

Parameter-Free Algorithms for Performative Regret Minimization under Decision-Dependent Distributions

Sungwoo Park¹

S.PARK@KAIST.AC.KR

Junyeop Kwon¹

JUNYEOPK@KAIST.AC.KR

Byeongnoh Kim²

B-N.KIM@SAMSUNG.COM

Suhyun Chae²

SUHYUN.CHAE@SAMSUNG.COM

Jeeyong Lee²

JIYONG.LEE@SAMSUNG.COM

Dabeen Lee^{1,†}

DABEENL@KAIST.AC.KR

¹*Department of Industrial and Systems Engineering, KAIST, Daejeon 34141, South Korea*

²*Device Solutions Research, Samsung Electronics, Hwaseong, Gyeonggi 18448, South Korea*

[†] *corresponding author*

Abstract

This paper studies performative risk minimization, a formulation of stochastic optimization under decision-dependent distributions. We consider the general case where the performative risk can be non-convex, for which we develop efficient parameter-free optimistic optimization-based methods. Our algorithms significantly improve upon the existing Lipschitz bandit-based method in many aspects. In particular, our framework does not require knowledge about the sensitivity parameter of the distribution map and the Lipschitz constant of the loss function. This makes our framework practically favorable, together with the efficient optimistic optimization-based tree-search mechanism. We provide experimental results that demonstrate the numerical superiority of our algorithms over the existing method and other black-box optimistic optimization methods.

Keywords: Decision-Dependent Distributions, Performative Risk Minimization, Optimistic Optimization, Black-Box Optimization, Stochastic Non-Convex Optimization

1 Introduction

In the realm of stochastic optimization, where navigating uncertainty is paramount, distributional shifts stand out as a significant challenge. Among the various sources of these shifts, one particularly intriguing phenomenon stems from feedback mechanisms intricately linked to decision-making processes. This feedback loop alters the distribution that governs the stochastic environment of the system, creating a dynamic landscape where decisions shape and are shaped by distributions. For example, the decisions made by a dynamic resource allocation algorithm for a renewable energy grid not only influence the immediate allocation of resources but also affect the underlying distribution of factors like energy demand and supply. Classifiers, such as insurance underwriting systems, often promote a shift in behavior within the population to improve their labels. Predictions of stock prices wield significant influence over trading decisions. Moreover, election predictions have the potential to shape and influence voter behavior, which in turn can impact voting results.

Decision-making processes under such phenomena can be formulated as stochastic optimization under *decision-dependent distributions*. [Perdomo et al. \(2020\)](#) proposed the notion of the *distribution map* to consider decision-dependent distributions for stochastic optimization models. That is, the distribution $\mathcal{D}(\theta)$ of the parameter z capturing the stochastic environment depends on the decision θ . Here, θ may encode the resource allocation decision for a renewable energy grid and the election prediction, in which case z corresponds to the energy demand and the voting results, respectively. For machine learning, we can associate θ with predictive models and z with data. Then the objective is to minimize the *performative risk* under a loss function f , defined as

$$\text{PR}(\theta) := \mathbb{E}_{z \sim \mathcal{D}(\theta)} [f(\theta, z)].$$

The expression *performative* comes from the term *performative prediction* ([Perdomo et al., 2020](#)), which implies the phenomenon where predictions influence the outcomes. The goal of this paper is to design an efficient algorithmic framework for minimizing the performative risk which models stochastic optimization under decision-dependent distributions.

1.1 Existing Methods for Performative Risk Minimization

Unlike the standard stochastic optimization problem, the decision θ may affect the underlying distribution $\mathcal{D}(\theta)$. Hence, a natural starting point to minimize the performative risk is to consider the following iterative algorithm, referred to as *repeated risk minimization* (RRM). Given an initial solution $\theta_0 \in \Theta$ where Θ is the domain, we apply

$$\theta_{t+1} \in \underset{\theta \in \Theta}{\operatorname{argmin}} \mathbb{E}_{z \sim \mathcal{D}(\theta_t)} [f(\theta, z)] \tag{RRM}$$

for $t \geq 0$. Here, computing the next iterate θ_{t+1} requires solving a stochastic optimization instance where the underlying distribution is fixed with $\mathcal{D}(\theta_t)$. Another approach is a gradient-based method such as

$$\theta_{t+1} = \theta_t - \eta_t \mathbb{E}_{z \sim \mathcal{D}(\theta_t)} [\nabla f(\theta_t, z)] \tag{RGD}$$

where η_t is a step size, and we refer to this procedure as *repeated gradient descent* (RGD). Note that running RGD, as well as RRM, is based on access to the distribution $\mathcal{D}(\theta_t)$ for every iteration $t \geq 0$, which may not be feasible in practice. A more sample-efficient method is to apply the standard *stochastic gradient descent* (SGD) update, given by

$$\theta_{t+1} = \theta_t - \eta_t \nabla f(\theta_t, z_t) \quad \text{where } z_t \sim \mathcal{D}(\theta_t). \tag{SGD}$$

[Drusvyatskiy and Xiao \(2023\)](#) analyzed variants of SGD such as stochastic proximal gradient, proximal point, clipped gradient, and accelerated gradient methods.

Convergence of these iterative methods has been established; ([Perdomo et al., 2020](#)) for RRM, ([Perdomo et al., 2020](#); [Mendler-Dünner et al., 2020](#)) for RGD, and ([Mendler-Dünner et al., 2020](#)) for SGD. They showed convergence to a *performatively stable* solution, under some strong convexity and smoothness assumptions on the loss function f . Here, we say that a solution θ_{PS} is performatively stable if it satisfies

$$\theta_{\text{PS}} \in \underset{\theta \in \Theta}{\operatorname{argmin}} \mathbb{E}_{z \sim \mathcal{D}(\theta_{\text{PS}})} [f(\theta, z)]$$

In particular, θ_{PS} is a fixed point of RRM. However, let alone the validity of the structural assumptions on the loss functions, the performatively stable solution θ_{PS} is in general not a minimizer of the performative risk (Perdomo et al., 2020; Miller et al., 2021). Let θ_{PO} denote a minimizer of the performative risk, i.e.,

$$\theta_{\text{PO}} \in \operatorname{argmin}_{\theta \in \Theta} \mathbb{E}_{z \sim \mathcal{D}(\theta)} [f(\theta, z)].$$

It turns out that $\text{PR}(\theta_{\text{PS}})$ can be arbitrarily large compared to $\text{PR}(\theta_{\text{PO}})$ (Miller et al., 2021).

Derivative-free zeroth-order optimization methods have been proposed to minimize the performative risk directly (Izzo et al., 2021; Miller et al., 2021; Izzo et al., 2022; Ray et al., 2022). For the derivative-free methods to minimize the performative risk $\text{PR}(\theta)$, the requirement, however, is that $\text{PR}(\theta)$ is a convex function of the decision θ . Miller et al. (2021) provided some sufficient conditions on the distribution map to guarantee the convexity of the performative risk. They argued that if the distribution map satisfies a certain stochastic dominance condition, which is related to *stochastic orders* (Shaked and Shanthikumar, 2007), then convexity of the loss function f leads to a convex performative risk. Nevertheless, as noted by Perdomo et al. (2020), the performative risk is non-convex in general even if the loss function is convex.

To tackle the general case of non-convex performative risk, Jagadeesan et al. (2022) developed a bandit optimization-based algorithm. The problem of minimizing the performative risk is indeed a bandit optimization problem because until the decision θ is deployed it is difficult to estimate the distribution $\mathcal{D}(\theta)$ and thus the performative risk $\text{PR}(\theta)$. That said, the framework of Jagadeesan et al. (2022) is inspired by the zooming algorithm (Zooming) for Lipschitz bandits due to Kleinberg et al. (2008). The core idea is to adaptively discretize the solution space Θ thereby narrowing down the location of the optimal decision θ_{PO} . In fact, if the loss function f is L_θ -Lipschitz continuous in θ and L_z -Lipschitz continuous in z , the ε -sensitivity of the distribution map (defined formally in Section 3) implies that $\text{PR}(\theta)$ is $(L_\theta + L_z\varepsilon)$ -Lipschitz continuous (Jagadeesan et al., 2022). Here, the ε -sensitivity measures how much the distributions $\mathcal{D}(\theta)$ and $\mathcal{D}(\theta')$ can differ for two distinct decisions θ and θ' . Then applying Zooming directly on $\text{PR}(\theta)$ would guarantee a sublinear regret.

Although this lays down a good starting point, direct application of Zooming fails to utilize the fact that the feedback obtained after deploying decision θ is $\mathcal{D}(\theta)$, based on which the learner can evaluate $\text{PR}(\theta)$ but also infer the distribution $\mathcal{D}(\theta')$ of other solutions θ' using the ε -sensitivity. Jagadeesan et al. (2022) referred to this as *performative feedback*. Building on this idea, they developed a variant of Zooming, and they provided a regret upper bound that is parameterized by not $L_\theta + L_z\varepsilon$ but $L_z\varepsilon$. Here, note that $L_z\varepsilon$ vanishes as $\varepsilon \rightarrow 0$ while $L_\theta + L_z\varepsilon$ does not. Moreover, $L_z\varepsilon$ does not depend on L_θ , so the algorithm works even when L_θ is not bounded. The two main components of their algorithm are adaptive discretization and sequential elimination based on *performative confidence bounds* which we explain in Section 3.

The algorithm of Jagadeesan et al. (2022) solves performative risk minimization, but several issues hinder its practical implementation. First, to implement the adaptive discretization procedure, we need to know the Rademacher complexity $\mathfrak{C}^*(f)$ of learning the performative risk $\text{PR}(\theta)$ under the loss function f based on data samples from distribution

$\mathcal{D}(\theta)$. The Rademacher complexity parameter $\mathfrak{C}^*(f)$ can be very high depending on the structure of f . Second, to build a performative confidence bound, we need the Lipschitz constant L_z and the sensitivity parameter ε . One may argue that there is a way of estimating the Lipschitz constant L_z for a known class of loss functions, but the sensitivity parameter ε determines the global landscape of the distribution map, which means that it would be difficult to measure ε in advance. Third, one iteration of the algorithm is computationally expensive. This is because each time a decision θ is deployed, we need to compute the performative confidence bound for every solution $\theta' \in \Theta$ remaining in the search space. Such an issue is inherent in Lipschitz bandit-based methods. Although [Jagadeesan et al. \(2022\)](#) did not demonstrate an implementation of their algorithm, our numerical results in Section 6 show that the algorithm is not efficient and incurs a high regret in practice.

The aforementioned limitations of the existing method due to [Jagadeesan et al. \(2022\)](#) for performative risk minimization motivate the following question.

Can we design a practical algorithm for performative risk minimization that relies on minimal knowledge about the problem parameters?

In this paper, we devise efficient parameter-free algorithms for performative risk minimization. We not only demonstrate strong theoretical performance guarantees but also show experimental results to highlight their numerical effectiveness.

1.2 Our Contributions

As in ([Jagadeesan et al., 2022](#)), we study the problem of minimizing the performative risk with performative feedback. The algorithm of [Jagadeesan et al. \(2022\)](#) is an adaptation of **Zooming** by [Kleinberg et al. \(2008\)](#), and as a result, it requires knowledge of problem parameters such as the Rademacher complexity $\mathfrak{C}^*(f)$, the Lipschitz constant L_z , and the sensitivity parameter ε .

Our main contribution is to design practical algorithms that do not assume knowledge of the problem parameters. To develop such parameter-free algorithms, we build upon the idea of *optimistic optimization* methods that may adapt to unknown smoothness of the objective function. Here, parameter-free optimistic optimization methods originate from the simultaneous optimistic optimization (S00) algorithm ([Munos, 2011](#)) and the stochastic extension of S00 (StoS00) algorithm ([Valko et al., 2013](#)), and they are devised to optimize black-box objective functions that are possibly non-convex.

Our algorithms are inspired by two more recent optimistic optimization-based parameter-free methods due to [Bartlett et al. \(2019\)](#), **Sequ00L** for the deterministic evaluation case and **Stroqu00L** for the noisy case. We start by considering the conceptual setting where the distribution $\mathcal{D}(\theta)$ associated with the deployed decision θ can be fully observed. We call this case the *full-feedback* setting. Next, we study the more practically relevant setting where we obtain a few samples from $\mathcal{D}(\theta)$ after deploying decision θ , and we refer to this case as the *data-driven* setting. We develop our algorithms for the full-feedback setting and the data-driven setting based on **Sequ00L** and **Stroqu00L**, respectively.

To highlight our results early in this paper, let us provide an informal summary of our main theorems. The following states a performance guarantee for the full-feedback case.

Theorem 1 (Full-Feedback Case, Informal) *Suppose that the distribution map satisfies the ε -sensitivity condition and the loss function $f(\theta, z)$ is L_z -Lipschitz continuous in z for any θ . Let d denote the $L_z\varepsilon$ -near-optimality dimension. For the full-feedback setting, Algorithm 1 after T decision deployments for a sufficiently large T finds a solution θ with*

$$\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) = \begin{cases} L_z\varepsilon \cdot 2^{O\left(-\frac{T}{\log T}\right)}, & \text{if } d = 0, \\ \tilde{O}\left(L_z\varepsilon \cdot T^{-\frac{1}{d}}\right), & \text{if } d > 0. \end{cases}$$

Here, the optimality gap bounds hide dependence on the ambient dimension D of the decision domain Θ . The notion of *near-optimality dimension* was first introduced by Bubeck et al. (2011a), and they argued that the near-optimality dimension and the *zooming dimension* due to Kleinberg et al. (2008) are closely related. In this paper, we use a more refined definition of the near-optimality dimension due to Grill et al. (2015).

Note that the optimality gap bounds as well as the near-optimality dimension depend on parameters L_z and ε but not on L_θ . In fact, direct application of Sequ00L to minimize the performative risk $\text{PR}(\theta)$ would result in dependence on $L_\theta + L_z\varepsilon$ as the Lipschitz constant of $\text{PR}(\theta)$ is $L_\theta + L_z\varepsilon$ (Jagadeesan et al., 2022). More precisely, we would need the $(L_\theta + L_z\varepsilon)$ -near-optimality dimension, and the resulting bounds would be $(L_\theta + L_z\varepsilon)2^{O(-T/\log T)}$ and $\tilde{O}((L_\theta + L_z\varepsilon)T^{-1/d})$. The important distinction is that our optimality gap bounds vanish as ε becomes arbitrarily small, which setting corresponds to the standard stochastic optimization problem with decision-agnostic distributions. Moreover, the $L_z\varepsilon$ -near-optimality dimension is always less than or equal to the $(L_\theta + L_z\varepsilon)$ -near-optimality dimension. Another aspect to highlight in Theorem 6 is that when $d = 0$, we show that the optimality gap decays at an exponentially fast rate, which was not discovered by Jagadeesan et al. (2022).

