute angle $\theta_{12} \in (0, \frac{\pi}{2})$. Let $v_3 = \lambda_1 v_1 + \lambda_2 v_2$ be a strict conic combination of v_1 and v_2, where $\lambda_1, \lambda_2 > 0$. Then, the angle θ_{23} between v_3 and v_2 is smaller than θ_{12}.*

Proof Without loss of generality, we consider v_1 and v_2 of unit norm. If not, we can normalize these vectors and perform a similar analysis. Let $\bar{v}_3 = \bar{\lambda}_1 v_1 + \bar{\lambda}_2 v_2$ be the normalized vector of v_3, where $\bar{\lambda}_1 = \lambda_1/\|v_3\|$, and $\bar{\lambda}_2 = \lambda_2/\|v_3\|$. It follows that

$$\|\bar{v}_3\|^2 = \bar{\lambda}_1^2 + \bar{\lambda}_2^2 + 2\bar{\lambda}_1\bar{\lambda}_2 v_1^\top v_2 = 1. \tag{84}$$

Since unit vectors v_1 and v_2 form an acute angle, we know $0 < v_1^\top v_2 < 1$. Therefore,

$$\bar{\lambda}_1^2 + \bar{\lambda}_2^2 < \bar{\lambda}_1^2 + \bar{\lambda}_2^2 + 2\bar{\lambda}_1\bar{\lambda}_2 v_1^\top v_2 < \left(\bar{\lambda}_1 + \bar{\lambda}_2\right)^2,$$

which leads to $\bar{\lambda}_1 + \bar{\lambda}_2 > 1$ and $0 < \bar{\lambda}_1, \bar{\lambda}_2 < 1$. Then,

$$\begin{aligned} \cos(\theta_{23}) - \cos(\theta_{12}) &\overset{\text{(a.1)}}{=} \left(\bar{\lambda}_1 v_1 + \bar{\lambda}_2 v_2\right)^\top v_2 - v_1^\top v_2 \overset{\text{(a.2)}}{=} \bar{\lambda}_2 + (\bar{\lambda}_1 - 1) v_1^\top v_2 \\ &\overset{\text{(a.3)}}{=} \bar{\lambda}_2 + (\bar{\lambda}_1 - 1)\frac{1-\bar{\lambda}_1^2-\bar{\lambda}_2^2}{2\bar{\lambda}_1\bar{\lambda}_2} = \frac{(\bar{\lambda}_1+1)(\bar{\lambda}_1+\bar{\lambda}_2-1)(\bar{\lambda}_2+1-\bar{\lambda}_1)}{2\bar{\lambda}_1\bar{\lambda}_2} \overset{\text{(a.4)}}{>} 0, \end{aligned}$$

where (a.1) and (a.2) hold because v_1, v_2, and $\bar{v}_3$ are unit vectors; (a.3) leverages (84) to rewrite $v_1^\top v_2$; (a.4) uses the fact that $\bar{\lambda}_1$, $\bar{\lambda}_1 + \bar{\lambda}_2 - 1$, and $\bar{\lambda}_2 + 1 - \bar{\lambda}_1$ are positive, because $\bar{\lambda}_2 + 1 - \bar{\lambda}_1 > \bar{\lambda}_2 > 0$. It follows from the monotonicity of the cosine function on $[0, \pi]$ that $\theta_{23} < \theta_{12}$. ■

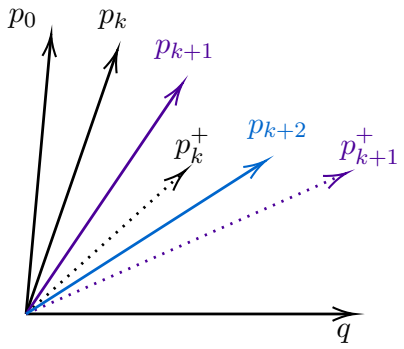


Figure 9: This figure illustrates the relative positions of p_k, p_{k+1}, and p_k^+ and the trend that p_k becomes closer to q as k increases, given $p_0^\top q > 0$.

Lemma 23 formalizes the geometric intuition that if a vector lies in a conic region bounded by two vectors forming an acute angle, then this vector is closer to one of the boundary vectors than the two boundary vectors are to each other. We provide the proof of Theorem 12 as follows.

