

# Delay-Tolerant Constrained OCO with Application to Network Resource Allocation

Juncheng Wang\*, Ben Liang\*, Min Dong<sup>†</sup>, Gary Boudreau<sup>‡</sup>, and Hatem Abou-zeid<sup>‡</sup>

\*Department of Electrical and Computer Engineering, University of Toronto, Canada,

<sup>†</sup>Department of Electrical, Computer and Software Engineering, Ontario Tech University, Canada, <sup>‡</sup>Ericsson Canada, Canada

**Abstract**—We consider online convex optimization (OCO) with multi-slot feedback delay, where an agent makes a sequence of online decisions to minimize the accumulation of time-varying convex loss functions, subject to short-term and long-term constraints that are possibly time-varying. The current convex loss function and the long-term constraint function are revealed to the agent only after the decision is made, and they may be delayed for multiple time slots. Existing work on OCO under this general setting has focused on the static regret, which measures the gap of losses between the online decision sequence and an offline benchmark that is fixed over time. In this work, we consider both the static regret and the more practically meaningful dynamic regret, where the benchmark is a time-varying sequence of per-slot optimizers. We propose an efficient algorithm, termed Delay-Tolerant Constrained-OCO (DTC-OCO), which uses a novel constraint penalty with double regularization to tackle the asynchrony between information feedback and decision updates. We derive upper bounds on its dynamic regret, static regret, and constraint violation, proving them to be sublinear under mild conditions. We further apply DTC-OCO to a general network resource allocation problem, which arises in many systems such as data networks and cloud computing. Simulation results demonstrate substantial performance gain of DTC-OCO over the known best alternative.

## I. INTRODUCTION

Online convex optimization (OCO) has emerged as a promising solution to many machine learning, signal processing, and resource allocation problems in the presence of uncertainty [1], [2]. Under the standard OCO setup, at the beginning of each time slot, an agent makes a decision from a known convex set. At the end of each slot, the system reveals information of the current convex loss function to the agent, and then the agent’s loss is realized. Due to the lack of in-time information of the current convex loss function, it is impossible for an agent to make an optimal decision at each slot. Instead, the agent aims at minimizing the *regret*, *i.e.*, the performance gap between the online decision sequence and some performance benchmark.

Most of the early works on OCO studied the *static* regret, which compares the online decision sequence with a static offline benchmark that has a priori information of all the convex loss functions. In the seminal work of OCO [3], an online projected gradient descent algorithm was shown to achieve  $\mathcal{O}(T^{\frac{1}{2}})$  static regret, where  $T$  is the time horizon. The static regret was further reduced to  $\mathcal{O}(\log T)$  in [4] for strongly

convex loss functions. However, when the optimum of the underlying system is time-varying, the static offline benchmark itself may perform poorly and achieving sublinear static regret may not be meaningful anymore. In [3], a more useful *dynamic* regret was also presented, where the benchmark was a time-varying solution sequence. This dynamic benchmark has been recognized as a more attractive but harder-to-track performance benchmark for OCO [5], [6].

The above works all focused on OCO with short-term constraints that must be strictly satisfied. Subsequently, OCO with *long-term* constraints was considered in [7]. In this case, a desired OCO algorithm should provide sublinear *constraint violation*, so that the time-averaged violation of each long-term constraint tends to zero as time approaches infinity. A trade-off between the regret and constraint violation bounds was shown in [8], and the constraint violation bound was further improved in [9]. These works assumed that the long-term constraints are time-invariant and known in advance, while recent works [10]–[12] studied OCO with *time-varying* long-term constraints.

In practical systems, the decision maker often gains access to the system information only after some delay. For example, in wireless communications, data transmission relies on channel state information, which is usually delayed for multiple transmission frames after channel estimation, quantization, and feedback. In mobile computing, feedback delay can be caused by offloading tasks from remote devices and communicating through wireless channels. Under standard OCO, the decision maker receives information on the current loss function (and if applicable the long-term constraint functions) at the end of each slot when the decision is made, *i.e.*, the feedback information is delayed for only one slot. This is the model used in, *e.g.*, [3]–[12], but it can be too restrictive for many practical applications. Therefore, [13] initiated a study on OCO with multi-slot feedback delay. Additional delay-adaptive OCO algorithms were proposed in [14] and [15], while a static regret bound was derived in recent work [16] for time-varying constrained OCO with multi-slot feedback delay.

However, to the best of our knowledge, all existing works on OCO with multi-slot feedback delay have focused on the static regret, which may not be a meaningful performance metric for some inherently time-varying systems. In fact, the gap between the static and dynamic regrets can be as large as  $\mathcal{O}(T)$  [17]. In the presence of multi-slot feedback delay, whether sublinear dynamic regret is achievable for OCO is an open problem. The need to consider (time-varying and time-invariant) long-term

TABLE I  
SUMMARY OF RELATED WORKS ON OCO

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	DTC-OCO
Long-term time-invariant constraints							Y	Y	Y								Y
Long-term time-varying constraints										Y	Y	Y				Y	Y
Multi-slot feedback delay													Y	Y	Y	Y	Y
Static performance benchmark	Y	Y	Y	Y			Y	Y	Y	Y			Y	Y	Y	Y	Y
Dynamic performance benchmark			Y		Y	Y					Y	Y					Y

constraints further adds to this challenge. In this context, the main contributions of this paper are as follows:

- We propose an efficient algorithm, termed Delay-Tolerant Constrained-OCO (DTC-OCO), for OCO with multi-slot feedback delay and both short-term and long-term constraints. While all existing OCO algorithms use either no regularization or single regularization for decision updates, in DTC-OCO, we propose a novel constraint penalty with double regularization to tackle the asynchrony between information feedback and decision updates, which improves tolerance to multi-slot delay and facilitates the performance bounding of DTC-OCO.
- We analyze the special structure of DTC-OCO due to double regularization and establish that it provides  $\mathcal{O}(\max\{\tau^{\frac{1}{2}}T^{\frac{1+\delta}{2}}, T^\nu\})$  dynamic regret and  $\mathcal{O}(\max\{\tau^{\frac{1}{2}}T^{\frac{1}{2}}, T^\nu\})$  static regret, where  $\tau$  is the feedback delay, and  $\delta$  and  $\nu$  respectively represent the growth rates of the accumulated variation of a dynamic benchmark and constraint functions.
- We derive  $\mathcal{O}(\max\{T^{\frac{1-\delta}{2}}, \tau T^\kappa\})$  constraint violation bound for DTC-OCO, where  $\kappa$  represents the growth rate of the accumulated squared variation of constraint functions. In the special case of time-invariant constraints, even with multi-slot feedback delay, DTC-OCO recovers the known best constraint violation bound for constrained OCO with one-slot feedback delay.
- We apply DTC-OCO to a general network resource allocation problem, which arises in many systems such as data networks and cloud computing. Simulation based on an example cloud computing system demonstrates that DTC-OCO is tolerant to feedback delay and substantially outperforms the known best alternative from [16].

The rest of this paper is organized as follows. In Section II, we present the related work. Section III describes the problem formulation and performance metrics. We present DTC-OCO, provide its performance bounds, and discuss its performance merits in Section IV. The application of DTC-OCO to network resource allocation is presented in Section V, followed by concluding remarks in Section VI.

## II. RELATED WORK

In this section, we survey existing works on OCO with either long-term constraints or multi-slot feedback delay. The differences between these works and our work are summarized in Table. I.

Among existing works on OCO with long-term constraints, a saddle-point-typed algorithm was proposed in [7] and achieved  $\mathcal{O}(T^{\frac{1}{2}})$  static regret and  $\mathcal{O}(T^{\frac{3}{4}})$  constraint violation. A follow-up work [8] provided  $\mathcal{O}(T^{\max\{\mu, 1-\mu\}})$  static regret

and  $\mathcal{O}(T^{1-\frac{\mu}{2}})$  constraint violation, where  $\mu \in (0, 1)$  is some trade-off parameter. A virtual-queue-based algorithm was proposed in [9] and established the known best  $\mathcal{O}(1)$  constraint violation. Another virtual-queue-based algorithm was proposed in [10] for independent and identically distributed (i.i.d.) constraints. The saddle-point-typed and virtual-queue-based algorithms were respectively modified in [11] and [12] to deal with time-varying constraints. However, the above works were all confined to the standard OCO setting where the feedback information is delayed for only one slot.

Most existing works on OCO with multi-slot feedback delay focused on online problems with short-term constraints [13]-[15]. The standard online gradient descent algorithm [3] was extended in [13] to provide  $\mathcal{O}(\tau^{\frac{1}{2}}T^{\frac{1}{2}})$  static regret. Delay-adaptive online gradient descent algorithms were proposed in [14] and [15] to accommodate adversarial feedback delay. The impact of multi-slot feedback delay on OCO with long-term constraints was considered only in [16]. However, [16] only studied the static regret, which may not be an attractive performance metric for an inherently time-varying system. Furthermore, the constraint violation bound provided in [16] was no less than  $\mathcal{O}(T^{\frac{3}{4}})$  even for time-invariant constraints, inherited from the saddle-point-typed algorithm [7]. Thus, the impact of multi-slot feedback delay on constrained OCO is still not well understood. Different from [16], we propose a novel constraint penalty with double regularization to achieve sublinear dynamic and static regret bounds, as well as a stronger constraint violation bound.

Constrained OCO is related to Lyapunov optimization [18], which makes use of the system state and queueing information to implicitly learn and adapt to changes in the system with unknown statistics. However, under this framework, the system states are commonly assumed to be i.i.d. or Markovian, while OCO solutions do not have such restriction. Furthermore, standard Lyapunov optimization relies on the current and accurate system state for decision updates. In the presence of feedback delay on the system state, one can apply Lyapunov optimization by leveraging historical information to predict the current system state with some errors [19]. However, this way of dealing with feedback delay is equivalent to extending the standard Lyapunov optimization to inaccurate system states in [20] and [21]. In this case, the optimality gap would be  $\mathcal{O}(\sigma T)$  with  $\sigma$  being some measure of system inaccuracy, so that such an approach cannot lead to the sublinear dynamic regret bound that we seek.