Next, we state our performance guarantee for the data-driven setting where we receive a finite number of samples from $\mathcal{D}(\theta)$ after deploying decision θ . The optimality gap bounds on Algorithm 2 hide dependence on the dimension D , the number of samples received after each decision deployment, the Rademacher complexity of learning the performative risk.

Theorem 2 (Data-Driven Case, Informal) *Assume the same conditions on the distribution map and the loss function. Let d denote the $L_z\varepsilon$ -near-optimality dimension. For the data-driven setting, Algorithm 2 after T decision deployments for a sufficiently large T finds a solution θ such that with high probability,*

$$\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) = \begin{cases} L_z\varepsilon \cdot 2^{O\left(-\frac{T}{\log T}\right)}, & \text{low-noise regime with } d = 0, \\ \tilde{O}\left(L_z\varepsilon \cdot T^{-\frac{1}{d}}\right), & \text{low-noise regime with } d > 0, \\ \tilde{O}\left(T^{-\frac{1}{2}} + (L_z\varepsilon)^{\frac{d}{d+2}} \cdot T^{-\frac{1}{d+2}}\right), & \text{high-noise regime.} \end{cases}$$

The *low-noise* and *high-noise* regimes are defined in Section 5. In particular, for the high-noise regime, the bound is $\tilde{O}(T^{-1/2} + L_z\varepsilon T^{-1/d})$ which incurs the additional term $T^{-1/2}$ due to errors in estimating the performative risk through noisy feedback. In particular, the case of $\varepsilon = 0$ is under the high-noise regime, in which case the bound reduces to $\tilde{O}(T^{-1/2})$.

Lastly, we test the numerical performance of our framework on instances in which the associated performative risk is non-convex. The experimental results show that our algorithms outperform the existing methods that include not only the sequential zooming

algorithm of Jagadeesan et al. (2022) but also S00 (Munos, 2011), StoS00 (Valko et al., 2013), Sequ00L, and Stroqu00L (Bartlett et al., 2019) applied directly to the performative risk as a black-box function without utilizing the performative feedback.

2 Related Work

This section summarizes prior work on performative prediction and optimistic optimization.

2.1 Performative Prediction and Performative Risk Minimization

Previous work on performative prediction has mainly focused on first-order and zeroth-order gradient-based optimization methods (Perdomo et al., 2020; Mendler-Dünner et al., 2020; Drusvyatskiy and Xiao, 2023; Brown et al., 2022; Miller et al., 2021; Izzo et al., 2021; Maheshwari et al., 2022; Li and Wai, 2022; Ray et al., 2022; Dong et al., 2023; Izzo et al., 2022). Convergence of these gradient-based methods to performatively stable solutions is studied, and Miller et al. (2021); Izzo et al. (2021) discovered some convexity conditions under which some gradient-based methods converge to a performative optimal solution. Although a performatively stable solution provides a good proxy for a performatively optimal solution, its performance can be arbitrarily worse than the optimum. Moreover, in general, the performative risk is non-convex and does not satisfy the convexity conditions. For the general case, Jagadeesan et al. (2022) developed a variant of Zooming for minimizing the performative risk. Mofakhami et al. (2023) studied RRM for training non-convex neural networks, but they considered a different setting in terms of defining the ε -sensitivity of the distribution map. For a comprehensive survey on performative prediction, we refer the reader to Hardt and Mendler-Dünner (2023) and references therein.

One of the most closely related application domains is *strategic classification* (Dalvi et al., 2004; Brückner et al., 2012; Hardt et al., 2016), which models a game between an institution deploying a classifier and an agent who adapts its features to increase its likelihood of being positively labeled. Recent work in this area includes (Dong et al., 2018; Chen et al., 2020; Milli et al., 2019; Bechavod et al., 2021; Zrnic et al., 2021).

2.2 Optimistic Optimization

Black-box optimization and *continuum-armed bandits* aim to optimize an objective function under minimal knowledge about the function. Some early work provides algorithms that assume some weak or local smoothness conditions around a global optimal solution, such as Zooming (Kleinberg et al., 2008), H00 (Bubeck et al., 2011a), D00 (Munos, 2011), HCT (azar et al., 2014). Here, H00, D00, and HCT are optimistic optimization-based methods, which means that these algorithms use some optimistic estimates of the black-box objective function when running a global search of the solution space. However, Zooming, H00, D00, and HCT require the knowledge of the local smoothness parameter. Then Munos (2011) presented S00 that works even when the local smoothness parameter is unknown. Valko et al. (2013) developed StoS00 which extends S00 for the case of stochastic function evaluation, but its convergence guarantee holds for the limited case of the near-optimality dimension being 0. P00 due to Grill et al. (2015) and GP0, PCT developed by Shang et al. (2019) work for more general families of objective functions. Later, Bartlett et al. (2019) presented Sequ00L for

the deterministic function evaluation case and **Stroqu00L** for the stochastic case, which work for general families of functions and exhibit state-of-the-art numerical performance.

Recently, [Li et al. \(2023b\)](#) provided **VHCT**, which does not require the budget on the number of decision deployments beforehand but needs the knowledge of the smoothness parameter. There exist more algorithms that work under more specific assumptions on smoothness. For example, **DiRect** ([Jones et al., 1993](#)) and methods for continuum-armed bandits due to ([Slivkins, 2011](#); [Bubeck et al., 2011b](#); [Malherbe and Vayatis, 2017](#)) can take Lipschitz-continuous objective functions.

3 Preliminaries: Optimization with Performative Feedback

In this section, we introduce the basics of performative prediction. Then we explain how to make use of performative feedback for performative risk minimization as established by [Jagadeesan et al. \(2022\)](#). In addition, we elaborate briefly on some limitations of the performative confidence bound-based zooming algorithm by [Jagadeesan et al. \(2022\)](#).

As mentioned in the introduction, the ε -sensitivity measures how much the distribution $\mathcal{D}(\theta)$ can change with changes in decision θ . Formally, we assume that the distribution map satisfies the following. Recall that Θ denotes the decision domain.

Assumption 1 (ε -sensitivity) *A distribution map $\mathcal{D}(\cdot)$ is ε -sensitive with $\alpha > 0$ if for any $\theta, \theta' \in \Theta$ we have*

$$\mathcal{W}(\mathcal{D}(\theta), \mathcal{D}(\theta')) \leq \varepsilon \|\theta - \theta'\|^\alpha,$$

where \mathcal{W} denotes the 1-Wasserstein distance.

The original definition due to [Perdomo et al. \(2020\)](#) considers the case $\alpha = 1$, while our framework allows arbitrary positive values of α . We remark that our framework is parameter-free in that we do not require knowledge of the parameters ε and α in advance. In theory, as we build upon optimistic optimization methods, we may take any *semi-metric* ℓ , satisfying $\ell(\theta, \theta') = \ell(\theta', \theta)$ and $\ell(\theta, \theta') = 0$ if and only if $\theta = \theta'$ for $\theta, \theta' \in \Theta$. That being said, we may run our algorithms regardless of the sensitivity structure of the distribution map, but we derive the theoretical performance guarantees based on the sensitivity structure given in Assumption 1.

Next, we define the notion of performative feedback used to infer the distribution $\mathcal{D}(\theta)$ as well as the performative risk $\text{PR}(\theta)$ after deploying decision θ .

Assumption 2 (performative feedback) *Deploying decision θ once, we receive feedback about the distribution as follows.*

- (Full-Feedback Setting) distribution $\mathcal{D}(\theta)$ itself.
- (Data-Driven Setting) m_0 i.i.d. samples $z_\theta^{(1)}, \dots, z_\theta^{(m_0)}$ from distribution $\mathcal{D}(\theta)$.

For the data-driven setting, we may deploy the same decision θ multiple times, say n . Then we may construct the empirical distribution $\widehat{\mathcal{D}}(\theta)$ with nm_0 i.i.d. samples from $\mathcal{D}(\theta)$. Using the performative feedback for θ , which provides $\mathcal{D}(\theta)$ or $\widehat{\mathcal{D}}(\theta)$, we may compute the performative risk $\text{PR}(\theta)$ or its empirical estimate

$$\widehat{\text{PR}}(\theta) = \mathbb{E}_{z \sim \widehat{\mathcal{D}}(\theta)} [f(\theta, z)].$$

Moreover, based on the performative feedback for θ , we may infer the performative risk of other decisions θ' . To be specific, we use the notion of *decoupled performative risk* (Perdomo et al., 2020) defined as follows.

$$\text{DPR}(\theta, \theta') = \mathbb{E}_{z \sim \mathcal{D}(\theta)} [f(\theta', z)] \quad \text{and} \quad \widehat{\text{DPR}}(\theta, \theta') = \mathbb{E}_{z \sim \widehat{\mathcal{D}}(\theta)} [f(\theta', z)]$$

for any $\theta, \theta' \in \Theta$ where $\text{DPR}(\theta, \theta')$ is the decoupled performative risk of decision θ' under distribution $\mathcal{D}(\theta)$ and $\widehat{\text{DPR}}(\theta, \theta')$ is its empirical estimate. The decoupled performative risk offers a good approximation of the performative risk, which we elaborate on below.

Assumption 3 *There is some $L_z > 0$ such that $f(\theta, \cdot)$ for any fixed $\theta \in \Theta$ is L_z -Lipschitz continuous.*

Under Assumptions 1 and 3, the Kantorovich-Rubinstein duality theorem (Kantorovich and Rubinstein, 1958; Villani, 2008) implies the following statement (Jagadeesan et al., 2022).

Lemma 3 *Under Assumptions 1 and 3, for $\theta, \theta' \in \Theta$,*

$$|\text{PR}(\theta') - \text{DPR}(\theta, \theta')| \leq L_z \varepsilon \|\theta - \theta'\|^\alpha.$$

Therefore, as long as decisions θ and θ' are close, the decoupled performative risk $\text{DPR}(\theta, \theta')$ deduced based on the performative feedback $\mathcal{D}(\theta)$ for θ would be a good proxy for the performative risk $\text{PR}(\theta')$ of decision θ' . By Lemma 3,

$$\text{DPR}(\theta, \theta') - L_z \varepsilon \|\theta - \theta'\|^\alpha \leq \text{PR}(\theta') \leq \text{DPR}(\theta, \theta') + L_z \varepsilon \|\theta - \theta'\|^\alpha$$

is a valid confidence interval for the performative risk of $\theta' \in \Theta$. Note that the confidence interval is tighter than the interval

$$\text{PR}(\theta) - L_\theta \|\theta - \theta'\| - L_z \varepsilon \|\theta - \theta'\|^\alpha \leq \text{PR}(\theta') \leq \text{PR}(\theta) + L_\theta \|\theta - \theta'\| + L_z \varepsilon \|\theta - \theta'\|^\alpha$$

which holds under the assumption that $f(\cdot, z)$ is L_θ -Lipschitz continuous for any fixed $z \in \mathcal{Z}$ where \mathcal{Z} denotes the domain of the stochastic parameter z (Jagadeesan et al., 2022). Here, the latter interval can be deduced by a black-box evaluation of $\text{PR}(\theta)$ while we derived the former using the performative feedback.

When we have a set \mathcal{S} of multiple decisions θ with known $\mathcal{D}(\theta)$, then for any $\theta' \in \Theta$,

$$\max_{\theta \in \mathcal{S}} \{\text{DPR}(\theta, \theta') - L_z \varepsilon \|\theta - \theta'\|^\alpha\} \leq \text{PR}(\theta') \leq \min_{\theta \in \mathcal{S}} \{\text{DPR}(\theta, \theta') + L_z \varepsilon \|\theta - \theta'\|^\alpha\}$$

is also valid, and we refer to the bounds as *performative confidence bounds*. The zooming algorithm of Jagadeesan et al. (2022) updates the performative confidence bounds whenever a new decision is deployed, based on which suboptimal decisions are sequentially deleted. This approach, however, has two key limitations. First, we need to know L_z and ε to derive performative confidence bounds. Second, the computational complexity of computing the performative confidence bounds associated with \mathcal{S} is $O(|\mathcal{S}| \cdot |\Theta|)$, which is an expensive per-time cost. Later, our experimental results reveal that the algorithm turns out to be not numerically efficient.

4 Optimistic Optimization-Based Parameter-Free Framework

Motivated by the challenges of the existing method, our goal is to design an efficient parameter-free framework for performative risk minimization. For simple presentation, we assume that the loss function and the decision domain.

Assumption 4 (bounded domain and objective) $\Theta \subseteq [0, 1]^D$ where D is the ambient dimension. Moreover, $f(\theta, z) \in [0, 1]$ for all $\theta \in \Theta$ and $z \in \mathcal{Z}$.

Let us explain the basic setup of our optimistic optimization-based framework as follows. We assume that a *hierarchical partitioning* (Bubeck et al., 2011a; Munos, 2011) of the decision domain is given. Basically, a hierarchical partitioning \mathcal{P} of Θ is given by $\{\mathcal{P}_{h,i} : 0 \leq h \leq h_{\max}, i \in [I_h]\}$ where h_{\max} is the deepest depth, I_h is the width at depth h with $I_1 = 1$, $\{\mathcal{P}_{h,i} : i \in [I_h]\}$ is a partition of Θ , and $\mathcal{P}_{h,i}$ is partitioned into $\{\mathcal{P}_{h+1,j} : j \in J\}$ for some $J \subseteq [I_{h+1}]$. Throughout the paper, we refer to $\mathcal{P}_{h,i}$ as a cell of depth h . The hierarchical partitioning naturally corresponds to a tree structure. When a cell $\mathcal{P}_{h,i}$ is partitioned into $\{\mathcal{P}_{h+1,j} : j \in J\}$, $\mathcal{P}_{h,i}$ is the parent cell of its child cells $\mathcal{P}_{h+1,j}$ for $j \in J$. Moreover, we assume that the partition at each depth level consists of cells of uniform size.

Assumption 5 (uniform partition) $\sup\{\|\theta - \theta'\| : \theta, \theta' \in \mathcal{P}_{h,i}\} \leq \sqrt{D}2^{-h}$ for $0 \leq h \leq h_{\max}$ and $1 \leq i \leq I_h$. Moreover, $I_h \leq 2^{Dh}$.