Proof We divide the proof into three parts. First, we prove the existence and uniqueness of the trajectory limit $p_{\text{ss}} = h_{\text{ss}}(q, p_0)$ of the dynamics (32). We then derive the steady-state sensitivity associated with $h_{\text{ss}}(q, p_0)$. Finally, we quantify how the coefficients λ and σ in (32) influence the closeness between p_{ss} and q (or $-q$), an indicator of the steering ability of the decision-maker.

Existence and uniqueness of the limit p_{ss}: Note that the initial state p_0 is nonzero, because $\|p_0\| = 1$. We first address the special case in Fig. 2a, i.e., $p_0^\top q = 0$. The dynamics (32) indicate that $p_k = p_0 \neq 0, \forall k \in \mathbb{N}$. Therefore, the limiting state exists and is unique: it equals p_0. In another special case where p_0 and q are parallel, the unique limit is either $q/\|q\|$ or $-q/\|q\|$, depending on whether p_0 and q point in the same or opposite directions.

We proceed to analyze the case in Fig. 2b, i.e., $p_0^\top q > 0$ and p_0 and q are not parallel. By (32), for any $k \in \mathbb{N}$, p_k lies in the linear span of p_0 and q. Let $\operatorname{sgn}(\cdot)$ denote the sign function, i.e., $\operatorname{sgn}(x)$ equals -1, 0, or 1 if $x < 0$, $x = 0$, or $x > 0$, respectively. The initial condition satisfies $\operatorname{sgn}(p_0^\top q) = 1$. Suppose that for a particular $k \in \mathbb{N}$, $p_k^\top q > 0$ holds. It follows from the dynamics (32) that

$$\operatorname{sgn}(p_{k+1}^\top q) = \operatorname{sgn}(\tilde{p}_{k+1}^\top q) = \operatorname{sgn}\left((\lambda + \sigma\|q\|^2) p_k^\top q + (1-\lambda) p_0^\top q\right) \overset{\text{(a.1)}}{=} 1, \tag{85}$$

where (a.1) uses the base case $p_0^\top q > 0$ and the induction hypothesis $p_k^\top q > 0$. The above derivation is similar to Dean and Morgenstern (2022, Proof of Proposition 1). Hence, $p_{k+1}^\top q > 0$ also holds. Therefore, for any $k \in \mathbb{N}$, $p_k^\top q > 0$, implying p_k is nonzero and that p_k and q always form an acute angle. Let $p_k^+ \triangleq (1-\lambda)p_0 + \sigma \cdot (p_k^\top q) q$ be an intermediate vector. We now prove the following proposition on the angles related to p_k, p_{k+1}, and p_k^+ illustrated by Fig. 9.

Proposition 24. *If $p_0^\top q > 0$ and p_0 and q are not parallel, then the dynamics* (32) *ensure that for $k \in \mathbb{N}$, the angle between p_k^+ and q is smaller than that between p_{k+1} and q, which in turn is smaller than that between p_k and q.*

In the base case when $k = 0$, p_0^+ is a strict conic combination of p_0 and q, and p_1 is a strict conic combination of p_0 and p_0^+. Therefore, Lemma 23 ensures that Proposition 24 holds. Suppose that

Proposition 24 is true for a particular $k \in \mathbb{N}$. Because p_k and p_{k+1} are of unit norm and the angle between p_{k+1} and q is smaller than that between p_k and q, we know that $0 < p_k^\top q < p_{k+1}^\top q$. Hence,

$$p_{k+1}^+ = (1-\lambda)p_0 + \sigma \cdot (p_{k+1}^\top q)q \overset{\text{(a.1)}}{=} p_k^+ + \underbrace{\sigma \cdot (p_{k+1}^\top q - p_k^\top q)}_{>0} q,$$

where (a.1) uses the expression of p_k^+. Since p_{k+1}^+ is a strict conic combination of p_k^+ and q, we know from Lemma 23 that the angle between p_{k+1}^+ and q is smaller than that between p_k^+ and q, which is in turn smaller than that between p_{k+1} and q (cf. the induction hypothesis). Finally, p_{k+2} lies in the convex cone formed by p_{k+1} and p_{k+1}^+. Hence, Proposition 24 also holds true for $k+1$. Therefore, Proposition 24 is proved.