In this work, we focus on centralized OCO. Distributed OCO is beyond our scope, and we refer interested readers to [22]-[26].

### III. CONSTRAINED OCO WITH MULTI-SLOT DELAY

#### A. Problem Formulation

We consider a time-slotted system with time indexed by  $t$ . Let  $\{f_t(\mathbf{x})\}$ , where  $f_t(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ , be a sequence of convex loss functions. Let  $\{\mathbf{g}_t(\mathbf{x})\}$  be a sequence of convex long-term constraint functions, where  $\mathbf{g}_t(\mathbf{x}) = [g_t^1(\mathbf{x}), \dots, g_t^C(\mathbf{x})]^T : \mathbb{R}^n \rightarrow \mathbb{R}^C$  with  $C$  being the number of constraints. Let  $\mathcal{C} = \{1, \dots, C\}$ . The loss function  $f_t(\mathbf{x})$  and the constraint function  $\mathbf{g}_t(\mathbf{x})$  can vary over time. We further consider short-term constraints represented by a compact convex set  $\mathcal{X}_0 \subseteq \mathbb{R}^n$ . The goal of constrained OCO is to select a sequence of decisions  $\{\mathbf{x}_t\}$  from  $\mathcal{X}_0$  that minimizes the accumulated loss while ensuring that the long-term constraints are satisfied [11], [12]. This leads to the following optimization problem:

$$\begin{aligned} \mathbf{P1} : \quad & \min_{\{\mathbf{x}_t \in \mathcal{X}_0\}} \sum_{t=1}^T f_t(\mathbf{x}_t) \\ & \text{s.t.} \quad \sum_{t=1}^T \mathbf{g}_t(\mathbf{x}_t) \preceq \mathbf{0}. \end{aligned} \quad (1)$$

Note that if the constraint functions are time-invariant, *i.e.*,  $\mathbf{g}_t(\mathbf{x}) = \mathbf{g}(\mathbf{x}), \forall t$ , then **P1** is simplified to the time-invariant constrained OCO problem considered in [7]-[9].

Under the standard constrained OCO setting [7]-[12], feedback information on the loss function  $f_t(\mathbf{x})$  and the long-term constraint function  $\mathbf{g}_t(\mathbf{x})$  is assumed to be delayed for only one slot, when it can be leveraged to make the new decision  $\mathbf{x}_{t+1}$ .<sup>1</sup> However, in many practical applications, *e.g.*, wireless transmission and mobile computing mentioned in Section I, this assumption is rarely satisfied since the feedback information can be severely delayed. Therefore, in this work, we consider a general scenario where the feedback information of  $f_t(\mathbf{x})$  and  $\mathbf{g}_t(\mathbf{x})$  is delayed for  $\tau \geq 1$  slots, so that the decision maker receives it at the end of slot  $t + \tau - 1$  [13]-[16]. Different from similar works with only short-term constraints [13]-[15], the additional long-term constraints in (1) of **P1** lead to a more complicated online problem, especially since the underlying system varies over time while the feedback information is delayed.

#### B. Performance Metrics

Due to the lack of in-time information of the current loss and constraint functions under the OCO setting, it is impossible to obtain an optimal solution to **P1**.<sup>2</sup> Instead, a time-varying constrained OCO algorithm aims at selecting an online solution sequence  $\{\mathbf{x}_t\}$  that is asymptotically no worse than some performance benchmark.

One common offline benchmark is

$$\mathbf{x}^* \in \arg \min_{\mathbf{x} \in \mathcal{X}_0} \left\{ \sum_{t=1}^T f_t(\mathbf{x}) \mid \mathbf{g}_t(\mathbf{x}) \preceq \mathbf{0}, \forall t \right\} \quad (2)$$

<sup>1</sup>We note that, to have a well-posed problem, information on the short-term constraints must be current. Furthermore, obviously delay is irrelevant to time-invariant long-term constraints [7]-[9].

<sup>2</sup>In fact, even for the most basic OCO problem [3] (*i.e.*, without long-term constraints (1)), an optimal solution cannot be found [4].

which is computed assuming all information of  $\{f_t(\mathbf{x})\}$  and  $\{\mathbf{g}_t(\mathbf{x})\}$  is known in advance. The performance gap between  $\{\mathbf{x}_t\}$  and  $\mathbf{x}^*$  is termed the static regret:

$$\text{RE}_s(T) \triangleq \sum_{t=1}^T [f_t(\mathbf{x}_t) - f_t(\mathbf{x}^*)]. \quad (3)$$

For example, this static regret was used in [16], while [7]-[9] used a special case of it with time-invariant constraints.<sup>3</sup> However, as a rather coarse performance metric, the static regret may not be a strong indicator for the actual algorithm performance especially when the underlying system is inherently time-varying.

A more attractive performance benchmark for time-varying constrained OCO is the dynamic benchmark  $\{\mathbf{x}_t^*\}$ , given by

$$\mathbf{x}_t^* \in \arg \min_{\mathbf{x} \in \mathcal{X}_0} \{f_t(\mathbf{x}) \mid \mathbf{g}_t(\mathbf{x}) \preceq \mathbf{0}\} \quad (4)$$

which is computed using the in-time information of  $f_t(\mathbf{x})$  and  $\mathbf{g}_t(\mathbf{x})$  at each slot  $t$ . The dynamic benchmark was originally proposed for OCO with short-term constraints [3] and was modified in [11] and [12] to incorporate long-term constraints. The corresponding dynamic regret is defined as

$$\text{RE}_d(T) \triangleq \sum_{t=1}^T [f_t(\mathbf{x}_t) - f_t(\mathbf{x}_t^*)]. \quad (5)$$

In some cases, the gap between  $\text{RE}_s(T)$  and  $\text{RE}_d(T)$  can be as large as  $\mathcal{O}(T)$  [17]. In this paper, we provide upper bounds on both  $\text{RE}_s(T)$  and  $\text{RE}_d(T)$  for a comprehensive performance study.

To measure the accumulated violation of the long-term constraints, the constraint violation<sup>4</sup> is defined for any  $c \in \mathcal{C}$  as [12], [16]

$$\text{VO}^c(T) \triangleq \sum_{t=1}^T g_t^c(\mathbf{x}_t). \quad (6)$$

Note that the constraint violation defined in [7]-[9] for time-invariant constraint  $\mathbf{g}(\mathbf{x})$  is a special case of (6). Our study accommodates both time-varying and time-invariant constraints.

A constrained OCO algorithm is desired to provide both sublinear regrets, *i.e.*,  $\text{RE}_d(T) = \mathbf{o}(T)$  and  $\text{RE}_s(T) = \mathbf{o}(T)$ , and sublinear constraint violation, *i.e.*,  $\text{VO}^c(T) = \mathbf{o}(T)$ . Sublinearity is important since it implies that the online decision is asymptotically no worse than the benchmark in terms of its time-averaged performance, and the long-term constraints are satisfied in the time-averaged sense. Unfortunately, with multi-slot feedback delay, existing literature only achieves sublinear static regret and  $\mathcal{O}(T^{\frac{3}{4}})$  constraint violation [16]. In this work, we propose an efficient algorithm DTC-OCO to provide sublinear dynamic and static regrets and a stronger constraint violation bound.

<sup>3</sup>Even with one-slot delay, [27] showed that it is impossible to achieve sublinear static regret w.r.t.  $\mathbf{x}^o \in \arg \min_{\mathbf{x} \in \mathcal{X}_0} \{\sum_{t=1}^T f_t(\mathbf{x}) \mid \sum_{t=1}^T \mathbf{g}_t(\mathbf{x}) \preceq \mathbf{0}\}$  and sublinear constraint violations simultaneously.

<sup>4</sup>The constraint violation is referred to as dynamic fit  $\text{Fit}(T) \triangleq \|[\sum_{t=1}^T \mathbf{g}_t(\mathbf{x}_t)]^+\|_2$  in [11], where  $[\mathbf{x}]^+ \triangleq \max\{\mathbf{x}, \mathbf{0}\}$  is the entry-wise positive projection operator. One can easily verify that the sublinearity of  $\text{VO}^c(T), \forall c \in \mathcal{C}$  implies  $\text{Fit}(T)$  being sublinear, and vice versa.

#### IV. DELAY-TOLERANT CONSTRAINED ONLINE CONVEX OPTIMIZATION (DTC-OCO)

In this section, we present details of DTC-OCO and study the impact of multi-slot feedback delay on the performance guarantees of DTC-OCO to provide regret and constraint violation bounds. We further give sufficient conditions under which DTC-OCO yields sublinear regret and constraint violation. Finally, the performance merits of DTC-OCO over existing constrained OCO algorithms are discussed.

##### A. DTC-OCO Algorithm

DTC-OCO introduces a novel virtual queue vector  $\mathbf{Q}_t = [Q_t^1, \dots, Q_t^c]^T$  for the long-term constraints (1), with the following updating rule for any  $c \in \mathcal{C}$ :

$$Q_t^c = \max \{ -\gamma g_{t-\tau}^c(\mathbf{x}_t), Q_{t-1}^c + \gamma g_{t-\tau}^c(\mathbf{x}_t) \} \quad (7)$$

where  $\gamma > 0$  is a step-size parameter. The role of  $\mathbf{Q}_t$  is similar to a Lagrangian multiplier vector for  $\mathbf{P1}$  or a backlog queue for the constraint violation, which are concepts used in [9], [10], and [12]. However, unique to our proposed approach,  $g_{t-\tau}^c(\mathbf{x}_t)$  is the  $\tau$ -slot delayed constraint violation caused by the *current* decision, and it needs to be scaled by an appropriate  $\gamma$  factor. We then convert  $\mathbf{P1}$  to solving a per-slot problem at each slot  $t > \tau$  with short-term constraints only, given by

$$\begin{aligned} \mathbf{P2}: \min_{\mathbf{x} \in \mathcal{X}_0} & [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T (\mathbf{x} - \mathbf{x}_{t-\tau}) \\ & + [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T [\gamma \mathbf{g}_{t-\tau}(\mathbf{x})] \\ & + \alpha \|\mathbf{x} - \mathbf{x}_{t-\tau}\|_2^2 + \eta \|\mathbf{x} - \mathbf{x}_{t-1}\|_2^2 \end{aligned}$$

where  $\alpha, \eta > 0$  are two step-size parameters.