There exists a hierarchical partitioning that satisfies Assumption 5 as we may take 2^D subsets of box $[0, 1]^D$ by dividing each coordinate direction equally and repeat the process for each subset.

Our framework adopts the tree-search mechanism as done for many optimistic optimization algorithms such as S00, StoS00, Sequ00L, and Stroqu00L. These algorithms select an arbitrary decision $\theta_{h,i}$ for each cell $\mathcal{P}_{h,i}$ as its representative in advance, and evaluating cell $\mathcal{P}_{h,i}$ means deploying decision $\theta_{h,i}$. If $f(\cdot, z)$ is L_θ -Lipschitz continuous for any z , then Assumption 5 implies that for any $\theta_{h,i} \in \mathcal{P}_{h,i}$,

$$\text{PR}(\theta_{h,i}) - \inf_{\theta \in \mathcal{P}_{h,i}} \text{PR}(\theta) \leq L_\theta \sqrt{D}2^{-h} + L_z \varepsilon D^{\alpha/2} 2^{-\alpha h}.$$

However, this bound may be too weak for our setting because the term $L_\theta \sqrt{D}2^{-h}$ can be much larger than the other term $L_z \varepsilon D^{\alpha/2} 2^{-\alpha h}$ when the sensitivity parameter ε is small. In contrast, based on performative feedback, we use a specific rule for choosing a representative decision $\theta_{h,i}$ given by

$$\theta_{h,i} \in \underset{\theta \in \mathcal{P}_{h,i}}{\text{argmin}} \text{DPR}(\theta_{h-1,j}, \theta).$$

Here, $\theta_{h-1,j}$ is the representative of the parent cell $\mathcal{P}_{h-1,j}$ of depth $h-1$ containing $\mathcal{P}_{h,i}$. Based on the performative feedback about decision $\theta_{h-1,j}$, we may compute the decoupled performative risk $\text{DPR}(\theta_{h-1,j}, \theta)$. Note that the procedure of choosing $\theta_{h,i}$ is much cheaper than computing performative confidence bounds because the former requires evaluating decisions in a local cell $\mathcal{P}_{h,i}$ whereas the latter considers the entire domain Θ .

Explaining the important components of our framework, we present Algorithm 1 for the full-feedback setting. $\overline{\log T}$ denotes the T -th harmonic number, that is, $\overline{\log T} = \sum_{t=1}^T 1/t$.

Algorithm 1 Deterministic Optimistic Optimization with Performative Feedback (DOOP)

Input: test budget T , hierarchical partitioning \mathcal{P} , $h_{\max} = \lfloor T/2^D \overline{\log T} \rfloor$
 Set $\mathcal{L}_0 \leftarrow \{\mathcal{P}_{0,1}\}$ and initialize $\mathcal{L}_1 \leftarrow \emptyset$
 Take a solution $\theta_{0,1} \in \mathcal{P}_{0,1}$
Run Open($\mathcal{P}_{0,1}$)
for $h = 1$ **to** h_{\max} **do**
 Initialize $\mathcal{L}_{h+1} \leftarrow \emptyset$
 Take $\lfloor h_{\max}/h \rfloor$ cells with the $\lfloor h_{\max}/h \rfloor$ smallest values in $\{\text{PR}(\theta_{h,i}) : \mathcal{P}_{h,i} \in \mathcal{L}_h\}$
 for each $\mathcal{P}_{h,i}$ of the $\lfloor h_{\max}/h \rfloor$ cells **do**
 Run Open($\mathcal{P}_{h,i}$)
 end for
end for
 Take $(h, i) \in \text{argmin}_{(h,i)} \{\text{PR}(\theta_{h,i}) : h \in [0 : h_{\max} + 1], \mathcal{P}_{h,i} \in \mathcal{L}_h\}$
Return $\theta_T \leftarrow \theta_{h,i}$

Subroutine Open($\mathcal{P}_{h,i}$)

Input: cell $\mathcal{P}_{h,i}$
for each child cell $\mathcal{P}_{h+1,j}$ of $\mathcal{P}_{h,i}$ **do**
 Take a solution $\theta_{h+1,j} \in \text{argmin}_{\theta \in \mathcal{P}_{h+1,j}} \text{DPR}(\theta_{h,i}, \theta)$ and deploy it
 Receive $\mathcal{D}(\theta_{h+1,j})$ to compute $\text{DPR}(\theta_{h+1,j}, \theta)$ for $\theta \in \mathcal{P}_{h+1,j}$
 Update $\mathcal{L}_{h+1} \leftarrow \mathcal{L}_{h+1} \cup \{\mathcal{P}_{h+1,j}\}$
end for

We use notation $[a : b]$ to denote the set $\{a, a + 1, \dots, b\}$ for integers a, b with $a < b$. We adopt SequOOL by Bartlett et al. (2019) as the backbone of Algorithm 1. As SequOOL, our algorithm explores the depth sequentially by testing multiple cells at the same depth level before going down to the next level. As going deeper, fewer cells are tested, thus focusing on a narrower area. This can be viewed as an exploration-exploitation procedure. Moreover, opening a cell $\mathcal{P}_{h,i}$ of depth h means considering its child cells $\{\mathcal{P}_{h+1,j} : j \in J\}$ at $h + 1$ by deploying their representative decisions $\theta_{h+1,j}$. In Algorithm 1, \mathcal{L}_h denotes the set of cells $\mathcal{P}_{h,i}$ of depth h whose representative decision $\theta_{h,i}$ has been deployed. Then $\mathcal{L}_0, \dots, \mathcal{L}_{h_{\max}+1}$ naturally form a tree whose vertices correspond to cells.

To analyze the performance of Algorithm 1, we use the notion of near-optimality dimension, as mentioned in the introduction. Its definition has been refined, and we adopt the version considered by Grill et al. (2015); Bartlett et al. (2019), that is, the near-optimality dimension associated with a given hierarchical partitioning.

Definition 4 (near-optimality dimension) For any $\nu > 0$, $C \geq 1$, and $\rho \in (0, 1)$, the (ν, ρ, C) -near-optimality dimension, denoted $d(\nu, \rho, C)$, of f with respect to the hierarchical partitioning \mathcal{P} is defined as

$$d(\nu, \rho, C) = \inf \left\{ d \in \mathbb{R}_+ : N_h(6\nu\rho^h) \leq C\rho^{-dh} \quad \forall h \geq 0 \right\}$$

where $N_h(\epsilon)$ is the number of cells $\mathcal{P}_{h,i}$ of depth h such that $\inf_{\theta \in \mathcal{P}_{h,i}} \text{PR}(\theta) \leq \text{PR}(\theta_{\text{PO}}) + \epsilon$.

In particular, we will use the $((2\sqrt{D})^\alpha L_z \varepsilon, 2^{-\alpha}, 1)$ -near-optimality dimension. It gets large as the sensitivity parameter ε increases. Note that by Assumption 5, the number of cells of depth h is at most 2^{Dh} . This gives rise to a global upper bound on $d(\nu, 2^{-\alpha}, 1)$ that holds for any $\nu > 0$, that is, $d(\nu, 2^{-\alpha}, 1) \leq D/\alpha$. Hence, when ε is small and $\text{PR}(\cdot)$ has sufficient curvature around the performative-optimal solution θ_{PO} , the $((2\sqrt{D})^\alpha L_z \varepsilon, 2^{-\alpha}, 1)$ -near-optimality dimension is supposed to be much smaller than D/α . When the ambient dimension D is fixed, one may regard the factor $(2\sqrt{D})^\alpha$ as a fixed constant and hide it by replacing $\mathcal{N}_h(6\nu\rho^h)$ with $\mathcal{N}_h(6(2\sqrt{D})^\alpha\nu\rho^h)$ in the definition of $d(\nu, \rho, C)$.

The following lemma is the key to analyzing the performance of Algorithm 1. Following Bartlett et al. (2019), we define \perp_h as the depth of the deepest cell containing θ_{PO} opened until Algorithm 1 finishes opening cells of depth h .

Lemma 5 *Let d denote the $((2\sqrt{D})^\alpha L_z \varepsilon, 2^{-\alpha}, 1)$ -near-optimality dimension. Then θ_T returned by Algorithm 1 satisfies the following bound.*

$$\text{PR}(\theta_T) - \text{PR}(\theta_{\text{PO}}) \leq 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha(\perp_{h_{\max}} + 1)}.$$

Note that the bound on the optimality gap scales with $L_z \varepsilon$, not $L_\theta + L_z \varepsilon$. Based on this, we prove the following theorem which provides a theoretical guarantee on the performance of Algorithm 1. As in the analysis of Sequ00L by Bartlett et al. (2019), we use the Lambert W function. The function is to describe the solution h to the equation $x = h \cdot e^h$ as $h = W(x)$.

Theorem 6 *Let d denote the $((2\sqrt{D})^\alpha L_z \varepsilon, 2^{-\alpha}, 1)$ -near-optimality dimension. For the full-feedback setting, Algorithm 1 after T decision deployments finds a solution θ with*

$$\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) \leq \begin{cases} 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h_{\max}}, & \text{if } d = 0, \\ 2(2\sqrt{D})^\alpha L_z \varepsilon e^{-(1/d)W(h_{\max}\alpha d \log 2)}, & \text{if } d > 0 \end{cases}$$

where $h_{\max} = \lfloor T/2^D \overline{\log T} \rfloor$. Moreover, if $d > 0$ and $h_{\max}\alpha d \log 2 \geq e$, then θ satisfies

$$\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) \leq 2(2\sqrt{D})^\alpha L_z \varepsilon \left(\frac{h_{\max}\alpha d \log 2}{\log(h_{\max}\alpha d \log 2)} \right)^{-\frac{1}{d}}.$$

As $h_{\max} = \Omega(T/\log T)$ and $h_{\max} = O(T/\log T)$, we deduce from Theorem 6 with $\alpha = 1$ the optimality gap bounds in Theorem 1. We provide the proof of the theorem in Appendix A. We follow the proof outline of Bartlett et al. (2019) for Sequ00L, but we need to adapt the analysis to our specific design of the procedure of opening a cell based on performative feedback.

The last remark is that the optimality gap $\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}})$ is the *simple regret* whereas Jagadeesan et al. (2022) studies the *cumulative regret* incurred by their algorithm. Although Theorem 6 characterizes an upper bound on the simple regret only, we later report our numerical results on the cumulative regret of Algorithm 1.

5 Data-Driven Setting

For the data-driven setting, we receive a few data samples as performative feedback, which provides an estimation of the distribution. Through the data samples, we obtain the empirical distribution $\widehat{\mathcal{D}}(\theta)$ after deploying decision θ . Then we may compute the estimator

$\widehat{\text{DPR}}(\theta, \theta')$ of the decoupled performative risk $\text{DPR}(\theta, \theta')$ for other decisions θ' . Here, controlling the estimation error $|\text{DPR}(\theta, \theta') - \widehat{\text{DPR}}(\theta, \theta')|$ is crucial to achieve a better performance. To reduce the error, we evaluate a cell multiple times to obtain enough data samples from the distribution of the representative decision. Following `Stroqu00L` by [Bartlett et al. \(2019\)](#), Algorithm 2 implements this idea, extending Algorithm 1 to the data-driven setting.

As Algorithm 1, Algorithm 2 takes fewer cells at deeper depth levels, thereby implementing the exploration-exploitation trade-off principle. On top of this, the algorithm keeps track of the number of times each cell has been evaluated. When exploring cells at a certain depth, the algorithm distributes the evaluation budget over cells based on how many times they have been evaluated. Among the cells that have been evaluated many times, we focus on a few that have a low performative risk. For the cells that have not been evaluated many times, we distribute the evaluation budget over more cells, among which we encourage exploration. Furthermore, as in `Stroqu00L`, Algorithm 2 has the cross-validation phase, in which we focus on cells whose representative decision has a low estimated performative risk.

Let $n_{h,i}^{\text{open}}$ denote the number of times cell $\mathcal{P}_{h,i}$ is opened, and let $n_{h,i}^{\text{deploy}}$ denote the number of times its representative decision $\theta_{h,i}$ is deployed. Note that if $\mathcal{P}_{h+1,j}$ is a child cell of $\mathcal{P}_{h,i}$, then we have $n_{h,i}^{\text{open}} = n_{h+1,j}^{\text{deploy}}$. Recall that $[a : b]$ denotes the set $\{a, a+1, \dots, b\}$ for integers a, b with $a < b$. For a positive integer a , let $[a]$ denote the set $\{1, \dots, a\}$. Moreover, as in Algorithm 1, $\mathcal{L}_0, \dots, \mathcal{L}_{h_{\max}+1}$ represent the tree search structure of Algorithm 2.

In what follows, we analyze the performance of Algorithm 2. Note that we compute $\widehat{\text{DPR}}(\theta, \theta')$ for many pairs of θ and θ' , and at the same time, we need the estimation error $|\text{DPR}(\theta, \theta') - \widehat{\text{DPR}}(\theta, \theta')|$ uniformly bounded for all pairs. To achieve this, we introduce the Rademacher complexity associated with the loss function f .

Definition 7 (Rademacher complexity) *Given an objective function f , the Rademacher complexity $\mathfrak{C}^*(f)$ is defined as*

$$\mathfrak{C}^*(f) = \sup_{\theta \in \Theta} \sup_{n \in \mathbb{N}} \sqrt{n} \cdot \mathbb{E}_{\epsilon, z^\theta} \left[\sup_{\theta' \in \Theta} \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j f(\theta', z_j^\theta) \right| \right],$$

where $\epsilon_j \sim \text{Rademacher}$ and $z_j^\theta \sim \mathcal{D}(\theta)$ for $j \in [n]$, which are all independent of each other.

Given the Rademacher complexity of the loss function, we may provide a uniform upper bound on the estimation error. Let us define the *clean event* under which the estimation error $|\text{DPR}(\theta, \theta') - \widehat{\text{DPR}}(\theta, \theta')|$ is uniformly bounded over all pairs.

Definition 8 (Clean event) *We define the clean event, denoted $E_{\text{clean}, \delta}$, as the event that*

$$\sup_{\theta \in \mathcal{P}_{h,i}} \left| \widehat{\text{DPR}}(\theta_{h,i}, \theta) - \text{DPR}(\theta_{h,i}, \theta) \right| \leq \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_{h,i}^{\text{deploy}} m_0}} \quad \forall \mathcal{P}_{h,i} \in \mathcal{L}_h, \quad \forall h \in [h_{\max} + 1]$$

and

$$\left| \widehat{\text{PR}}(\theta_T(p)) - \text{PR}(\theta_T(p)) \right| \leq \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}} \quad \forall p \in [0 : p_{\max}].$$

We may prove that the clean event holds with high probability, parameterized by δ .

