RISK-AVERSE LEARNING WITH NON-STATIONARY DISTRIBUTIONS

Siyi Wang, Zifan Wang, Karl H. Johansson
Department of Decision and Control Systems
KTH Royal Institute of Technology
Stockholm, SE 11423, Sweden
{siyiw,zifanw,kallej}@kth.se

Xinlei Yi
College of Electronics and Information Engineering
Tongji University
Shanghai, China
xinleiyi@tongji.edu.cn

Michael M. Zavlanos
Mechanical Engineering and Material Science
Duke University
Durham, USA
michael.zavlanos@duke.edu

Sandra Hirche
Chair of Information-oriented Control
Technical University of Munich
Munich, Germany
hirche@tum.de

ABSTRACT

Considering non-stationary environments in online optimization enables decision-maker to effectively adapt to changes and improve its performance over time. In such cases, it is favorable to adopt a strategy that minimizes the negative impact of change to avoid potentially risky situations. In this paper, we investigate risk-averse online optimization where the distribution of random costs changes over time. The Conditional Value at Risk (CVaR) is employed as risk measure. Due to the difficulty in obtaining the exact CVaR gradient, we employ a zeroth-order optimization approach that queries the cost values multiple times per iteration and estimates the CVaR gradient based on these samples. In regret analysis, the varying distributions are captured by a novel variation metric based on the Wasserstein distance. Given that the distribution variation is sublinear in the iteration horizon, we show that the developed learning algorithm achieves sublinear dynamic regret with high probability for both convex and strongly convex functions. Moreover, theoretical results suggest dynamic regret bounds decrease with the increasing sampling number until it reaches a specific limit. Finally, we provide numerical experiments of dynamic pricing in a parking lot to illustrate the efficacy of the designed algorithm.

1 INTRODUCTION

Online convex optimization is a powerful framework that deals with decision-making problems in dynamic and uncertain environments (Hazan, 2006). It has many applications, including traffic routing (Sessa et al., 2019), resource allocation (Chen et al., 2017), and online marketing (Gordon et al., 2008). In online optimization, the decision maker sequentially updates its decision in a changing environment, relying on historical information such as observations of previous actions and costs. The decisions generated by the optimization algorithm induce a sequence of associated cost values. The performance of the algorithm is evaluated using regret, which is the accumulated loss generated by the algorithm against the optimal actions in hindsight.

Non-stationary environments describe scenarios where the underlying conditions of the system change over time. The reason for environmental changes can be variations in the distribution of the stochastic cost function. For instance, in dynamic pricing for vehicle parking (Ray et al., 2022), the pricing depends on real-time changes in demand and supply; conversely, price adjustments influence the distribution of the occupancy rate. In non-cooperative games, the objective function of each agent follows a distribution, which may evolve over time in response to action updates of other agents (Narang et al., 2022). Compared with the standard regret analysis assuming a stationary environment, dynamic regret accounts for the impact of fluctuations in non-stationary environments

(Besbes et al., 2015; Zhao & Guan, 2018). Metrics used for analyzing dynamic regret include variations in the cost functions (Besbes et al., 2015), in the optimal actions (Zhao et al., 2021), and in the distributions Jiang et al. (2025); Shames & Farokhi (2020). When making decisions under uncertainty, it is often essential to consider the entire distribution of potential outcomes rather than focusing on a single optimal point. Specifically, (Jiang et al., 2025; Shames & Farokhi, 2020) use the Wasserstein distance, which measures the dissimilarity between probability distributions, to quantify changes in non-stationary distributions. In this paper, we employ the distribution variation metric proposed in (Jiang et al., 2025). It is meaningful to investigate risk-averse learning under nonstationarity, as it is inherently sensitive to environmental changes.

When the decision-maker is sensitive to potential negative consequences, its primary consideration is not minimizing the expected cost, but rather reducing the risk of a catastrophe. For example, in the financial market, it is unfavorable to pursue a strategy that entails high risks despite offering the highest expected reward. Some measures are proposed to model the potential risk, such as Value at Risk (VaR) (Linsmeier & Pearson, 2000) and Conditional Value at Risk (CVaR) (Rockafellar & Uryasev, 2002). Given a risk level $\alpha \in (0, 1]$, the CVaR value describes the average value of the α-tail distribution of the stochastic cost. It has a coherent risk measure property, which offers some mathematical properties that facilitate theoretical analysis. The classical paper (Rockafellar et al., 2000) formulates the computation of the CVaR value as an optimization problem by introducing an additional decision variable to construct an augmented objective function. It enables the application of CVaR for the optimization problem in bandit optimization (Cardoso & Xu, 2019), online games (Wang et al., 2022a;b) and safe control (Chapman & Lessard, 2021; Kishida & Cetinkaya, 2022; Chapman et al., 2019). However, since the computation of the CVaR gradient relies on the distribution of the stochastic cost, CVaR optimization problems rarely enjoy a closed-form expression. To handle this problem, a common approach is to estimate the CVaR gradient using zeroth-order optimization algorithms, see (Cardoso & Xu, 2019; Wang et al., 2022b).

1.1 Our Contributions

In this paper, we study online CVaR optimization with time-varying distributions. The risk-averse learning algorithm is designed for both convex and strongly convex cost functions. The algorithm is periodically restarted to ensure that the decision-maker effectively adapts to changing distributions. At each iteration, multiple samples are queried to construct an empirical distribution of random costs. Based on it, we estimate CVaR values and construct the CVaR gradient estimate via the zeroth-order optimization approach. The main contributions are summarized below:

Table 1: Order of Regret

Sampling parameter	Convex	Strongly convex
$0 < a \leq 1$	$T^{\frac{4}{4+a}} V_D^{\frac{a}{4+a}}$	$T^{\frac{4}{4+a}} V_D^{\frac{a}{4+a}}$
$1 < a \leq \frac{4}{3}$	$T^{\frac{4}{5}} V_D^{\frac{1}{5}}$	$T^{\frac{4}{4+a}} V_D^{\frac{a}{4+a}}$
$a > \frac{4}{3}$	$T^{\frac{4}{5}} V_D^{\frac{1}{5}}$	$T^{\frac{3}{4}} V_D^{\frac{1}{4}}$

1. We analyze the dynamic regret bound for both convex and strongly convex losses using distribution variations measured by the Wasserstein distance. Compared to [24], which uses variations of the optimal points to measure distribution changes, our approach better captures environmental variations.

2. The inherently complex nature of CVaR functions poses challenges for regret analysis. To address it, we theoretically bound the variation in the CVaR value induced by distribution changes. Moreover, we establish theoretical properties of the CVaR function, including strong convexity when the random cost is strongly convex.

3. We propose a risk-averse learning algorithm with a restarting procedure. We show that this algorithm achieves sublinear regret with high probability, provided the distribution variation is sublinear in the total number of iterations. As shown in Table 1, our convergence results explicitly show the regret bound decreases as the total sampling number increases, which is reflected by the parameter a. This reduction ceases when reaching a limit even if the number of samples continues to increase.

1.2 RELATED WORKS

Non-stationary optimization There is a growing interest in non-stationary online convex optimization, see (Besbes et al., 2015; Zhao & Guan, 2018; Ray et al., 2022). To the best of our knowledge, only a few works address environmental changes that result in non-stationary distributions, e.g., (Cao et al., 2020; Jiang et al., 2025; Shames & Farokhi, 2020). For instance, Cao et al. (2020) and Shames & Farokhi (2020) investigate time-varying distributions by quantifying the variations in optimal points. To deal with evolving distributions, recent works have turned to Wasserstein-based metrics (Jiang et al., 2025). Specifically, Jiang et al. (2025) employs a Wasserstein-based non-stationarity budget (WBNB) that measures cumulative deviations from a global average.

Risk-averse learning Another related research line is risk-averse learning using CVaR as the risk measure (Tamkin et al., 2019; Soma & Yoshida, 2020; Urpí et al., 2021; Cardoso & Xu, 2019; Wang et al., 2022b;a). For instance, (Tamkin et al., 2019) develops a risk-averse multi-armed bandit algorithm and provides an upper confidence bound for CVaR. In the context of online convex games, Wang et al. (2022b) proposes a learning scheme that achieves sublinear regret with high probability. However, all the above works focus on stationary environments. A notable exception is Liang et al. (2021), which explores risk-averse online optimization specifically for portfolio selection with linear costs.

Notations: Let $\|\cdot\|$ denote the l_2 norm. Let $\lceil\cdot\rceil$ denote the ceiling function. Let $\mathbf{1}(\cdot)$ denote the indicator function. For a random variable X, let $X \sim \mathcal{D}_X$ denote that X is distributed according to the distribution $\mathcal{D}_X$. Let the notation $\mathcal{O}$ hide the constant and $\tilde{\mathcal{O}}$ hide constant and polylogarithmic factors of the number of iterations T, respectively. Let $A \oplus B = \{a + b | a \in A, b \in B\}$ denote the Minkowski sum of two sets of position vectors A and B in Euclidean space.

