

# Risk-Averse Learning with Non-Stationary Distributions<sup>★</sup>

Siyi Wang<sup>a</sup>, Zifan Wang<sup>b</sup>, Xinlei Yi<sup>c</sup>, Michael M. Zavlanos<sup>d</sup>, Karl H. Johansson<sup>b</sup>,  
Sandra Hirche<sup>a</sup>

<sup>a</sup>*Chair of Information-oriented Control (ITR), Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany*

<sup>b</sup>*Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 10044 Stockholm, Sweden*

<sup>c</sup>*Lab for Information & Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

<sup>d</sup>*Mechanical Engineering and Material Science, Duke University, Durham, NC 27708, USA*

---

## 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 the random cost changes over time. We minimize risk-averse objective function using the Conditional Value at Risk (CVaR) as risk measure. Due to the difficulty in obtaining the exact CVaR gradient, we employ a zeroth-order optimization approach that queries the cost function values multiple times at each iteration and estimates the CVaR gradient using the sampled values. To facilitate the regret analysis, we use a variation metric based on Wasserstein distance to capture time-varying distributions. Given that the distribution variation is sub-linear in the total number of episodes, we show that our designed learning algorithm achieves sub-linear dynamic regret with high probability for both convex and strongly convex functions. Moreover, theoretical results suggest that increasing the number of samples leads to a reduction in the dynamic regret bounds until the sampling number reaches a specific limit. Finally, we provide numerical experiments of dynamic pricing in a parking lot to illustrate the efficacy of the designed algorithm.

*Key words:* Dynamic regret, online convex optimization, risk-averse, time-varying distribution.

---

## 1 Introduction

Online convex optimization is a powerful framework that deals with decision-making problems in dynamic and un-

certain environments [1]. It has many applications, including traffic routing [2], resource allocation [3], and online marketing [4]. 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 the notion of regret [5], 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 [6], the pricing depends on real-time changes in demand and supply; conversely, price adjustments influence the

---

<sup>★</sup> This work was supported by the European Research Council (ERC) Consolidator Grant "Safe data-driven control for human-centric 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. Corresponding author: Zifan Wang.

*Email addresses:* siyi.wang@tum.de (Siyi Wang), zifanw@kth.se (Zifan Wang), xinleiyi@mit.edu (Xinlei Yi), michael.zavlanos@duke.edu (Michael M. Zavlanos), kallej@kth.se (Karl H. Johansson), hirche@tum.edu (Sandra Hirche).

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 [7]. A further example is performative prediction, in which the data distribution evolves with the decisions over time. It has gained significant attention in the machine learning community recently [8, 9].

Compared with the standard regret analysis assuming a stationary environment, dynamic regret provides a more relevant performance evaluation by considering the impact of fluctuations in non-stationary environments [10, 11]. Variation metrics are introduced to analyze the dynamic regret, for example, the variations in the cost functions [10] and the variations in the optimal actions [12], which is also known as the path length of the comparators. When making decisions under uncertainty, it is often essential to consider the entire distribution of potential outcomes rather than focusing on specific optimal points. Specifically, [13, 14] use the Wasserstein distance metric, which defines the dissimilarity between probability distributions, to quantify the changes of non-stationary distributions. In this paper, we employ the distribution variation metric proposed in [13].

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) [15] and Conditional Value at Risk (CVaR) [16]. Given a risk level  $\alpha \in (0, 1]$ , the CVaR value describes the average value of the  $\alpha$ -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 [17] 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 [18], online games [19, 20] and safe control [21–23]. 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 [18, 20].

### 1.1 Our Contributions

In this paper, we investigate online CVaR optimization, when the distribution of the stochastic cost changes over time. To the best of our knowledge, such risk-averse learning with non-stationary distributions has not been explored in the literature. As mentioned above, the exact

CVaR gradient is generally unknown. Hence, we use a zeroth-order optimization algorithm to estimate it. However, it is not possible to efficiently estimate the CVaR gradient with only a single sample of the random cost per iteration, especially when the distribution of the cost changes over time. To address this issue, we propose a sampling strategy motivated by [19], which queries function values multiple times at each iteration. Then we use the sampled function values to construct the empirical distribution function of the random cost and estimate the CVaR values. Based on these CVaR values, we use the zeroth-order optimization approach to construct the CVaR gradient estimate.

Additionally, we introduce the concept of distribution variation based on the Wasserstein distance metric to measure the variation of the non-stationary distributions. To ensure that the decision-maker is able to adapt to the changing distributions, the learning algorithm is periodically restarted. Then we analyze the dynamic regret of the designed risk-averse learning algorithm in terms of the distribution variation metric for both convex and strongly convex cost functions. Under mild condition on the learning rate and the smoothing parameter, the regret upperbound of the algorithm is minimized. We show that the algorithm achieves sub-linear dynamic regret with high probability, given that the distribution variation is sub-linear in the total number of iterations. A tuning parameter  $a > 0$  regulates the number of samples, where a higher value of  $a$  indicates a larger number of samples. Our results suggest that the regret bound achieved by the algorithm decreases with the increasing sampling number until it reaches a certain limit. Denote  $V_D$  as the distribution variation budget. Table 1 summarizes the dynamic regret bounds for the convex and strongly convex cases with various values of  $a$ . It can be observed that when  $0 < a \leq 1$ , the strongly convex problem has the same regret bound as the convex problem. When  $a > 1$ , the regret bound of the strongly convex problem is strictly lower than that of the convex problem. Finally, we illustrate our algorithm using the example of dynamic pricing in a parking lot.

Table 1  
Order of Regret

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

### 1.2 Related Works

There is a growing interest in non-stationary online convex optimization, see [6, 10, 11]. For example, [10] investigates non-stationary stochastic optimization for both

convex and strongly convex functions, and analyzes the dynamic regret using the variations in the cost functions as the measure. In [12], the authors investigate bandit convex optimization in non-stationary environments. To the best of our knowledge, only few works address environmental changes that results in non-stationary distributions, e.g., [13, 14, 24]. The authors in [24] investigate online stochastic optimization with time-varying distributions using the variations of optimal points as the measure. Moreover, [14] investigates online stochastic optimization in the strongly convex case with the variations of the optimal points, while the dynamic regret is not strictly sub-linear due to the fluctuation of gradient estimates. However, computing the distribution variations is generally more convenient than computing the variations in the optimal solutions, as the latter often necessitates multiple iterations of gradient descent updates. In [13], the dynamic regret of online constrained optimization is analyzed in terms of the Wasserstein-based non-stationarity budget (WBNB), which upper-bounds the cumulative deviation of all the distributions from their average distribution. However, even when the distribution changes only once throughout all the iterations, the WBNB can still be large and thus not applicable to our problem.