DTC-OCO uses the  $\tau$ -slot delayed gradient  $\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})$  in  $\mathbf{P2}$  for accumulated loss minimization. Compared with the original  $\mathbf{P1}$ , the long-term constraints (1) are converted to penalizing  $\mathbf{g}_{t-\tau}(\mathbf{x})$  for queue stabilities as one part of the objective in  $\mathbf{P2}$ . Note that DTC-OCO uses a novel constraint penalty with double regularization  $\alpha \|\mathbf{x} - \mathbf{x}_{t-\tau}\|_2^2$  and  $\eta \|\mathbf{x} - \mathbf{x}_{t-1}\|_2^2$  in the per-slot objective in  $\mathbf{P2}$  to handle the asynchrony between information feedback and decision updates. This will be shown, analytically in Sections IV-B and IV-C and numerically in Section V, to give DTC-OCO substantial performance advantage over existing works in terms of regret bounds and average performance.

To summarize, DTC-OCO is comprised of three major steps: 1) Initialize  $\mathbf{x}_t \in \mathcal{X}_0, \mathbf{Q}_t = \mathbf{0}, \forall t \in [1, \tau]$  and  $\mathbf{g}_0(\mathbf{x}) \equiv \mathbf{0}$ ; 2) At the beginning of each slot  $t > \tau$ , obtain the current decision  $\mathbf{x}_t$  by solving  $\mathbf{P2}$ ; 3) At the end of each slot  $t > \tau$ , update the virtual queue  $\mathbf{Q}_t$  via (7).<sup>5</sup> Note that DTC-OCO has three step-size parameters  $\alpha, \eta, \gamma > 0$ . Their choice depends on our knowledge of the system and will be discussed in Section IV-C, after we derive the regret and constraint violation bounds in Section IV-B. We will clarify the impact of these step-sizes on those bounds.

<sup>5</sup>If information feedback of  $f_t(\mathbf{x})$  and  $\mathbf{g}_t(\mathbf{x})$  are respectively delayed for  $\tau_1$  and  $\tau_2$  slots with  $\tau_1 \neq \tau_2$ , DTC-OCO can still be applied by setting  $\tau = \max\{\tau_1, \tau_2\}$ .

The main difference between DTC-OCO and the saddle-point-typed OCO algorithms [7], [8], [11], [16] is that DTC-OCO uses a virtual queue to track the constraint violation. The virtual queue was also used in Lyapunov optimization [18], and later extended to OCO in [9], [10], and [12]. Although a small part of our performance bounding borrows some techniques from Lyapunov drift analysis, as explained in Section II, DTC-OCO is structurally different from Lyapunov optimization. Furthermore, the virtual-queue-based OCO algorithms in [9] and [10] were designed for time-invariant and i.i.d. constraints, respectively, while obtaining only static regret bounds. Compared with [12], DTC-OCO only uses the gradient of the loss functions *at the past decision points*, instead of the complete information of the past loss functions. Furthermore, [9], [10], and [12] are limited to one-slot feedback delay. Therefore, the virtual queue construction, algorithm design, and performance bound derivation for DTC-OCO are all substantially different from those of [9], [10], and [12].

##### B. Regret and Constraint Violation Bounds

In this section, we present new techniques to derive the performance bounds of DTC-OCO, particularly to account for its constraint penalty with double regularization.

We make the following assumptions that are common in the literature for constrained OCO [7]-[12], [16].

**Assumption 1.** The gradient  $\nabla f_t(\mathbf{x})$  is bounded:  $\exists D > 0, s.t.,$

$$\|\nabla f_t(\mathbf{x})\|_2 \leq D, \quad \forall \mathbf{x} \in \mathcal{X}_0, \quad \forall t. \quad (8)$$

**Assumption 2.** For any  $t, \mathbf{g}_t(\mathbf{x})$  satisfies the following:

2.1)  $\mathbf{g}_t(\mathbf{x})$  is Lipschitz continuous on  $\mathcal{X}_0$ :  $\exists \beta > 0, s.t.,$

$$\|\mathbf{g}_t(\mathbf{x}) - \mathbf{g}_t(\mathbf{y})\|_2 \leq \beta \|\mathbf{x} - \mathbf{y}\|_2, \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}_0, \quad \forall t. \quad (9)$$

2.2)  $\mathbf{g}_t(\mathbf{x})$  is bounded:  $\exists G > 0, s.t.,$

$$\|\mathbf{g}_t(\mathbf{x})\|_2 \leq G, \quad \forall \mathbf{x} \in \mathcal{X}_0, \quad \forall t. \quad (10)$$

2.3) Existence of an interior point:  $\exists \epsilon > 0$  and  $\tilde{\mathbf{x}}_t \in \mathcal{X}_0, s.t.,$

$$\mathbf{g}_t(\tilde{\mathbf{x}}_t) \preceq -\epsilon \mathbf{I}, \quad \forall t. \quad (11)$$

**Assumption 3.** The radius of  $\mathcal{X}_0$  is bounded:  $\exists R > 0, s.t.,$

$$\|\mathbf{x} - \mathbf{y}\|_2 \leq R, \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}_0. \quad (12)$$

We first provide bounds on the virtual queue vector in the following lemma. The proof follows from the virtual queue definition in (7) and is omitted due to page limitation.

**Lemma 1.** The virtual queue vector generated by DTC-OCO is bounded for any  $t > \tau$  by the following inequalities:

$$\mathbf{Q}_t + \gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t) \succeq \mathbf{0}, \quad (13)$$

$$\|\mathbf{Q}_t\|_2 \geq \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|_2, \quad (14)$$

$$\|\mathbf{Q}_t\|_2 \leq \|\mathbf{Q}_{t-1}\|_2 + \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|_2. \quad (15)$$

Define  $L_t \triangleq \frac{1}{2} \|\mathbf{Q}_t\|_2^2$  as a quadratic Lyapunov function and  $\Delta_t \triangleq L_{t+1} - L_t$  as the corresponding Lyapunov drift [18]. Leveraging results in Lemma 1, we provide an upper bound on  $\Delta_t$  in the following lemma.

**Lemma 2.** The Lyapunov drift is upper bounded for any  $t > \tau$  as follows:

$$\Delta_{t-1} \leq \gamma \mathbf{Q}_{t-1}^T \mathbf{g}_{t-\tau}(\mathbf{x}_t) + \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|_2^2. \quad (16)$$

*Proof:* (Proof outline due to page limitation) For any  $c \in \mathcal{C}$  and  $t > \tau$ , we can verify that

$$\frac{1}{2}(Q_t^c)^2 - \frac{1}{2}(Q_{t-1}^c)^2 \leq \gamma Q_{t-1}^c g_{t-\tau}^c(\mathbf{x}_t) + [\gamma g_{t-\tau}^c(\mathbf{x}_t)]^2$$

by considering the two cases from (7): 1)  $Q_{t-1}^c + \gamma g_{t-\tau}^c(\mathbf{x}_t) \geq -\gamma g_{t-\tau}^c(\mathbf{x}_t)$  and 2)  $-\gamma g_{t-\tau}^c(\mathbf{x}_t) > Q_{t-1}^c + \gamma g_{t-\tau}^c(\mathbf{x}_t)$ . Summing the above inequalities over  $c \in \mathcal{C}$  yields (16). ■

We also require the following lemma, which is reproduced from Lemma 2.8 in [1].

**Lemma 3.** Let  $\mathcal{S} \in \mathbb{R}^n$  be a nonempty convex set. Let  $h(\mathbf{v}) : \mathbb{R}^n \rightarrow \mathbb{R}$  be a  $\rho$ -strongly-convex function over  $\mathcal{S}$  w.r.t. a norm  $\|\cdot\|$ . Let  $\mathbf{w} = \arg \min_{\mathbf{v} \in \mathcal{S}} h(\mathbf{v})$ . Then, for any  $\mathbf{u} \in \mathcal{S}$ , we have  $h(\mathbf{w}) \leq h(\mathbf{u}) - \frac{\rho}{2} \|\mathbf{u} - \mathbf{w}\|^2$ .

A main goal of this paper is to examine the impact of multi-slot feedback delay on the dynamic regret bound for constrained OCO, which has not been addressed in the existing literature. To this end, we need to quantify the accumulated variations of the underlying time-varying system. We define the accumulated variation of the dynamic benchmark  $\{\mathbf{x}_t^*\}$  (commonly termed the path length [3]) as

$$\Delta_{\mathbf{x}^*} \triangleq \sum_{t=1}^T \|\mathbf{x}_t^* - \mathbf{x}_{t-1}^*\|_2. \quad (17)$$

Furthermore, we define the accumulated squared variation of the constraint function sequence  $\{\mathbf{g}_t(\mathbf{x})\}$  as

$$\Delta_{\mathbf{g}} \triangleq \sum_{t=1}^T \max_{\mathbf{x} \in \mathcal{X}_0} \{\|\mathbf{g}_t(\mathbf{x}) - \mathbf{g}_{t-1}(\mathbf{x})\|_2^2\}. \quad (18)$$

Another related quantity regarding the accumulated variation of  $\{\mathbf{g}_t(\mathbf{x})\}$  is defined as

$$\tilde{\Delta}_{\mathbf{g}} \triangleq \sum_{t=1}^T \max_{\mathbf{x} \in \mathcal{X}_0} \{\|\mathbf{g}_t(\mathbf{x}) - \mathbf{g}_{t-1}(\mathbf{x})\|_2\}. \quad (19)$$

In the order sense,  $\Delta_{\mathbf{g}}$  is usually smaller than  $\tilde{\Delta}_{\mathbf{g}}$  for a constraint function sequence  $\{\mathbf{g}_t(\mathbf{x})\}$  that varies sublinearly [12].