2 PROBLEM FORMULATION

Consider the cost function $J(x, \xi) : \mathcal{X} \times \Xi \to \mathbb{R}$, where $\xi \subseteq \Xi$ denotes random noise and $x \in \mathcal{X}$ denotes the decision variable with $\mathcal{X} \subseteq \mathbb{R}^d$ being the admissible set. Without loss of generality, we assume that $\mathcal{X}$ contains the ball of radius r centered at the origin, which is denoted as $r\mathbb{B} \subseteq \mathbb{R}^d$. Denote the diameter of the admissible set $\mathcal{X}$ as $D_x = \sup_{x,y\in\mathcal{X}} \|x - y\|$.

2.1 CVAR

We use CVaR as risk measure. Suppose $J(x, \xi)$ has the cumulative distribution function $F(y) = P(J(x, \xi) \leq y)$, and is bounded by $U > 0$, i.e., $|J(x, \xi)| \leq U$. Given a confidence level $\alpha \in (0, 1]$, the α-VaR is $J^\alpha = F^{-1}(\alpha) := \inf\{y : F(y) \geq \alpha\}$. The α-CVaR describes the expectation of the α-fraction of the worst outcomes of $J(x, \xi)$ (Rockafellar et al., 2000), which is defined as

$$C(x) := \mathrm{CVaR}_\alpha [J(x, \xi)] = \mathbb{E}_F [J(x, \xi) | J(x, \xi) \geq J^\alpha] .$$

2.2 NON-STATIONARY DISTRIBUTION

Non-stationary stochastic optimization necessitates some metric to capture temporal uncertainties of the dynamic environment. A classic metric is the cumulative variations of the cost functions over time. In practice, environmental fluctuations might be more intuitively modeled as the time-varying distribution. To account for this, we employ the Wasserstein distance to quantify distributional variations, as it effectively captures a broad range of distributional shifts, including changes in shape and support (Shen et al., 2023). The Wasserstein distance measures dissimilarity between probability distributions defined over a given metric space, as detailed in (Kantorovich, 1960; Edwards, 2011).

Let (Ω, d) be a probability space, where Ω is a set and d is a metric on Ω. Let $\mathcal{D}_x$ and $\mathcal{D}_y$ be two probability distribution on Ω, the dual form of the Wasserstein distance is given as follows.

Lemma 1. *(Edwards, 2011) For any fixed $K > 0$, $W_1(\mathcal{D}_x, \mathcal{D}_y) = \frac{1}{K} \sup_{\|f\|_L \leq K} \{\mathbb{E}_{x\sim\mathcal{D}_x}[f(x)] - \mathbb{E}_{y\sim\mathcal{D}_y}[f(y)]\}$, where $\|\cdot\|_L$ is the Lipschitz norm, The right-hand side is the Kantorovich–Rubenstein dual form of the Wasserstein distance metric.*

As mentioned in Section 1, some works quantify regret achieved by algorithm using the function variation, i.e., $\sum_{t=2}^{T} \sup_{x\in\mathcal{X}} |f_t(x) - f_{t-1}(x)|$, which measures the cumulative changes of the func-

tion values over time. Inspired by this definition, we introduce the concept of distribution variation as follows.

Definition 1. ***(Distribution variation)*** *Let $\{\mathcal{D}_t\}_{t=1}^T \in \mathcal{D}$ be the distribution on metric space Ω, with $\mathcal{D}$ being the admissible set of distribution sequences. The distribution variation along iterations $\{1, \ldots, T\}$ is $\sum_{t=2}^T W_1(\mathcal{D}_{t-1}, \mathcal{D}_t)$.*

2.3 Problem statement

Consider the time-varying random noise $\xi_t \sim \mathcal{D}_t$ and the corresponding cumulative distribution function F_t, the α-CVaR of the function $J(x, \xi_t)$ is written as:

$$C_t(x) := \mathrm{CVaR}_\alpha \left[J(x, \xi_t)\right] = \mathbb{E}_{F_t}\left[J(x, \xi_t) | J(x, \xi_t) \geq J^\alpha\right],$$

where ξ_t is independent of past actions $\{x_0, \ldots, x_{t-1}\}$ and $\{\xi_t\}_t$ is an independent and identically distributed (i.i.d.) random variable. Additionally, we make the following assumptions on the cost function, which are common in the online learning literature, see (Hazan et al., 2016; Besbes et al., 2015).

Assumption 1. *The cost function $J(x, \xi_t)$ is convex in x for every $\xi_t \in \Xi$.*

Assumption 2. *The cost function $J(x, \xi_t)$ is Lipschitz continuous in x for every $\xi_t \in \Xi$. That is, there exists a positive real constant L_0 such that, for all $x, y \in \mathcal{X}$, we have $|J(x, \xi_t) - J(y, \xi_t)| \leq L_0 \|x - y\|$.*

It follows the lemma for the CVaR function.

Lemma 2. *(Cardoso & Xu, 2019) Given Assumption 1, we have that $C_t(x)$ is convex in x.*

Assumption 3. *The distribution sequence $\{\mathcal{D}_t\}_{t=1}^T \in \mathcal{D}$ satisfies the variation budget V_D over the iteration horizon T: $\sum_{t=2}^T W_1(\mathcal{D}_{t-1}, \mathcal{D}_t) \leq V_D$, where the distribution variation budget is sublinear in the iteration horizon T, i.e., $V_D = \mathcal{O}(T^\beta)$ with $\beta \in [0, 1)$.*

Assumption 3 constrains the cumulative variations in distribution over the time horizon T. Assuming that the distribution variation budget is sublinear is essential for developing a no-regret learning algorithm in stochastic, non-stationary environments.

We use the dynamic regret to measure the performance of the designed algorithm, which is defined as the cumulative loss under the performed actions against the best actions in hindsight:

$$\mathrm{DR}(T) = \sum_{t=1}^T C_t(\hat{x}_t) - \sum_{t=1}^T C_t(x_t^*), \tag{1}$$

where the action $\hat{x}_t$ is selected according to the designed algorithm at iteration t, and $x_t^* = \arg\min_{x_t \in \mathcal{X}} C_t(x_t)$ denotes the one-step optimal action at iteration t, for $t = 1, \ldots, T$. Specifically, this paper aims to design a risk-averse learning algorithm such that the dynamic regret of this algorithm is bounded in terms of the distribution variation, i.e., $\lim_{T \to \infty} \frac{\mathrm{DR}(T)}{T} = 0$.

3 Main Result

In this section, we design a risk-averse learning algorithm for both convex and strongly convex cost functions. Then we analyze the dynamic regret using the distribution variation metric.

Since the exact CVaR gradient is generally unavailable, we use the zeroth-order optimization algorithm to estimate the CVaR gradient. To begin with, we construct a smoothed approximation of the CVaR function. Given a point $x \in \mathcal{X}$, define the perturbed action by $\hat{x} = x + \delta u$, where u is the direction vector sampled from a unit sphere $\mathbb{S}^d \in \mathbb{R}^d$ and δ is the perturbation radius, also known as the smoothing parameter. Then, the smoothed version of the CVaR function (Cardoso & Xu, 2019) is given as

$$C_t^\delta(x) = \mathbb{E}_{u \sim \mathbb{S}^d}[C_t(x + \delta u)]. \tag{2}$$

In the following, we present lemmas regarding properties of smoothed approximation of the CVaR function.

Lemma 3. *(Cardoso & Xu, 2019) Given Assumption 1, we have that $C_t^\delta(x)$ is convex in x.*

Lemma 4. *(Cardoso & Xu, 2019) Given Assumption 2, we have that $C_t^\delta(x)$ is L_0-Lipschitz in x and $|C_t^\delta(x) - C_t(x)| \leq \delta L_0$.*

Non-stationary environments require decision makers to continuously adapt to changing conditions. The generic idea of the *restarting procedure* is to let the learning algorithm reset its internal state and update its parameters, and therefore capture the new dynamics in the environment. The risk-averse learning algorithm is summarized in Algorithm 1.

Let $\mathcal{A}$ be an online optimization algorithm. We employ the restarting procedure to refresh the parameters and restart $\mathcal{A}$ every Δ_T iterations. Suppose that there are totally T iterations and they are divided into s interval with a length of $\Delta_T \in (1, T)$, where $s = \left\lceil \frac{T}{\Delta_T} \right\rceil$. Each interval is formally defined as $\mathcal{T}_j = \left\{t : (j-1)\Delta_T < t \leq \min\{T, j\Delta_T\}\right\}$, for $j = 1, \ldots, s$. For each iteration t, inside interval j, it holds that $j = \left\lceil \frac{t}{\Delta_T} \right\rceil$ and its timestamp within the interval is slot τ, i.e.,

$$\tau = t - (j-1)\Delta_T. \tag{3}$$

Let the algorithm query the cost function values multiple times at each iteration to improve the estimation accuracy under the changing distribution. Then, the sampling strategy depending on the slot τ is given as $n_t = \phi(\tau)$, which satisfies

$$\sum_{\tau=1}^{\Delta_T} \frac{1}{\sqrt{\phi(\tau)}} \leq c\Delta_T^{1-\frac{a}{2}} \tag{4}$$

with the tuning parameter $a > 0$ and some constant $c > 0$. Specifically, a higher value of a corresponds to a larger sampling number over the iteration horizon. Similarly, the learning rate at each interval is designed according to $\eta_t = \sigma(\tau)$, where the function σ will be designed later. In Algorithm 1, we refresh the evolution of the sampling number n_t and the learning rate η_t every Δ_T iterations.