Another related research line is risk-averse learning using CVaR as the risk measure [18–20, 25–27]. For instance, [25] proposes a risk-averse learning algorithm for the multi-armed bandit problem and provides an upper confidence bound of the CVaR value. Moreover, [20] investigates the risk-averse learning algorithm in online convex games, which achieves sub-linear regret for each agent with high probability. Specifically, compared to [20], where the sampling numbers are designed to decrease with time, we only require the total samples over all iterations to be over a certain threshold. However, all the above works focus on the stationary environment, i.e., both the cost function and distribution do not change over time. With only a few exception, e.g., [28] explores the risk-averse online optimization in a non-stationary environment and in the context of the online portfolio selection problem with linear costs. This paper addresses a more generalize setting that can be applied to a broader range of applications.

### 1.3 Outlines

The remainder of this paper is structured as follows: Section 2 introduces preliminaries and problem formulation. Section 3 presents the main result on risk-averse learning under non-stationary distributions. Section 4 demonstrates the efficacy of the designed algorithm by numerical simulations. Section 5 draws conclusions.

**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  $J(x, \xi) : \mathcal{X} \times \Xi \rightarrow \mathbb{R}$ , 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  $\alpha$ -VaR is

$$J^\alpha = F^{-1}(\alpha) := \inf\{y : F(y) \geq \alpha\}.$$

The  $\alpha$ -CVaR describes the expectation of the  $\alpha$ -fraction of the worst outcomes of  $J(x, \xi)$ , and is defined as

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

### 2.2 Non-stationary distribution

Non-stationary stochastic optimization requires some measure to model temporal uncertainties of the dynamic environment. A classic metric is the sum of the variations of the cost functions over time. In practice, environmental fluctuations might be more intuitively modeled as the time-varying distribution. Hence, we use the Wasserstein distance metric to quantify distribution variations.

Wasserstein distance is a distance function that describes the dissimilarity between probability distributions on a certain metric space, see [29, 30]. Let  $(\Omega, d)$  be a probability space, where  $\Omega$  is a set and  $d$  is a metric on  $\Omega$ . Let  $\mathcal{D}_x$  and  $\mathcal{D}_y$  be two probability distribution on  $\Omega$ , the dual form of the Wasserstein distance is given as follows.

**Lemma 1** [30] *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–Rubinstein 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 temporal changes of the function 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  $\Omega$ , with  $\mathcal{D}$  being the admissible set of distribution sequences. The distribution variation along iterations  $\{1, \dots, T\}$  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  $\alpha$ -CVaR of the function  $J(x, \xi_t)$  is written as:

$$\begin{aligned} C_t(x) &:= \text{CVaR}_\alpha [J(x, \xi_t)] \\ &= \mathbb{E}_{F_t} [J(x, \xi_t) | J(x, \xi_t) \geq J^\alpha]. \end{aligned}$$

We make the following assumptions on the cost function, which are common in the online learning literature, see [5, 10].

**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** [18] 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  $T$ :

$$\sum_{t=2}^T W_1(\mathcal{D}_{t-1}, \mathcal{D}_t) \leq V_D.$$

**Assumption 4** Assume that the distribution variation budget is sub-linear in the iteration horizon  $T$ , i.e.,  $V_D = \mathcal{O}(T^\beta)$  with  $\beta \in [0, 1)$ .

Assumption 4 enables us to obtain a no-regret, i.e., the upperbound of the regret is sublinear in  $T$ , learning policy in stochastic non-stationary environments. We provide an example with the non-stationary distribution satisfying Assumptions 3 and 4 as follows.

**Example 1** [31] Brownian motion is the random motion of particles suspended in a medium. Assuming that

$N$  particles start from the origin at the initial time  $t = 0$ , the density of Brownian particles  $\rho$  at point  $x$  at time  $t$  is given as  $\rho(x, t) = \frac{N}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}}$  with  $D$  being the mass diffusivity. It can be observed that the distribution flattens with the increasing  $t$  and ultimately becomes uniform when time goes to infinity. Then the variation of the distribution of random variables  $x$  is sub-linear in iteration horizon  $T$ .

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:

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

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  $t = 1, \dots, T$ . 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.,  $\lim_{T \rightarrow \infty} \frac{\text{DR}(T)}{T} = 0$ .

## 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.

Before presenting the algorithm, we provide some results for the zeroth-order optimization that lay a foundation for the estimation of CVaR gradient. 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  $\delta$  is the perturbation radius, also known as the smoothing parameter. Then, the smoothed version of the CVaR function is given as

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

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

**Lemma 3** [18] Given Assumption 1, we have that  $C_t^\delta(x)$  is convex in  $x$ .

**Lemma 4** [18] Given Assumption 2, we have that  $C_t^\delta(x)$  is  $L_0$ -Lipschitz in  $x$  and  $|C_t^\delta(x) - C_t(x)| \leq \delta L_0$ .

---

**Algorithm 1** Risk-averse learning with restarting procedure

---

**Require:** Initial value  $x_0$ , iteration horizon  $T$ , batch size  $\Delta_T$ , smoothing parameter  $\delta$ , risk level  $\alpha$ .

- 1: **for** iteration  $t = 1, \dots, T$  **do**
  - 2:   Identify batch  $j = \left\lceil \frac{t}{\Delta_T} \right\rceil$
  - 3:   Identify epoch in batch  $\tau = t - (j - 1)\Delta_T$
  - 4:   Select sampling number  $n_t = \phi(\tau)$  and learning rate  $\eta_t = \sigma(\tau)$
  - 5:   Sample  $u_t \in \mathbb{S}^d$
  - 6:   Play  $\hat{x}_t = x_t + \delta u_t$
  - 7:   **for**  $i = 1, \dots, n_t$  **do**
  - 8:     Play  $\hat{x}_t$  and obtain  $J_t(\hat{x}_t, \xi_t^i)$
  - 9:   **end for**
  - 10:   Build empirical distribution function  $\hat{F}_t(y)$  given in (6)
  - 11:   Estimate CVaR:  $\text{CVaR}_\alpha[\hat{F}_t]$
  - 12:   Construct gradient estimate  $\hat{g}_t = \frac{d}{\delta} \text{CVaR}_\alpha[\hat{F}_t] u_t$
  - 13:   Update  $x: x_{t+1} \leftarrow \mathcal{P}_{\mathcal{X}^\delta}(x_t - \eta_t \hat{g}_t)$
  - 14: **end for**
- 

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. Let  $\mathcal{A}$  be an online optimization algorithm. We employ the restarting procedure to refresh the parameters and restart  $\mathcal{A}$  every  $\Delta_T$  iterations.

The risk-averse learning algorithm is given as Algorithm 1. Suppose that there are totally  $T$  iterations and they are divided into  $s$  batches with a length of  $\Delta_T \in (1, T)$ , where  $s = \left\lceil \frac{T}{\Delta_T} \right\rceil$ . Each batch is formally defined as

$$\mathcal{T}_j = \{t : (j-1)\Delta_T < t \leq \min\{T, j\Delta_T\}\},$$

for  $j = 1, \dots, s$ .