Algorithm 2 Stochastic Optimistic Optimization with Performative Feedback (SOOP)

Input: test budget T , hierarchical partitioning \mathcal{P} ,

$$h_{\max} = \left\lfloor \frac{T}{2^{D+1}(\log_2 T + 1)^2} \right\rfloor, \quad p_{\max} = \lfloor \log_2 h_{\max} \rfloor$$

◀ *Initialization Phase* ▶

Set $\mathcal{L}_0 \leftarrow \{\mathcal{P}_{0,1}\}$ and initialize $\mathcal{L}_1 \leftarrow \emptyset$

Take a solution $\theta_{0,1} \in \mathcal{P}_{0,1}$, deploy it h_{\max} times, and set $n_{0,1}^{\text{deploy}} \leftarrow h_{\max}$

◀ *Exploration Phase* ▶

Run Open($\mathcal{P}_{0,1}, h_{\max}$)

for $h = 1$ **to** h_{\max} **do**

Initialize $\mathcal{L}_{h+1} \leftarrow \emptyset$

for $p = \lfloor \log_2(h_{\max}/h) \rfloor$ **down to** 0 **do**

Take $\lfloor h_{\max}/h2^p \rfloor$ cells that correspond to the $\lfloor h_{\max}/h2^p \rfloor$ smallest values in $\left\{ \widehat{\text{PR}}(\theta_{h,i}) : \mathcal{P}_{h,i} \in \mathcal{L}_h, n_{h,i}^{\text{open}} = 0, n_{h,i}^{\text{deploy}} \geq 2^p \right\}$

for each $\mathcal{P}_{h,i}$ of the $\lfloor h_{\max}/h2^p \rfloor$ cells **do**

Run Open($\mathcal{P}_{h,i}, 2^p$)

end for

end for

end for

◀ *Cross-validation Phase* ▶

for $p = 0$ **to** p_{\max} **do**

Take $(h, i) \in \operatorname{argmin}_{(h,i)} \left\{ \widehat{\text{PR}}(\theta_{h,i}) : h \in [0 : h_{\max} + 1], \mathcal{P}_{h,i} \in \mathcal{L}_h, n_{h,i}^{\text{deploy}} \geq 2^p \right\}$

Set $\theta_T(p) \leftarrow \theta_{h,i}$

Deploy h_{\max} times solution $\theta_T(p)$ to form $\widehat{\text{PR}}(\theta_T(p))$

end for

Return $\theta_T \leftarrow \theta_T(p)$ with $p \in \operatorname{argmin}_{p \in [0:p_{\max}]} \widehat{\text{PR}}(\theta_T(p))$

Subroutine Open($\mathcal{P}_{h,i}, n$)

Input: cell $\mathcal{P}_{h,i}$, number n

Set $n_{h,i}^{\text{open}} \leftarrow n$

for each child cell $\mathcal{P}_{h+1,j}$ of $\mathcal{P}_{h,i}$ **do**

Take $\theta_{h+1,j} \in \operatorname{argmin}_{\theta \in \mathcal{P}_{h+1,j}} \widehat{\text{DPR}}(\theta_{h,i}, \theta)$, deploy it n times, and set $n_{h+1,j}^{\text{deploy}} \leftarrow n$

Form $\widehat{\text{DPR}}(\theta_{h+1,j}, \theta)$ for $\theta \in \mathcal{P}_{h+1,j}$

Initialize $n_{h+1,j}^{\text{open}} \leftarrow 0$

Update $\mathcal{L}_{h+1} \leftarrow \mathcal{L}_{h+1} \cup \{\mathcal{P}_{h+1,j}\}$

end for

Lemma 9 *The clean event holds with probability at least $1 - \delta$, i.e., $\mathbb{P}[E_{\text{clean},\delta}] \geq 1 - \delta$.*

Next, we present the key lemma for our analysis. Following [Bartlett et al. \(2019\)](#), we define $\perp_{h,p}$ as the depth of the deepest cell containing the performative optimal solution θ_{PO} opened for at least 2^p times until Algorithm 2 finishes opening cells of depth h .

Lemma 10 *Assume that the clean event $E_{\text{clean},\delta}$ holds for some $\delta \in (0, 1)$, and let d denote the $((2\sqrt{D})^\alpha L_z \varepsilon, 2^{-\alpha}, 1)$ -near-optimality dimension $d((2\sqrt{D})^\alpha L_z \varepsilon, 2^{-\alpha}, 1)$. Then for any $p \in [0 : p_{\max}]$, the following bound on the regret holds.*

$$\begin{aligned} & \text{PR}(\theta_T) - \text{PR}(\theta_{\text{PO}}) \\ & \leq 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha(\perp_{h_{\max}, p} + 1)} + \frac{8\mathfrak{C}^*(f) + 8\sqrt{\log(T/\delta)}}{\sqrt{2^p m_0}} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}. \end{aligned}$$

Therefore, to provide an upper bound on the simple regret, it is sufficient to provide upper bounds on the three terms on the right-hand side. In particular, the second term did not appear in the analysis of `Stroqu00L` by [Bartlett et al. \(2019\)](#). Nevertheless, we show that under Algorithm 2, the three terms are controlled, thereby leading to the desired performance guarantees.

We saw that the simple regret of Algorithm 1 behaves differently depending on whether the near-optimality dimension d satisfies $d = 0$ or $d > 0$. Similarly, the simple regret of Algorithm 2 varies depending on problem parameters. To illustrate, let us define the *low-noise* and *high-noise* regimes. For simplicity, we use notations ν and B defined as

$$\nu = (2\sqrt{D})^\alpha L_z \varepsilon \quad \text{and} \quad B = \frac{2\sqrt{2} \left(\mathfrak{C}^*(f) + \sqrt{\log(T/\delta)} \right)}{\sqrt{m_0}}.$$

Here, ν captures the term $L_z \varepsilon$, and B is related to the estimation error. Intuitively, if B is high, then the estimation error is large. We define \tilde{h} as

$$\begin{aligned} \tilde{h} &= \frac{1}{\alpha(d+2)\log 2} W \left(\frac{h_{\max} \nu^2 \alpha(d+2) \log 2}{B^2} \right) \\ &= \frac{1}{\alpha(d+2)\log 2} \left(\log \left(\frac{h_{\max} \nu^2 \alpha(d+2) \log 2}{B^2} \right) - \log \log \left(\frac{h_{\max} \nu^2 \alpha(d+2) \log 2}{B^2} \right) \right) + o(1) \\ &= \Omega \left(\log \left(L_z^2 \varepsilon^2 \frac{T}{\log T} \right) \right) \end{aligned}$$

We refer to the case $B < L_z \varepsilon \cdot 2^{-\alpha \tilde{h}}$ as the low-noise regime and the case $B \geq L_z \varepsilon \cdot 2^{-\alpha \tilde{h}}$ as the high-noise regime.

Theorem 11 *Let d denote the $((2\sqrt{D})^\alpha L_z \varepsilon, 2^{-\alpha}, 1)$ -near-optimality dimension. For the data-driven setting, Algorithm 2 after T decision deployments finds a solution θ that satisfies the following with probability at least $1 - \delta$. Under the low-noise regime with $d = 0$,*

$$\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) \leq (2 + 2\sqrt{2})(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h_{\max}} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}$$

where $h_{\max} = \lfloor T/2^{D+1}(\log_2 T + 1)^2 \rfloor$. Under the low-noise regime with $d > 0$,

$$\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) \leq (2 + 2\sqrt{2})(2\sqrt{D})^\alpha L_z \varepsilon e^{-(1/d)W(h_{\max} \alpha d \log 2)} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}.$$

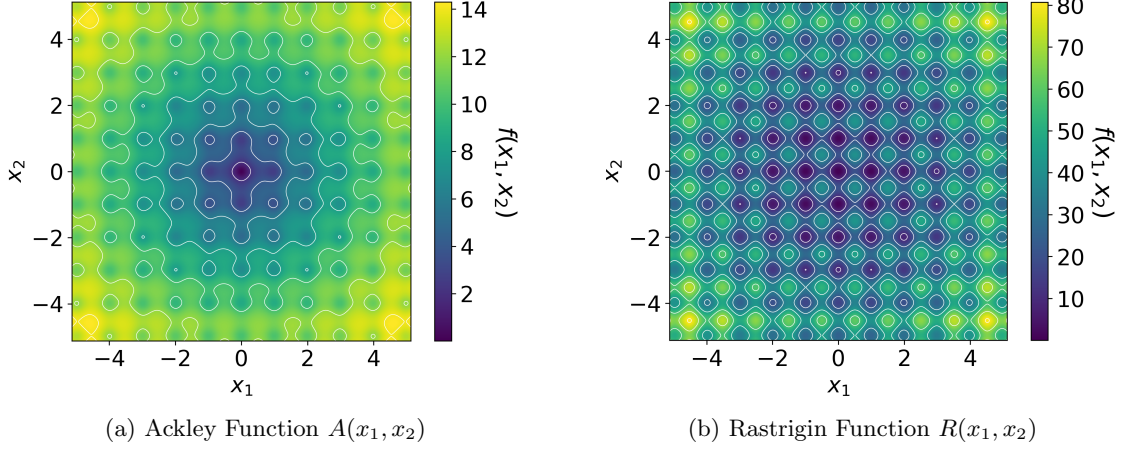


Figure 1: Contour plots of the Ackley and Rastrigin functions on $[-5.12, 5.12]^2$

Under the high-noise regime,

$$\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) \leq 6(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha \tilde{h}} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}.$$

Moreover, if $h_{\max} \geq \max\{1, e/\alpha d \log 2, B^2 e/\nu^2 \alpha(d+2) \log 2\}$, then under the low-noise regime,

$$\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) \leq \begin{cases} (2 + 3\sqrt{2})(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h_{\max}}, & \text{if } d = 0, \\ (2 + 3\sqrt{2})(2\sqrt{D})^\alpha L_z \varepsilon \left(\frac{h_{\max} \alpha d \log 2}{\log(h_{\max} \alpha d \log 2)} \right)^{-1/d}, & \text{if } d > 0. \end{cases}$$

Lastly, if $h_{\max} \geq B^2 e/\nu^2 \alpha(d+2) \log 2$, then under the high-noise regime,

$$\begin{aligned} & \text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) \\ & \leq 6(2\sqrt{D})^\alpha L_z \varepsilon \left(\frac{h_{\max} \nu^2 \alpha(d+2) \log 2 / B^2}{\log(h_{\max} \nu^2 \alpha(d+2) \log 2 / B^2)} \right)^{-\frac{1}{d+2}} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}. \end{aligned}$$

As $h_{\max} = \Omega(T/\log T)$, $h_{\max} = O(T/\log T)$, and $\tilde{h} = \Omega(\log(L_z^2 \varepsilon^2 T / \log T))$, we deduce from Theorem 11 with $\alpha = 1$ the simple regret bounds in Theorem 2. We provide the proof of the theorem in Appendix B.

6 Experiments

In this section, we empirically demonstrate how our algorithms, DOOP for the full-feedback case and SOOP for the data-driven setting, numerically perform for performative regret minimization. We compare DOOP and the existing methods S00, Sequ00L, and SZooming for the full-feedback case, and we test SOOP against StoS00, Stroqu00L, and SZooming for the data-driven setting. Here, SZooming indicates the variant of the zooming algorithm

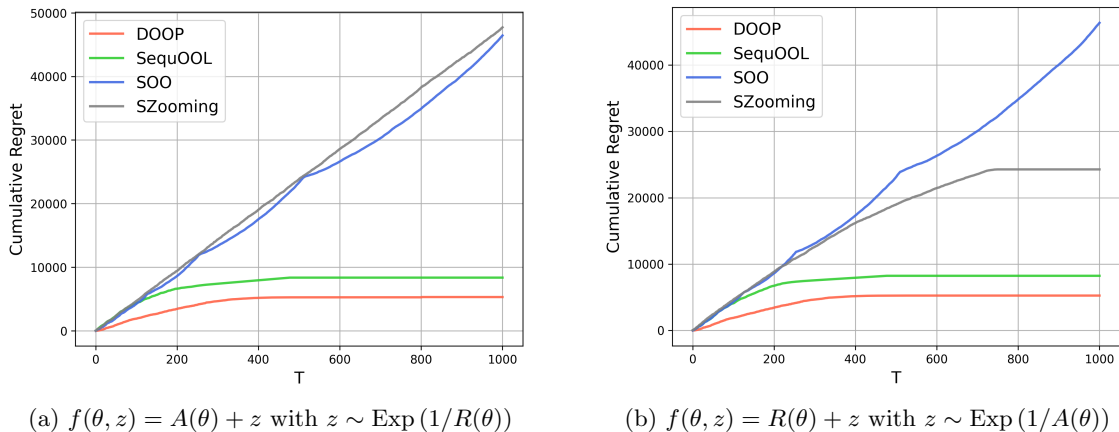


Figure 2: Cumulative regret comparison in the full-feedback case

by Jagadeesan et al. (2022). For SOO, StoSOO, SequOOL, and StroquOOL, we used the package developed by Li et al. (2023a).

We tested the algorithms for synthetic objectives on a bounded two-dimensional domain for optimization. In our experiments, we used two multi-modal functions as shown in Figure 1 to express our loss function $f(\theta, z)$ and the distribution map $\mathcal{D}(\theta)$; the first is the *Ackley function* given by

$$A(x_1, x_2) = -20 \cdot \exp \left[-0.2 \sqrt{0.5(x_1^2 + x_2^2)} \right] - \exp [0.5 (\cos(2\pi x_1) + \cos(2\pi x_2))]$$

and the second is the *Rastrigin function* given by

$$R(x_1, x_2) = 20 + (x_1^2 - 10 \cos(2\pi x_1)) + (x_2^2 - 10 \cos(2\pi x_2)).$$

Note that both functions have a global minimum at $A(0, 0) = R(0, 0) = 0$, and their domains are both $[-5.12, 5.12]^2$. With the Ackley and Rastrigin functions, we define two types of the loss function.