Algorithm 1 Risk-averse learning with restarting procedure

Require: Initial value x_0, iteration horizon T, risk level α, interval Δ_T, smoothing parameter δ.

for iteration $t = 1, \ldots, T$ **do**
  Identify interval $j = \left\lceil \frac{t}{\Delta_T} \right\rceil$
  Identify slot in interval $\tau = t - (j-1)\Delta_T$
  Select sampling number $n_t = \phi(\tau)$ and learning rate $\eta_t = \sigma(\tau)$
  Sample $u_t \in \mathbb{S}^d$
  Play $\hat{x}_t = x_t + \delta u_t$
  for $i = 1, \ldots, n_t$ **do**
    Play $\hat{x}_t$ and obtain $J_t(\hat{x}_t, \xi_t^i)$
  end for
  Build empirical distribution function $\hat{F}_t(y)$ given in (5)
  Estimate CVaR: $\mathrm{CVaR}_\alpha[\hat{F}_t]$
  Construct gradient estimate $\hat{g}_t = \frac{d}{\delta}\mathrm{CVaR}_\alpha[\hat{F}_t]u_t$
  Update x: $x_{t+1} \leftarrow \mathcal{P}_{\mathcal{X}_t^\delta}(x_t - \eta_\tau \hat{g}_t)$
end for

At iteration t, we implement the perturbed action $\hat{x}_t = x_t + \delta u_t$ for n_t times and obtain the cost function values denoted by $J(\hat{x}_t, \xi^i),\ i = 1, \ldots, n_t$. Then, we use the queried function values to construct the empirical distribution function, which is given as

$$\hat{F}_t(y) = \frac{1}{n_t}\sum_{i=1}^{n_t} \mathbf{1}\{J(\hat{x}_t, \xi_t^i) \leq y\}. \tag{5}$$

With (5), we construct the CVaR estimate $\mathrm{CVaR}_\alpha[\hat{F}_t]$ and further the CVaR gradient estimate

$$\hat{g}_t = \frac{d}{\delta}\mathrm{CVaR}_\alpha\big[\hat{F}_t\big]u_t. \tag{6}$$

The gradient descent update for the risk-averse learning process proceeds as:

$$x_{t+1} = \mathcal{P}_{\mathcal{X}^\delta}(x_t - \eta_t \hat{g}_t), \tag{7}$$

where $\mathcal{P}_{\mathcal{X}^\delta}(x) := \arg\min_{y \in \mathcal{X}^\delta} \|x - y\|^2$ denotes the projection operator with $\mathcal{X}^\delta = \{x \in \mathcal{X} | \frac{1}{1-\delta/r} x \in \mathcal{X}\}$ being the projection set. The projection keeps the sampled actions $\hat{x}_t$ inside of the admissible set $\mathcal{X}$, which establishes as $\big(1-\frac{\delta}{r}\big)\mathcal{X} \oplus \delta\mathbb{B} = \big(1-\frac{\delta}{r}\big)\mathcal{X} \oplus \frac{\delta}{r}r\mathbb{B} \subseteq \big(1-\frac{\delta}{r}\big)\mathcal{X} \oplus \frac{\delta}{r}\mathcal{X} = \mathcal{X}$. Without loss of generality, we let the initial action at the restarting slot be based on action learned

from the previous interval. As shown in Algorithm 1, each iteration involves sampling, CVaR computation, and projection operations. The complexity of these operations is generally modest, scaling primarily with the dimension d as well as the number of samples specified in (4).

Note that we construct the empirical distribution function of the cost function using finite samples, which induces the CVaR estimate error:

$$\hat{\epsilon}_t := \mathrm{CVaR}_\alpha[\hat{F}_t] - \mathrm{CVaR}_\alpha[F_t]. \tag{8}$$

In the following, we present two lemmas to bound the estimate error. The first lemma shows that the CVaR values with two cumulative distribution functions can be bounded by the sup difference of these two cumulative distribution functions, which is presented below.

Lemma 5. *(Wang et al., 2022a) Let F and G be two cumulative distribution functions of two random variables and the random variables are bounded by U. Then we have that $|\mathrm{CVaR}_\alpha[F] - \mathrm{CVaR}_\alpha[G]| \leq \frac{U}{\alpha} \sup_x |F(x) - G(x)|$.*

The following lemma shows the fluctuation of CVaR values is bounded in terms of the distribution shift.

Lemma 6. *Suppose $f(x)$ is L_0-Lipschitz in x. For two random variables X and Y with distributions $\mathcal{D}_X$ and $\mathcal{D}_Y$, respectively, we have $|\underset{X\sim\mathcal{D}_X}{\mathrm{CVaR}_\alpha}[f(X)] - \underset{Y\sim\mathcal{D}_Y}{\mathrm{CVaR}_\alpha}[f(Y)]| \leq \frac{L_0}{\alpha} W_1(\mathcal{D}_X, \mathcal{D}_Y)$.*

The proof of Lemma 6 is provided in the Appendix.

3.1 Convex case

In this section, we investigate the dynamic regret of Algorithm 1 for the convex case. The main result is presented in the following theorem.

Theorem 1. *Let Assumptions 1–3 hold. Suppose that the sampling numbers over iteration horizon T satisfy (4) with a constant $a > 0$.*

1. *When $a \in (0, 1]$, select $\delta = \left(\frac{V_D}{T}\right)^{\frac{a}{4+a}}$, $\eta_t = \left(\frac{V_D}{T}\right)^{\frac{3a}{4+a}}$, $\Delta_T = \left(\frac{T}{V_D}\right)^{\frac{4}{4+a}}$. Then, Algorithm 1 achieves $\mathrm{DR}(T) = \tilde{\mathcal{O}}(T^{\frac{4}{4+a}} V_D^{\frac{a}{4+a}})$ with high probability.*
2. *When $a > 1$, select $\delta = \left(\frac{V_D}{T}\right)^{\frac{1}{5}}$, $\eta_t = \left(\frac{V_D}{T}\right)^{\frac{3}{5}}$, $\Delta_T = \left(\frac{T}{V_D}\right)^{\frac{4}{5}}$. Then, Algorithm 1 achieves $\mathrm{DR}(T) = \tilde{\mathcal{O}}(T^{\frac{4}{5}} V_D^{\frac{1}{5}})$ with high probability.*

3.2 Strongly convex case

In the section, we further investigate Algorithm 1 for the strongly convex case. We first provide the following assumption and lemma related to the strongly convex condition.

Assumption 4. *The function $J(x, \xi_t) : \mathcal{X} \times \Xi \to \mathbb{R}$ is m-strongly convex in x for every $\xi_t \in \Xi$. That is, for all $x, y \in \mathcal{X}$ and every $\xi_t \in \Xi$, we have $J(y, \xi_t) \geq J(x, \xi_t) + \nabla_x J(x, \xi_t)^\top (y - x) + \frac{m}{2}\|x - y\|_2^2$.*

Lemma 7. *Given Assumption 4, we have that: (1) $C_t(x)$ is m-strongly convex in x; (2) $C_t^\delta(x)$ is m-strongly convex in x.*

The proof of Lemma 7 is provided in the Appendix. It follows the main result for the strongly convex case.