For each iteration  $t$ , inside batch  $j$ , it holds that  $j = \left\lceil \frac{t}{\Delta_T} \right\rceil$  and its timestamp within the batch is epoch  $\tau$ , i.e.,

$$\tau = t - (j-1)\Delta_T. \quad (3)$$

Then, we design the sampling strategy depending on the epoch  $\tau$ , which is given as  $n_t = \phi(\tau)$ . We allow the algorithm to sample the cost function values multiple times at each iteration, therefore improving the estimation accuracy under the changing distribution. The sampling strategy function  $\phi$  satisfies

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

with the tuning parameter  $a > 0$  and some constant  $c > 0$ . The sampling strategy proposed in [19] is an example that satisfies (4), where

$$\phi(\tau) = \lceil b(\Delta_T - \tau + 1)^a \rceil \quad (5)$$

with parameters  $a, b > 0$ . Specifically, the tuning parameter  $a$  plays the same role in (4) and (5), where a higher value of  $a$  corresponds to a larger number of sampling numbers over the iteration horizon. Similarly, the learning rate at each batch is designed according to  $\eta_t = \sigma(\tau)$ , where the function  $\sigma$  will be designed later. In Algorithm 1, we refresh the evolution of the sampling number  $n_t$  and the learning rate  $\eta_t$  every  $\Delta_T$  iterations.

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  $J(\hat{x}_t, \xi^i)$ ,  $i = 1, \dots, n_t$ . Then, we use the queried function values to construct the empirical distribution function, which is given as

$$\hat{F}_t(y) = \frac{1}{n_t} \sum_{i=1}^{n_t} \mathbf{1}\{J(\hat{x}_t, \xi_t^i) \leq y\}. \quad (6)$$

With this empirical distribution function, we construct the CVaR estimate  $\text{CVaR}_\alpha[\hat{F}_t]$  and further the CVaR gradient estimate

$$\hat{g}_t = \frac{d}{\delta} \text{CVaR}_\alpha[\hat{F}_t] u_t. \quad (7)$$

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

$$x_{t+1} = \mathcal{P}_{\mathcal{X}^\delta}(x_t - \eta_t \hat{g}_t), \quad (8)$$

where  $\mathcal{P}_{\mathcal{X}^\delta}(x) := \arg \min_{y \in \mathcal{X}^\delta} \|x - y\|^2$  denotes the projection operator with  $\mathcal{X}^\delta = \{x \in \mathcal{X} \mid \frac{1}{1-\delta/r} x \in \mathcal{X}\}$  being the projection set. The projection keeps the sampled actions  $\hat{x}_t$  inside of the admissible set  $\mathcal{X}$ , which establishes as

$$\begin{aligned} \left(1 - \frac{\delta}{r}\right) \mathcal{X} \oplus \delta \mathbb{B} &= \left(1 - \frac{\delta}{r}\right) \mathcal{X} \oplus \frac{\delta}{r} r \mathbb{B} \\ &\subseteq \left(1 - \frac{\delta}{r}\right) \mathcal{X} \oplus \frac{\delta}{r} \mathcal{X} = \mathcal{X}. \end{aligned}$$

Without loss of generality, we let the initial action at the restarting epoch be based on action learned from the previous batch.

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

$$\hat{\epsilon}_t := \text{CVaR}_\alpha[\hat{F}_t] - \text{CVaR}_\alpha[F_t]. \quad (9)$$

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** [19] *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*

$$|CVaR_\alpha[F] - CVaR_\alpha[G]| \leq \frac{U}{\alpha} \sup_x |F(x) - G(x)|. \quad (10)$$

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*

$$|CVaR_\alpha[f(X)] - CVaR_\alpha[f(Y)]| \leq \frac{L_0}{\alpha} W_1(\mathcal{D}_X, \mathcal{D}_Y). \quad (11)$$

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

*Proof.* For  $t = 1, \dots, T$ , we have

$$\begin{aligned} \min_{x_t \in \mathcal{X}^\delta} C_t^\delta(x_t) &= \min_{x_t \in \mathcal{X}} C_t^\delta((1 - \delta/r)x_t) \\ &\leq \min_{x_t \in \mathcal{X}} (\delta/r)C_t^\delta(0) + (1 - \delta/r)C_t^\delta(x_t) \\ &\leq \min_{x_t \in \mathcal{X}} C_t^\delta(x_t) + (\delta/r)L_0 \|x_t\| \\ &\leq \min_{x_t \in \mathcal{X}} C_t^\delta(x_t) + D_x L_0 \delta/r. \end{aligned} \quad (12)$$

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

$$\begin{aligned} DR(T) &\leq \sum_{t=1}^T C_t^\delta(\hat{x}_t) - \sum_{t=1}^T C_t^\delta(x_t^*) + 2\delta L_0 T \\ &\leq \sum_{t=1}^T C_t^\delta(x_t) - \sum_{t=1}^T C_t^\delta(x_t^*) + 3\delta L_0 T \\ &\leq \sum_{t=1}^T C_t^\delta(x_t) - \sum_{t=1}^T C_t^\delta(x_t^{\delta,*}) + (3 + D_x/r)\delta L_0 T, \end{aligned} \quad (13)$$

where the first inequality follows from the definition of the function  $C_t^\delta$  as in (2), the second inequality follows from the Lipschitzness of the function  $C_t^\delta$ , and the third inequality establishes by substituting (12) into the  $C_t^\delta(x_t^*)$ . Denote  $x_j^{\delta,*} = \arg \min_{x \in \mathcal{X}^\delta} \sum_{t \in \mathcal{T}_j} C_t^\delta(x)$  as the single best action over batch  $j$ , for  $j = 1, \dots, s$ . It follows that

$$\begin{aligned} &\sum_{t=1}^T C_t^\delta(x_t) - \sum_{t=1}^T C_t^\delta(x_t^{\delta,*}) \\ &= \sum_{j=1}^s \sum_{t \in \mathcal{T}_j} (C_t^\delta(x_t) - C_t^\delta(x_j^{\delta,*})) \\ &\quad + \sum_{j=1}^s \sum_{t \in \mathcal{T}_j} (C_t^\delta(x_j^{\delta,*}) - C_t^\delta(x_t^{\delta,*})) \\ &= \sum_{j=1}^s \mathcal{R}_1^j + \mathcal{R}_2, \end{aligned} \quad (14)$$

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

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

In combination of (7) and (9), we obtain that

$$\nabla C_t^\delta(x_t) = \mathbb{E}[\hat{g}_t - \frac{d}{\delta} \hat{\epsilon}_t u_t]. \quad (15)$$