Leveraging results in Lemmas 1 and 2, and the tuning freedom brought by the double regularization for constructing telescoping terms, the following theorem provides an upper bound on the dynamic regret  $\text{RE}_d(T)$  for DTC-OCO with  $\tau$ -slot feedback delay.

**Theorem 4.** For any  $\alpha, \gamma > 0$  and  $\eta \geq \gamma^2 \beta^2$ , the dynamic regret of DTC-OCO is upper bounded by

$$\begin{aligned} \text{RE}_d(T) &\leq \frac{D^2}{4\alpha} T + \frac{\gamma^2 G^2}{2} + \gamma^2 \Delta_{\mathbf{g}} \\ &\quad + (\alpha\tau + \eta)(R^2 + 2R\Delta_{\mathbf{x}^*}) + DR\tau. \end{aligned} \quad (20)$$

*Proof:* The objective function of  $\mathbf{P2}$  is  $2(\alpha + \eta)$ -strongly-convex over  $\mathcal{X}_0$  w.r.t. Euclidean norm  $\|\cdot\|_2$  due to the double

regularization. Since  $\mathbf{x}_t$  minimizes  $\mathbf{P2}$  over  $\mathcal{X}_0$  for any  $t > \tau$ , we have

$$\begin{aligned} &[\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T (\mathbf{x}_t - \mathbf{x}_{t-\tau}) + \alpha \|\mathbf{x}_t - \mathbf{x}_{t-\tau}\|_2^2 \\ &\quad + [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T [\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)] + \eta \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 \\ &\stackrel{(a)}{\leq} [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T (\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-\tau}) + \alpha \|\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-\tau}\|_2^2 \\ &\quad + [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T [\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_{t-\tau}^*)] \\ &\quad + \eta \|\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-1}\|_2^2 - (\alpha + \eta) \|\mathbf{x}_t - \mathbf{x}_{t-\tau}^*\|_2^2 \end{aligned} \quad (21)$$

$$\begin{aligned} &\stackrel{(b)}{\leq} [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T (\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-\tau}) \\ &\quad + \alpha (\|\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-\tau}\|_2^2 - \|\mathbf{x}_t - \mathbf{x}_{t-\tau}^*\|_2^2) \\ &\quad + \eta (\|\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-1}\|_2^2 - \|\mathbf{x}_t - \mathbf{x}_{t-\tau}^*\|_2^2), \end{aligned} \quad (22)$$

where (a) follows from Lemma 3; and (b) is because  $\mathbf{Q}_\tau = \mathbf{0}$ ,  $\mathbf{g}_0(\mathbf{x}) \equiv \mathbf{0}$  by initialization,  $\mathbf{Q}_t + \gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t) \succeq \mathbf{0}, \forall t > \tau$ , in (13),  $\gamma > 0$ , and  $\mathbf{g}_{t-\tau}(\mathbf{x}_{t-\tau}^*) \leq \mathbf{0}, \forall t > \tau$ , in (4), such that  $[\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T [\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_{t-\tau}^*)] \leq 0, \forall t > \tau$ .

Now, we bound the second and third terms in (22). From  $\|\mathbf{a} + \mathbf{b}\|_2^2 \geq \|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2 - 2\|\mathbf{a}\|_2 \|\mathbf{b}\|_2$ , we have

$$\begin{aligned} &\|\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-\tau}\|_2^2 - \|\mathbf{x}_t - \mathbf{x}_{t-\tau}^*\|_2^2 \\ &\leq \|\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-\tau}\|_2^2 - \|\mathbf{x}_t^* - \mathbf{x}_t\|_2^2 - \|\mathbf{x}_{t-\tau}^* - \mathbf{x}_t^*\|_2^2 \\ &\quad + 2\|\mathbf{x}_t^* - \mathbf{x}_t\|_2 \|\mathbf{x}_{t-\tau}^* - \mathbf{x}_t^*\|_2 \leq \Phi_{t-\tau} + 2R\phi_{t-\tau}, \end{aligned} \quad (23)$$

where  $\Phi_{t-\tau} \triangleq \|\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-\tau}\|_2^2 - \|\mathbf{x}_t^* - \mathbf{x}_t\|_2^2$  and  $\phi_{t-\tau} \triangleq \|\mathbf{x}_{t-\tau}^* - \mathbf{x}_t^*\|_2$ . Similarly, we can show that

$$\|\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-1}\|_2^2 - \|\mathbf{x}_t - \mathbf{x}_{t-\tau}^*\|_2^2 \leq \Psi_{t-\tau} + 2R\psi_{t-\tau}, \quad (24)$$

where  $\Psi_{t-\tau} \triangleq \|\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-1}\|_2^2 - \|\mathbf{x}_{t-\tau+1}^* - \mathbf{x}_t\|_2^2$  and  $\psi_{t-\tau} \triangleq \|\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-\tau+1}^*\|_2$ .

Substituting (23) and (24) into (22), adding  $f_{t-\tau}(\mathbf{x}_{t-\tau})$  on both sides, applying the first-order condition of convexity

$$f_{t-\tau}(\mathbf{x}_{t-\tau}) + [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T (\mathbf{x}_{t-\tau}^* - \mathbf{x}_{t-\tau}) \leq f_{t-\tau}(\mathbf{x}_{t-\tau}^*)$$

to its RHS, and rearranging terms, we have

$$\begin{aligned} &f_{t-\tau}(\mathbf{x}_{t-\tau}) - f_{t-\tau}(\mathbf{x}_{t-\tau}^*) \\ &\leq -[\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T (\mathbf{x}_t - \mathbf{x}_{t-\tau}) - \alpha \|\mathbf{x}_t - \mathbf{x}_{t-\tau}\|_2^2 \\ &\quad - [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T [\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)] - \eta \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 \\ &\quad + \alpha (\Phi_{t-\tau} + 2R\phi_{t-\tau}) + \eta (\Psi_{t-\tau} + 2R\psi_{t-\tau}). \end{aligned} \quad (25)$$

We now bound the right-hand side of (25). Note that

$$\begin{aligned} & - [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T [\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)] \\ &\stackrel{(a)}{\leq} -\Delta_{t-1} + \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|_2^2 - \gamma^2 \mathbf{g}_{t-\tau-1}^T(\mathbf{x}_{t-1}) \mathbf{g}_{t-\tau}(\mathbf{x}_t) \\ &\stackrel{(b)}{=} -\Delta_{t-1} + \frac{\gamma^2}{2} (\|\mathbf{g}_{t-\tau}(\mathbf{x}_t)\|_2^2 - \|\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|_2^2) \\ &\quad + \frac{\gamma^2}{2} \|\mathbf{g}_{t-\tau}(\mathbf{x}_t) - \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|_2^2 \\ &\stackrel{(c)}{\leq} -\Delta_{t-1} + \gamma^2 \left( \frac{1}{2} \varphi_{t-\tau} + \beta^2 \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 + \varpi_{t-\tau} \right), \end{aligned} \quad (26)$$

where  $\varphi_{t-\tau} \triangleq \|\mathbf{g}_{t-\tau}(\mathbf{x}_t)\|_2^2 - \|\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|_2^2$  and  $\varpi_{t-\tau} \triangleq \|\mathbf{g}_{t-\tau}(\mathbf{x}_{t-1}) - \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|_2^2$ . Here, (a) follows from rearranging terms of (16) in Lemma 2 such that

$-\gamma \mathbf{Q}_{t-1}^T \mathbf{g}_{t-\tau}(\mathbf{x}_t) \leq -\Delta_{t-1} + \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|_2^2$ , (b) is because  $\mathbf{a}^T \mathbf{b} = \frac{1}{2}(\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2 - \|\mathbf{a} - \mathbf{b}\|_2^2)$ , and (c) follows from  $\mathbf{g}_t(\mathbf{x})$  being Lipschitz continuous in (9) and the fact that  $\frac{1}{2}\|\mathbf{a} + \mathbf{b}\|_2^2 \leq \|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2$ .

Substituting (26) into (25), we have

$$\begin{aligned} & f_{t-\tau}(\mathbf{x}_{t-\tau}) - f_{t-\tau}(\mathbf{x}_{t-\tau}^*) \\ & \leq -[\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T (\mathbf{x}_t - \mathbf{x}_{t-\tau}) - \alpha \|\mathbf{x}_t - \mathbf{x}_{t-\tau}\|_2^2 \\ & \quad + (\gamma^2 \beta^2 - \eta) \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 - \Delta_{t-1} + \frac{\gamma^2}{2} \varphi_{t-\tau} + \gamma^2 \varpi_{t-\tau} \\ & \quad + \alpha(\Phi_{t-\tau} + 2R\phi_{t-\tau}) + \eta(\Psi_{t-\tau} + 2R\psi_{t-\tau}) \\ & \stackrel{(a)}{\leq} \frac{D^2}{4\alpha} - \Delta_{t-1} + \frac{\gamma^2}{2} \varphi_{t-\tau} + \gamma^2 \varpi_{t-\tau} \\ & \quad + \alpha(\Phi_{t-\tau} + 2R\phi_{t-\tau}) + \eta(\Psi_{t-\tau} + 2R\psi_{t-\tau}), \end{aligned} \quad (27)$$

where (a) follows from  $\eta \geq \gamma^2 \beta^2$ , the bound on  $\nabla f_t(\mathbf{x})$  in (8), and completing the square such that

$$\begin{aligned} & -[\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T (\mathbf{x}_t - \mathbf{x}_{t-\tau}) - \alpha \|\mathbf{x}_t - \mathbf{x}_{t-\tau}\|_2^2 \\ & = -\left\| \frac{\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})}{2\sqrt{\alpha}} + \sqrt{\alpha}(\mathbf{x}_t - \mathbf{x}_{t-\tau}) \right\|_2^2 + \frac{1}{4\alpha} \|\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})\|_2^2 \\ & \leq \frac{1}{4\alpha} \|\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})\|_2^2 \leq \frac{D^2}{4\alpha}. \end{aligned}$$

Summing (27) over  $t \in [\tau + 1, T]$ , we have

$$\begin{aligned} & \sum_{t=\tau+1}^T f_{t-\tau}(\mathbf{x}_{t-\tau}) - f_{t-\tau}(\mathbf{x}_{t-\tau}^*) = \sum_{t=1}^{T-\tau} f_t(\mathbf{x}_t) - f_t(\mathbf{x}_t^*) \\ & \stackrel{(a)}{\leq} \frac{D^2}{4\alpha} T + \frac{\gamma^2 G^2}{2} + \gamma^2 \Delta_{\mathbf{g}} + (\alpha\tau + \eta)(R^2 + 2R\Delta_{\mathbf{x}^*}), \end{aligned} \quad (28)$$

where (a) follows from  $\Delta_{t-1}$ ,  $\varphi_{t-\tau}$ ,  $\varpi_{t-\tau}$ ,  $\Phi_{t-\tau}$ ,  $\phi_{t-\tau}$ ,  $\Psi_{t-\tau}$  and  $\psi_{t-\tau}$  all being telescoping terms such that their sums over  $t \in [\tau + 1, T]$  are upper bounded by 0,  $G^2$ ,  $\Delta_{\mathbf{g}}$ ,  $\tau R^2$ ,  $\tau \Delta_{\mathbf{x}^*}$ ,  $R^2$ , and  $\Delta_{\mathbf{x}^*}$ , respectively.