Theorem 2. *Let Assumptions 2–4 hold. Suppose that the sampling numbers over iteration horizon T satisfy (4) with a constant $a > 0$. Select the learning rate as $\eta_t = \sigma(\tau) = \frac{1}{m\tau}$, where the slot τ is determined by (3).*

1. *When $a \in (0, \frac{4}{3}]$, select $\delta = \left(\frac{V_D}{T}\right)^{\frac{a}{4+a}}$ and $\Delta_T = \left(\frac{T}{V_D}\right)^{\frac{4}{4+a}}$. Then, Algorithm 1 achieves dynamic regret $\mathrm{DR}(T) = \tilde{\mathcal{O}}(T^{\frac{4}{4+a}} V_D^{\frac{a}{4+a}})$ with high probability;*

2. *When* $a > \frac{4}{3}$*, select* $\delta = \left(\frac{V_D}{T}\right)^{\frac{1}{4}}$ *and* $\Delta_T = \left(\frac{T}{V_D}\right)^{\frac{3}{4}}$*. Then, Algorithm 1 achieves dynamic regret* $\mathrm{DR}(T) = \tilde{\mathcal{O}}(T^{\frac{3}{4}} V_D^{\frac{1}{4}})$ *with high probability.*

Remark 1. *Comparing Theorems 1 and 2, we observe that Algorithm 1 achieves the same regret bound in both the strongly convex and convex cases when* $a \in (0, 1]$*. However, for* $a > 1$*, it achieves a smaller regret bound in the strongly convex case compared to the convex case. This arises due to the cumulative error of* CVaR *gradient estimates in the strongly convex case, i.e.,* $\frac{d^2U^2}{\delta^2 m}(1+\ln T)\frac{T}{\Delta_T} + \frac{\sqrt{2}dUD_x\sqrt{\ln(2\Delta_T/\gamma)}}{\alpha\delta} T\Delta_T^{-\frac{a}{2}}$*, decreases as the tuning parameter* a *increases. The cumulative estimation error is one of the dominant terms in the regret bound (30) when* $a \in (0, 4/3]$ *and becomes negligible when* $a > 4/3$*. As a consequence, this regret bound reduction ceases when* $a > 4/3$ *and thus results in a smaller order compared to the convex case.*

4 SIMULATIONS

In this section, we consider the parking lot dynamic pricing problem, see (Ray et al., 2022). Factors such as parking prices, availability, and locations generally influence driving decisions. This encourages us to dynamically adjust the parking price according to real-time demand. Denote $r_t \in [0, 1]$ as curb occupancy rate. Let the occupancy rate be influenced by the price x_t and environmental uncertainties ξ_t, which is $r_t = \xi_t + Ax_t$, where $A = -0.15$ is the estimated price elasticity, which is determined by (Ray et al., 2022) through analysis of the real-world data. The uncertainty ξ_t is distributed according to the time-varying distribution $\mathcal{D}_t$, which will change periodically according to environmental effects such as climate conditions and dates. Specifically, to make it easy to find a parking space, it is desirable to maintain an occupancy rate of 70%. Hence, the loss function is defined as

$$J(x_t, \xi_t) = \|\xi_t + Ax_t - 0.7\|^2 + \frac{\nu}{2}\|x_t\|^2, \tag{9}$$

where $\nu = 0.001$ is the regularization parameter. To avoid the overcrowded situation, we aim at minimizing the risk-averse objective function

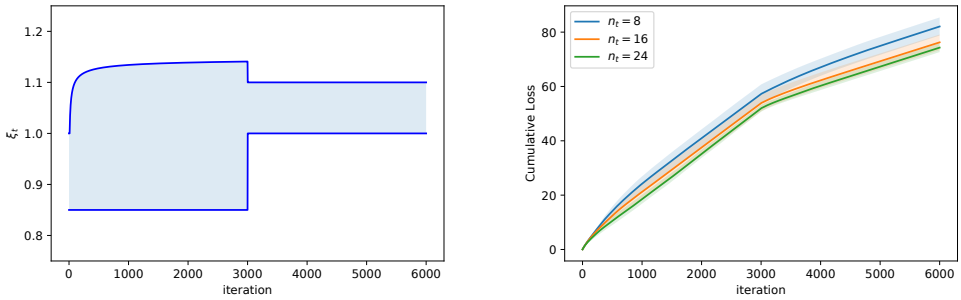
$$C_t(x_t) = \underset{\xi_t \sim \mathcal{D}_t}{\mathrm{CVaR}_\alpha}[J(x_t, \xi_t)], \tag{10}$$

where the risk level is selected as $\alpha = 0.5$. We assume that the random variable ξ_t has a continuous uniform distribution and lies in the time-varying distribution range $[L_t, R_t]$, which is selected as

$$[L_t, R_t] = \begin{cases} [0.85, 1.15 - 0.5t^{-0.5}] & \text{if } t < T/2 \\ [0.85 + 0.5t^{-0.1}, 1.1] & \text{if } t \geq T/2. \end{cases}$$

Fig. 1a depicts the distribution range of the random variable ξ_t, where the time horizon is selected as $T = 6000$. We use Algorithm 1 to update the parking prices in the dynamic environment with changing distributions. From Theorems 1 and 2, we observe that the algorithm design for convex and strongly convex cases mainly differs in the selection of the learning rate η_t and the batch size Δ_T. For the objective function (9), when the strongly convexity property is not utilized, the step size is set as $\eta_t = 0.01$, as in Theorem 1. In contrast, when leveraging the strongly convex property, the restarting period is set to $\Delta_T = 3$, and the learning rate parameter in Theorem 2 is chosen as $m = 20$, as in Theorem 2. The smoothing parameter for the zeroth-order optimization is selected as $\delta = 0.05$. We set the initial price to $x_0 = 1$ and restrict the potential prices to $[1, 5]$.

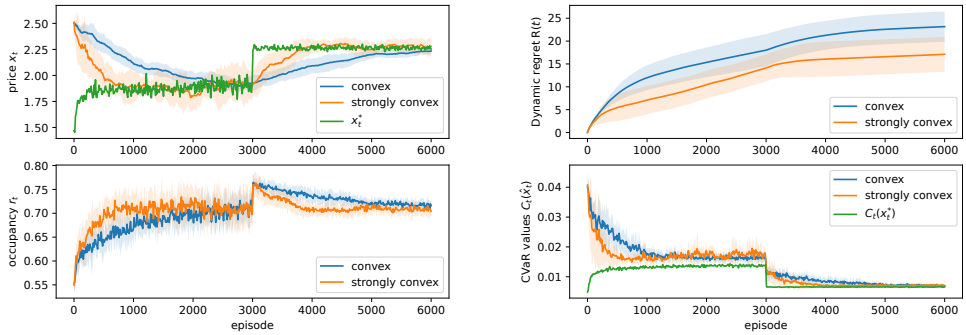
The simulation results of Algorithm 1 are presented in Figs. 1b–2b, where shaded areas represent $\pm$ one standard deviation over 10 runs. We first evaluate Algorithm 1 under the strongly convex setting for different sample sizes, using the cumulative loss $\sum_{t=1}^{T} C_t(\hat{x}_t)$ as the evaluation metric. We adopt the sampling strategies with parameters $c = 10$ and $a \in \{\frac{2}{3}, 1, \frac{4}{3}\}$. The corresponding sampling number is $n_t \in \{8, 16, 24\}$, respectively. It is shown in Fig. 1b that, more samples lead to a smaller cumulative loss. Figs. 2a and 2b compare the performance of Algorithm 1 under convex and strongly convex settings. We fix the sampling number as $n_t = 8$ and run the algorithm for 10 trials. The top subfigure of Fig. 2a shows the parking price x_t generated by the algorithm for both convex and strongly convex cases, alongside the optimal price x_t^*, which minimizes the CVaR function (10). The bottom subfigure displays the corresponding occupancy levels r_t. Since the CVaR function lacks a closed-form expression, the optimal prices are determined by sampling. Specifically, at each



(a) Distribution range of the random variable ξ_t.

(b) Impact of sample size n_t.

Figure 1: Experimental setup and sensitivity analysis of Algorithm 1.



(a) Price and Occupancy

(b) Regret and CVaR

Figure 2: Performance comparison of Algorithm 1 under convex and strongly convex settings. The optimal benchmark x_t^* is obtained via brute-force search.

iteration, 100 points are uniformly sampled from the continuous set $\mathcal{X}$, and the point that minimizes the CVaR value is selected. The results demonstrate that the prices x_t generated by Algorithm 1 closely track the optimal prices x_t^*, and the occupancy r_t converge to desired rate quickly, for both convex and strongly convex cases. Notably, the algorithm designed for the strongly convex case exhibits greater adaptability and converges to the optimal prices more quickly. Fig. 2b illustrates the CVaR values corresponding to the prices generated by Algorithm 1 for convex and strongly convex cases, denoted as $C_t(\hat{x}_t)$, alongside the CVaR values for the optimal prices, $C_t(x_t^*)$, and the dynamic regret $\mathrm{DR}(T)$. Fig. 2b shows that Algorithm 1 achieves sublinear regret in both convex and strongly convex settings. Moreover, the algorithm tailored for the strongly convex case achieves lower regret than the convex case.

5 Conclusions

In this paper, we developed a risk-averse learning algorithm that is able to handle time-varying distributions. The algorithm restarts periodically and queries the function values multiple times per iteration to estimate the CVaR gradient, where the cumulative error of the CVaR gradient is shown to be bounded with high probability. We theoretically bound the variation in the CVaR function induced by distributional shifts. Then, the dynamic regret is analyzed in terms of the distribution variations and iteration horizons for both convex and strongly convex cases. The regret bound of the designed algorithm decreases as the sampling number increases, and this reduction ceases after the total sampling number reaches a certain threshold. Moreover, using the strongly convexity property yields a smaller regret bound than in the convex case.