By the update rule (8), we have

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

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

$$\begin{aligned}
& \langle \hat{g}_t, x_t - x_j^{\delta,*} \rangle \\
&\leq \frac{\|x_t - x_j^{\delta,*}\|^2 - \|x_{t+1} - x_j^{\delta,*}\|^2}{2\eta_t} + \frac{\eta_t}{2} \|\hat{g}_t\|^2. \quad (16)
\end{aligned}$$

Substituting (15) and (16) into  $\mathcal{R}_1^j$ , we obtain

$$\begin{aligned}
\mathcal{R}_1^j &\leq \sum_{t \in \mathcal{T}_j} \mathbb{E}[\langle \hat{g}_t - \frac{d}{\delta} \hat{\epsilon}_t, x_t - x_j^{\delta,*} \rangle] \\
&\leq \sum_{t \in \mathcal{T}_j} \left( \frac{1}{2\eta_t} \mathbb{E}[\|x_t - x_j^{\delta,*}\|^2 - \|x_{t+1} - x_j^{\delta,*}\|^2] \right. \\
&\quad \left. + \frac{\eta_t}{2} \mathbb{E}[\|\hat{g}_t\|^2] + \frac{d}{\delta} \mathbb{E}[\|\hat{\epsilon}_t\| \|x_t - x_j^{\delta,*}\|] \right) \\
&= \frac{1}{2\eta_t} \left( \|x_{(j-1)\Delta_T+1} - x_j^{\delta,*}\|^2 - \|x_{j\Delta_T} - x_j^{\delta,*}\|^2 \right) \\
&\quad + \mathcal{R}_{12}^j + \mathcal{R}_{13}^j \\
&\leq \frac{D_x^2}{\eta_t} + \mathcal{R}_{12}^j + \mathcal{R}_{13}^j, \quad (17)
\end{aligned}$$

with  $\mathcal{R}_{12}^j = \sum_{t \in \mathcal{T}_j} \frac{\eta_t}{2} \mathbb{E}[\|\hat{g}_t\|^2]$  and  $\mathcal{R}_{13}^j = \sum_{t \in \mathcal{T}_j} \frac{d}{\delta} \mathbb{E}[\|\hat{\epsilon}_t\| \|x_t - x_j^{\delta,*}\|]$ . Regarding  $\mathcal{R}_{12}^j$ , for  $i = 1, \dots, s$ , it writes,

$$\begin{aligned}
\mathcal{R}_{12}^j &= \sum_{t \in \mathcal{T}_j} \frac{\eta_t}{2} \left\| \frac{d}{\delta} \text{CVaR}_\alpha[\hat{F}_t]_{u_t} \right\|^2 \\
&\leq \sum_{t \in \mathcal{T}_j} \frac{\eta_t}{2} \left( \frac{dU}{\delta} \right)^2 \leq \frac{d^2 U^2 \eta_t}{2\delta^2} \Delta_T. \quad (18)
\end{aligned}$$

The first inequality establishes as  $\text{CVaR}_\alpha[\hat{F}_t] \leq U$ . Regarding  $\mathcal{R}_{13}^j$ , we first analyze the bound of  $\hat{\epsilon}_t$  given in (9). By leveraging the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality [32], we have that

$$\mathbb{P} \left\{ \sup_y |\hat{F}_t(y) - F_t(y)| \geq \sqrt{\frac{\ln(2/\bar{\gamma})}{2n_t}} \right\} \leq \bar{\gamma}. \quad (19)$$

Denote the event in (19) as  $A_t$ , and  $\mathbb{P}\{A_t\}$  denotes the occurrence probability of event  $A_t$ , for  $t = 1, \dots, T$ . By

Lemma 5, we bound the error of CVaR estimate by

$$|\hat{\epsilon}_t| = \frac{U}{\alpha} \sup |\hat{F}_t - F_t| \leq \frac{U}{\alpha} \sqrt{\frac{\ln(2/\bar{\gamma})}{2n_t}} \quad (20)$$

with probability at least  $1 - \bar{\gamma}$ , for  $t = 1, \dots, T$ . Let  $\gamma = \bar{\gamma}T$ . Then, by substituting (20) into  $\mathcal{R}_{13}^j$ , we obtain that

$$\begin{aligned}
\sum_{j=1}^s \mathcal{R}_{13}^j &\leq \frac{dD_x}{\delta} \sum_{j=1}^s \sum_{t \in \mathcal{T}_j} \mathbb{E}[\|\hat{\epsilon}_t\|] \\
&\leq \frac{dU D_x}{\alpha \delta} \sum_{j=1}^s \sum_{t \in \mathcal{T}_j} \sqrt{\frac{\ln(2T/\gamma)}{2n_t}} \\
&\leq \frac{cdU D_x}{\alpha \delta} \left\lceil \frac{T}{\Delta_T} \right\rceil \sqrt{\frac{\ln(2T/\gamma)}{2}} \Delta_T^{1-\frac{\alpha}{2}} \quad (21)
\end{aligned}$$

with probability at least  $1 - \gamma$ , which establishes as  $1 - \mathbb{P}\{\bigcup_{t=1}^T A_t\} \geq 1 - \sum_{t=1}^T \mathbb{P}\{A_t\} \geq 1 - T\bar{\gamma} \geq 1 - \gamma$ . As shown in (20), when the number of samples increases, the empirical density function approaches the true one. Namely, a sufficient number of samples enable the event  $\{\bigcup_{t=1}^T A_t\}$  to occur with high probability. Regarding  $\mathcal{R}_2$ , we have that

$$\mathcal{R}_2 \leq \frac{2L_0}{\alpha} \Delta_T \sum_{t=2}^T W_1(\mathcal{D}_t, \mathcal{D}_{t-1}) \leq \frac{2L_0}{\alpha} \Delta_T V_D, \quad (22)$$

where the first inequality follows from Lemma 8, respectively. The second inequality follows from Lemma 6, and the last inequality follows from Assumption 3. Moreover, Lemma 6 indicates that when batch size  $\Delta_T$  approaches 1, the batch-optimal actions  $x_j^{\delta,*}$  will be closer to one-step optimal action  $x_t^{\delta,*}$ , for  $t \in \mathcal{T}_j$ , and the accumulated loss defined by (22) will be smaller. However, restarting also induces errors such as  $\frac{D_x^2}{\eta_t}$ . Thus, it is necessary to select an optimal batch size  $\Delta_T$  to minimize the regret. Substituting (17), (18), (21) and (22) into (14), and combining it with (13), it results