- $f(\theta, z) = A(\theta) + z$ with $z \sim \mathcal{D}(\theta) = \text{Exp}(1/R(\theta))$ and $\theta \in [-5.12, 5.12]^2$,
- $f(\theta, z) = R(\theta) + z$ with $z \sim \mathcal{D}(\theta) = \text{Exp}(1/A(\theta))$ and $\theta \in [-5.12, 5.12]^2$

where $\text{Exp}(1/\lambda)$ denotes the exponential distribution with mean λ . In both cases, we have

$$\text{PR}(\theta) = \mathbb{E}_{z \sim \mathcal{D}(\theta)}[f(\theta, z)] = A(\theta) + R(\theta).$$

For the full-feedback case, we test DOOP with SOO, SequOOL, and SZooming, and the performative risk is constructed based on combining the *Ackley function* and the *Rastrigin function*. Recall that the tree search-based algorithms, not including SZooming, require a hierarchical partitioning, and for them, we used the binary partitioning. For the tree search-based algorithms, we used the same maximum level of depth h_{\max} . For SZooming, the decision domain Θ is set to be a finite set of 3,025 discrete points on domain $[-5.12, 5.12]^2$.

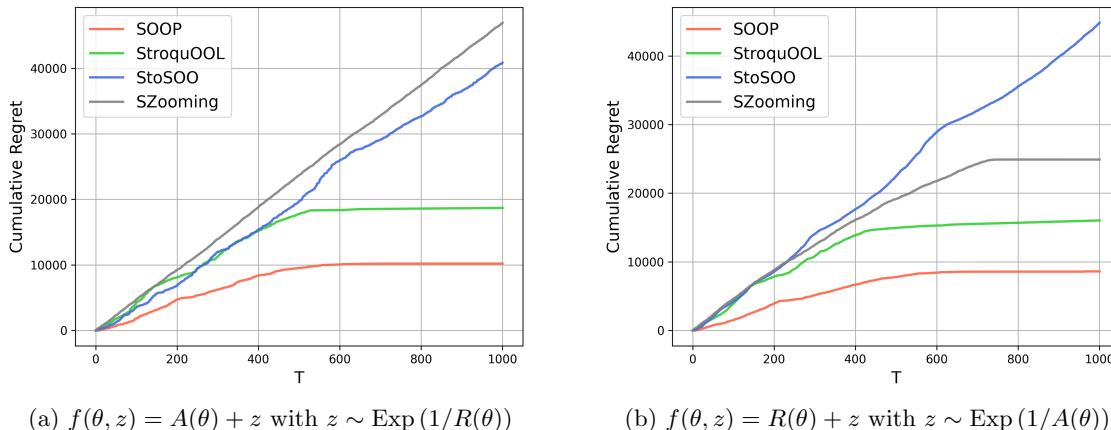


Figure 3: Cumulative regret comparison in the data-driven setting

In addition, the sensitivity parameter ϵ and the Lipschitz constant L_z are chosen according to the shape of the objective functions on the decision domain Θ ; both ϵ and L_z for $z \sim \text{Exp}(1/f(\theta))$ is calculated to be $\sup (|f(\theta_1) - f(\theta_2)| / \|\theta_1 - \theta_2\|)$ for any $\theta_1, \theta_2 \in \Theta$.

For the data-driven setting, we test **SOOP** with **StoS00**, **StroquOOL**, and **SZooming**. The performative feedback consists of $m_0 = 10$ samples drawn randomly from $\mathcal{D}(\theta)$. While both $A(\theta)$ and $R(\theta)$ are multi-modal on the given domain, each has different sets of local optima and range. In particular, $R(\theta)$ yields values in a broader range, thus the associated distribution $\text{Exp}(1/R(\theta))$ has a larger variance since the variance of $\text{Exp}(1/\lambda)$ is λ^2 . The other components of the experimental setup are the same as those for the full-feedback case.

Figures 2 and 3 summarize our experimental results. As shown in Figures 2 and 3, **DOOP** and **SOOP** outperform the other methods in terms of cumulative regret. We have noticed that **SZooming** incurs a very high cumulative regret in the first phase, and this is because there exist high estimation errors in the first phase of **SZooming** and it turns out that the majority of exploration of **SZooming** occurs during the first phase. Moreover, **SZooming** is not computationally efficient, as it takes a huge amount of time to find an optimal decision. For the full-feedback case, **SZooming** takes 73691 seconds for $f(\theta, z) = A(\theta) + z$ with $z \sim \mathcal{D}(\theta) = \text{Exp}(1/R(\theta))$ and 3737.0 seconds for $f(\theta, z) = R(\theta) + z$ with $z \sim \mathcal{D}(\theta) = \text{Exp}(1/A(\theta))$. In contrast, **DOOP** takes 4.2390 seconds and 2.3609 seconds, respectively. For the data-driven case, **SZooming** takes 73112 seconds for $f(\theta, z) = A(\theta) + z$ with $z \sim \mathcal{D}(\theta) = \text{Exp}(1/R(\theta))$ and 3999.4 seconds for $f(\theta, z) = R(\theta) + z$ with $z \sim \mathcal{D}(\theta) = \text{Exp}(1/A(\theta))$. In contrast, **SOOP** takes 2.5896 seconds and 1.4877 seconds, respectively.

Acknowledgments and Disclosure of Funding

This research is supported, in part, by KAIST Starting Fund (KAIST-G04220016), FOUR Brain Korea 21 Program (NRF-5199990113928), and National Research Foundation of Korea (NRF-2022M3J6A1063021).

Appendix A. Regret Analysis of D00P

A.1 Approximation Bounds under the Hierarchical Partitioning Scheme

In this section, we prove two lemmas that are related to the quality of the representative decision of each cell.

Lemma 12 *For $\theta \in \mathcal{P}_{h,i}$ and $\mathcal{P}_{h,i} \in \mathcal{L}_h$, we have*

$$|\text{DPR}(\theta_{h,i}, \theta) - \text{PR}(\theta)| \leq (\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h}.$$

Proof Note that

$$\text{DPR}(\theta_{h,i}, \theta) \geq \text{PR}(\theta) - L_z \varepsilon \|\theta_{h,i} - \theta\|^\alpha \geq \text{PR}(\theta) - (\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h}$$

where the first inequality is from Lemma 3 and the second inequality follows from Assumption 5 with $\theta_{h,i}, \theta \in \mathcal{P}_{h,i}$. Similarly, we can argue that $\text{DPR}(\theta_{h,i}, \theta) \leq \text{PR}(\theta) + L_z \varepsilon \|\theta_{h,i} - \theta\|^\alpha \leq \text{PR}(\theta) + (\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h}$, as required. \blacksquare

Lemma 13 *Let $\mathcal{P}_{h,i} \in \mathcal{L}_h$. Then*

$$\text{PR}(\theta_{h,i}) \leq \inf_{\theta \in \mathcal{P}_{h,i}} \text{PR}(\theta) + 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h}.$$

Proof Let $\theta_{h,i}^* \in \text{argmin}_{\theta \in \mathcal{P}_{h,i}} \text{PR}(\theta)$, and let $\mathcal{P}_{h-1,j}$ be the parent cell of $\mathcal{P}_{h,i}$. Note that

$$\text{PR}(\theta_{h,i}) \leq \text{DPR}(\theta_{h-1,j}, \theta_{h,i}) + (2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h} \leq \text{DPR}(\theta_{h-1,j}, \theta_{h,i}^*) + (2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h}$$

where the first inequality follows from Lemma 12 and the second inequality holds due to our choice of $\theta_{h,i}$ minimizing $\text{DPR}(\theta_{h-1,j}, \theta)$ over $\theta \in \mathcal{P}_{h,i}$. Lastly, by Lemma 12, we have

$$\text{DPR}(\theta_{h-1,j}, \theta_{h,i}^*) \leq \text{PR}(\theta_{h,i}^*) + (2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h}.$$

Consequently, it follows that $\text{PR}(\theta_{h,i}) \leq \text{PR}(\theta_{h,i}^*) + 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h}$, as required. \blacksquare

A.2 Proof of Theorem 6

Recall that \perp_h is defined on the depth of the deepest cell containing θ_{PO} opened until Algorithm 1 finishes opening cells of depth h .

Lemma 14 *Let d denote the $((2\sqrt{D})^\alpha L_z \varepsilon, 2^{-\alpha}, 1)$ -near-optimality dimension. For any $h \in [h_{\max}]$, if $h_{\max}/h \geq 2^{\alpha h d}$, then $\perp_h = h$ with $\perp_0 = 0$.*

Proof Let $h \in [h_{\max}]$, and assume that h satisfies the condition of the lemma. Then we will argue by induction that $\perp_{h'} = h'$ for all $h' \in [h]$, thereby proving that $\perp_h = h$.

Note that $\mathcal{P}_{0,1} = \Theta$ contains θ_{PO} and $\mathcal{P}_{0,1}$ is opened, so $\perp_0 = 0$. Next, we assume that $\perp_{h'-1} = h' - 1$ for some $h' \in [h]$. Then it is sufficient to show that $\perp_{h'} = h'$. Let $i_{h'-1}^*$

denote the index such that $\mathcal{P}_{h'-1, i_{h'-1}^*}$ is the cell containing θ_{PO} at depth $h' - 1$. Algorithm 1 opens $\lfloor h_{\max}/h' \rfloor$ cells from depth h' cells. Suppose for a contradiction that cell $\mathcal{P}_{h', i_{h'}^*}$ is not one of them. This implies that for each solution $\theta_{h', i}$ of the $\lfloor h_{\max}/h' \rfloor$ cells from depth h' , we have $\text{PR}(\theta_{h', i}) \leq \text{PR}(\theta_{h', i_{h'}^*})$. Consequently, it follows that

$$\text{PR}(\theta_{h', i}) \leq \text{PR}(\theta_{h', i_{h'}^*}) \leq \text{PR}(\theta_{\text{PO}}) + 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'}$$

where the second inequality follows from Lemma 13 as θ_{PO} is contained in cell $\mathcal{P}_{h', i_{h'}^*}$. This implies that

$$\mathcal{N}_h(6(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'}) \geq \left\lfloor \frac{h_{\max}}{h'} \right\rfloor + 1 \geq \left\lfloor \frac{h_{\max}}{h} \right\rfloor + 1 \geq 2^{\alpha h d} + 1 \geq 2^{\alpha h' d} + 1$$

where $\lfloor h_{\max}/h' \rfloor$ comes from cells $\mathcal{P}_{h', i}$ and 1 is due to cell $\mathcal{P}_{h', i_{h'}^*}$ in the first inequality, the second and the fourth inequalities hold because $h' \leq h$, and the third inequality comes from the condition of the lemma. This in turn implies that $\mathcal{N}_h(6(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'}) > 2^{\alpha h' d}$, a contradiction. Therefore, it follows that $\perp_{h'} = h'$. Then the induction argument shows that $\perp_h = h$, as required. \blacksquare

Based on Lemma 14, we prove Lemma 5 that shows

$$\text{PR}(\theta_T) - \text{PR}(\theta_{\text{PO}}) \leq 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha(\perp_{h_{\max}} + 1)}.$$

Proof [Proof of Lemma 5] Let $\mathcal{P}_{\perp_{h_{\max}} + 1, i^*}$ be the cell at depth $\perp_{h_{\max}} + 1$ containing θ_{PO} . Note that

$$\text{PR}(\theta_T) \leq \text{PR}(\theta_{\perp_{h_{\max}} + 1, i^*}) \leq \text{PR}(\theta_{\text{PO}}) + 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha(\perp_{h_{\max}} + 1)}$$

where the first inequality holds due to the choice of θ_T and the second inequality follows from Lemma 13. \blacksquare

For simplicity, we introduce notations ρ and ν defined as

$$\rho = 2^{-\alpha} \quad \text{and} \quad \nu = (2\sqrt{D})^\alpha L_z \varepsilon.$$

Moreover, we define \bar{h} as the number satisfying

$$\frac{h_{\max}}{\bar{h}} = \rho^{-d\bar{h}}.$$

Note that if $d = 0$, then $\bar{h} = h_{\max}$. If $d > 0$, then

$$\bar{h} = \frac{1}{d \log(1/\rho)} W(h_{\max} d \log(1/\rho))$$

where $W(\cdot)$ denotes the Lambert W function.

Lemma 15 (Bartlett et al. (2019)) *Let d denote the $(\nu, \rho, 1)$ -near-optimality dimension. Then $\perp_{h_{\max}} + 1 \geq \bar{h}$.*

Combining Lemmas 5 and 15, we are ready to provide the desired regret bounds on Algorithm 1.

Proof [Proof of Theorem 6] As $\bar{h} = h_{\max}$ when $d = 0$ and $\bar{h} = W(h_{\max}\alpha d \log 2)/\alpha d \log 2$, it follows directly from Lemmas 5 and 15 that

$$\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) \leq \begin{cases} 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h_{\max}}, & \text{if } d = 0, \\ 2(2\sqrt{D})^\alpha L_z \varepsilon e^{-(1/d)W(h_{\max}\alpha d \log 2)}, & \text{if } d > 0. \end{cases}$$

Lastly, [Hoorfar and Hassani \(2008\)](#) showed that if $x \geq e$, then $W(x) \geq \log(x/\log(x))$. Hence, if $d > 0$ and $h_{\max}\alpha d \log 2 \geq e$, then θ satisfies

$$\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) \leq 2(2\sqrt{D})^\alpha L_z \varepsilon \left(\frac{h_{\max}\alpha d \log 2}{\log(h_{\max}\alpha d \log 2)} \right)^{-\frac{1}{d}},$$

as required. ■

Appendix B. Regret Analysis of S00P

B.1 Total Number of Solution Deployments

Recall that

$$h_{\max} = \left\lfloor \frac{T}{2^{D+1}(\log_2 T + 1)^2} \right\rfloor \quad \text{and} \quad p_{\max} = \lfloor \log_2 h_{\max} \rfloor.$$

Lemma 16 *The total number of solution deployments before the cross-validation phase is at most $3T/4$, and in the cross-validation phase, the total number of solution deployments is at most $T/4$.*

Proof Note that as $T \geq 2$, we have $\log_2 T + 1 \geq 2$, in which case $h_{\max} \leq T/2^{D+3}$. Hence, we deploy solution $\theta_{0,1}$ at most $T/2^{D+3} \leq T/8$ times. Moreover, we open $\mathcal{P}_{0,1}$ at most $T/2^{D+3}$ times, and since $\mathcal{P}_{0,1}$ has 2^D child cells, it corresponds to at most $T/8$ solution deployments. Next, during the exploration phase, we make $\sum_{h=1}^{h_{\max}} \sum_{p=0}^{p_{\max}} \lfloor h_{\max}/h2^p \rfloor 2^p$ openings. Here,

$$\sum_{h=1}^{h_{\max}} \sum_{p=0}^{p_{\max}} \left\lfloor \frac{h_{\max}}{h2^p} \right\rfloor \cdot 2^p \leq \sum_{h=1}^{h_{\max}} \sum_{p=0}^{p_{\max}} \frac{h_{\max}}{h} = (p_{\max} + 1)h_{\max} \sum_{h=1}^{h_{\max}} \frac{1}{h} \leq h_{\max}(p_{\max} + 1)^2 \leq \frac{T}{2^{D+1}}$$

where the last inequality holds due to $p_{\max} \leq \log_2 T$. Since each opening requires 2^D solution deployments, it incurs $T/2$ solution deployments. In total, before the cross-validation phase, we make $T/8 + T/8 + T/2 = 3T/4$ solution deployments.