ACKNOWLEDGMENTS

This work was supported in part by the European Research Council (ERC) Consolidator Grant "Safe data-driven control for humancentric systems (CO-MAN)" under grant agreement number 864686, by the Swedish Research Council Distinguished Professor Grant 2017-01078, Knut and Alice Wallenberg Foundation, Wallenberg Scholar Grant, the Swedish Strategic Research Foundation CLAS Grant RIT17-0046, AFOSR under award #FA9550-19-1-0169, and NSF under award CNS-1932011, and by the National Natural Science Foundation of China under Grant 62503365.

REFERENCES

Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015.

Xuanyu Cao, Junshan Zhang, and H Vincent Poor. Online stochastic optimization with time-varying distributions. *IEEE Transactions on Automatic Control*, 66(4):1840–1847, 2020.

Adrian Rivera Cardoso and Huan Xu. Risk-averse stochastic convex bandit. In *International Conference on Artificial Intelligence and Statistics*, pp. 39–47, 2019.

Margaret P Chapman and Laurent Lessard. Toward a scalable upper bound for a CVaR-lq problem. *IEEE Control Systems Letters*, 6:920–925, 2021.

Margaret P Chapman, Jonathan Lacotte, Aviv Tamar, Donggun Lee, Kevin M Smith, Victoria Cheng, Jaime F Fisac, Susmit Jha, Marco Pavone, and Claire J Tomlin. A risk-sensitive finite-time reachability approach for safety of stochastic dynamic systems. In *American Control Conference*, pp. 2958–2963, 2019.

Tianyi Chen, Qing Ling, and Georgios B Giannakis. An online convex optimization approach to proactive network resource allocation. *IEEE Transactions on Signal Processing*, 65(24):6350–6364, 2017.

Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics*, pp. 642–669, 1956.

David A Edwards. On the Kantorovich–Rubinstein theorem. *Expositiones Mathematicae*, 29(4): 387–398, 2011.

Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. *arXiv preprint cs/0408007*, 2004.

Geoffrey J Gordon, Amy Greenwald, and Casey Marks. No-regret learning in convex games. In *International Conference on Machine Learning*, pp. 360–367, 2008.

Elad Hazan. *Efficient algorithms for online convex optimization and their applications*. Princeton University, 2006.

Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.

Jiashuo Jiang, Xiaocheng Li, and Jiawei Zhang. Online stochastic optimization with wasserstein-based nonstationarity. *Management Science*, 2025.

Leonid V Kantorovich. Mathematical methods of organizing and planning production. *Management Science*, 6(4):366–422, 1960.

Masako Kishida and Ahmet Cetinkaya. Risk-aware linear quadratic control using conditional value-at-risk. *IEEE Transactions on Automatic Control*, 68(1):416–423, 2022.

Qianqiao Liang, Mengying Zhu, Xiaolin Zheng, and Yan Wang. An adaptive news-driven method for CVaR-sensitive online portfolio selection in non-stationary financial markets. In *International Joint Conference on Artificial Intelligence*, pp. 2708–2715, 2021.

Thomas J Linsmeier and Neil D Pearson. Value at risk. *Financial Analysts Journal*, 56(2):47–67, 2000.

Adhyyan Narang, Evan Faulkner, Dmitriy Drusvyatskiy, Maryam Fazel, and Lillian Ratliff. Learning in stochastic monotone games with decision-dependent data. In *International Conference on Artificial Intelligence and Statistics*, pp. 5891–5912, 2022.

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

R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.

R Tyrrell Rockafellar, Stanislav Uryasev, et al. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

Pier Giuseppe Sessa, Ilija Bogunovic, Maryam Kamgarpour, and Andreas Krause. No-regret learning in unknown games with correlated payoffs. *Advances in Neural Information Processing Systems*, 32(24):13624–13633, 2019.

Iman Shames and Farhad Farokhi. Online stochastic convex optimization: Wasserstein distance variation. *arXiv preprint arXiv:2006.01397*, 2020.

Yi Shen, Pan Xu, and Michael M Zavlanos. Wasserstein distributionally robust policy evaluation and learning for contextual bandits. *arXiv preprint arXiv:2309.08748*, 2023.

Tasuku Soma and Yuichi Yoshida. Statistical learning with conditional value at risk. *arXiv preprint arXiv:2002.05826*, 2020.

Alex Tamkin, Ramtin Keramati, Christoph Dann, and Emma Brunskill. Distributionally-aware exploration for CVaR bandits. In *NeurIPS 2019 Workshop on Safety and Robustness on Decision Making*, 2019.

Núria Armengol Urpí, Sebastian Curi, and Andreas Krause. Risk-averse offline reinforcement learning. *arXiv preprint arXiv:2102.05371*, 2021.

Zifan Wang, Yi Shen, Zachary I Bell, Scott Nivison, Michael M Zavlanos, and Karl H Johansson. A zeroth-order momentum method for risk-averse online convex games. In *IEEE Conference on Decision and Control*, pp. 5179–5184. IEEE, 2022a.

Zifan Wang, Yi Shen, and Michael Zavlanos. Risk-averse no-regret learning in online convex games. In *International Conference on Machine Learning*, pp. 22999–23017, 2022b.

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

Peng Zhao, Guanghui Wang, Lijun Zhang, and Zhi-Hua Zhou. Bandit convex optimization in non-stationary environments. *Journal of Machine Learning Research*, 22(1):5562–5606, 2021.

A APPENDIX

A.1 PROOF OF LEMMA 6

Define the augmented functions

$$L_{\mathcal{D}_X}(v) = v + \frac{1}{\alpha}\mathbb{E}_{X\sim\mathcal{D}_X}[f(X) - v]_+, \quad L_{\mathcal{D}_Y}(v) = v + \frac{1}{\alpha}\mathbb{E}_{Y\sim\mathcal{D}_Y}[f(Y) - v]_+,$$

where $\mathcal{D}_X$ and $\mathcal{D}_Y$ are the distributions of the random variables X and Y, respectively. According to (Rockafellar et al., 2000), we have

$$\underset{X\sim\mathcal{D}_X}{\text{CVaR}_\alpha}[f(X)] = \min_v L_{\mathcal{D}_X}(v), \quad \underset{Y\sim\mathcal{D}_Y}{\text{CVaR}_\alpha}[f(Y)] = \min_v L_{\mathcal{D}_Y}(v).$$

We assume that $v_x = \arg\min_v L_{\mathcal{D}_X}(v)$ and $v_y = \arg\min_v L_{\mathcal{D}_Y}(v)$. Then, we have

$$L_{\mathcal{D}_X}(v_x) = \underset{X\sim\mathcal{D}_X}{\text{CVaR}_\alpha}[f(X)], \quad L_{\mathcal{D}_Y}(v_y) = \underset{Y\sim\mathcal{D}_Y}{\text{CVaR}_\alpha}[f(Y)].$$

Define $g(x) = [f(x) - v_y]_+$, we have

$$\begin{aligned}
&\underset{X\sim\mathcal{D}_X}{\text{CVaR}_\alpha}[f(X)] - \underset{Y\sim\mathcal{D}_Y}{\text{CVaR}_\alpha}[f(Y)] \\
=&L_{\mathcal{D}_X}(v_x) - L_{\mathcal{D}_Y}(v_y) \leq L_{\mathcal{D}_X}(v_y) - L_{\mathcal{D}_Y}(v_y) \\
=&v_y + \frac{1}{\alpha}\mathbb{E}_{X\sim\mathcal{D}_X}[f(X) - v_y]_+ - v_y - \frac{1}{\alpha}\mathbb{E}_{Y\sim\mathcal{D}_Y}[f(Y) - v_y]_+ \\
=&\frac{1}{\alpha}\mathbb{E}_{X\sim\mathcal{D}_X}[g(X)] - \frac{1}{\alpha}\mathbb{E}_{Y\sim\mathcal{D}_Y}[g(Y)],
\end{aligned}$$

where the first inequality is from $v_x = \arg\min_v L_{\mathcal{D}_X}(v)$. According to the definition of $g(x)$, we have

$$\begin{aligned}
|g(x) - g(y)| &= [f(x) - v_y]_+ - [f(y) - v_y]_+ \\
&\leq [f(x) - f(y)]_+ \leq |f(x) - f(y)| \leq L_0 \|x - y\|, \qquad (11)
\end{aligned}$$

where the first inequality follows from the fact that $a_+ - b_+ \leq [a - b]_+$, for $\forall a, b \in \mathbb{R}$. From (11), we conclude that the function $g(x)$ is L_0-Lipschitz continuous. Hence,

$$\begin{aligned}
&\underset{X\sim\mathcal{D}_X}{\text{CVaR}_\alpha}[f(X)] - \underset{Y\sim\mathcal{D}_Y}{\text{CVaR}_\alpha}[f(Y)] \\
\leq&\frac{1}{\alpha}\mathbb{E}_{X\sim\mathcal{D}_X}[g(X)] - \frac{1}{\alpha}\mathbb{E}_{Y\sim\mathcal{D}_Y}[g(Y)] \leq \frac{L_0}{\alpha}W_1(\mathcal{D}_X, \mathcal{D}_Y),
\end{aligned}$$

where the last inequality follows from the Kantorovich-Rubinstein Duality of the Wasserstein distance, see (Edwards, 2011).