Finally, adding  $\sum_{t=\tau+1}^T f_t(\mathbf{x}_t) - f_t(\mathbf{x}_t^*)$  on both sides of (28), and noting that the convexity of  $f_t(\mathbf{x})$  implies

$$f_t(\mathbf{x}_t) - f_t(\mathbf{x}_t^*) \leq \|\nabla f_t(\mathbf{x}_t)\|_2 \|\mathbf{x}_t^* - \mathbf{x}_t\|_2 \leq DR, \quad (29)$$

we complete the proof.  $\blacksquare$

Next, leveraging the proof techniques for the dynamic regret  $\text{RE}_d(T)$  in Theorem 4, we provide an upper bound on the static regret  $\text{RE}_s(T)$  yielded by DTC-OCO.

**Theorem 5.** For any  $\alpha, \gamma > 0$  and  $\eta \geq \gamma^2 \beta^2$ , the static regret of DTC-OCO is upper bounded by

$$\text{RE}_s(T) \leq \frac{D^2}{4\alpha} T + \frac{\gamma^2 G^2}{2} + \gamma^2 \Delta_{\mathbf{g}} + (\alpha\tau + \eta)R^2 + DR\tau. \quad (30)$$

*Proof:* (Proof outline due to page limitation) Replacing all the per-slot optimizers with the offline benchmark  $\mathbf{x}^*$  in the proof of Theorem 4, we can show that (27) still holds by redefining  $\Phi_{t-\tau} \triangleq \|\mathbf{x}^* - \mathbf{x}_{t-\tau}\|_2^2 - \|\mathbf{x}^* - \mathbf{x}_t\|_2^2$ ,  $\phi_{t-\tau} \triangleq 0$ ,  $\Psi_{t-\tau} \triangleq \|\mathbf{x}^* - \mathbf{x}_{t-1}\|_2^2 - \|\mathbf{x}^* - \mathbf{x}_t\|_2^2$ , and  $\psi_{t-\tau} \triangleq 0$ . Summing the above version of (27) over  $t \in [\tau + 1, T]$ , noting that  $\Phi_{t-\tau}$  and  $\Psi_{t-\tau}$  are still telescoping, and leveraging (29), we complete the proof.  $\blacksquare$

We now proceed to provide an upper bound on the constraint violation  $\text{VO}^c(T)$  for DTC-OCO. We first relate the virtual queue vector  $\mathbf{Q}_T$  to  $\text{VO}^c(T)$  in the following lemma.

**Lemma 6.** The virtual queue vector produced by DTC-OCO satisfies the following inequality for any  $c \in \mathcal{C}$ :

$$\text{VO}^c(T) \leq \frac{1}{\gamma} \|\mathbf{Q}_T\|_2 + \tau \tilde{\Delta}_{\mathbf{g}} + G\tau. \quad (31)$$

*Proof:* From (7), we have  $Q_t^c \geq Q_{t-1}^c + \gamma g_{t-\tau}^c(\mathbf{x}_t)$ ,  $\forall t > \tau$ . Summing it over  $t \in [\tau + 1, T]$  and rearranging terms, we have

$$\sum_{t=\tau+1}^T g_{t-\tau}^c(\mathbf{x}_t) \leq \frac{1}{\gamma} \sum_{t=\tau+1}^T Q_t^c - Q_{t-1}^c = \frac{1}{\gamma} Q_T^c.$$

From the above inequality and the definition of  $\text{VO}^c(T)$  in (6), we have

$$\text{VO}^c(T) \leq \frac{1}{\gamma} Q_T^c + \sum_{t=1}^{T-\tau} [g_{t+\tau}^c(\mathbf{x}_{t+\tau}) - g_t^c(\mathbf{x}_{t+\tau})] + \sum_{t=1}^{\tau} g_t^c(\mathbf{x}_t).$$

Noting  $\|\mathbf{a}\|_{\infty} \leq \|\mathbf{a}\|_2$ , the bound on  $\mathbf{g}_t(\mathbf{x})$  in (10), and the definition of  $\tilde{\Delta}_{\mathbf{g}}$  in (19), we complete the proof.  $\blacksquare$

From Lemma 6, we see that one can bound the constraint violation  $\text{VO}^c(T)$  through an upper bound on the virtual queue vector  $\mathbf{Q}_T$ . This is used to obtain an upper bound on the constraint violation for DTC-OCO in the following theorem.

**Theorem 7.** For any  $\alpha, \eta, \gamma > 0$ , the constraint violation of DTC-OCO for any  $c \in \mathcal{C}$  is upper bounded by

$$\text{VO}^c(T) \leq 2G + \frac{2\gamma^2 G^2 + DR + (\alpha + \eta)R^2}{\epsilon\gamma^2} + \tau \tilde{\Delta}_{\mathbf{g}} + G\tau. \quad (32)$$

*Proof:* From Lemma 3, we can show that inequality (21) still holds for any  $t > \tau$  after replacing the per-slot optimizer  $\mathbf{x}_{t-\tau}^*$  with the interior point  $\tilde{\mathbf{x}}_{t-\tau}$ . We have

$$\begin{aligned} & [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T [\gamma \mathbf{g}_{t-\tau}(\tilde{\mathbf{x}}_{t-\tau})] \\ & \stackrel{(a)}{\leq} -\epsilon\gamma [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T \mathbf{1} \\ & \stackrel{(b)}{\leq} -\epsilon\gamma \|\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|_2 \\ & \stackrel{(c)}{\leq} -\epsilon\gamma (\|\mathbf{Q}_{t-1}\|_2 - \|\gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|_2), \end{aligned} \quad (33)$$

where (a) follows from the existence of interior point in (11) and the virtual queue bound in (13), (b) is because  $\|\mathbf{a}\|_2 \leq \|\mathbf{a}\|_1$ , and (c) follows from  $\|\|\mathbf{a}\|_2 - \|\mathbf{b}\|_2\| \leq \|\mathbf{a} - \mathbf{b}\|_2$ . Applying (33) to the aforementioned version of (21) with  $\tilde{\mathbf{x}}_{t-\tau}$  and rearranging terms, we have

$$\begin{aligned} & \gamma \mathbf{Q}_{t-1}^T \mathbf{g}_{t-\tau}(\mathbf{x}_t) \\ & \leq -\epsilon\gamma (\|\mathbf{Q}_{t-1}\|_2 - \|\gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|_2) - \alpha \|\mathbf{x}_t - \mathbf{x}_{t-\tau}\|_2^2 \\ & \quad - [\gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T [\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)] - \eta \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2 \\ & \quad + [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T (\tilde{\mathbf{x}}_{t-\tau} - \mathbf{x}_t) + \alpha \|\tilde{\mathbf{x}}_{t-\tau} - \mathbf{x}_{t-\tau}\|_2^2 \\ & \quad + \eta \|\tilde{\mathbf{x}}_{t-\tau} - \mathbf{x}_{t-1}\|_2^2 - (\alpha + \eta) \|\mathbf{x}_t - \tilde{\mathbf{x}}_{t-\tau}\|_2^2 \\ & \stackrel{(a)}{\leq} -\epsilon\gamma \|\mathbf{Q}_{t-1}\|_2 + \epsilon\gamma^2 G + \gamma^2 G^2 + DR + (\alpha + \eta)R^2, \end{aligned} \quad (34)$$

where (a) follows from the Cauchy-Schwartz inequality  $|\mathbf{a}^T \mathbf{b}| \leq \|\mathbf{a}\|_2 \|\mathbf{b}\|_2$ , the bound on  $\nabla f_t(\mathbf{x})$  in (8), the bound on  $\mathbf{g}_t(\mathbf{x})$  in (10), and the bound on  $\mathcal{X}_0$  in (12). Substituting (34) into (16) in Lemma 2 and noting that  $\|\mathbf{g}_{t-\tau}(\mathbf{x}_t)\|_2^2 \leq G^2$  from (10) yields

$$\Delta_{t-1} \leq -\epsilon\gamma \|\mathbf{Q}_{t-1}\|_2 + \epsilon\gamma^2 G + 2\gamma^2 G^2 + DR + (\alpha + \eta)R^2.$$

Thus, a sufficient condition for  $\Delta_{t-1} < 0$  is

$$\|\mathbf{Q}_{t-1}\|_2 > \gamma G + \frac{2\gamma^2 G^2 + DR + (\alpha + \eta)R^2}{\epsilon\gamma}. \quad (35)$$

If (35) holds, we have  $\|\mathbf{Q}_t\|_2 < \|\mathbf{Q}_{t-1}\|_2$ , i.e., the virtual queue decreases; otherwise, from the virtual queue bound in (15), there is a maximum increase from  $\|\mathbf{Q}_{t-1}\|_2$  to  $\|\mathbf{Q}_t\|_2$  since  $\|\mathbf{Q}_t\|_2 - \|\mathbf{Q}_{t-1}\|_2 \leq \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|_2 \leq \gamma G$ . Therefore, the virtual queue is upper bounded for any  $t > \tau$  by

$$\|\mathbf{Q}_t\|_2 \leq 2\gamma G + \frac{2\gamma^2 G^2 + DR + (\alpha + \eta)R^2}{\epsilon\gamma}. \quad (36)$$

Substituting (36) into (31), we complete the proof.  $\blacksquare$

### C. Discussion on the Regret and Constraint Violation Bounds

In this section, we discuss the sufficient conditions for DTC-OCO to yield sublinear regret and constraint violation, and highlight several prominent advantages of DTC-OCO over existing constrained OCO algorithms.