In the cross-validation phase, the number of solution deployments is given by

$$h_{\max}(p_{\max} + 1) \leq \frac{T}{2^{D+1}(\log_2 T + 1)} \leq \frac{T}{2^{D+2}} \leq \frac{T}{4},$$

as required. ■

B.2 Rademacher Complexity-Based Concentration Bounds for Estimating the Performative Risk

In this section, we prove Lemma 9 which shows that the clean event holds with probability at least $1 - \delta$, i.e., $\mathbb{P}[E_{\text{clean},\delta}] \geq 1 - \delta$.

Proof [Proof of Lemma 9] By Lemma 16, Algorithm 2 deploys at most T solutions. Let n_{distinct} denote the number of distinct solutions deployed by Algorithm 2. As new solutions are deployed during the exploration phase, we have $n_{\text{distinct}} \leq 3T/4$ by Lemma 16. Among the n_{distinct} solutions, we use notation $(h(s), i(s))$ to indicate the cell $\mathcal{P}_{h(s),i(s)}$ containing the s th deployed solution for $1 \leq s \leq n_{\text{distinct}}$. As h_{\max} is fixed, n_{distinct} , $h(s)$, and $n_{h(s),i(s)}^{\text{deploy}}$ are all deterministic functions of s . In particular, we use notation $n_s^{\text{deploy}} := n_{h(s),i(s)}^{\text{deploy}}$ to emphasize that $n_{h(s),i(s)}^{\text{deploy}}$ is deterministic in s . Then we maintain a *virtual tape of samples* for each solution θ . Basically, for each solution θ , we maintain $\{z_j^\theta : j = 1, \dots, Tm_0\}$, and if θ becomes the s th solution deployed, then we use $n_s^{\text{deploy}}m_0$ samples in $\{z_j^\theta : j = (n_1^{\text{deploy}} + \dots + n_{s-1}^{\text{deploy}})m_0 + 1, \dots, (n_1^{\text{deploy}} + \dots + n_s^{\text{deploy}})m_0\}$ to estimate $\widehat{D}(\theta)$. For $1 \leq s \leq n_{\text{distinct}}$, let us define $E_{\text{clean},\delta}^s$ as the event that

$$\sup_{\theta \in \mathcal{P}_s} \left| \widehat{\text{DPR}}(\theta_s, \theta) - \text{DPR}(\theta_s, \theta) \right| \leq \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_s^{\text{deploy}} m_0}}$$

where θ_s denotes the s th solution deployed $\theta_{h(s),i(s)}$, \mathcal{P}_s denotes $\mathcal{P}_{h(s),i(s)}$ containing the s th solution deployed, and n_s^{deploy} denotes the number of times solution $\theta_{h(s),i(s)}$ is deployed. Moreover, for $p \in [0 : p_{\max}]$, let us define $E_{\text{clean},\delta}^{T,p}$ as the event that

$$\left| \widehat{\text{PR}}(\theta_T(p)) - \text{PR}(\theta_T(p)) \right| \leq \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}.$$

Then we know that

$$\begin{aligned} \mathbb{P}[E_{\text{clean},\delta}] &= \mathbb{P}[E_{\text{clean},\delta}^1 \cap \dots \cap E_{\text{clean},\delta}^{n_{\text{distinct}}} \cap E_{\text{clean},\delta}^{T,0} \cap \dots \cap E_{\text{clean},\delta}^{T,p_{\max}}] \\ &\geq 1 - \sum_{s=1}^{n_{\text{distinct}}} \mathbb{P}[\neg E_{\text{clean},\delta}^s] - \sum_{p=0}^{p_{\max}} \mathbb{P}[\neg E_{\text{clean},\delta}^{T,p}] \end{aligned}$$

where the inequality is the union bound. For simplicity, let J denote $J = \{(n_1^{\text{deploy}} + \dots + n_{s-1}^{\text{deploy}})m_0 + 1, \dots, (n_1^{\text{deploy}} + \dots + n_s^{\text{deploy}})m_0\}$. Note that

$$\begin{aligned} \mathbb{P}[\neg E_{\text{clean},\delta}^s] &= \mathbb{P} \left[\sup_{\theta \in \mathcal{P}_s} \left| \widehat{\text{DPR}}(\theta_s, \theta) - \text{DPR}(\theta_s, \theta) \right| > \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_s^{\text{deploy}} m_0}} \right] \\ &= \mathbb{P} \left[\sup_{\theta \in \mathcal{P}_s} \left| \frac{1}{n_s^{\text{deploy}} m_0} \sum_{j \in J} f(\theta, z_j^{\theta_s}) - \text{DPR}(\theta_s, \theta) \right| > \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_s^{\text{deploy}} m_0}} \right] \\ &\leq \mathbb{P} \left[\sup_{\theta \in \Theta} \left| \frac{1}{n_s^{\text{deploy}} m_0} \sum_{j \in J} f(\theta, z_j^{\theta_s}) - \text{DPR}(\theta_s, \theta) \right| > \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_s^{\text{deploy}} m_0}} \right] \end{aligned}$$

where the inequality holds because $\mathcal{P}_s \subseteq \Theta$. Here, the right-most side of this inequality is equal to

$$\mathbb{E}_{\bar{\theta} \sim \theta_s} \left[\mathbb{P} \left[\sup_{\theta \in \Theta} \left| \frac{1}{n_s^{\text{deploy}} m_0} \sum_{j \in J} f(\theta, z_j^{\bar{\theta}}) - \text{DPR}(\bar{\theta}, \theta) \right| > \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_s^{\text{deploy}} m_0}} \mid \theta_s = \bar{\theta} \right] \right].$$

Therefore, to provide an upper bound on $\mathbb{P}[-E_{\text{clean}, \delta}^s]$, it suffices to provide an upper bound on

$$\mathbb{P} \left[\sup_{\theta \in \Theta} \left| \frac{1}{n_s^{\text{deploy}} m_0} \sum_{j \in J} f(\theta, z_j^{\bar{\theta}}) - \text{DPR}(\bar{\theta}, \theta) \right| > \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_s^{\text{deploy}} m_0}} \mid \theta_s = \bar{\theta} \right] \quad (1)$$

for every $\bar{\theta} \in \Theta$. Note that data samples in $\{z_j^{\bar{\theta}} : j = (n_1^{\text{deploy}} + \dots + n_{s-1}^{\text{deploy}})m_0 + 1, \dots, (n_1^{\text{deploy}} + \dots + n_s^{\text{deploy}})m_0\}$ are independent of the event that $\theta_s = \bar{\theta}$ because the samples are obtained after the s th solution for deployment is chosen. Therefore, the probability term (1) is equal to

$$\mathbb{P} \left[\sup_{\theta \in \Theta} \left| \frac{1}{n_s^{\text{deploy}} m_0} \sum_{j \in J} f(\theta, z_j^{\bar{\theta}}) - \text{DPR}(\bar{\theta}, \theta) \right| > \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_s^{\text{deploy}} m_0}} \right]. \quad (2)$$

What remains is to bound this probability term for every $\bar{\theta} \in \Theta$. By the bounded differences inequality and Assumption 4, with probability at least $1 - (\delta/T)$, we have

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \frac{1}{n_s^{\text{deploy}} m_0} \sum_{j \in J} f(\theta, z_j^{\bar{\theta}}) - \text{DPR}(\bar{\theta}, \theta) \right| \\ & \leq \mathbb{E} \left[\sup_{\theta \in \Theta} \left| \frac{1}{n_s^{\text{deploy}} m_0} \sum_{j \in J} f(\theta, z_j^{\bar{\theta}}) - \text{DPR}(\bar{\theta}, \theta) \right| \right] + \sqrt{\frac{2 \log(T/\delta)}{n_s^{\text{deploy}} m_0}}. \end{aligned} \quad (3)$$

Let ϵ_j denote i.i.d. Rademacher random variables. Then by a symmetrization argument, the right-hand side of (3) is at most

$$\begin{aligned} & \mathbb{E} \left[\sup_{\theta \in \Theta} \left| \frac{1}{n_s^{\text{deploy}} m_0} \sum_{j \in J} f(\theta, z_j^{\bar{\theta}}) - \text{DPR}(\bar{\theta}, \theta) \right| \right] + \sqrt{\frac{2 \log(T/\delta)}{n_s^{\text{deploy}} m_0}} \\ & \leq 2 \cdot \mathbb{E} \left[\sup_{\theta \in \Theta} \left| \frac{1}{n_s^{\text{deploy}} m_0} \sum_{j \in J} f(\theta, z_j^{\bar{\theta}}) \cdot \epsilon_j \right| \right] + \sqrt{\frac{2 \log(T/\delta)}{n_s^{\text{deploy}} m_0}} \\ & \leq \frac{2}{\sqrt{n_s^{\text{deploy}} m_0}} \cdot \sup_{n \geq 1} \sqrt{n} \cdot \mathbb{E} \left[\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{j=1}^n f(\theta, z_j^{\bar{\theta}}) \cdot \epsilon_j \right| \right] + \sqrt{\frac{2 \log(T/\delta)}{n_s^{\text{deploy}} m_0}} \\ & \leq \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_s^{\text{deploy}} m_0}} \end{aligned} \quad (4)$$

where $\{z_j^\theta\}_{j \in \mathbb{N}}$ denotes an infinite sequence of samples from $\mathcal{D}(\theta)$ for $\theta \in \Theta$. By (3) and (4), the probability term (2) as well as (1) is at most δ/T . Therefore, it follows that $\mathbb{P}[\neg E_{\text{clean},\delta}^s] \leq \delta/T$.

Next, we consider $\mathbb{P}[\neg E_{\text{clean},\delta}^{T,p}]$. Note that the total number of solution deployments during the exploration phase, denoted n_{total} is deterministic. Note that $\theta_T(p)$ is deployed for h_{max} times and obtain samples $\{z_j^{\theta_T(p)} : j \in J'\}$ where $J' = \{(n_{\text{total}} + ph_{\text{max}})m_0 + 1, \dots, (n_{\text{total}} + (p+1)h_{\text{max}})m_0\}$. Then we have

$$\begin{aligned} & \mathbb{P}[\neg E_{\text{clean},\delta}^{T,p}] \\ &= \mathbb{P}\left[\left|\widehat{\text{PR}}(\theta_T(p)) - \text{PR}(\theta_T(p))\right| > \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{h_{\text{max}}m_0}}\right] \\ &= \mathbb{P}\left[\left|\frac{1}{h_{\text{max}}m_0} \sum_{j \in J'} f(\theta_T(p), z_j^{\theta_T(p)}) - \text{DPR}(\theta_T(p), \theta_T(p))\right| > \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{h_{\text{max}}m_0}}\right] \\ &\leq \mathbb{P}\left[\sup_{\theta \in \Theta} \left|\frac{1}{h_{\text{max}}m_0} \sum_{j \in J'} f(\theta, z_j^{\theta_T(p)}) - \text{DPR}(\theta_T(p), \theta)\right| > \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{h_{\text{max}}m_0}}\right]. \end{aligned}$$

As before, we can argue that $\mathbb{P}[\neg E_{\text{clean},\delta}^{T,p}] \leq \delta/T$. Since $n_{\text{distinct}} + (p_{\text{max}} + 1)h_{\text{max}} \leq T$ by Lemma 16, it follows that $\mathbb{P}[E_{\text{clean},\delta}] \geq 1 - \delta$ as $n_{\text{distinct}} \leq T$. \blacksquare

B.3 Approximation Bounds under the data-driven setting

In this section, we prove the following lemma analyzing the quality of the representative decision of each cell under the data-driven setting.

Lemma 17 *Assume that $E_{\text{clean},\delta}$ holds for some $\delta \in (0, 1)$. Let $\mathcal{P}_{h,i} \in \mathcal{L}_h$, and let $\mathcal{P}_{h-1,j}$ be the parent cell of $\mathcal{P}_{h,i}$. Then*

$$\text{PR}(\theta_{h,i}) \leq \inf_{\theta \in \mathcal{P}_{h,i}} \text{PR}(\theta) + 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{n_{h-1,j}^{\text{deploy}} m_0}}.$$

Proof Let $\theta_{h,i}^* \in \text{argmin}_{\theta \in \mathcal{P}_{h,i}} \text{PR}(\theta)$. By Lemma 12,

$$\text{PR}(\theta_{h,i}) \leq \text{DPR}(\theta_{h-1,j}, \theta_{h,i}) + (2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h}.$$

As $E_{\text{clean},\delta}$ holds, we have

$$\text{DPR}(\theta_{h-1,j}, \theta_{h,i}) \leq \widehat{\text{DPR}}(\theta_{h-1,j}, \theta_{h,i}) + \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_{h-1,j}^{\text{deploy}} m_0}}.$$

Moreover,

$$\widehat{\text{DPR}}(\theta_{h-1,j}, \theta_{h,i}) \leq \widehat{\text{DPR}}(\theta_{h-1,j}, \theta_{h,i}^*) \leq \text{DPR}(\theta_{h-1,j}, \theta_{h,i}^*) + \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_{h-1,j}^{\text{deploy}} m_0}}$$

where the first inequality is due to our choice of $\theta_{h,i}$ minimizing $\widehat{\text{DPR}}(\theta_{h-1,j}, \theta)$ over $\theta \in \mathcal{P}_{h,i}$ and the second inequality holds because $E_{\text{clean},\delta}$ holds. Lastly, by Lemma 12,

$$\text{DPR}(\theta_{h-1,j}, \theta_{h,i}^*) \leq \text{PR}(\theta_{h,i}^*) + (2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h}.$$

Consequently, it follows that

$$\text{PR}(\theta_{h,i}) \leq \text{PR}(\theta_{h,i}^*) + 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{n_{h-1,j}^{\text{deploy}} m_0}},$$

as required. ■

B.4 Proof of Theorem 11

Recall that $\perp_{h,p}$ is defined as the depth of the deepest cell containing the performative optimal solution θ_{PO} opened for at least 2^p times until Algorithm 2 finishes opening cells of depth h .