A consequence of Proposition 24 is that the angle between p_k and q is monotonically decreasing as k increases. Additionally, this angle is bounded from below by zero, because p_k always lies in the convex cone formed by p_0 and q. The monotone convergence theorem, together with the fact that p_k stays on the unit sphere inside the cone spanned by p_0 and q, ensures that there exists a unique limit p_{ss} given a fixed decision q and an initial position p_0.

The analysis for the case in Fig. 2c (i.e., $p_0^\top q < 0$) is similar to the above analysis for Fig. 2b (i.e., $p_0^\top q > 0$) by considering the convex cone formed by p_0 and $-q$, or put differently, the angle between p_k and $-q$ for different k.

Derivation of the steady-state sensitivity: The trajectory limit p_{ss} of the dynamics (32) satisfies

$$\widetilde{p}_{\text{ss}} = \lambda p_{\text{ss}} + (1-\lambda)p_0 + \sigma \cdot (p_{\text{ss}}^\top q)q \triangleq \tilde{f}(p_{\text{ss}}, q, p_0), \tag{86a}$$

$$p_{\text{ss}} = \frac{\widetilde{p}_{\text{ss}}}{\|\widetilde{p}_{\text{ss}}\|} = \frac{\tilde{f}(p_{\text{ss}}, q, p_0)}{\|\tilde{f}(p_{\text{ss}}, q, p_0)\|} \triangleq f(p_{\text{ss}}, q, p_0). \tag{86b}$$

We focus on the general case and consider a pair (q, p_0) with $p_0^\top q \neq 0$. We have proved that p_{ss} and $\widetilde{p}_{\text{ss}}$ are nonzero; see the reasoning about (85). Further, f in (86b) is well-defined and differentiable in an open neighborhood of (q, p_0). Hence,

$$\nabla_q f(p_{\text{ss}}, q, p_0) = \sigma \left[(p_{\text{ss}}^\top q)I + p_{\text{ss}} q^\top\right] \left(\frac{1}{\|\widetilde{p}_{\text{ss}}\|} I - \frac{\widetilde{p}_{\text{ss}} \widetilde{p}_{\text{ss}}^\top}{\|\widetilde{p}_{\text{ss}}\|^3}\right) = \frac{\sigma}{\|\widetilde{p}_{\text{ss}}\|} \left[(p_{\text{ss}}^\top q)I + p_{\text{ss}} q^\top\right] \left(I - p_{\text{ss}} p_{\text{ss}}^\top\right),$$

$$\nabla_p f(p_{\text{ss}}, q, p_0) = (\lambda I + \sigma q q^\top) \frac{I - p_{\text{ss}} p_{\text{ss}}^\top}{\|\widetilde{p}_{\text{ss}}\|}.$$

We further apply (5) and obtain the steady-state sensitivity as follows

$$\begin{aligned}\nabla_q h_{\text{ss}}(q, p_0) &= -\nabla_q f(p_{\text{ss}}, q, p_0) \left[\nabla_p f(p_{\text{ss}}, q, p_0) - I\right]^{-1} \\ &= -\sigma \left(p_{\text{ss}}^\top q I + p_{\text{ss}} q^\top\right) (I - p_{\text{ss}} p_{\text{ss}}^\top) \left[(\lambda I + \sigma q q^\top)(I - p_{\text{ss}} p_{\text{ss}}^\top) - \|\widetilde{p}_{\text{ss}}\| I\right]^{-1}.\end{aligned}$$

The influence of coefficients: We quantify how the coefficients λ and σ in (32) determine the steering ability of the decision-maker, which is reflected by the angle $\bar{\theta}$ between p_{ss} and q (or $-q$). We consider the case where $p_0^\top q > 0$ and p_0 and q are not parallel, and similar reasoning applies to the case of $p_0^\top q < 0$. Let $\phi \in (0, 90^\circ)$ denote the angle between p_0 and q. Note that p_0 and p_{ss} are of unit norm. Since p_{ss} satisfies the fixed-point equation (86), the vector $(1-\lambda)p_0 + \sigma \cdot (p_{\text{ss}}^\top q)q$ (i.e., $(1-\lambda)p_0 + \sigma \|q\| \cos(\bar{\theta}) q$) is collinear with p_{ss}; see Fig. 10. While this illustration and the following