Following similar arguments, we can obtain the other side of the inequality. Here completes the proof. □

The following lemma, together with Lemma 5, analyzes the cumulative variations in the CVaR function due to changes in the underlying distribution.

Lemma 8. *Consider the function $C_t^\delta(x) :\to \mathbb{R}$, define the function sequences over iterations horizon T as $\{C_t^\delta(x_t)\}_{t=1}^T$, we have that*

$$\sum_{j=1}^{s}\sum_{t\in\mathcal{T}_j}\left(C_t^\delta(\tilde{x}_j^{\delta,*}) - C_t^\delta(x_t^{\delta,*})\right) \leq \frac{2L_0\Delta_T}{\alpha}\sum_{t=2}^{T}W_1(\mathcal{D}_t, \mathcal{D}_{t-1}). \qquad (12)$$

Proof of Lemma 8: This proof is adopted from Proposition 2 of Besbes et al. (2015). First we denote

$$V_j = \sum_{t\in\mathcal{T}_j}\sup_{x\in\mathcal{X}^\delta}|C_t^\delta(x) - C_{t-1}^\delta(x)|$$

as the function variation over batch $\mathcal{T}_j$, it is straightforward to write $\sum_{j=1}^{s} V_j = \sum_{t=2}^{T} \sup_{x\in\mathcal{X}^\delta} |C_t^\delta(x) - C_{t-1}^\delta(x)|$. Let $\bar{\tau}_j$ be the first epoch of batch $\mathcal{T}_j$, for $j = 1, \ldots, s$, we have

$$\begin{aligned}\sum_{t\in\mathcal{T}_j} C_t^\delta(\tilde{x}_j^{\delta,*}) - C_t^\delta(x_t^{\delta,*}) &\le \sum_{t\in\mathcal{T}_j} C_t^\delta(x_{\bar{\tau}_j}^{\delta,*}) - C_t^\delta(x_t^{\delta,*}) \\ &\le \Delta_T \cdot \max_{t\in\mathcal{T}_j}\{C_t^\delta(x_{\bar{\tau}_j}^{\delta,*}) - C_t^\delta(x_t^{\delta,*})\},\end{aligned}$$

where the first inequality is from the definition of $\tilde{x}_j^{\delta,*}$. In the following we will prove $\max_{t\in\mathcal{T}_j}\{C_t^\delta(x_{\bar{\tau}_j}^{\delta,*}) - C_t^\delta(x_t^{\delta,*})\} \le 2V_j$ by contraction. Suppose otherwise, there exists an iteration $\tilde{t} \in \mathcal{T}_j$ such that $C_{\tilde{t}}^\delta(x_{\bar{\tau}_j}^{\delta,*}) - C_{\tilde{t}}^\delta(x_{\tilde{t}}^{\delta,*}) > 2V_j$. It implies that

$$C_t^\delta(x_{\tilde{t}}^{\delta,*}) \le C_{\tilde{t}}^\delta(x_{\tilde{t}}^{\delta,*}) + V_j < C_{\tilde{t}}^\delta(x_{\bar{\tau}_j}^{\delta,*}) - V_j \le C_t^\delta(x_{\bar{\tau}_j}^{\delta,*}),$$

where the first and the last inequality results from the fact that V_j is the maximal variation over batch $\mathcal{T}_j$. Hence, we have $\sum_{t\in\mathcal{T}_j} C_t^\delta(\tilde{x}_j^{\delta,*}) - \sum_{t\in\mathcal{T}_j} C_t^\delta(x_t^{\delta,*}) \le 2\Delta_T V_j$. Summarize the variation along batches $\{\mathcal{T}_1, \ldots, \mathcal{T}_s\}$, it results

$$\sum_{j=1}^{s}\left(\sum_{t\in\mathcal{T}_j} C_t^\delta(\tilde{x}_j^{\delta,*}) - \sum_{t\in\mathcal{T}_j} C_t^\delta(x_t^{\delta,*})\right) \le \sum_{j=1}^{s} 2\Delta_T V_j \le \frac{2L_0}{\alpha}\Delta_T \sum_{t=2}^{T} W_1(\mathcal{D}_t, \mathcal{D}_{t-1}). \tag{13}$$

where the last inequality follows from Lemma 6. □

A.2 Proof of Theorem 1

For $t = 1, \ldots, T$, we have

$$\begin{aligned}\min_{x_t\in\mathcal{X}^\delta} C_t^\delta(x_t) &= \min_{x_t\in\mathcal{X}} C_t^\delta((1-\delta/r)x_t) \\ &\le \min_{x_t\in\mathcal{X}} (\delta/r)C_t^\delta(0) + (1-\delta/r)C_t^\delta(x_t) \\ &\le \min_{x_t\in\mathcal{X}} C_t^\delta(x_t) + (\delta/r)L_0\|x_t\| \\ &\le \min_{x_t\in\mathcal{X}} C_t^\delta(x_t) + D_x L_0\delta/r. \end{aligned} \tag{14}$$

The first inequality is from the convexity of $C_t^\delta(x)$ as shown in Lemma 3, and the second inequality is from Lipschitzness of $C_t^\delta(x)$ as shown in Lemma 4. To simplify notations, we denote $x_t^{\delta,*} = \arg\min_{x\in\mathcal{X}^\delta} C_t^\delta(x)$. The dynamic regret defined in (1) is further written as

$$\begin{aligned}\mathrm{DR}(T) &\le \sum_{t=1}^{T} C_t^\delta(\hat{x}_t) - \sum_{t=1}^{T} C_t^\delta(x_t^*) + 2\delta L_0 T \\ &\le \sum_{t=1}^{T} C_t^\delta(x_t) - \sum_{t=1}^{T} C_t^\delta(x_t^*) + 3\delta L_0 T \\ &\le \sum_{t=1}^{T} C_t^\delta(x_t) - \sum_{t=1}^{T} C_t^\delta(x_t^{\delta,*}) + (3 + D_x/r)\delta L_0 T, \end{aligned} \tag{15}$$

where the first inequality follows from the definition of the function C_t^δ, defined in (2), the second inequality follows from the Lipschitzness of the function C_t^δ, and the third inequality establishes by substituting (14) into the $C_t^\delta(x_t^*)$. Denote $\tilde{x}_j^{\delta,*} = \arg\min_{x\in\mathcal{X}^\delta} \sum_{t\in\mathcal{T}_j} C_t^\delta(x)$ as the single best action over interval j, for $j = 1, \ldots, s$. It follows that

$$\begin{aligned}&\sum_{t=1}^{T} C_t^\delta(x_t) - \sum_{t=1}^{T} C_t^\delta(x_t^{\delta,*}) \\ &= \sum_{j=1}^{s}\sum_{t\in\mathcal{T}_j}\left(C_t^\delta(x_t) - C_t^\delta(\tilde{x}_j^{\delta,*})\right) + \sum_{j=1}^{s}\sum_{t\in\mathcal{T}_j}\left(C_t^\delta(\tilde{x}_j^{\delta,*}) - C_t^\delta(x_t^{\delta,*})\right) = \sum_{j=1}^{s}\mathcal{R}_1^j + \mathcal{R}_2, \end{aligned} \tag{16}$$

with $\mathcal{R}_1^j = \sum_{t\in\mathcal{T}_j}\Big(C_t^\delta(x_t) - C_t^\delta(\tilde{x}_j^{\delta,*})\Big)$, for $j = 1,\ldots,s$, and $\mathcal{R}_2 = \sum_{j=1}^s \sum_{t\in\mathcal{T}_j}\Big(C_t^\delta(\tilde{x}_j^{\delta,*}) - C_t^\delta(x_t^{\delta,*})\Big)$. We first bound the term $\mathcal{R}_1^j$. By the convexity of the function C_t^δ, we have

$$\mathcal{R}_1^j \le \sum_{t\in\mathcal{T}_j}\langle \nabla C_t^\delta(x_t), x_t - \tilde{x}_j^{\delta,*}\rangle.$$

In combination of (6) and (8), we obtain that

$$\nabla C_t^\delta(x_t) = \mathbb{E}[\hat{g}_t - \frac{d}{\delta}\hat{\epsilon}_t u_t]. \tag{17}$$