1) *Sublinear Regret and Constraint Violation:* From Theorems 4, 5, and 7, we can derive the following corollaries regarding the regret and constraint violation bounds. The proofs are omitted for brevity. Here, we define parameters  $\delta, \nu, \kappa \geq 0$  to represent the time variability of the per-slot optimizers and constraint functions, such that  $\Delta_{\mathbf{x}^*} = \mathcal{O}(T^\delta)$ ,  $\Delta_{\mathbf{g}} = \mathcal{O}(T^\nu)$ , and  $\hat{\Delta}_{\mathbf{g}} = \mathcal{O}(T^\kappa)$  [11], [12]. These corollaries provide two sets of performance bounds depending on whether the value of  $\delta$  is known.

**Corollary 8.** *Step-sizes with knowledge of  $\delta$ :* Let  $\alpha = \tau^{-\frac{1}{2}} T^{\frac{1-\delta}{2}}$ ,  $\eta = \beta^2 \gamma^2$ , and  $\gamma = 1$  in DTC-OCO. We have

$$\text{RE}_d(T) = \mathcal{O}\left(\max\left\{\tau^{\frac{1}{2}} T^{\frac{1+\delta}{2}}, T^\nu\right\}\right), \quad (37)$$

$$\text{RE}_s(T) = \mathcal{O}\left(\max\left\{\tau^{\frac{1}{2}} T^{\frac{1}{2}}, T^\nu\right\}\right), \quad (38)$$

$$\text{VO}^c(T) = \mathcal{O}\left(\max\left\{T^{\frac{1-\delta}{2}}, \tau T^\kappa\right\}\right). \quad (39)$$

In particular, if  $\tau = \mathcal{O}(1)$ ,  $\delta < 1$ ,  $\nu < 1$ , and  $\kappa < 1$ , both the dynamic and static regrets are sublinear, and the constraint violation is sublinear.

**Corollary 9.** *Step-sizes without knowledge of  $\delta$ :* Let  $\alpha = \tau^{-\frac{1}{2}} T^{\frac{1}{2}}$ ,  $\eta = \beta^2 \gamma^2$ , and  $\gamma = 1$  in DTC-OCO. We have

$$\text{RE}_d(T) = \mathcal{O}\left(\max\left\{\tau^{\frac{1}{2}} T^{\frac{1}{2}+\delta}, T^\nu\right\}\right), \quad (40)$$

$$\text{RE}_s(T) = \mathcal{O}\left(\max\left\{\tau^{\frac{1}{2}} T^{\frac{1}{2}}, T^\nu\right\}\right), \quad (41)$$

$$\text{VO}^c(T) = \mathcal{O}\left(\max\left\{T^{\frac{1}{2}}, \tau T^\kappa\right\}\right). \quad (42)$$

From Corollaries 8 and 9, a sufficient condition for DTC-OCO to yield sublinear dynamic and static regrets and sublinear constraint violation is that the accumulated variations of the per-slot optimizers and constraints evolve sufficiently slowly. Otherwise, if the system varies too drastically, no online algorithm can track it due to the lack of in-time information [11], [12], [17]. However, in many online applications, the system tends to stabilize over time and sublinear regret and constraint violation can be achieved by DTC-OCO.

**Remark 1.** If the feedback delay  $\tau$  is unknown, by setting  $\alpha = T^{\frac{1-\delta}{2}}$  in DTC-OCO, we can show that  $\text{RE}_d(T) = \mathcal{O}(\max\{\tau T^{\frac{1+\delta}{2}}, T^\nu\})$  and  $\text{RE}_s(T) = \mathcal{O}(\max\{\tau T^{\frac{1}{2}}, T^\nu\})$ , both still being sublinear under the same conditions in Corollary 8. If both  $\tau$  and  $\delta$  are unknown, by setting  $\alpha = T^{\frac{1}{2}}$  in DTC-OCO, we can show that  $\text{RE}_d(T) = \mathcal{O}(\max\{\tau T^{\frac{1}{2}+\delta}, T^\nu\})$  and  $\text{RE}_s(T) = \mathcal{O}(\max\{\tau T^{\frac{1}{2}}, T^\nu\})$ .

**Remark 2.** With unknown time horizon  $T$ , the standard doubling trick [1], [6], [28] can be applied to extend DTC-OCO into one that has similar regret bounds.

**Remark 3.** The performance analysis in [16] assumes  $T$  is large enough such that  $\frac{(1+C)\max\{D,\beta\}^2+2}{\sqrt{\tau T}} + [(5C+1)\max\{D,\beta\}^2+2]\sqrt{\frac{\tau}{T}} \leq \sqrt{\frac{1}{3}}$ , and therefore it provides no performance guarantee for a mid-range value of  $T$ . Furthermore, only a static regret bound is provided in [16]. In addition, the optimal step-sizes in [16] require knowing the values of  $C$  and  $D$ , which we do not need for DTC-OCO.

2) *Special Case of One-Slot Feedback Delay:* Since no other algorithm has provided a dynamic regret bound for constrained OCO with multi-slot feedback delay, we next consider the special case of one-slot feedback delay. In the following two remarks, we compare DTC-OCO with [11] and [12] under this setting.

**Remark 4.** The dynamic regret and constraint violation bounds achieved by [11] rely on the key assumption that the slack constant  $\epsilon$  is larger than the maximum variation of the constraints, i.e.,  $\epsilon > \max_{t \in [1, T]} \{\max_{\mathbf{x} \in \mathcal{X}_0} \{\|\mathbf{g}_t(\mathbf{x}) - \mathbf{g}_{t-1}(\mathbf{x})\|_2\}\}$ , which may be difficult to satisfy in general. In contrast, DTC-OCO only assumes  $\epsilon > 0$  as indicated in (11). Furthermore, the optimal step-sizes used in [11] require knowledge of the accumulated variation of the constraint function sequence  $\kappa$ , while DTC-OCO only needs an upper bound on its gradient  $\beta$ , which is much easier to acquire than  $\kappa$ . When  $\delta$  is unknown, [11] achieves  $\mathcal{O}(\max\{T^{\frac{1}{3}+\delta}, T^{\frac{1}{3}+\kappa}, T^{\frac{2}{3}}\})$  dynamic regret and  $\mathcal{O}(T^{\frac{2}{3}})$  constraint violation, both being at least  $\mathcal{O}(T^{\frac{2}{3}})$ . In contrast, the performance bounds of DTC-OCO decreases smoothly to  $\mathcal{O}(T^{\frac{1}{2}})$  if the system variation is sufficiently small.

**Remark 5.** To achieve sublinear dynamic regret and constraint violation, [12] relies on additional assumptions that the accumulated variation of the convex loss functions  $\{f_t(\mathbf{x})\}$  is sublinear regardless of the trajectory of the online deci-

sion sequence, *i.e.*,  $\sum_{t=1}^T \max_{\mathbf{x} \in \mathcal{X}_0} \{\|f_t(\mathbf{x}) - f_{t-1}(\mathbf{x})\|_2\} = \mathbf{o}(T)$ , and the accumulated variation of the optimal dual points  $\{\lambda_t^*\}$  of the optimization problem **P1** is sublinear, *i.e.*,  $\sum_{t=1}^T \|\lambda_{t+1}^* - \lambda_t^*\|_2 = \mathbf{o}(T)$ . Neither assumption is required for DTC-OCO. Furthermore, [12] requires complete information feedback of the convex loss function  $f_t(\mathbf{x})$ , and therefore has the gradient information at any point and can even penalize the loss function directly. In contrast, DTC-OCO requires only the gradient information of the convex loss function at the online decision points  $\nabla f_t(\mathbf{x}_t)$ .

3) *Special Case of Time-invariant Constraints*: When the constraints are time-invariant, the following corollary suggests that DTC-OCO recovers the known best  $\mathcal{O}(\tau^{\frac{1}{2}} T^{\frac{1}{2}})$  static regret for unconstrained OCO with multi-slot delay [13], and its  $\mathcal{O}(\tau)$  constraint violation recovers the known best  $\mathcal{O}(1)$  constraint violation under the standard OCO setting of one-slot feedback delay [9] as a special case.

**Corollary 10.** *Time-invariant constraints*: Let  $\alpha = \gamma^2 = \tau^{-\frac{1}{2}} T^{\frac{1-\delta}{2}}$  and  $\eta = \beta^2 \gamma^2$  in DTC-OCO. We have  $\text{RE}_d(T) = \mathcal{O}(\tau^{\frac{1}{2}} T^{\frac{1+\delta}{2}})$ ,  $\text{RE}_s(T) = \mathcal{O}(\tau^{\frac{1}{2}} T^{\frac{1}{2}})$ , and  $\text{VO}^c(T) = \mathcal{O}(\tau)$ .

In particular, if  $\tau = \mathcal{O}(1)$  and  $\delta < 1$ , both the dynamic and static regrets are sublinear, and the constraint violation is upper bounded by a constant.