$$\begin{aligned}
& \text{DR}(T) \\
&\leq (3 + D_x/r) \delta L_0 T + \mathcal{R}_2 + \sum_{j=1}^s \left( \frac{D_x^2}{\eta_t} + \mathcal{R}_{12}^j + \mathcal{R}_{13}^j \right) \\
&\leq (3 + D_x/r) \delta L_0 T + \frac{2L_0}{\alpha} \Delta_T V_D + \frac{D_x^2}{\eta_t} \left\lceil \frac{T}{\Delta_T} \right\rceil \\
&\quad + \left( \frac{d^2 U^2 \eta_t \Delta_T}{2\delta^2} + \frac{cdU D_x}{\alpha \delta} \sqrt{\frac{\ln(2T/\gamma)}{2}} \Delta_T^{1-\frac{\alpha}{2}} \right) \left\lceil \frac{T}{\Delta_T} \right\rceil \\
&\leq (3 + D_x/r) \delta L_0 T + \frac{2L_0}{\alpha} \Delta_T V_D + \frac{2D_x^2 T}{\eta_t \Delta_T} \\
&\quad + \frac{d^2 U^2 \eta_t T}{\delta^2} + \frac{\sqrt{2}cdU D_x \sqrt{\ln(2T/\gamma)}}{\alpha \delta} T \Delta_T^{-\frac{\alpha}{2}}. \quad (23)
\end{aligned}$$

with probability at least  $1 - \gamma$ . The last inequality results from  $\left\lceil \frac{T}{\Delta_T} \right\rceil \leq \frac{T}{\Delta_T} + 1 \leq \frac{2T}{\Delta_T}$ . When  $a \in (0, 1]$ , we select  $\delta = \left(\frac{V_D}{T}\right)^{\frac{a}{4+a}}$ ,  $\eta_t = \left(\frac{V_D}{T}\right)^{\frac{3a}{4+a}}$  and  $\Delta_T = \left(\frac{T}{V_D}\right)^{\frac{4}{4+a}}$  to minimize the dynamic regret, it achieves  $\text{DR}(T) = \tilde{O}(T^{\frac{4}{4+a}} V_D^{\frac{a}{4+a}})$  with probability at least  $1 - \gamma$  (see [33] 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  $\text{DR}(T) = \tilde{O}(T^{\frac{4}{5}} V_D^{\frac{1}{5}})$  with probability at least  $1 - \gamma$ . The proof is complete.  $\square$

**Remark 1** When  $a \in (0, 1]$ , the regret bound achieved by Algorithm 1 decreases with the increasing tuning parameter  $a$  and ceases increasing when  $a > 1$ . It is because the dynamic regret contains the accumulative error of CVaR gradient estimates, i.e.,  $\frac{d^2 U^2 \eta_t T}{\delta^2} + \frac{\sqrt{2cdUD_x} \sqrt{\ln(2T/\gamma)}}{\alpha \delta} T \Delta_T^{-\frac{a}{2}}$ , which is induced by using finite samples to estimate the CVaR gradients in the zeroth-order algorithm. Since a larger tuning parameter  $a$  implies more queries of cost values and a more accurate estimate of CVaR gradients, the order of the accumulative estimation error decreases with the increasing  $a$ . Furthermore, we select the smoothing parameter  $\delta$ , the batch size  $\Delta_T$  and the step size  $\eta_t$  that minimizes the regret defined in (23). When  $a \in (0, 1]$ , this accumulative estimation error term is one of the dominant terms in (23). While when  $a > 1$ , it becomes negligible compared with the rest terms in (23).

### 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, which are common in risk-averse learning, see [18, 34]

**Assumption 5** The function  $J(x, \xi_t) : \mathcal{X} \times \Xi \rightarrow \mathbb{R}$  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

$$J(y, \xi_t) \geq J(x, \xi_t) + \nabla_x J(x, \xi_t)^\top (y - x) + \frac{m}{2} \|x - y\|_2^2.$$

It follows the lemma for the CVaR function and the smoothed version of the CVaR function.

**Lemma 7** Given Assumption 5, 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. Now we are ready to present the main result for the strongly convex case.

**Theorem 2** Let Assumptions 2, 3 and 5 hold. Suppose that the sampling numbers over iteration horizon  $T$  satisfy (4) with a constant  $a > 0$ . Select the learning rate as

$$\eta_t = \sigma(\tau) = \frac{1}{m\tau}, \quad (24)$$

where the epoch  $\tau$  is obtained from (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  $\text{DR}(T) = \tilde{O}(T^{\frac{4}{4+a}} V_D^{\frac{a}{4+a}})$  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  $\text{DR}(T) = \tilde{O}(T^{\frac{3}{4}} V_D^{\frac{1}{4}})$  with high probability.

*Proof.* Following the derivation of (12)-(14) in the proof of Theorem 1, the dynamic regret under Algorithm 1 is written as

$$\begin{aligned} \text{DR}(T) &\leq (3 + D_x/r) \delta L_0 T + \sum_{j=1}^s \sum_{t \in \mathcal{T}_j} \left( C_t^\delta(x_t) - C_t^\delta(x_j^{\delta,*}) \right) \\ &\quad + \sum_{j=1}^s \sum_{t \in \mathcal{T}_j} \left( C_t^\delta(x_j^{\delta,*}) - C_t^\delta(x_t^{\delta,*}) \right) \\ &\leq (3 + D_x/r) \delta L_0 T + \sum_{j=1}^s \tilde{\mathcal{R}}_1^j + \mathcal{R}_2 \end{aligned} \quad (25)$$

with  $\tilde{\mathcal{R}}_1^j = \sum_{t \in \mathcal{T}_j} C_t^\delta(x_t) - C_t^\delta(x_j^{\delta,*})$  and the definition of  $\mathcal{R}_2$  is as in (14). By the strong convexity of the function  $C_t^\delta$ , we have

$$\begin{aligned} \tilde{\mathcal{R}}_1 &\leq \sum_{j=1}^s \left( \sum_{t \in \mathcal{T}_j} \langle \nabla C_t^\delta(x_t), x_t - x_j^{\delta,*} \rangle - \frac{m}{2} \|x_t - x_j^{\delta,*}\|^2 \right) \\ &\leq \sum_{j=1}^s \sum_{t \in \mathcal{T}_j} \left( \frac{1}{2\eta_t} (\|x_t - x_j^{\delta,*}\|^2 - \|x_{t+1} - x_j^{\delta,*}\|^2) \right. \\ &\quad \left. - \frac{m}{2} \|x_t - x_j^{\delta,*}\|^2 + \frac{\eta_t}{2} \mathbb{E}[\|\hat{g}_t\|^2] \right. \\ &\quad \left. + \frac{d}{\delta} \mathbb{E}[\|\hat{\epsilon}_t\| \|x_t - x_j^{\delta,*}\|] \right) \\ &\leq \sum_{j=1}^s \left( \tilde{\mathcal{R}}_{11}^j + \tilde{\mathcal{R}}_{12}^j + \mathcal{R}_{13}^j \right) \end{aligned} \quad (26)$$