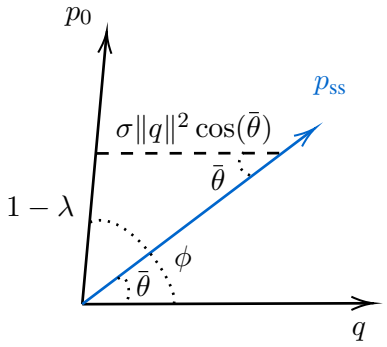


Figure 10: This figure demonstrates the relative positions of p_0, p_{ss}, and q and shows that $(1-\lambda)p_0 + \sigma \cdot (p_{\text{ss}}^\top q)q$ is collinear with p_{ss}.

analysis are for two-dimensional vectors, extensions can be derived for higher-dimensional settings. We analyze the triangle in Fig. 10 and use the law of sines to obtain

$$\frac{\sigma\|q\|^2\cos(\bar{\theta})}{\sin(\phi-\bar{\theta})} = \frac{1-\lambda}{\sin(\bar{\theta})} \quad \overset{\text{(a.1)}}{\Longrightarrow} \quad \frac{\sin(2\bar{\theta})}{\sin(\phi-\bar{\theta})} = \frac{2(1-\lambda)}{\sigma\|q\|^2},$$

where (a.1) invokes the double angle formula $\sin(2\bar{\theta}) = 2\sin(\bar{\theta})\cos(\bar{\theta})$. Note that $\sin(2\bar{\theta})/\sin(\phi-\bar{\theta})$ is an increasing function of $\bar{\theta}$. Increases in $\lambda \in (0,1)$ and $\sigma > 0$ result in a decrease of $\bar{\theta}$, implying a stronger steering ability of the decision-maker. ■

G.2 Effect of approximate sensitivity matrices

In Section 5.1, we construct an approximate sensitivity matrix as per (38) based on samples from the current transient distribution (instead of the limiting steady-state distribution) and the analytical formula (5) derived from the implicit function theorem (Dontchev and Rockafellar, 2009, Theorem 1B.1). When incorporated into the online stochastic update (e.g., (9)), this approximate sensitivity can introduce an additional gradient error. In this subsection, we characterize this effect and show that the optimality and generalization analysis in the main text still applies.

We denote the generic steady-state map by h and write $p_{\text{ss}} = h(u,d)$; in Section 5.1, $h_{\text{ss}}(q,p_0)$ plays a similar role. Similar to (5), we consider an open neighborhood of (u,d) and define the sensitivity matrix based on the random variable p as $\nabla_u h(u,d;p) \triangleq -\nabla_u f(p,u,d)\,[\nabla_p f(p,u,d) - I]^{-1}$. When evaluated at the steady state $p_{\text{ss}} = h(u,d)$, the matrix $\nabla_u h(u,d;p_{\text{ss}})$ is the exact sensitivity at (u,d). We assume that $\nabla_u h(u,d;p)$ is M_h^p-Lipschitz in p with respect to the weighted norm $\|\cdot\|_P$ and M_h^d-Lipschitz in d with respect to $\|\cdot\|$. These conditions hold when the relevant first and second derivatives of the dynamics function f in (3), together with the inverse $[\nabla_p f(p,u,d) - I]^{-1}$, are uniformly bounded in the aforementioned open neighborhood.

We analyze the setting where we use the approximate sensitivity $\nabla_u h(u_k,d;p_k)$ with p_k at time k in place of the exact sensitivity $\nabla_u h(u_k,d)$ in iterative updates. For a given decision $u_k \in \mathbb{R}^n$, let

$$\overline{\nabla}\tilde{\Phi}(u_k) \triangleq \mathbb{E}_{(p,d)\sim\gamma_k}[\nabla_u \Phi(u_k,p) + \nabla_u h(u_k,d;p)\nabla_p \Phi(u_k,p)] \tag{87}$$