**Remark 6.** The constraint violation bounds in [11] and [16] are no less than  $\mathcal{O}(T^{\frac{2}{3}})$  and  $\mathcal{O}(T^{\frac{3}{4}})$ , respectively, even if the constraint function is time-invariant. In contrast, when  $\tilde{\Delta}_{\mathbf{g}}$  is small, the constraint violation  $\text{VO}^c(T)$  yielded by DTC-OCO decreases smoothly to  $\mathcal{O}(T^{\frac{1}{2}})$ . Particularly, if the constraint function is time-invariant, *i.e.*,  $\tilde{\Delta}_{\mathbf{g}} = 0$ , DTC-OCO provides  $\mathcal{O}(\tau)$  constraint violation.

## V. APPLICATION TO NETWORK RESOURCE ALLOCATION

We apply DTC-OCO to a general network resource allocation problem [11], [12], [18], [29], [30]. We present numerical results to demonstrate the performance advantage of DTC-OCO over the known best alternative from [16].

### A. Online Network Resource Allocation

Fig. 1 shows a general network consisting of  $J$  scheduling nodes and  $K$  processing nodes. For example, in a wired or wireless network, the scheduling nodes may be relays, and the processing nodes may be sink nodes. In a cloud computing network, the scheduling nodes may be mappers, and the processing nodes may be computing servers.

At each time slot  $t$ , the amount of data arriving at scheduling node  $j$  is denoted by  $d_t^j$ , and we define an extended data arrival vector denoted by  $\mathbf{d}_t = [d_t^1, \dots, d_t^J, \mathbf{0}_{1 \times K}]^T$ . A central controller decides the transmission rate  $y_t^{jk}$  of the link  $(j, k)$  connecting scheduling node  $j$  and processing node  $k$ , and the processing rate  $z_t^k$  at processing node  $k$ . In compact form, the decision vector at time  $t$  is  $\mathbf{x}_t = [y_t^{11}, \dots, y_t^{JK}, z_t^1, \dots, z_t^K]^T$ . Denote the maximum data transmission rate of link  $(j, k)$  by  $y_{\max}^{jk}$ , and the maximum data processing rate of processing node  $k$  by  $z_{\max}^k$ . The data rate limits are compactly expressed

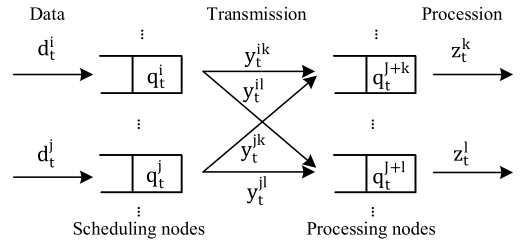


Fig. 1. An illustration of general online network resource allocation.

in the convex set as  $\mathcal{X}_0 \triangleq \{\mathbf{x} | \mathbf{0} \preceq \mathbf{x} \preceq \mathbf{x}_{\max}\}$ , where  $\mathbf{x}_{\max} = [y_{\max}^{11}, \dots, y_{\max}^{JK}, z_{\max}^1, \dots, z_{\max}^K]^T$  is the maximum data rate vector. Each scheduling node  $j$  and processing node  $k$  has a local data queue backlog at time  $t$  denoted by  $q_t^j$  and  $q_t^{j+k}$ , respectively. The queue backlog vector is  $\mathbf{q}_t = [q_t^1, \dots, q_t^J, q_t^{J+1}, \dots, q_t^{J+K}]^T$ , and we can express its update as  $\mathbf{q}_{t+1} = [\mathbf{q}_t + \mathbf{C}\mathbf{x}_t + \mathbf{d}_t]^+$ , where  $\mathbf{C} \in \mathbb{R}^{(J+K) \times (JK+K)}$  represents the network topology and is given by

$$\mathbf{C} = \left[ \begin{array}{c|c} \text{blkdiag}\{-\mathbf{1}_{1 \times K}, \dots, -\mathbf{1}_{1 \times K}\} & \mathbf{0}_{J \times K} \\ \hline \mathbf{I}_{K \times K}, \dots, \mathbf{I}_{K \times K} & -\mathbf{I}_{K \times K} \end{array} \right].$$

The goal for a network controller is to minimize the network cost while controlling the long-term averaged data outgoing rate to be no less than the incoming data rate for queue stability. However, since the controller can receive only delayed feedback on system parameters  $\mathbf{d}_t$ ,  $\mathbf{q}_t$ , and  $f_t(\mathbf{x})$  over time, it must employ an online solution. This online network resource allocation problem is a special case of the OCO problem **P1**, with the convex set  $\mathcal{X}_0$  define above, for any convex loss function  $f_t(\mathbf{x})$ , and the convex constraint function  $\mathbf{g}_t(\mathbf{x}) \triangleq \mathbf{C}\mathbf{x} + \mathbf{d}_t$ , which represents the difference between incoming and outgoing data at time  $t$ . Note that achieving sub-linear constraint violation, *i.e.*,  $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{g}_t(\mathbf{x}_t) \rightarrow \mathbf{0}$ , is equivalent to queue stability.

A special case of this problem where there is no delay in system information has been addressed with Lyapunov optimization techniques [18], [29], [30]. Furthermore, solutions under the standard OCO setting with one-slot feedback delay are given in [11] and [12]. However, due to the intermittence and crowdedness of the communication links between the nodes and central controller, it is likely that the central controller experiences multi-slot feedback delay of the system parameters. The proposed DTC-OCO algorithm provides a suitable online solution to this problem.

### B. Numerical Performance Evaluation

When applying DTC-OCO to the above online network resource allocation problem, the sublinear regret bounds derived in Section IV imply that it will provide an efficient solution. Furthermore, the sublinear constraint violation bound guarantees queue stability. In this section, we further study the numerical performance of DTC-OCO in a practical setting, compare its performance with [16], and verify the benefit of the proposed double regularization.

As a tangible example, we consider a mobile cloud computing system consisting of  $J = 10$  scheduling nodes, and



$K = 10$  processing nodes. Following the typical long-term evolution (LTE) specifications [31], we set the noise power spectral density  $N_0 = -174$  dBm/Hz, noise figure  $N_F = 10$  dB, and channel bandwidth  $B_W = 10$  MHz as default system parameters. We set the time slot duration to be 1 ms and assume  $d_t$  kB of data arrive at each time  $t$ . The maximum data transmission and processing rates, in MB/s, are randomly uniformly distributed as  $y_{\max}^{jk} \sim \mathcal{U}(10, 100)$  and  $z_{\max}^k \sim \mathcal{U}(100, 250)$ , respectively. According to the Shannon bound, we consider the transmission power of each link  $(j, k)$  as exponential w.r.t. its transmission rate  $y^{jk}$  given by  $\frac{\sigma_n^2}{L_t^{jk}} (2^{\frac{y^{jk}}{B_W}} - 1)$ , where  $\sigma_n^2[\text{dBm}] = N_0 B_W + N_F$  is the noise power and  $L_t^{jk}$  represents the integrated impact of path-loss, interference, and capacity gap, and it is time-varying due to the fluctuation of wireless channels. We assume each processing node  $k$  follows a quadratic power-frequency relationship given by  $\theta(\xi_t^k z_t^k)^2$ , where  $\theta = 120 \text{ W}/(\text{GHz})^2$  [32] and  $\xi_t^k$ , in cycles per byte, depends on the computational complexity of the computing tasks processed by node  $k$  at time  $t$  [33], [34] and thus is time varying [35]. We consider both the data transmission and processing power by defining the following cost function at time  $t$ :

$$f_t(\mathbf{x}) \triangleq \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} \frac{\sigma_n^2}{L_t^{jk}} (2^{\frac{y^{jk}}{B_W}} - 1) + \sum_{k \in \mathcal{K}} \theta(\xi_t^k z_t^k)^2.$$

We remark here that our proposed OCO solution can be applied to more general cost functions, as long as they are convex w.r.t. the decision variables  $\mathbf{x}$ .

For performance comparison, we consider the online algorithm proposed in [16], which is the known best time-varying constrained OCO algorithm that accommodates multi-slot feedback delay. We also consider the dynamic benchmark in (4). To verify the performance advantage brought by the double regularization in DTC-OCO, we also study two simplified versions of DTC-OCO with single regularization on  $\mathbf{x}_{t-\tau}$  and  $\mathbf{x}_{t-1}$ , respectively.

We assume both  $\tau$  and  $\delta$  are unknown, and thus set  $\alpha = T^{\frac{1}{2}}$ ,  $\eta = \|\mathbf{C}\|_2^2$ , and  $\gamma = 1$  in DTC-OCO (see Remark 1). Our performance metrics are the time-averaged network cost  $\bar{f}(T) \triangleq \frac{1}{T} \sum_{t=1}^T f_t(\mathbf{x}_t)$  and the time-averaged constraint violation  $\bar{g}(T) \triangleq \frac{1}{T} \sum_{t=1}^T \mathbf{1}^T \mathbf{g}_t(\mathbf{x}_t)$ . For fair comparison on  $\bar{f}(T)$ , the step-sizes in [16] are selected such that the algorithm has a similar steady state value of  $\bar{g}(T)$  as DTC-OCO. The time-varying system parameters  $\{d_t^j\}$ ,  $\{L_t^{jk}\}$ ,  $\{\xi_t^k\}$  are modeled by the following two cases.

1) *I.i.d. parameters.* The system parameters are i.i.d.:  $d_t^j \sim \mathcal{U}(10, 100)$ ,  $L_t^{jk}[\text{dB}] \sim \mathcal{U}(-126, -120)$ , and  $\xi_t^k \sim \mathcal{U}(1, 3)$ .

2) *Noisy periodic parameters.* The system parameters vary periodically with noise:  $d_t^j = 30 \sin(\frac{\pi t}{20}) + n_t^{j,d}$ ,  $L_t^{jk}[\text{dB}] = -120 - 3 \sin(\frac{\pi t}{20}) - n_t^{j,k,L}$ , and  $\xi_t^k = 0.5 \sin(\frac{\pi t}{20}) + n_t^{k,\xi}$ , where  $n_t^{j,d} \sim \mathcal{U}(40, 70)$ ,  $n_t^{j,k,L} \sim \mathcal{U}(6, 9)$ , and  $n_t^{k,\xi} \sim \mathcal{U}(1, 3)$ .