where  $\tilde{\mathcal{R}}_{11}^j = \sum_{t \in \mathcal{T}_j} \left( \frac{1}{2\eta_t} (\|x_t - x_j^{\delta,*}\|^2 - \|x_{t+1} - x_j^{\delta,*}\|^2) - \frac{m}{2} \|x_t - x_j^{\delta,*}\|^2 \right)$ ,  $\tilde{\mathcal{R}}_{12}^j = \sum_{t \in \mathcal{T}_j} \frac{\eta_t}{2} \mathbb{E}[\|\hat{g}_t\|^2]$ , and the definition of  $\mathcal{R}_{13}^j$  is the same as (21). The second inequality follows the derivation of (15)-(17) in Theorem 1. For notational simplicity, denote the first epoch of  $\mathcal{T}_j$  as  $\bar{\tau}_j$ . By



expanding the sequences of  $\tilde{\mathcal{R}}_{11}^j$ , we obtain

$$\begin{aligned} \sum_{j=1}^s \tilde{\mathcal{R}}_{11}^j &= \sum_{j=1}^s \sum_{t \in \{\mathcal{T}_j \setminus \bar{\tau}_j\}} \left( \frac{1}{2\eta_t} - \frac{1}{2\eta_{t-1}} - \frac{m}{2} \right) \|x_t - x_j^{\delta,*}\|^2 \\ &\quad + \frac{1}{2\eta_{\bar{\tau}_j}} \|x_{\bar{\tau}_j} - x_j^{\delta,*}\|^2 - \frac{1}{2\eta_{j\Delta_T}} \|x_{j\Delta_T} - x_j^{\delta,*}\|^2 \\ &\quad - \frac{m}{2} \|x_{\bar{\tau}_j} - x_j^{\delta,*}\|^2 \\ &\leq \sum_{j=1}^s \frac{1}{2\eta_{\bar{\tau}_j}} \|x_{\bar{\tau}_j} - x_j^{\delta,*}\|^2 \leq mD_x^2 \left\lceil \frac{T}{\Delta_T} \right\rceil. \end{aligned} \quad (27)$$

The first inequality establishes by substituting the learning rate given by (24) 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

$$\begin{aligned} \sum_{j=1}^s \tilde{\mathcal{R}}_{12}^j &\leq \frac{d^2 U^2}{2\delta^2} \sum_{j=1}^s \sum_{t \in \mathcal{T}_j} \eta_t \leq \frac{d^2 U^2}{2\delta^2 m} \sum_{j=1}^s \sum_{t=1}^{\Delta_T} \frac{1}{t} \\ &\leq \frac{d^2 U^2}{2\delta^2 m} (1 + \ln \Delta_T) \left\lceil \frac{T}{\Delta_T} \right\rceil. \end{aligned} \quad (28)$$

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 (21), (27) and (28) into (26), combine them with (25), and obtain

$$\begin{aligned} \text{DR}(T) &= (3 + D_x/r)\delta L_0 T + \mathcal{R}_2 + \sum_{j=1}^s \tilde{\mathcal{R}}_{11}^j + \tilde{\mathcal{R}}_{12}^j + \mathcal{R}_{13}^j \\ &\leq (3 + D_x/r)\delta L_0 T + \frac{2L_0}{\alpha} \Delta_T V_D + mD_x^2 \left\lceil \frac{T}{\Delta_T} \right\rceil \\ &\quad + \frac{d^2 U^2}{2\delta^2 m} (1 + \ln \Delta_T) \left\lceil \frac{T}{\Delta_T} \right\rceil \\ &\quad + \frac{cdUD_x}{\alpha\delta} \sqrt{\frac{\ln(2T/\gamma)}{2}} \Delta_T^{1-\frac{a}{2}} \left\lceil \frac{T}{\Delta_T} \right\rceil \\ &\leq (3 + D_x/r)\delta L_0 T + \frac{2L_0}{\alpha} \Delta_T V_D + 2mD_x^2 \frac{T}{\Delta_T} \\ &\quad + \frac{d^2 U^2}{\delta^2 m} (1 + \ln T) \frac{T}{\Delta_T} \\ &\quad + \frac{\sqrt{2}cdUD_x \sqrt{\ln(2T/\gamma)}}{\alpha\delta} T \Delta_T^{-\frac{a}{2}}, \end{aligned} \quad (29)$$

with probability at least  $1 - \gamma$ . The remaining claim is as in Theorem 2. The proof completes here.  $\square$

**Remark 2** Comparing the results of Theorems 1 and 2, we observe that Algorithm 1 achieves the same regret order in the strongly convex and convex cases when  $a \in (0, 1]$ , and achieves smaller regret in the strongly convex case than in the convex case when  $a > 1$ . Additionally, in the strongly convex case, this regret order reduction ceases

when  $a > \frac{4}{3}$ . This is because the accumulative error of CVaR gradient estimates in the strongly convex case is  $\frac{d^2 U^2}{\delta^2 m} (1 + \ln T) \frac{T}{\Delta_T} + \frac{\sqrt{2}cdUD_x \sqrt{\ln(2\Delta_T/\gamma)}}{\alpha\delta} T \Delta_T^{-\frac{a}{2}}$ , of which order decreases with the increasing tuning parameter  $a$ . In this case, we select the smoothing parameter  $\delta$  and the batch size  $\Delta_T$  to minimize the regret defined in (29). The accumulative estimation error is one of the dominant terms in the regret defined in (29) when  $a \in (0, \frac{4}{3}]$  and becomes negligible when  $a > \frac{4}{3}$ , which results in a smaller regret order than in the convex case.

## 4 SIMULATION

In this section, we consider the parking lot dynamic pricing problem, see [6]. 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  $\xi_t$ , which is

$$r_t = \xi_t + Ax_t,$$

where  $A = -0.15$  is the estimated price elasticity, which is determined by [6] through analysis of the real-world data. The uncertainty  $\xi_t$  is distributed according to the time-varying distribution  $\mathcal{D}_t$ , which will change periodically according to environmental effects such as climate condition 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

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

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

$$C_t(x_t) = \text{CVaR}_{\alpha}[J(x_t, \xi_t)], \quad (30)$$

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

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

The time horizon is selected as  $T = 6000$ . Fig. 1 depicts the distribution range of the random variable  $\xi_t$ .

We use Algorithm 1 to update the parking prices in the non-stationary environment with changing distributions. We select the restarting period as  $\Delta_T = 200$ .

We set the initial price value as  $x_0 = 1$  and restrict the potential prices in the range of  $[1, 5]$ . The smoothing parameter for the zeroth-order optimization is set as  $\delta = 0.05$ . We set  $a = \frac{2}{3}$  and  $c = 10$  for the sampling requirement (4). Accordingly, we select  $n_t = 8$  and run the algorithm for 10 trials. Shaded areas represent  $\pm$  one standard deviation over 10 runs.