be another approximate expected gradient at u_k. In (87), p and d follow the joint distribution γ_k, and the sensitivity is constructed from the random variable at time k. Recall from (8) that $\nabla\tilde{\Phi}(u_k)$ is the exact gradient of the objective $\tilde{\Phi}$ at u_k. We extend Lemma 4 and show that the resulting gradient difference $\bar{e}_k \triangleq \overline{\nabla}\tilde{\Phi}(u_k) - \nabla\tilde{\Phi}(u_k)$ admits an upper bound proportional to the Wasserstein distance $W_1(\gamma_k, \gamma_{\text{ss}}(u_k))$ between the joint distributions γ_k and $\gamma_{\text{ss}}(u_k)$, albeit with a different multiplicative constant. The following extension provides *local* characterizations, since the sensitivity formula (5) involving p_{ss} holds in an open neighborhood of (u, d).

Lemma 25. *Consider an open neighborhood of (u, d). Suppose that Assumptions 1 to 3 are satisfied. We further assume that $\nabla_u h(u, d; p)$ is M_h^p-Lipschitz in p with respect to the weighted norm $\|\cdot\|_P$ and M_h^d-Lipschitz in d with respect to $\|\cdot\|$ in this neighborhood. Then,*

$$\|\bar{e}_k\| = \|\overline{\nabla}\tilde{\Phi}(u_k) - \nabla\tilde{\Phi}(u_k)\| \leq \bar{L} W_1(\gamma_k, \gamma_{\text{ss}}(u_k)), \tag{88}$$

where $\bar{L} = \max\left(L_\Phi^p M_h^d, \frac{M_\Phi^u + L_h^u M_\Phi^p}{\sqrt{\lambda_{\min}(P)}} + L_\Phi^p M_h^p\right)$.

Proof The proof is similar to that of Lemma 4 in Appendix C.3. However, we additionally address the error of $\nabla_u h(u, d; p)$ relative to the exact sensitivity. Suppose that $\overline{\Psi}(p, d) \triangleq \nabla_u \Phi(u_k, p) + \nabla_u h(u_k, d; p)\nabla_p \Phi(u_k, p)$. Then, $\mathbb{E}_{(p,d)\sim\gamma_k}[\overline{\Psi}(p, d)] = \overline{\nabla}\tilde{\Phi}(u_k)$, and $\mathbb{E}_{(p,d)\sim\gamma_{\text{ss}}}[\overline{\Psi}(p, d)] = \nabla\tilde{\Phi}(u_k)$. For any $p_1, p_2 \in \mathbb{R}^m$ and $d_1, d_2 \in \mathbb{R}^r$,

$$\begin{aligned}
\|\overline{\Psi}(p_1, d_1) - \overline{\Psi}(p_2, d_2)\| &\leq \|\overline{\Psi}(p_1, d_1) - \overline{\Psi}(p_1, d_2)\| + \|\overline{\Psi}(p_1, d_2) - \overline{\Psi}(p_2, d_2)\| \\
&\leq \|\nabla_u h(u_k, d_1; p_1) - \nabla_u h(u_k, d_2; p_1)\| \, \|\nabla_p \Phi(u_k, p_1)\| + \|\nabla_u \Phi(u_k, p_1) - \nabla_u \Phi(u_k, p_2)\| \\
&\quad + \|\nabla_u h(u_k, d_2; p_1)\nabla_p \Phi(u_k, p_1) - \nabla_u h(u_k, d_2; p_2)\nabla_p \Phi(u_k, p_2)\| \\
&\overset{\text{(a.1)}}{\leq} L_\Phi^p M_h^d \|d_1 - d_2\| + M_\Phi^u \|p_1 - p_2\| + \|\nabla_u h(u_k, d_2; p_1)\| \|\nabla_p \Phi(u_k, p_1) - \nabla_p \Phi(u_k, p_2)\| \\
&\quad + \|\nabla_u h(u_k, d_2; p_1) - \nabla_u h(u_k, d_2; p_2)\| \|\nabla_p \Phi(u_k, p_2)\| \\
&\leq L_\Phi^p M_h^d \|d_1 - d_2\| + M_\Phi^u \|p_1 - p_2\| + L_h^u M_\Phi^p \|p_1 - p_2\| + L_\Phi^p M_h^p \|p_1 - p_2\|_P \\
&\leq \bar{L}(\|p_1 - p_2\|_P + \|d_1 - d_2\|),
\end{aligned} \tag{89}$$

where $\bar{L}$ is specified in Lemma 25. In (a.1) of (89), we leverage the Lipschitz continuity of $\nabla_u h(u_k, d; p)$, $\Phi(u_k, p)$, and $\nabla_u \Phi(u_k, p)$ and apply the triangle inequality to decompose the last term in the preceding inequality. Furthermore, we exploit Lemma 16 in Appendix A to obtain (88). Hence, Lemma 25 is proved. ■