By the update rule (7), we have

$$\begin{aligned}\|x_{t+1} - \tilde{x}_j^{\delta,*}\|^2 &= \|\mathcal{P}_{\mathcal{X}^\delta}(x_t - \eta_t\hat{g}_t) - \tilde{x}_j^{\delta,*}\|^2\\ &\le \|x_t - \eta_t\hat{g}_t - \tilde{x}_j^{\delta,*}\|^2\\ &= \|x_t - \tilde{x}_j^{\delta,*}\|^2 + \eta_t^2\|\hat{g}_t\|^2 - 2\eta_t\langle\hat{g}_t, x_t - \tilde{x}_j^{\delta,*}\rangle,\end{aligned}$$

where the inequality follows from the fact that $\tilde{x}_j^{\delta,*} \in \mathcal{X}^\delta$. Then, we obtain

$$\langle\hat{g}_t, x_t - \tilde{x}_j^{\delta,*}\rangle \le \frac{\|x_t - \tilde{x}_j^{\delta,*}\|^2 - \|x_{t+1} - \tilde{x}_j^{\delta,*}\|^2}{2\eta_t} + \frac{\eta_t}{2}\|\hat{g}_t\|^2. \tag{18}$$

For notational simplicity, denote the first slot of $\mathcal{T}_j$ as $\bar{\tau}_j$. Substituting (17) and (18) into $\mathcal{R}_1^j$, we obtain

$$\begin{aligned}\mathcal{R}_1^j &\le \sum_{t\in\mathcal{T}_j}\mathbb{E}[\langle\hat{g}_t - \frac{d}{\delta}\hat{\epsilon}_t u_t, x_t - \tilde{x}_j^{\delta,*}\rangle]\\ &\le \sum_{t\in\mathcal{T}_j}\left(\frac{1}{2\eta_t}\mathbb{E}[\|x_t - \tilde{x}_j^{\delta,*}\|^2 - \|x_{t+1} - \tilde{x}_j^{\delta,*}\|^2] + \frac{\eta_t}{2}\mathbb{E}[\|\hat{g}_t\|^2] + \frac{d}{\delta}\mathbb{E}[\|\hat{\epsilon}_t\|\|x_t - \tilde{x}_j^{\delta,*}\|]\right)\\ &= \frac{1}{2\eta_{\bar{\tau}_j}}\left(\|x_{(j-1)\Delta_T+1} - \tilde{x}_j^{\delta,*}\|^2 - \|x_{j\Delta_T} - \tilde{x}_j^{\delta,*}\|^2\right) + \mathcal{R}_{12}^j + \mathcal{R}_{13}^j\\ &\le \frac{D_x^2}{2\eta_{\bar{\tau}_j}} + \mathcal{R}_{12}^j + \mathcal{R}_{13}^j,\end{aligned} \tag{19}$$

with $\mathcal{R}_{12}^j = \sum_{t\in\mathcal{T}_j}\frac{\eta_t}{2}\mathbb{E}[\|\hat{g}_t\|^2]$ and $\mathcal{R}_{13}^j = \sum_{t\in\mathcal{T}_j}\frac{d}{\delta}\mathbb{E}[\|\hat{\epsilon}_t\|\|x_t - \tilde{x}_j^{\delta,*}\|]$. The last inequality of (19) is from the definition $D_x = \sup_{x,y\in\mathcal{X}}\|x - y\|$. Regarding $\mathcal{R}_{12}^j$, for $i = 1,\ldots,s$, it writes,

$$\mathcal{R}_{12}^j = \sum_{t\in\mathcal{T}_j}\frac{\eta_t}{2}\left\|\frac{d}{\delta}\mathrm{CVaR}_\alpha[\hat{F}_t]u_t\right\|^2 \le \sum_{t\in\mathcal{T}_j}\frac{\eta_t}{2}\left(\frac{dU}{\delta}\right)^2. \tag{20}$$

The first inequality establishes as $\mathrm{CVaR}_\alpha[\hat{F}_t] \le U$. Regarding $\mathcal{R}_{13}^j$, we first analyze the bound of $\hat{\epsilon}_t$ given in (8). By leveraging the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality (Dvoretzky et al., 1956), we have that

$$\mathbb{P}\left\{\sup_y|\hat{F}_t(y) - F_t(y)| \ge \sqrt{\frac{\ln(2/\bar{\gamma})}{2n_t}}\right\} \le \bar{\gamma}. \tag{21}$$

Denote the event in (21) as A_t, and $\mathbb{P}\{A_t\}$ denotes the occurrence probability of event A_t, for $t = 1,\ldots,T$. According to Lemma 5, the error of CVaR estimate is bounded by

$$|\hat{\varepsilon}_t| = \frac{U}{\alpha}\sup\left|\hat{F}_t - F_t\right| \le \frac{U}{\alpha}\sqrt{\frac{\ln(2/\bar{\gamma})}{2n_t}} \tag{22}$$

with probability at least $1-\bar{\gamma}$, for $t=1,\ldots,T$. Let $\gamma=\bar{\gamma}T$. Then, by substituting (22) into $\mathcal{R}_{13}^j$, we obtain that

$$\begin{aligned}\sum_{j=1}^{s}\mathcal{R}_{13}^j &\le \frac{dD_x}{\delta}\sum_{j=1}^{s}\sum_{t\in\mathcal{T}_j}\mathbb{E}[\|\hat{\varepsilon}_t\|] \\ &\le \frac{dUD_x}{\alpha\delta}\sum_{j=1}^{s}\sum_{t\in\mathcal{T}_j}\sqrt{\frac{\ln(2T/\gamma)}{2n_t}} \le \frac{cdUD_x}{\alpha\delta}\left\lceil\frac{T}{\Delta_T}\right\rceil\sqrt{\frac{\ln(2T/\gamma)}{2}}\Delta_T^{1-\frac{a}{2}} \end{aligned} \tag{23}$$

with probability at least $1-\gamma$, which establishes as $1-\mathbb{P}\{\bigcup_{t=1}^{T}A_t\}\ge 1-\sum_{t=1}^{T}\mathbb{P}\{A_t\}\ge 1-T\frac{\gamma}{T}\ge 1-\gamma$. As shown in (22), as the number of samples increases, the empirical density function approaches the true one. In other words, a sufficiently large sample size ensures that the event $\{\bigcup_{t=1}^{T}A_t\}$ occurs with high probability. Regarding $\mathcal{R}_2$, we have that

$$\mathcal{R}_2 \le \frac{2L_0}{\alpha}\Delta_T\sum_{t=2}^{T}W_1(\mathcal{D}_t,\mathcal{D}_{t-1}) \le \frac{2L_0}{\alpha}\Delta_T V_D, \tag{24}$$

where the first inequality follows from Lemma 8 and the second inequality follows from Assumption 3. From the definition of $\mathcal{R}_2$ as in (16), when interval size Δ_T approaches 1, the interval-optimal actions $\tilde{x}_j^{\delta,*}$ will be closer to one-step optimal action $x_t^{\delta,*}$, for $t\in\mathcal{T}_j$, and the accumulated loss (24) will be smaller. However, restarting also induces errors such as $\frac{D_x^2}{\eta_{\mathcal{T}_j}}$. Thus, it is necessary to select an optimal interval size Δ_T to minimize the regret bound. Let $\eta_t=\eta$ for all t. Substituting (19), (20), (23) and (24) into (16), and combining it with (15), it results

$$\begin{aligned}\mathrm{DR}(T) &\le (3+D_x/r)\delta L_0 T + \mathcal{R}_2 + \sum_{j=1}^{s}\left(\frac{D_x^2}{2\eta}+\mathcal{R}_{12}^j+\mathcal{R}_{13}^j\right) \\ &\le (3+D_x/r)\delta L_0 T + \frac{2L_0}{\alpha}\Delta_T V_D + \frac{D_x^2}{2\eta}\left\lceil\frac{T}{\Delta_T}\right\rceil \\ &+ \left(\frac{d^2U^2\eta\Delta_T}{2\delta^2}+\frac{cdUD_x}{\alpha\delta}\sqrt{\frac{\ln(2T/\gamma)}{2}}\Delta_T^{1-\frac{a}{2}}\right)\left\lceil\frac{T}{\Delta_T}\right\rceil \\ &\le (3+D_x/r)\delta L_0 T + \frac{2L_0}{\alpha}\Delta_T V_D + \frac{D_x^2 T}{\eta\Delta_T} + \frac{d^2U^2\eta T}{\delta^2} + \frac{\sqrt{2}cdUD_x\sqrt{\ln(2T/\gamma)}}{\alpha\delta}T\Delta_T^{-\frac{a}{2}}. \end{aligned} \tag{25}$$

with probability at least $1-\gamma$. The last inequality results from $\left\lceil\frac{T}{\Delta_T}\right\rceil\le\frac{T}{\Delta_T}+1\le\frac{2T}{\Delta_T}$. When $a\in(0,1]$, we select $\delta=\left(\frac{V_D}{T}\right)^{\frac{a}{4+a}}$, $\eta=\left(\frac{V_D}{T}\right)^{\frac{3a}{4+a}}$ and $\Delta_T=\left(\frac{T}{V_D}\right)^{\frac{4}{4+a}}$ to minimize the regret bound, it achieves $\mathrm{DR}(T)=\tilde{\mathcal{O}}(T^{\frac{4}{4+a}}V_D^{\frac{a}{4+a}})$ with probability at least $1-\gamma$ (see (Flaxman et al., 2004) for parameters selection details). When $a>1$, we select $\delta=\left(\frac{V_D}{T}\right)^{\frac{1}{5}}$, $\eta_t=\left(\frac{V_D}{T}\right)^{\frac{3}{5}}$ and $\Delta_T=\left(\frac{T}{V_D}\right)^{\frac{4}{5}}$, it achieves $\mathrm{DR}(T)=\tilde{\mathcal{O}}(T^{\frac{4}{5}}V_D^{\frac{1}{5}})$ with probability at least $1-\gamma$. □