The simulation results of Algorithm 1 are presented in Figs. 2-4. The top subfigure of Fig. 2 depicts the parking price  $x_t$  and the optimal price  $x_t^*$ , which minimizes the CVaR function defined in (30). The bottom subfigure of Fig. 2 depicts the resulting occupancy  $r_t$ . Since the CVaR function does not have a closed-form expression, we search for optimal prices by drawing a sufficiently large number of samples. More specifically, at each iteration, we select 100 distinct points uniformly in the continuous set  $\mathcal{X}$  and identify the point that corresponds to the minimum CVaR values. It can be observed that the price  $x_t$  generated by Algorithm 1 catches up with the optimal parking price. Fig. 3 shows the CVaR value under the prices generated by Algorithm 1, i.e.,  $C_t(\hat{x}_t)$  and the CVaR value under the optimal prices, i.e.,  $C_t(x_t^*)$ , and the dynamic regret  $DR(T)$ . Moreover, to explore the effect of different sampling numbers on the algorithm, we use 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. As the optimal prices  $x_t^*$  as well as the minimum CVaR value  $\text{CVaR}(x_t^*)$  are the same in these three cases, we use the accumulated loss under the designed algorithm, i.e.,  $\sum_{t=1}^T C_t(\hat{x}_t)$  as the measure. It is shown in Fig 4 that, more samples lead to a smaller cumulative loss, which illustrates our theoretical results.

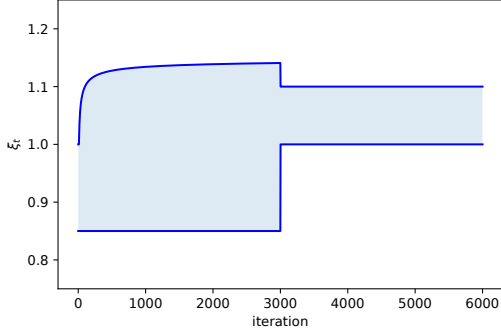


Fig. 1. Distribution range of the uniform random variable  $\xi_t$ .

## 5 CONCLUSIONS

In this paper, we investigated risk-averse learning with time-varying distributions. We employed a risk-averse learning algorithm that queries the function values for multiple times to estimate the gradient of CVaR and

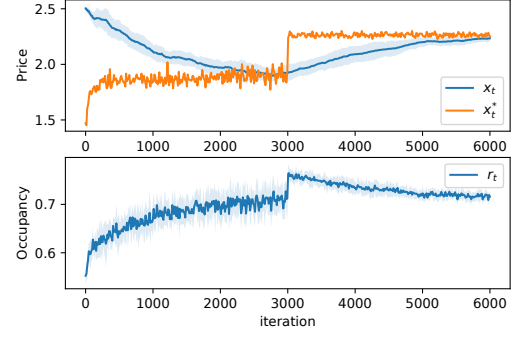


Fig. 2. From top to bottom: the parking price  $x_t$  under Algorithm 1 and the optimal parking price  $x_t^*$ ; the resulted occupancy  $r_t$  under Algorithm 1.

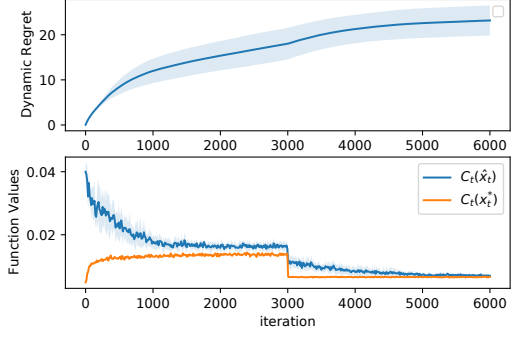


Fig. 3. From top to bottom: dynamic regret; the CVaR values achieved by Algorithm 1 and the minimum CVaR values.

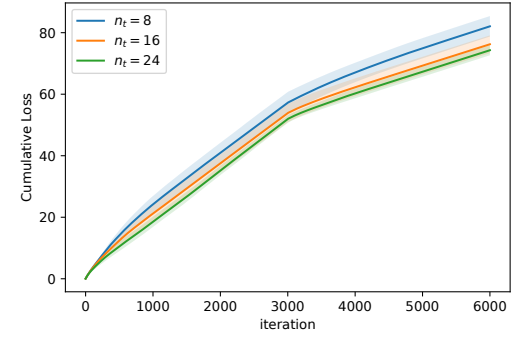


Fig. 4. Accumulated loss achieved by Algorithm 1 under sampling strategies with constant number  $n_t = 8, 16, 24$ , respectively.

proved that the accumulative error of the CVaR gradient is bounded with high probability. By leveraging the restarting procedure, we bound the dynamic regret in terms of the distribution variations for both convex and strongly convex cases. The theoretical results suggest that the increasing sampling numbers reduce the order bound achieved by the designed algorithm, and this re-

duction ceases after the total sampling number reaches a certain threshold. Moreover, the strongly convex assumption leads to a smaller regret order bound than in the convex case.

## 6 Appendix

*Proof of Lemma 6:* Define the augmented functions

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

where  $\mathcal{D}_X$  and  $\mathcal{D}_Y$  are the distributions of the random variables  $X$  and  $Y$ , respectively. According to [17], we have

$$\begin{aligned} \text{CVaR}_\alpha[f(X)] &= \min_v L_{\mathcal{D}_X}(v), \\ \text{CVaR}_\alpha[f(Y)] &= \min_v L_{\mathcal{D}_Y}(v). \end{aligned}$$

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

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

It follows that

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

where we define  $g(x) = [f(x) - v_y]_+$ . The first inequality is due to the fact that  $v_x = \arg \min_v L_{\mathcal{D}_X}(v)$ . According to the definition of  $g(x)$ , we have

$$\begin{aligned} |g(x) - g(y)| &= [f(x) - v_y]_+ - [f(y) - v_y]_+ \\ &\leq [f(x) - f(y)]_+ \\ &\leq [f(x) - f(y)] \leq L_0 \|x - y\|, \end{aligned} \quad (31)$$

where the first inequality follows from the fact that  $a_+ - b_+ \leq [a - b]_+$ , for  $\forall a, b \in \mathbb{R}$ . From (31), we conclude that

the function  $g(x)$  is  $L_0$ -Lipschitz continuous. Hence,

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

where the last inequality follows from the Kantorovich-Rubinstein Duality of the Wasserstein distance, see [30].