Lemma 25 characterizes the difference between the approximate gradient $\overline{\nabla}\tilde{\Phi}(u_k)$ and the exact gradient $\nabla\tilde{\Phi}(u_k)$ arising from the use of real-time samples to evaluate the sensitivity matrix based on (5). This characterization enables the subsequent theoretical analysis in Sections 4.2 to 4.4 to be applied in a similar manner.

References

Liviu Aolaritei, Nicolas Lanzetti, Hongruyu Chen, and Florian Dörfler. Distributional uncertainty propagation via optimal transport. *IEEE Transactions on Automatic Control*, 70(9):6080–6095, 2025.

Beste Basciftci, Shabbir Ahmed, and Siqian Shen. Distributionally robust facility location problem under decision-dependent stochastic demand. *European Journal of Operational Research*, 292 (2):548–561, 2021.

Anastasia Bizyaeva, Alessio Franci, and Naomi Ehrich Leonard. Nonlinear opinion dynamics with tunable sensitivity. *IEEE Transactions on Automatic Control*, 68(3):1415–1430, 2023.

Léon Bottou, Frank E Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2):223–311, 2018.

Gavin Brown, Shlomi Hod, and Iden Kalemaj. Performative prediction in a stateful world. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 6045–6061, 2022.

Songfu Cai, Fei Han, and Xuanyu Cao. Performative control for linear dynamical systems. *Advances in Neural Information Processing Systems*, 37, 2024.

Peter E Caines. Mean field games. In *Encyclopedia of Systems and Control*, pages 1197–1202. Springer, 2021.

Xuanyu Cao, Junshan Zhang, and H Vincent Poor. Online stochastic optimization with time-varying distributions. *IEEE Transactions on Automatic Control*, 66(4):1840–1847, 2021.

Sanjay Chandrasekaran, Giulia De Pasquale, Giuseppe Belgioioso, and Florian Dörfler. Network-aware recommender system via online feedback optimization. *IEEE Transactions on Automatic Control*, 71(3):1708–1723, 2026.

Ngoc Huy Chau, Éric Moulines, Miklos Rásonyi, Sotirios Sabanis, and Ying Zhang. On stochastic gradient Langevin dynamics with dependent data streams: The fully nonconvex case. *SIAM Journal on Mathematics of Data Science*, 3(3):959–986, 2021.

Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Optimal transport in systems and control. *Annual Review of Control, Robotics, and Autonomous Systems*, 4:89–113, 2021.

Lauren Conger, Franca Hoffmann, Eric Mazumdar, and Lillian Ratliff. Strategic distribution shift of interacting agents via coupled gradient flows. *Advances in Neural Information Processing Systems*, 36:45971–46006, 2023.

Kai Cui and Heinz Koeppl. Approximately solving mean field games via entropy-regularized deep reinforcement learning. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 1909–1917, 2021.

Joshua Cutler, Dmitriy Drusvyatskiy, and Zaid Harchaoui. Stochastic optimization under distributional drift. *Journal of Machine Learning Research*, 24(147):1–56, 2023.

Alexander Davydov, Veronica Centorrino, Anand Gokhale, Giovanni Russo, and Francesco Bullo. Time-varying convex optimization: A contraction and equilibrium tracking approach. *IEEE Transactions on Automatic Control*, 70(11):7446–7460, 2025.

Sarah Dean and Jamie Morgenstern. Preference dynamics under personalized recommendations. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 795–816, 2022.

Sarah Dean, Evan Dong, Meena Jagadeesan, and Liu Leqi. Accounting for AI and users shaping one another: The role of mathematical models. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856.

Asen L Dontchev and R Tyrrell Rockafellar. *Implicit Functions and Solution Mappings*, volume 543. Springer, New York, 2009.