Fig. 2 shows  $\bar{f}(T)$  and  $\bar{g}(T)$  versus  $T$  with different values of  $\tau$  for the above two cases. We observe that the network cost of DTC-OCO can approach that of the dynamic benchmark in (4), indicating that sublinear dynamic regret

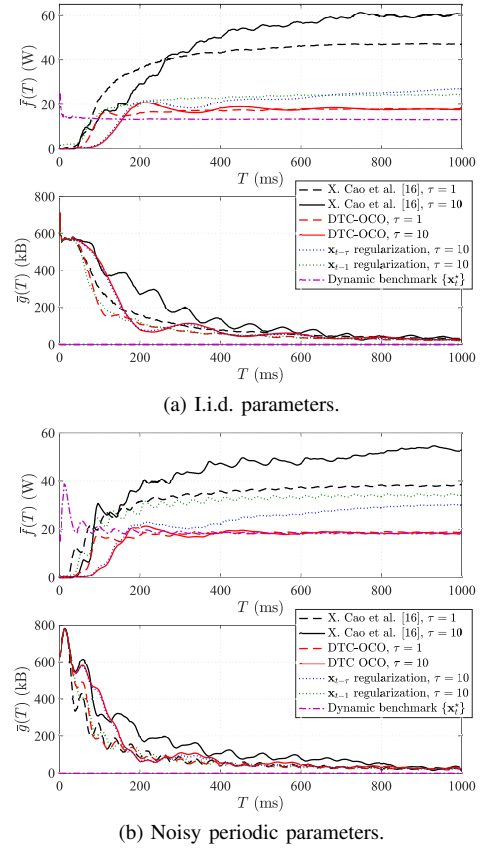


Fig. 2.  $\bar{f}(T)$  and  $\bar{g}(T)$  vs.  $T$  under different  $\tau$ .

is achieved. Compared with [16], DTC-OCO achieves lower network cost and is much more tolerant to the feedback delay  $\tau$ . The reason for this are two fold: first, DTC-OCO penalizes the constraint function  $\mathbf{g}_t(\mathbf{x})$  directly instead of its first-order approximation, which improves the control on constraint violation; second, the double regularization in DTC-OCO allows the online decision  $\mathbf{x}_t$  to be updated from  $\mathbf{x}_{t-\tau}$  and  $\mathbf{x}_{t-1}$ , which improves the efficacy of gradient descent for loss minimization. Comparison with the  $\mathbf{x}_{t-\tau}$  and  $\mathbf{x}_{t-1}$  single-regularization approaches further suggests that the proposed method of double regularization is essential to the performance advantage of DTC-OCO.

## VI. CONCLUSIONS

This paper considers OCO with multi-slot feedback delay and short-term and long-term constraints. We propose an efficient algorithm DTC-OCO, which uses a novel constraint penalty with double regularization to handle the asynchrony between information feedback and decision updates. Our analysis considers the impact of multi-slot feedback delay and the double regularization structure on the performance guarantees of DTC-OCO, to show sublinear dynamic and static regrets and sublinear constraint violation under mild conditions. We apply DTC-OCO to a general online network resource allocation problem, using mobile cloud computing for numerical example. Simulation results demonstrate the superior delay tolerance and substantial performance advantage of DTC-OCO over the known best alternative under different scenarios.

## REFERENCES

- [1] S. Shalev-Shwartz, "Online learning and online convex optimization," *Found. Trends Mach. Learn.*, vol. 4, pp. 107–194, Feb. 2012.
- [2] E. Hazan, "Introduction to online convex optimization," *Found. Trends Optim.*, vol. 2, pp. 157–325, Aug. 2016.
- [3] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proc. Intel. Conf. Mach. Learn. (ICML)*, 2003.
- [4] E. Hazan, A. Agarwal, and S. Kale, "Logarithmic regret algorithms for online convex optimization," *Mach. Learn.*, vol. 69, pp. 169–192, 2007.
- [5] E. C. Hall and R. M. Willett, "Online convex optimization in dynamic environments," *IEEE J. Sel. Topics Signal Process.*, vol. 9, pp. 647–662, Jun. 2015.
- [6] A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan, "Online optimization : Competing with dynamic comparators," in *Proc. Intel. Conf. Artif. Intell. Statist. (AISTATS)*, vol. 38, 2015.
- [7] M. Mahdavi, R. Jin, and T. Yang, "Trading regret for efficiency: Online convex optimization with long term constraints," *J. Mach. Learn. Res.*, vol. 13, pp. 2503–2528, Sep. 2012.
- [8] R. Jenatton, J. Huang, and C. Archambeau, "Adaptive algorithms for online convex optimization with long-term constraints," in *Proc. Intel. Conf. Mach. Learn. (ICML)*, 2016.
- [9] H. Yu and M. J. Neely, "A low complexity algorithm with  $O(\sqrt{T})$  regret and  $O(1)$  constraint violations for online convex optimization with long term constraints," *J. Mach. Learn. Res.*, vol. 21, pp. 1–24, Feb. 2020.
- [10] H. Yu, M. J. Neely, and X. Wei, "Online convex optimization with stochastic constraints," in *Proc. Adv. Neural Info. Proc. Sys. (NIPS)*, 2017.
- [11] T. Chen, Q. Ling, and G. B. Giannakis, "An online convex optimization approach to proactive network resource allocation," *IEEE Trans. Signal Process.*, vol. 65, pp. 6350–6364, Dec. 2017.
- [12] X. Cao, J. Zhang, and H. V. Poor, "A virtual-queue-based algorithm for constrained online convex optimization with applications to data center resource allocation," *IEEE J. Sel. Topics Signal Process.*, vol. 12, pp. 703–716, Aug. 2018.
- [13] J. Langford, A. J. Smola, and M. Zinkevich, "Slow learners are fast," in *Proc. Adv. Neural Info. Proc. Sys. (NIPS)*, 2009.
- [14] K. Quanrud and D. Khashabi, "Online learning with adversarial delays," in *Proc. Adv. Neural Info. Proc. Sys. (NIPS)*, 2015.
- [15] H. B. McMahan and M. Streeter, "Delay-tolerant algorithms for asynchronous distributed online learning," in *Proc. Adv. Neural Info. Proc. Sys. (NIPS)*, 2014.
- [16] X. Cao, J. Zhang, and H. V. Poor, "Impact of delays on constrained online convex optimization," in *Proc. Asilomar Conf. Signal Sys. Comput.*, Nov. 2019.
- [17] O. Besbes, Y. Gur, and A. Zeevi, "Non-stationary stochastic optimization," *Oper. Res.*, vol. 63, pp. 1227–1244, Sep. 2015.
- [18] M. J. Neely, *Stochastic Network Optimization with Application on Communication and Queueing Systems*. Morgan & Claypool, 2010.
- [19] M. Lotfinezhad, B. Liang, and E. S. Sousa, "Optimal control of constrained cognitive radio networks with dynamic population size," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, 2010.
- [20] H. Yu and M. J. Neely, "Dynamic transmit covariance design in MIMO fading systems with unknown channel distributions and inaccurate channel state information," *IEEE Trans. Wireless Commun.*, vol. 16, pp. 3996–4008, Jun. 2017.
- [21] J. Wang, M. Dong, B. Liang, and G. Boudreau, "Online precoding design for downlink MIMO wireless network virtualization with imperfect CSI," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, 2020.
- [22] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Trans. Automat. Contr.*, vol. 57, pp. 592–606, Mar. 2012.
- [23] M. Akbari, B. Ghahesifard, and T. Linder, "Distributed online convex optimization on time-varying directed graphs," *IEEE Trans. Control Netw. Syst.*, vol. 4, pp. 417–428, Sep. 2017.
- [24] S. Shahrampour and A. Jadbabaie, "Distributed online optimization in dynamic environments using mirror descent," *IEEE Trans. Automat. Contr.*, vol. 63, pp. 714–725, Mar. 2018.
- [25] N. Eshraghi and B. Liang, "Distributed online optimization over a heterogeneous network with any-batch mirror descent," in *Proc. Intel. Conf. Mach. Learn. (ICML)*, 2020.
- [26] X. Yi, X. Li, L. Xie, and K. H. Johansson, "Distributed online convex optimization with time-varying coupled inequality constraints," *IEEE Trans. Signal Process.*, vol. 68, pp. 731–746, Jan. 2020.
- [27] S. Mannor, J. N. Tsitsiklis, and J. Y. Yu, "Online learning with sample path constraints," *J. Mach. Learn. Res.*, vol. 10, pp. 569–590, Mar. 2009.
- [28] K.-S. Jun, F. Orabona, S. Wright, and R. Willett, "Improved strongly adaptive online learning using coin betting," in *Proc. Intel. Conf. Artif. Intell. Statist. (AISTATS)*, 2017.
- [29] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Found. Trends Netw.*, pp. 1–144, 2006.
- [30] T. Chen, A. Mokhtari, X. Wang, A. Ribeiro, and G. B. Giannakis, "Stochastic averaging for constrained optimization with application to online resource allocation," *IEEE Trans. Signal Process.*, vol. 65, pp. 3078–3093, Jun. 2017.
- [31] H. Holma and A. Toskala, *WCDMA for UMTS - HSPA evolution and LTE*. John Wiley & Sons, 2010.
- [32] R. Uргаonkar, U. C. Kozat, K. Igarashi, and M. J. Neely, "Dynamic resource allocation and power management in virtualized data centers," in *IEEE/IFIP Netw. Oper. Manage. Symp. (NOMS)*, 2010.
- [33] H. Topcuoglu, S. Hariri, and Min-You Wu, "Performance-effective and low-complexity task scheduling for heterogeneous computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 13, pp. 260–274, 2002.
- [34] S. Sundar and B. Liang, "Offloading dependent tasks with communication delay and deadline constraint," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, 2018.
- [35] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 4569–4581, 2013.