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

*Proof of Lemma 7:* 1) We define the function  $h(x, \xi_t) = J(x, \xi_t) - \frac{m}{2} \|x\|^2$ . As  $J(x, \xi_t)$  is  $m$ -strongly convex in  $x$  for every  $\xi_t \in \Xi$ , we have that  $h(x, \xi_t)$  is convex in  $x$  for every  $\xi_t \in \Xi$ . Using Lemma 2, we have that  $\text{CVaR}_\alpha[h(x, \xi_t)]$  is convex in  $x$ . According to the translation invariance property of CVaR, we obtain

$$\begin{aligned} \text{CVaR}_\alpha[h(x, \xi_t)] &= \text{CVaR}_\alpha[J(x, \xi_t)] - \frac{m}{2} \|x\|^2 \\ &= C_t(x) - \frac{m}{2} \|x\|^2. \end{aligned}$$

Using trivial arguments in [35], we conclude that  $C_t(x)$  is  $m$ -strongly convex in  $x$ .

2) As shown in Lemma 2.8 in [5], if the original function is strongly convex, the smoothed function is also strongly convex. The proof is complete.  $\square$

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

$$\begin{aligned} & \sum_{j=1}^s \sum_{t \in \mathcal{T}_j} \left( C_t^\delta(x_j^{\delta,*}) - C_t^\delta(x_t^{\delta,*}) \right) \\ &\leq \frac{2L_0\Delta_T}{\alpha} \sum_{t=2}^T W_1(\mathcal{D}_t, \mathcal{D}_{t-1}). \end{aligned} \quad (32)$$

*Proof of Lemma 8:* This proof is adopted from Proposition 2 of [10]. 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  $j = 1, \dots, s$ , we have

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

In the following we will prove  $\max_{t \in \mathcal{T}_j} \{C_t^\delta(x_{\bar{\tau}_j}^{\delta,*}) - C_t^\delta(x_t^{\delta,*})\} \leq 2V_j$  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

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

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

$$\sum_{t \in \mathcal{T}_j} C_t^\delta(x_j^{\delta,*}) - \sum_{t \in \mathcal{T}_j} C_t^\delta(x_t^{\delta,*}) \leq 2\Delta_T V_j.$$

Summarize the variation along batches  $\{\mathcal{T}_1, \dots, \mathcal{T}_s\}$ , it results

$$\begin{aligned} & \sum_{j=1}^s \left( \sum_{t \in \mathcal{T}_j} C_t^\delta(x_j^{\delta,*}) - \sum_{t \in \mathcal{T}_j} C_t^\delta(x_t^{\delta,*}) \right) \\ & \leq \sum_{j=1}^s 2\Delta_T V_j \leq \frac{2L_0}{\alpha} \Delta_T \sum_{t=2}^T W_1(\mathcal{D}_t, \mathcal{D}_{t-1}). \end{aligned} \quad (33)$$

Here is the proof.  $\square$

## References

- [1] Elad Hazan. *Efficient algorithms for online convex optimization and their applications*. Princeton University, 2006.
- [2] 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.
- [3] 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.
- [4] Geoffrey J Gordon, Amy Greenwald, and Casey Marks. No-regret learning in convex games. In *Proc. of the 25th International Conference on Machine Learning*, pages 360–367, 2008.
- [5] Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325, 2016.
- [6] Mitas Ray, Lillian J Ratliff, Dmitriy Drusvyatskiy, and Maryam Fazel. Decision-dependent risk minimization in geometrically decaying dynamic environments. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 8081–8088, 2022.
- [7] Adhyayan 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*, pages 5891–5912, 2022.
- [8] Juan Perdomo, Tijana Zrnic, Celestine Mendler-Dünner, and Moritz Hardt. Performative prediction. In *International Conference on Machine Learning*, pages 7599–7609, 2020.
- [9] John P Miller, Juan C Perdomo, and Tijana Zrnic. Outside the echo chamber: Optimizing the performative risk. In *International Conference on Machine Learning*, pages 7710–7720, 2021.
- [10] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015.
- [11] Chaoyue Zhao and Yongpei Guan. Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters*, 46(2):262–267, 2018.
- [12] Peng Zhao, Guanghui Wang, Lijun Zhang, and Zhi-Hua Zhou. Bandit convex optimization in non-stationary environments. *The Journal of Machine Learning Research*, 22(1):5562–5606, 2021.
- [13] Jiashuo Jiang, Xiaocheng Li, and Jiawei Zhang. Online stochastic optimization with Wasserstein based non-stationarity. *arXiv preprint arXiv:2012.06961*, 2020.
- [14] Iman Shames and Farhad Farokhi. Online stochastic convex optimization: Wasserstein distance variation. *arXiv preprint arXiv:2006.01397*, 2020.
- [15] Thomas J Linsmeier and Neil D Pearson. Value at risk. *Financial Analysts Journal*, 56(2):47–67, 2000.
- [16] R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.
- [17] R Tyrrell Rockafellar, Stanislav Uryasev, et al. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.
- [18] Adrian Rivera Cardoso and Huan Xu. Risk-averse stochastic convex bandit. In *Proc. of the 22nd International Conference on Artificial Intelligence and Statistics*, pages 39–47, 2019.
- [19] 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 *Proc. of the 61st IEEE Conference on Decision and Control*, pages 5179–5184. IEEE, 2022.
- [20] Zifan Wang, Yi Shen, and Michael Zavlanos. Risk-averse no-regret learning in online convex games. In *International Conference on Machine Learning*, pages 22999–23017, 2022.
- [21] Margaret P Chapman and Laurent Lessard. Toward a scalable upper bound for a CVaR-lq problem. *IEEE Control Systems Letters*, 6:920–925, 2021.
- [22] Masako Kishida and Ahmet Cetinkaya. Risk-aware linear quadratic control using conditional value-at-risk. *IEEE Transactions on Automatic Control*, 2022.
- [23] 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 *2019 American Control Conference*, pages 2958–2963. IEEE, 2019.

- [24] 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.
- [25] 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.
- [26] Tasuku Soma and Yuichi Yoshida. Statistical learning with conditional value at risk. *arXiv preprint arXiv:2002.05826*, 2020.
- [27] Núria Armengol Urpí, Sebastian Curi, and Andreas Krause. Risk-averse offline reinforcement learning. *arXiv preprint arXiv:2102.05371*, 2021.
- [28] 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 *Proc. of the 30th International Joint Conference on Artificial Intelligence*, pages 2708–2715, 2021.
- [29] Leonid V Kantorovich. Mathematical methods of organizing and planning production. *Management Science*, 6(4):366–422, 1960.
- [30] David A Edwards. On the Kantorovich–Rubinstein theorem. *Expositiones Mathematicae*, 29(4):387–398, 2011.
- [31] Joseph L Doob. The Brownian movement and stochastic equations. *Annals of Mathematics*, pages 351–369, 1942.
- [32] Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pages 642–669, 1956.
- [33] 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.
- [34] Edward Anderson, Huifu Xu, and Dali Zhang. Varying confidence levels for CVaR risk measures and minimax limits. *Mathematical Programming*, 180:327–370, 2020.
- [35] Stephen P Boyd and Lieven Vandenbergh. *Convex optimization*. Cambridge University Press, 2004.