Dmitriy Drusvyatskiy and Lin Xiao. Stochastic optimization with decision-dependent distributions. *Mathematics of Operations Research*, 48(2):954–998, 2023.

John C Duchi and Hongseok Namkoong. Learning models with uniform performance via distributionally robust optimization. *Annals of Statistics*, 49(3):1378–1406, 2021.

Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3):707–738, 2015.

Jason Gaitonde, Jon Kleinberg, and Éva Tardos. Polarization in geometric opinion dynamics. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 499–519, 2021.

Saeed Ghadimi and Guanghui Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

Moritz Hardt and Celestine Mendler-Dünner. Performative prediction: Past and future. *Statistical Science*, 40(3):417–436, 2025.

Nicholas JA Harvey, Christopher Liaw, Yaniv Plan, and Sikander Randhawa. Tight analyses for non-smooth stochastic gradient descent. In *Proceedings of the 32nd Conference on Learning Theory*, pages 1579–1613, 2019.

Adrian Hauswirth, Zhiyu He, Saverio Bolognani, Gabriela Hug, and Florian Dörfler. Optimization algorithms as robust feedback controllers. *Annual Reviews in Control*, 57, 2024.

Jan Hązła, Yan Jin, Elchanan Mossel, and Govind Ramnarayan. A geometric model of opinion polarization. *Mathematics of Operations Research*, 49(1):251–277, 2024.

Lars Hellemo, Paul I Barton, and Asgeir Tomasgard. Decision-dependent probabilities in stochastic programs with recourse. *Computational Management Science*, 15(3):369–395, 2018.

Jiawei Huang, Niao He, and Andreas Krause. Model-based RL for mean-field games is not statistically harder than single-agent RL. In *Proceedings of the International Conference on Machine Learning*, 2024.

Zachary Izzo, Lexing Ying, and James Zou. How to learn when data reacts to your model: Performative gradient descent. In *Proceedings of the International Conference on Machine Learning*, pages 4641–4650, 2021.

Saber Jafarpour, Alexander Davydov, Anton Proskurnikov, and Francesco Bullo. Robust implicit networks via non-Euclidean contractions. *Advances in Neural Information Processing Systems*, 34:9857–9868, 2021.

Meena Jagadeesan, Tijana Zrnic, and Celestine Mendler-Dünner. Regret minimization with performative feedback. In *Proceedings of the International Conference on Machine Learning*, pages 9760–9785, 2022.

Jiashuo Jiang, Xiaocheng Li, and Jiawei Zhang. Online stochastic optimization with Wasserstein based non-stationarity. *Management Science*, 71(11):9104–9122, 2025.

H.K. Khalil. *Nonlinear Systems*. Prentice Hall, 2002.

Daniel Kuhn, Soroosh Shafiee, and Wolfram Wiesemann. Distributionally robust optimization. *Acta Numerica*, 34:579–804, 2025.

Nicolas Lanzetti, Florian Dörfler, and Nicolò Pagan. The impact of recommendation systems on opinion dynamics: Microscopic versus macroscopic effects. In *Proceedings of the 62nd IEEE Conference on Decision and Control*, pages 4824–4829, 2023.

Mathieu Laurière, Sarah Perrin, Matthieu Geist, and Olivier Pietquin. Learning mean field games: A survey. *arXiv preprint arXiv:2205.12944*, 2022.

Naomi Ehrich Leonard, Keena Lipsitz, Anastasia Bizyaeva, Alessio Franci, and Yphtach Lelkes. The nonlinear feedback dynamics of asymmetric political polarization. *Proceedings of the National Academy of Sciences*, 118(50):e2102149118, 2021.

Qiang Li and Hoi-To Wai. State dependent performative prediction with stochastic approximation. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 3164–3186, 2022.

Qiang Li, Chung-Yiu Yau, and Hoi-To Wai. Multi-agent performative prediction with greedy deployment and consensus seeking agents. *Advances in Neural Information Processing Systems*, 35:38449–38460, 2022.

Shaojie Li and Yong Liu. High probability guarantees for nonconvex stochastic gradient descent with heavy tails. In *Proceedings of the International Conference on Machine Learning*, pages 12931–12963, 2022.