Lemma 18 *Assume that the clean event $E_{\text{clean},\delta}$ holds for some $\delta \in (0, 1)$, and let d denote the $((2\sqrt{D})^\alpha L_z \varepsilon, 2^{-\alpha}, 1)$ -near-optimality dimension $d((2\sqrt{D})^\alpha L_z \varepsilon, 2^{-\alpha}, 1)$. For any $h \in \llbracket h_{\text{max}}/2^p \rrbracket$ and $p \in [0 : \lfloor \log_2(h_{\text{max}}/h) \rrbracket]$, if the following condition holds, then $\perp_{h,p} = h$ with $\perp_{0,p} = 0$.*

$$\frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{2^p m_0}} \leq (2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h} \quad \text{and} \quad \frac{h_{\text{max}}}{h 2^p} \geq 2^{\alpha h d}.$$

Proof Let (h, p) with $h \in \llbracket h_{\text{max}}/2^p \rrbracket$ and $p \in [0 : \lfloor \log_2(h_{\text{max}}/h) \rrbracket]$ satisfy the condition of the lemma. Then we will argue by induction that $\perp_{h',p} = h'$ for all $h' \in [h]$, thereby proving that $\perp_{h,p} = h$.

Note that $\mathcal{P}_{0,1} = \Theta$ contains θ_{PO} and $\mathcal{P}_{0,1}$ is opened h_{max} times with $h_{\text{max}} \geq 2^{p_{\text{max}}}$, so $\perp_{0,p} = 0$. Next, we assume that $\perp_{h'-1,p} = h' - 1$ for some $h' \in [h]$. Then it is sufficient to show that $\perp_{h',p} = h'$. Let $i_{h'-1}^*$ denote the index such that $\mathcal{P}_{h'-1, i_{h'-1}^*}$ is the cell containing θ_{PO} at depth $h' - 1$. By the induction hypothesis, cell $\mathcal{P}_{h'-1, i_{h'-1}^*}$ is opened at least 2^p times, i.e., $n_{h'-1, i_{h'-1}^*}^{\text{open}} \geq 2^p$. This implies that $n_{h'-1, i_{h'-1}^*}^{\text{open}} \geq 2^p$ because $n_{h'-1, i_{h'-1}^*}^{\text{deploy}} \geq 2^{p'} = n_{h'-1, i_{h'-1}^*}^{\text{open}}$ for some p' according to the design of Algorithm 2. Let $i_{h'}^*$ denote the index such that $\mathcal{P}_{h', i_{h'}^*}$ is the cell containing θ_{PO} at depth h' . This means that $\mathcal{P}_{h', i_{h'}^*}$ is a child cell of $\mathcal{P}_{h'-1, i_{h'-1}^*}$ and $n_{h', i_{h'}^*}^{\text{deploy}} = n_{h'-1, i_{h'-1}^*}^{\text{open}} \geq 2^p$.

We open $\lfloor h_{\text{max}}/h' 2^p \rfloor$ cells from depth h' cells with at least 2^p deployments. Suppose for a contradiction that cell $\mathcal{P}_{h', i_{h'}^*}$ is not one of them. This implies that for each solution $\theta_{h',i}$ of the $\lfloor h_{\text{max}}/h' 2^p \rfloor$ cells with 2^p deployments from depth h' , we have $\widehat{\text{PR}}(\theta_{h',i}) \leq \widehat{\text{PR}}(\theta_{h', i_{h'}^*})$. Moreover, such $\theta_{h',i}$ satisfies the following.

$$\begin{aligned} \text{PR}(\theta_{h',i}) - (2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'} &\leq \text{PR}(\theta_{h',i}) - (2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h} \\ &\leq \text{PR}(\theta_{h',i}) - \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{2^p m_0}} \end{aligned} \quad (5)$$

where the first inequality holds because $h' \leq h$ and the second inequality holds due to the condition of the lemma. Furthermore,

$$\begin{aligned} \text{PR}(\theta_{h',i}) - \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{2^p m_0}} &\leq \text{PR}(\theta_{h',i}) - \frac{2\mathfrak{C}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_{h',i}^{\text{deploy}} m_0}} \\ &\leq \widehat{\text{PR}}(\theta_{h',i}) \\ &\leq \widehat{\text{PR}}(\theta_{h',i_{h'}^*}) \end{aligned} \quad (6)$$

where the first inequality holds because $n_{h',i}^{\text{deploy}} \geq 2^p$ and the second inequality holds due to the assumption that $E_{\text{clean},\delta}$ holds. Combining (5) and (6), we deduce that

$$\text{PR}(\theta_{h',i}) - (2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'} \leq \widehat{\text{PR}}(\theta_{h',i_{h'}^*}).$$

Similarly, we can argue that

$$\text{PR}(\theta_{h',i_{h'}^*}) + (2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'} \geq \widehat{\text{PR}}(\theta_{h',i_{h'}^*}).$$

Consequently, it follows that

$$\begin{aligned} \text{PR}(\theta_{h',i}) &\leq \text{PR}(\theta_{h',i_{h'}^*}) + 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'} \\ &\leq \inf_{\theta \in \Theta} \text{PR}(\theta) + 4(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{2^p m_0}} \end{aligned}$$

where the second inequality follows from Lemma 17, $n_{h'-1,i_{h'-1}^*}^{\text{deploy}} \geq 2^p$, and θ_{PO} is contained in cell $\mathcal{P}_{h',i_{h'}^*}$. Furthermore, by the condition of this lemma, it follows that

$$\text{PR}(\theta_{h',i}) \leq \text{PR}(\theta_{\text{PO}}) + 6(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'}.$$

In addition, since θ_{PO} is contained in cell $\mathcal{P}_{h',i_{h'}^*}$, Lemma 17 implies that

$$\text{PR}(\theta_{h',i_{h'}^*}) \leq \text{PR}(\theta_{\text{PO}}) + 4(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'}.$$

This implies that

$$\mathcal{N}_h(6(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'}) \geq \left\lfloor \frac{h_{\max}}{h'2^p} \right\rfloor + 1 \geq \left\lfloor \frac{h_{\max}}{h2^p} \right\rfloor + 1 \geq 2^{\alpha h d} + 1 \geq 2^{\alpha h' d} + 1$$

where $\lfloor h_{\max}/h'2^p \rfloor$ comes from cells $\mathcal{P}_{h',i}$ and 1 is due to cell $\mathcal{P}_{h',i_{h'}^*}$ in the first inequality, the second and the fourth inequalities hold because $h' \leq h$, and the third inequality comes from the condition of the lemma. This in turn implies that $\mathcal{N}_h(6(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha h'}) > 2^{\alpha h' d}$, a contradiction. Therefore, it follows that $\perp_{h',p} = h'$. Then the induction argument shows that $\perp_{h,p} = h$, as required. \blacksquare

Next, we prove Lemma 10 which shows that

$$\begin{aligned} & \text{PR}(\theta_T) - \text{PR}(\theta_{\text{PO}}) \\ & \leq 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha(\perp_{h_{\max}, p} + 1)} + \frac{8\mathfrak{E}^*(f) + 8\sqrt{\log(T/\delta)}}{\sqrt{2^p m_0}} + \frac{4\mathfrak{E}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}. \end{aligned}$$

Proof [Proof of Lemma 10] Let $p \in [0 : p_{\max}]$, and let

$$(h, i) \in \underset{(h, i)}{\text{argmin}} \left\{ \widehat{\text{PR}}(\theta_{h, i}) : h \in [h_{\max} + 1], \mathcal{P}_{h, i} \in \mathcal{L}_h, n_{h, i}^{\text{deploy}} \geq 2^p \right\}.$$

Recall that $\theta_T(p)$ is set to $\theta_{h, i}$ and that we obtain $h_{\max} m_0$ new samples from $\mathcal{D}(\theta_T(p))$ from which we construct $\widehat{\text{PR}}(\theta_T(p))$. Moreover, $\widehat{\text{PR}}(\theta_T) \leq \widehat{\text{PR}}(\theta_T(p))$. As $E_{\text{clean}, \delta}$ holds, it follows that

$$\begin{aligned} \text{PR}(\theta_T) - \frac{2\mathfrak{E}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}} & \leq \widehat{\text{PR}}(\theta_T) \\ & \leq \widehat{\text{PR}}(\theta_T(p)) \\ & \leq \text{PR}(\theta_T(p)) + \frac{2\mathfrak{E}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}. \end{aligned} \quad (7)$$

Again, as $E_{\text{clean}, \delta}$ holds and $\theta_T(p) = \theta_{h, i}$,

$$\text{PR}(\theta_T(p)) \leq \widehat{\text{PR}}(\theta_{h, i}) + \frac{2\mathfrak{E}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{n_{h, i}^{\text{deploy}} m_0}} \leq \widehat{\text{PR}}(\theta_{h, i}) + \frac{2\mathfrak{E}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{2^p m_0}}. \quad (8)$$

Recall that $\perp_{h_{\max}, p}$ is the depth of the deepest cell containing θ_{PO} opened for at least 2^p times until Algorithm 2 finishes opening cells of depth h_{\max} . Let $(\perp_{h_{\max}, p} + 1, i^*)$ denote the deepest cell containing θ_{PO} and a solution deployed at least 2^p times. By the choice of (h, i) , we have

$$\widehat{\text{PR}}(\theta_{h, i}) \leq \widehat{\text{PR}}(\theta_{\perp_{h_{\max}, p} + 1, i^*}) \leq \text{PR}(\theta_{\perp_{h_{\max}, p} + 1, i^*}) + \frac{2\mathfrak{E}^*(f) + 2\sqrt{\log(T/\delta)}}{\sqrt{2^p m_0}} \quad (9)$$

where the second inequality holds because $E_{\text{clean}, \delta}$ holds and $n_{\perp_{h_{\max}, p} + 1, i^*}^{\text{deploy}} \geq 2^p$. Moreover, since the parent cell of $\mathcal{P}_{\perp_{h_{\max}, p} + 1, i^*}$ is opened at least 2^p times, it means that the parent cell contains a solution deployed at least 2^p times. Then by Lemma 17, it follows that

$$\text{PR}(\theta_{\perp_{h_{\max}, p} + 1, i^*}) \leq \text{PR}(\theta_{\text{PO}}) + 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha(\perp_{h_{\max}, p} + 1)} + \frac{4\mathfrak{E}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{2^p m_0}}. \quad (10)$$

In summary, we deduce from (7) – (10) that

$$\begin{aligned} & \text{PR}(\theta_T) - \text{PR}(\theta_{\text{PO}}) \\ & \leq 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha(\perp_{h_{\max}, p} + 1)} + \frac{8\mathfrak{E}^*(f) + 8\sqrt{\log(T/\delta)}}{\sqrt{2^p m_0}} + \frac{4\mathfrak{E}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}, \end{aligned}$$

as required. ■

Note that the regret bound given by Lemma 10 holds for any $p \in [0 : p_{\max}]$. Hence, to prove an upper bound on the regret $\text{PR}(\theta_T) - \text{PR}(\theta_{\text{PO}})$, we need to choose an appropriate p that achieves a small value of

$$2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha(\perp_{h_{\max}, p} + 1)} + \frac{8\mathfrak{C}^*(f) + 8\sqrt{\log(T/\delta)}}{\sqrt{2^p m_0}}.$$

As in [Bartlett et al. \(2019\)](#), the strategy is to choose p under which there is a strong lower bound on $\perp_{h_{\max}, p} + 1$. In our case, however, we have the additional term $\tilde{O}(1/\sqrt{2^p})$. In fact, we will argue that the choice of p , under which $\perp_{h_{\max}, p} + 1$ is large, also makes the additional term small.

For simplicity, we use notations ρ , ν , and B defined as

$$\rho = 2^{-\alpha}, \quad \nu = (2\sqrt{D})^\alpha L_z \varepsilon, \quad B = \frac{2\sqrt{2} \left(\mathfrak{C}^*(f) + \sqrt{\log(T/\delta)} \right)}{\sqrt{m_0}}.$$

With these notations, Lemma 18 can be restated as follows.

Lemma 19 *Assume that $E_{\text{clean}, \delta}$ holds for some $\delta \in (0, 1)$. Let d denote the $(\nu, \rho, 1)$ -near-optimality dimension $d(\nu, \rho, 1)$. For any $h \in [\lfloor h_{\max}/2^p \rfloor]$ and $p \in [0 : \lfloor \log_2(h_{\max}/h) \rfloor]$, if the following condition holds, then $\perp_{h, p} = h$.*

$$\frac{B}{\sqrt{2^{p+1}}} \leq \nu \rho^h \quad \text{and} \quad \frac{h_{\max}}{h 2^p} \geq \rho^{-dh}.$$

Next, we define \tilde{h} and \tilde{p} as the numbers satisfying the following condition.

$$\frac{h_{\max} \nu^2 \rho^{2\tilde{h}}}{\tilde{h} B^2} = \rho^{-\tilde{h}d} \quad \text{and} \quad \frac{B}{\sqrt{2^{\tilde{p}}}} = \nu \rho^{\tilde{h}}.$$

Then by definition of the Lambert W function, we have

$$\tilde{h} = \frac{1}{(d+2) \log(1/\rho)} W \left(\frac{h_{\max} \nu^2 (d+2) \log(1/\rho)}{B^2} \right).$$

Here, $B \geq L_z \varepsilon \cdot 2^{-\alpha \tilde{h}}$ holds if and only if $2^{\tilde{p}} \geq 1$. Hence, the case when $2^{\tilde{p}} \geq 1$ corresponds to the high-noise regime and the setting where $2^{\tilde{p}} < 1$ corresponds to the low-noise regime. Next, as in [Bartlett et al. \(2019\)](#), we define \ddot{h} and \ddot{p} as follows.

- (High-noise regime) Set $\ddot{h} = \tilde{h}$ and $\ddot{p} = \tilde{p}$.
- (Low-noise regime) Set \ddot{h} as $\ddot{h} = \bar{h}$ that satisfies $h_{\max}/\bar{h} = \rho^{-d\bar{h}}$ and $\ddot{p} = 0$.

Note that for this choice of \ddot{h} and \ddot{p} , we have $h_{\max}/\ddot{h} 2^{\ddot{p}} = \rho^{-d\ddot{h}}$ under both regimes. Moreover, with Lemma 19, we may argue that the following statement holds.