A.3 Proof of Theorem 2

Following the derivation of (14)–(16) in the proof of Theorem 1, the dynamic regret under Algorithm 1 is written as

$$\begin{aligned}\mathrm{DR}(T) &\le (3+D_x/r)\delta L_0 T + \sum_{j=1}^{s}\sum_{t\in\mathcal{T}_j}\left(C_t^{\delta}(x_t)-C_t^{\delta}(\tilde{x}_j^{\delta,*})\right)+\sum_{j=1}^{s}\sum_{t\in\mathcal{T}_j}\left(C_t^{\delta}(\tilde{x}_j^{\delta,*})-C_t^{\delta}(x_t^{\delta,*})\right) \\ &\le (3+D_x/r)\delta L_0 T + \sum_{j=1}^{s}\tilde{\mathcal{R}}_1^j + \mathcal{R}_2 \end{aligned} \tag{26}$$

with $\tilde{\mathcal{R}}_1^j = \sum_{t\in\mathcal{T}_j} C_t^\delta(x_t) - C_t^\delta(\tilde{x}_j^{\delta,*})$ and the definition of $\mathcal{R}_2$ is as in (16). By the strongly convexity of the function C_t^δ, we have

$$\begin{aligned}
\tilde{\mathcal{R}}_1 &\leq \sum_{j=1}^{s}\Big(\sum_{t\in\mathcal{T}_j}\langle\nabla C_t^\delta(x_t), x_t - \tilde{x}_j^{\delta,*}\rangle - \frac{m}{2}\|x_t - \tilde{x}_j^{\delta,*}\|^2\Big)\\
&\leq \sum_{j=1}^{s}\sum_{t\in\mathcal{T}_j}\Big(\frac{1}{2\eta_t}\big(\|x_t - \tilde{x}_j^{\delta,*}\|^2 - \|x_{t+1} - \tilde{x}_j^{\delta,*}\|^2\big) - \frac{m}{2}\|x_t - \tilde{x}_j^{\delta,*}\|^2 + \frac{\eta_t}{2}\mathbb{E}[\|\hat{g}_t\|^2]\\
&\quad + \frac{d}{\delta}\mathbb{E}[\|\hat{\epsilon}_t\|\|x_t - \tilde{x}_j^{\delta,*}\|]\Big)\\
&\leq \sum_{j=1}^{s}\Big(\tilde{\mathcal{R}}_{11}^j + \tilde{\mathcal{R}}_{12}^j + \mathcal{R}_{13}^j\Big) \qquad (27)
\end{aligned}$$

where $\tilde{\mathcal{R}}_{11}^j = \sum_{t\in\mathcal{T}_j}\Big(\frac{1}{2\eta_t}\big(\|x_t - \tilde{x}_j^{\delta,*}\|^2 - \|x_{t+1} - \tilde{x}_j^{\delta,*}\|^2\big) - \frac{m}{2}\|x_t - \tilde{x}_j^{\delta,*}\|^2\Big)$, $\tilde{\mathcal{R}}_{12}^j = \sum_{t\in\mathcal{T}_j}\frac{\eta_t}{2}\mathbb{E}[\|\hat{g}_t\|^2]$, and the definition of $\mathcal{R}_{13}^j$ is the same as (23). The second inequality follows the derivation of (17)–(19) in Theorem 1. By expanding the sequences of $\tilde{\mathcal{R}}_{11}^j$, we obtain

$$\begin{aligned}
\sum_{j=1}^{s}\tilde{\mathcal{R}}_{11}^j &= \sum_{j=1}^{s}\sum_{t\in\{\mathcal{T}_j\setminus\bar{\tau}_j\}}\big(\frac{1}{2\eta_t} - \frac{1}{2\eta_{t-1}} - \frac{m}{2}\big)\|x_t - \tilde{x}_j^{\delta,*}\|^2\\
&\quad + \frac{1}{2\eta_{\bar{\tau}_j}}\|x_{\bar{\tau}_j} - \tilde{x}_j^{\delta,*}\|^2 - \frac{1}{2\eta_{j\Delta_T}}\|x_{j\Delta_T} - \tilde{x}_j^{\delta,*}\|^2 - \frac{m}{2}\|x_{\bar{\tau}_j} - \tilde{x}_j^{\delta,*}\|^2\\
&\leq \sum_{j=1}^{s}\frac{1}{2\eta_{\bar{\tau}_j}}\|x_{\bar{\tau}_j} - \tilde{x}_j^{\delta,*}\|^2 \leq mD_x^2\left\lceil\frac{T}{\Delta_T}\right\rceil. \qquad (28)
\end{aligned}$$

The first inequality establishes by substituting the learning rate η_t into $\tilde{\mathcal{R}}_{11}^j$, and $\frac{m}{2}, \frac{1}{2\eta_{j\Delta_T}} > 0$. Regarding $\sum_{j=1}^{s}\tilde{\mathcal{R}}_{12}^j$, it writes

$$\sum_{j=1}^{s}\tilde{\mathcal{R}}_{12}^j \leq \frac{d^2U^2}{2\delta^2}\sum_{j=1}^{s}\sum_{t\in\mathcal{T}_j}\eta_t \leq \frac{d^2U^2}{2\delta^2 m}\sum_{j=1}^{s}\sum_{t=1}^{\Delta_T}\frac{1}{t} \leq \frac{d^2U^2}{2\delta^2 m}(1+\ln\Delta_T)\left\lceil\frac{T}{\Delta_T}\right\rceil. \qquad (29)$$

The last inequality establishes as $\sum_{t=1}^{\Delta_T}\frac{1}{t} \leq 1+\ln\Delta_T$. Note that the upperbound of $\mathcal{R}_2$ only relates to the distribution variation budget, which applies here directly. Hence, we substitute (23), (28) and (29) into (27), combine them with (26), and obtain

$$\begin{aligned}
\mathrm{DR}(T) &= (3 + D_x/r)\delta L_0 T + \mathcal{R}_2 + \sum_{j=1}^{s}\tilde{\mathcal{R}}_{11}^j + \tilde{\mathcal{R}}_{12}^j + \mathcal{R}_{13}^j\\
&\leq (3 + D_x/r)\delta L_0 T + \frac{2L_0}{\alpha}\Delta_T V_D + mD_x^2\left\lceil\frac{T}{\Delta_T}\right\rceil + \frac{d^2U^2}{2\delta^2 m}(1+\ln\Delta_T)\left\lceil\frac{T}{\Delta_T}\right\rceil\\
&\quad + \frac{cdUD_x}{\alpha\delta}\sqrt{\frac{\ln(2T/\gamma)}{2}}\Delta_T^{1-\frac{a}{2}}\left\lceil\frac{T}{\Delta_T}\right\rceil\\
&\leq (3 + D_x/r)\delta L_0 T + \frac{2L_0}{\alpha}\Delta_T V_D + 2mD_x^2\frac{T}{\Delta_T} + \frac{d^2U^2}{\delta^2 m}(1+\ln T)\frac{T}{\Delta_T}\\
&\quad + \frac{\sqrt{2}cdUD_x\sqrt{\ln(2T/\gamma)}}{\alpha\delta}T\Delta_T^{-\frac{a}{2}}, \qquad (30)
\end{aligned}$$

with probability at least $1-\gamma$ The remaining claim is as in Theorem 2. □

Remark 2. *When $a \in (0, 1]$, the regret bound achieved by Algorithm 1 decreases with the increasing tuning parameter a until $a > 1$. The cumulative error of* CVaR *gradient estimates contained in the regret bound (25), i.e., $\frac{d^2U^2\eta_t T}{\delta^2} + \frac{\sqrt{2}cdUD_x\sqrt{\ln(2T/\gamma)}}{\alpha\delta}T\Delta_T^{-\frac{a}{2}}$, results from using finite samples to*

estimate the CVaR *gradients via the zeroth-order algorithm. This is because a larger tuning parameter* a *means more queries of cost values and a more accurate estimate of* CVaR *gradients. Thus, the order of the cumulative estimation error decreases with the increasing* a*. When choosing the parameters to minimize regret bound, this cumulative estimation error term is one of the dominant terms in (25) for* $a \in (0, 1]$*. When* $a > 1$*, this term becomes negligible compared to the remaining terms in (25).*