Xiaoyu Li and Francesco Orabona. A high probability analysis of adaptive SGD with momentum. In *Workshop on Beyond First-Order Methods in ML Systems at the International Conference on Machine Learning*, 2020.

Fengqiao Luo and Sanjay Mehrotra. Distributionally robust optimization with decision dependent ambiguity sets. *Optimization Letters*, 14(8):2565–2594, 2020.

Liam Madden, Emiliano Dall'Anese, and Stephen Becker. High probability convergence bounds for non-convex stochastic gradient descent with sub-Weibull noise. *Journal of Machine Learning Research*, 25(241):1–36, 2024.

Debmalya Mandal, Stelios Triantafyllou, and Goran Radanovic. Performative reinforcement learning. In *Proceedings of the International Conference on Machine Learning*, pages 23642–23680, 2023.

Wenjun Mei, Julien M Hendrickx, Ge Chen, Francesco Bullo, and Florian Dörfler. Convergence, consensus, and dissensus in the weighted-median opinion dynamics. *IEEE Transactions on Automatic Control*, 69(10):6700–6714, 2024.

Celestine Mendler-Dünner, Juan Perdomo, Tijana Zrnic, and Moritz Hardt. Stochastic optimization for performative prediction. *Advances in Neural Information Processing Systems*, 33:4929–4939, 2020.

John P Miller, Juan C Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. In *Proceedings of the International Conference on Machine Learning*, pages 7710–7720, 2021.

Adhyyan Narang, Evan Faulkner, Dmitriy Drusvyatskiy, Maryam Fazel, and Lillian J Ratliff. Multiplayer performative prediction: Learning in decision-dependent games. *Journal of Machine Learning Research*, 24(202):1–56, 2023.

Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *Proceedings of the International Conference on Machine Learning*, pages 7599–7609, 2020.

Georgios Piliouras and Fang-Yi Yu. Multi-agent performative prediction: From global stability and optimality to chaos. In *Proceedings of the 24th ACM Conference on Economics and Computation*, pages 1047–1074, 2023.

Anton V Proskurnikov and Roberto Tempo. A tutorial on modeling and analysis of dynamic social networks. Part I. *Annual Reviews in Control*, 43:65–79, 2017.

Mitas Ray, Lillian J Ratliff, Dmitriy Drusvyatskiy, and Maryam Fazel. Decision-dependent risk minimization in geometrically decaying dynamic environments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8081–8088, 2022.

Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2021.

Andrea Simonetto, Emiliano Dall'Anese, Santiago Paternain, Geert Leus, and Georgios B Giannakis. Time-varying convex optimization: Time-structured algorithms and applications. *Proceedings of the IEEE*, 108(11):2032–2048, 2020.

Antonio Terpin, Nicolas Lanzetti, and Florian Dörfler. Dynamic programming in probability spaces via optimal transport. *SIAM Journal on Control and Optimization*, 62(2):1183–1206, 2024.

Hiroyasu Tsukamoto, Soon-Jo Chung, and Jean-Jacques E Slotine. Contraction theory for nonlinear stability analysis and learning-based control: A tutorial overview. *Annual Reviews in Control*, 52: 135–169, 2021.

Cédric Villani. *Optimal Transport: Old and New*, volume 338. Springer, 2009.

Andreas Wächter and Lorenz T Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106:25–57, 2006.

Killian Wood and Emiliano Dall'Anese. Stochastic saddle point problems with decision-dependent distributions. *SIAM Journal on Optimization*, 33(3):1943–1967, 2023.

Killian Wood, Gianluca Bianchin, and Emiliano Dall'Anese. Online projected gradient descent for stochastic optimization with decision-dependent distributions. *IEEE Control Systems Letters*, 6: 1646–1651, 2022.

Wenjing Yan and Xuanyu Cao. Zero-regret performative prediction under inequality constraints. *Advances in Neural Information Processing Systems*, 36:1298–1308, 2023.

Vicky Chuqiao Yang, Daniel M Abrams, Georgia Kernell, and Adilson E Motter. Why are US parties so polarized? A "satisficing" dynamical model. *SIAM Review*, 62(3):646–657, 2020.

Yan Zhang, Yi Zhou, Kaiyi Ji, and Michael M Zavlanos. A new one-point residual-feedback oracle for black-box learning and control. *Automatica*, 136:110006, 2022.