Lemma 20 ([Bartlett et al. \(2019\)](#)) *Assume that $E_{\text{clean},\delta}$ holds for some $\delta \in (0, 1)$. Let d denote the $(\nu, \rho, 1)$ -near-optimality dimension $d(\nu, \rho, 1)$. Then $\ddot{h} \leq \tilde{h}$ and*

$$\lfloor h_{\max, \lfloor \ddot{p} \rfloor} \rfloor + 1 \geq \lfloor \ddot{h} \rfloor + 1 \geq \ddot{h}$$

under both the high-noise and low-noise regimes.

Now we are ready to complete the proof of Theorem 11.

Proof [Proof of Theorem 11] Under $E_{\text{clean},\delta}$, Lemmas 10 and 20 imply that

$$\text{PR}(\theta_T) - \text{PR}(\theta_{\text{PO}}) \leq 2(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha \ddot{h}} + \frac{8\mathfrak{C}^*(f) + 8\sqrt{\log(T/\delta)}}{\sqrt{2^{\lfloor \ddot{p} \rfloor} m_0}} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}$$

holds.

Let us first consider the low-noise regime. Since $2^{\ddot{p}} < 1$, we know that $B < \nu \rho^{\ddot{h}}$. By Lemma 20, we have $\ddot{h} \leq \tilde{h}$, which implies that $B < \nu \rho^{\tilde{h}} \leq \nu \rho^{\ddot{h}}$. Then as $\ddot{p} = 0$ under the low-noise regime, it follows that

$$\text{PR}(\theta_T) - \text{PR}(\theta_{\text{PO}}) \leq (2 + 2\sqrt{2})(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha \ddot{h}} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}.$$

Moreover, if $h_{\max} \geq 1$, then as $B \leq \nu \rho^{\ddot{h}}$,

$$\text{PR}(\theta_T) - \text{PR}(\theta_{\text{PO}}) \leq (2 + 3\sqrt{2})(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha \ddot{h}}.$$

When $d = 0$, we have $\ddot{h} = h_{\max}$. When $d > 0$, we have

$$\ddot{h} = \frac{1}{\alpha d \log 2} W(h_{\max} \alpha d \log 2).$$

For the high-noise regime, we have $\ddot{h} = \tilde{h}$ and

$$\frac{8\mathfrak{C}^*(f) + 8\sqrt{\log(T/\delta)}}{\sqrt{2^{\lfloor \ddot{p} \rfloor} m_0}} = \frac{4B}{\sqrt{2^{\lfloor \ddot{p} \rfloor + 1}}} \leq \frac{4B}{\sqrt{2^{\ddot{p}}}} = 4\nu \rho^{\tilde{h}}.$$

Therefore, under the high-noise regime, we have

$$\text{PR}(\theta_T) - \text{PR}(\theta_{\text{PO}}) \leq 6(2\sqrt{D})^\alpha L_z \varepsilon 2^{-\alpha \tilde{h}} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}.$$

Recall that \tilde{h} is given by

$$\tilde{h} = \frac{1}{\alpha(d+2) \log 2} W\left(\frac{(4D)^\alpha \alpha(d+2) \log 2}{8(\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)})^2} L_z^2 \varepsilon^2 m_0 h_{\max}\right).$$

Lastly, [Hoorfar and Hassani \(2008\)](#) showed that if $x \geq e$, then $W(x) \geq \log(x/\log(x))$. Hence, if $d > 0$ and $h_{\max} \alpha d \log 2 \geq e$ under the low-noise regime, then θ satisfies

$$\text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) \leq (2 + 3\sqrt{2})(2\sqrt{D})^\alpha L_z \varepsilon \left(\frac{h_{\max} \alpha d \log 2}{\log(h_{\max} \alpha d \log 2)}\right)^{-\frac{1}{d}}.$$

Moreover, if $B^2 h_{\max} \nu^2 \alpha (d+2) \log 2 \geq e$, then

$$\begin{aligned} & \text{PR}(\theta) - \text{PR}(\theta_{\text{PO}}) \\ & \leq 6(2\sqrt{D})^\alpha L_z \varepsilon \left(\frac{h_{\max} \nu^2 \alpha (d+2) \log 2 / B^2}{\log(h_{\max} \nu^2 \alpha (d+2) \log 2 / B^2)} \right)^{-\frac{1}{d+2}} + \frac{4\mathfrak{C}^*(f) + 4\sqrt{\log(T/\delta)}}{\sqrt{h_{\max} m_0}}, \end{aligned}$$

as required. ■

References

- M. G. Azar, A. Lazaric, and E. Brunskill. Online stochastic optimization under correlated bandit feedback. In E. P. Xing and T. Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1557–1565, Beijing, China, 22–24 Jun 2014. PMLR. URL <https://proceedings.mlr.press/v32/azar14.html>.
- P. L. Bartlett, V. Gabillon, and M. Valko. A simple parameter-free and adaptive approach to optimization under a minimal local smoothness assumption. In A. Garivier and S. Kale, editors, *Proceedings of the 30th International Conference on Algorithmic Learning Theory*, volume 98 of *Proceedings of Machine Learning Research*, pages 184–206. PMLR, 22–24 Mar 2019. URL <https://proceedings.mlr.press/v98/bartlett19a.html>.
- Y. Bechavod, K. Ligett, S. Wu, and J. Ziani. Gaming helps! learning from strategic interactions in natural dynamics. In A. Banerjee and K. Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 1234–1242. PMLR, 13–15 Apr 2021. URL <https://proceedings.mlr.press/v130/bechavod21a.html>.
- G. Brown, S. Hod, and I. Kalemaj. Performative prediction in a stateful world. In G. Camps-Valls, F. J. R. Ruiz, and I. Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 6045–6061. PMLR, 28–30 Mar 2022. URL <https://proceedings.mlr.press/v151/brown22a.html>.
- M. Brückner, C. Kanzow, and T. Scheffer. Static prediction games for adversarial learning problems. *Journal of Machine Learning Research*, 13(85):2617–2654, 2012. URL <http://jmlr.org/papers/v13/brueckner12a.html>.
- S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári. $|i_x|/i_x$ -armed bandits. *Journal of Machine Learning Research*, 12(46):1655–1695, 2011a. URL <http://jmlr.org/papers/v12/bubeck11a.html>.
- S. Bubeck, G. Stoltz, and J. Y. Yu. Lipschitz bandits without the Lipschitz constant. In J. Kivinen, C. Szepesvári, E. Ukkonen, and T. Zeugmann, editors, *Algorithmic Learning Theory*, pages 144–158, Berlin, Heidelberg, 2011b. Springer Berlin Heidelberg. ISBN 978-3-642-24412-4.

- Y. Chen, Y. Liu, and C. Podimata. Learning strategy-aware linear classifiers. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 15265–15276. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/ae87a54e183c075c494c4d397d126a66-Paper.pdf.
- N. Dalvi, P. Domingos, Mausam, S. Sanghai, and D. Verma. Adversarial classification. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, page 99–108, New York, NY, USA, 2004. Association for Computing Machinery. ISBN 1581138881. doi: 10.1145/1014052.1014066. URL <https://doi.org/10.1145/1014052.1014066>.
- J. Dong, A. Roth, Z. Schutzman, B. Waggoner, and Z. S. Wu. Strategic classification from revealed preferences. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, EC '18, page 55–70, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450358293. doi: 10.1145/3219166.3219193. URL <https://doi.org/10.1145/3219166.3219193>.
- R. Dong, H. Zhang, and L. Ratliff. Approximate regions of attraction in learning with decision-dependent distributions. In F. Ruiz, J. Dy, and J.-W. van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 11172–11184. PMLR, 25–27 Apr 2023. URL <https://proceedings.mlr.press/v206/dong23b.html>.
- D. Drusvyatskiy and L. Xiao. Stochastic optimization with decision-dependent distributions. *Mathematics of Operations Research*, 48(2):954–998, 2023. doi: 10.1287/moor.2022.1287. URL <https://doi.org/10.1287/moor.2022.1287>.
- J.-B. Grill, M. Valko, R. Munos, and R. Munos. Black-box optimization of noisy functions with unknown smoothness. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/ab817c9349cf9c4f6877e1894a1faa00-Paper.pdf.
- M. Hardt and C. Mendler-Dünner. Performative prediction: Past and future, 2023.
- M. Hardt, N. Megiddo, C. Papadimitriou, and M. Wootters. Strategic classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, ITCS '16, page 111–122, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450340571. doi: 10.1145/2840728.2840730. URL <https://doi.org/10.1145/2840728.2840730>.
- A. Hoorfar and M. Hassani. Inequalities on the lambert function and hyperpower function. *Journal of Inequalities in Pure & Applied Mathematics*, 9(2):5–9, 2008. URL <http://eudml.org/doc/130024>.
- Z. Izzo, L. Ying, and J. Zou. How to learn when data reacts to your model: Performative gradient descent. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning*

- Research*, pages 4641–4650. PMLR, 18–24 Jul 2021. URL <https://proceedings.mlr.press/v139/izzo21a.html>.
- Z. Izzo, J. Zou, and L. Ying. How to learn when data gradually reacts to your model. In G. Camps-Valls, F. J. R. Ruiz, and I. Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 3998–4035. PMLR, 28–30 Mar 2022. URL <https://proceedings.mlr.press/v151/izzo22a.html>.
- M. Jagadeesan, T. Zrnic, and C. Mendler-Dünner. Regret minimization with performative feedback. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 9760–9785. PMLR, 17–23 Jul 2022. URL <https://proceedings.mlr.press/v162/jagadeesan22a.html>.
- D. R. Jones, C. D. Perttunen, and B. E. Stuckman. Lipschitzian optimization without the lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1):157–181, 1993. doi: 10.1007/BF00941892. URL <https://doi.org/10.1007/BF00941892>.
- L. Kantorovich and G. S. Rubinstein. On a space of totally additive functions. *Vestnik Leningrad. Univ*, 13:52–59, 1958.
- R. Kleinberg, A. Slivkins, and E. Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC ’08, page 681–690, New York, NY, USA, 2008. Association for Computing Machinery. ISBN 9781605580470. doi: 10.1145/1374376.1374475. URL <https://doi.org/10.1145/1374376.1374475>.
- Q. Li and H.-T. Wai. State dependent performative prediction with stochastic approximation. In G. Camps-Valls, F. J. R. Ruiz, and I. Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 3164–3186. PMLR, 28–30 Mar 2022. URL <https://proceedings.mlr.press/v151/li22c.html>.
- W. Li, H. Li, J. Honorio, and Q. Song. Pyxab – a python library for \mathcal{X} -armed bandit and online blackbox optimization algorithms, 2023a. URL <https://arxiv.org/abs/2303.04030>.
- W. Li, C.-H. Wang, G. Cheng, and Q. Song. Optimum-statistical collaboration towards general and efficient black-box optimization. *Transactions on Machine Learning Research*, 2023b. ISSN 2835-8856. URL <https://openreview.net/forum?id=C1Icmwldlxn>.
- C. Maheshwari, C.-Y. Chiu, E. Mazumdar, S. Sastry, and L. Ratliff. Zeroth-order methods for convex-concave min-max problems: Applications to decision-dependent risk minimization. In G. Camps-Valls, F. J. R. Ruiz, and I. Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 6702–6734. PMLR, 28–30 Mar 2022. URL <https://proceedings.mlr.press/v151/maheshwari22a.html>.

- C. Malherbe and N. Vayatis. Global optimization of Lipschitz functions. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2314–2323. PMLR, 06–11 Aug 2017. URL <https://proceedings.mlr.press/v70/malherbe17a.html>.
- C. Mendler-Dünner, J. Perdomo, T. Zrnic, and M. Hardt. Stochastic optimization for performative prediction. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 4929–4939. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/33e75ff09dd601bbe69f351039152189-Paper.pdf.
- J. P. Miller, J. C. Perdomo, and T. Zrnic. Outside the echo chamber: Optimizing the performative risk. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7710–7720. PMLR, 18–24 Jul 2021. URL <https://proceedings.mlr.press/v139/miller21a.html>.
- S. Milli, J. Miller, A. D. Dragan, and M. Hardt. The social cost of strategic classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* '19*, page 230–239, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450361255. doi: 10.1145/3287560.3287576. URL <https://doi.org/10.1145/3287560.3287576>.
- M. Mofakhami, I. Mitliagkas, and G. Gidel. Performative prediction with neural networks. In F. Ruiz, J. Dy, and J.-W. van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 11079–11093. PMLR, 25–27 Apr 2023. URL <https://proceedings.mlr.press/v206/mofakhami23a.html>.
- R. Munos. Optimistic optimization of a deterministic function without the knowledge of its smoothness. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper_files/paper/2011/file/7e889fb76e0e07c11733550f2a6c7a5a-Paper.pdf.
- J. Perdomo, T. Zrnic, C. Mendler-Dünner, and M. Hardt. Performative prediction. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7599–7609. PMLR, 13–18 Jul 2020. URL <https://proceedings.mlr.press/v119/perdomo20a.html>.
- M. Ray, L. J. Ratliff, D. Drusvyatskiy, and M. Fazel. Decision-dependent risk minimization in geometrically decaying dynamic environments. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelfth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages

- 8081–8088. AAAI Press, 2022. doi: 10.1609/AAAI.V36I7.20780. URL <https://doi.org/10.1609/aaai.v36i7.20780>.
- M. Shaked and J. Shanthikumar. *Stochastic Orders*. Springer Series in Statistics. Springer New York, 2007. ISBN 9780387346755.
- X. Shang, E. Kaufmann, and M. Valko. General parallel optimization a without metric. In A. Garivier and S. Kale, editors, *Proceedings of the 30th International Conference on Algorithmic Learning Theory*, volume 98 of *Proceedings of Machine Learning Research*, pages 762–788. PMLR, 22–24 Mar 2019. URL <https://proceedings.mlr.press/v98/xuedong19a.html>.
- A. Slivkins. Multi-armed bandits on implicit metric spaces. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL https://proceedings.neurips.cc/paper_files/paper/2011/file/7634ea65a4e6d9041cfd3f7de18e334a-Paper.pdf.
- M. Valko, A. Carpentier, and R. Munos. Stochastic simultaneous optimistic optimization. In S. Dasgupta and D. McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 19–27, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR. URL <https://proceedings.mlr.press/v28/valko13.html>.
- C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008. ISBN 9783540710509. URL https://books.google.co.kr/books?id=hV8o5R7_5tkC.
- T. Zrnic, E. Mazumdar, S. Sastry, and M. Jordan. Who leads and who follows in strategic classification? In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 15257–15269. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/812214fb8e7066bfa6e32c626c2c688b-Paper.pdf.