Delay-Tolerant OCO with Long-Term Constraints: Algorithm and Its Application to Network Resource Allocation

Juncheng Wang, *Student Member, IEEE*, Min Dong, *Senior Member, IEEE*, Ben Liang, *Fellow, IEEE*, Gary Boudreau, *Senior Member, IEEE*, and Hatem Abou-Zeid, *Member, IEEE*

***Abstract*—We consider online convex optimization (OCO) with multi-slot feedback delay. An agent selects a sequence of online decisions to minimize the accumulation of time-varying convex loss functions, subject to short-term and long-term constraints that may be time-varying. Both the convex loss function and the long-term constraint function may experience multiple time slots of feedback delay to be received by the agent. Existing works on OCO under this general setting has focused on the static regret, which measures the gap of losses between an online decision sequence and a time-invariant static offline benchmark. In this work, besides the static regret, we also consider a more practically meaningful metric, the dynamic regret, where the benchmark is a time-varying online optimal decision sequence. We propose an efficient algorithm, termed Delay-Tolerant Constrained-OCO (DTC-OCO), which uses a novel double regularization together with a new penalty mechanism on the long-term constraint violation, to tackle the asynchrony between information feedback and decision updates. We obtain upper bounds for its static regret, dynamic regret, and constraint violation, proving that they are sublinear under mild conditions. Furthermore, we consider a variation of DTC-OCO with multi-step gradient descent, and show it provides improved dynamic regret and constraint violation bounds for strongly convex loss functions. For numerical demonstration, we apply DTC-OCO to a general network resource allocation problem. Our simulation results suggest substantial performance gain by DTC-OCO over the current best alternative.**

***Index Terms*—Online convex optimization, long-term constraint, multi-slot delay, dynamic regret, constraint violation, online network resource allocation.**

I. Introduction

Online convex optimization (OCO) is a promising solution to many system control, machine learning, and resource allocation problems, such as prediction with expert advice, spam filtering, target tracking, online regression, and network routing [2], [3]. Under the standard OCO setup, an agent selects a decision from a known convex set at the beginning of each time slot. At the end of each slot, the system reveals information of the current convex loss function to the agent. 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 time slot. Instead, the agent aims at minimizing the *regret*, *i.e.,* the accumulated performance gap between the online decision sequence and some performance benchmark over time.

Most of the early works on OCO studied the *static* regret, which is the performance gap between the online decision sequence and a static offline benchmark that is based on apriori information of all the convex loss functions over the entire time horizon. In the seminal work of OCO [4], 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 [5] for strongly convex loss functions. However, in a dynamic environment, the performance of the static offline benchmark may be far from optimal. As a result, achieving sublinear static regret may not be meaningful. In [4], a more useful metric, the *dynamic* regret, was introduced to measure the difference between the online decision sequence and a dynamic benchmark. Even the dynamic regret is more difficult to analyze, it has received increasing attention in recent works [6]-[11].

The above-mentioned works all focused on OCO with short-term constraints that must be strictly satisfied at each time slot. *Long-term* constraints are also common in many applications. For example, the energy budget of an electronic device may be viewed as a long-term constraint on its power usage. OCO with long-term constraints was first considered in [12]. With such constraints, constraint violation within a finite time period may occur. Thus, in addition to achieving sublinear regret, OCO algorithms should also provide sublinear *constraint violation*, which indicates that the time-averaged violation of each long-term constraint tends to zero as time approaches infinity. Early works on constrained OCO assumed that the long-term constraints are time-invariant [13], [14], while more recent works [15]-[18] 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 due to limited feedback resources. In machine learning, collecting training datasets and transmitting the learning models over wireless links may induce feedback delays from the mobile devices to the parameter server. In smart grid, the large number of renewable energy sources can

Juncheng Wang and Ben Liang are with the Department of Electrical and Computer Engineering, University of Toronto, ON M5S 1A1, Canada (e-mail: jcwang@ece.utoronto.ca; liang@ece.utoronto.ca).

Min Dong is with the Department of Electrical, Computer, and Software Engineering, Ontario Tech University, Oshawa, ON L1G 0C5, Canada (e-mail: min.dong@ontariotechu.ca).

Gary Boudreau is with Ericsson Canada, Ottawa, ON K2K 2V6, Canada (e-mail: gary.boudreau@ericsson.com).

Hatem Abou-Zeid was with Ericsson Canada and is now with the Department of Electrical and Software Engineering, University of Calgary, AB T2N 1N4, Canada (e-mail: hatem.abouzeid@ucalgary.ca).

This work was supported in part by Ericsson Canada, the Natural Sciences and Engineering Research Council of Canada, and the Ontario Centre of Innovation.

A preliminary version of this work was presented in IEEE International Conference on Computer Communications (INFOCOM), 2021 [1] (DOI: 10.1109/INFOCOM42981.2021.9488698).

lead to feedback delays on the amount of power supply at the central controller. In mobile computing, offloading tasks from remote devices to the cloud via wireless channels can cause feedback delay. 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 time slot when the decision is made, *i.e.,* the feedback information is delayed for only one slot [4]-[18]. However, such assumption is too restrictive for many practical applications. To address this, [19] initiated a study on OCO with multi-slot feedback delay. Additional delay-adaptive OCO algorithms were proposed in [20] and [21]. A recent work [22] considered OCO with both time-varying constraints and multi-slot feedback delay.

Despite the above research efforts, to the best of our knowledge, all existing works on OCO with multi-slot feedback delay focus on the static regret, which as explained above may not be a meaningful performance metric for inherently time-varying systems. In fact, the gap between the static regret and dynamic regret can be as large as $\mathcal{O}(T)$ [23]. In the presence of multi-slot feedback delay, whether sublinear dynamic regret is achievable for OCO is an open problem. Adding to this challenge is our limited understanding on system performance under long-term constraints (either time-varying or time-invariant). In this work, we aim to address these challenges. Our main contributions are as follows:

- We propose an efficient algorithm, termed Delay-Tolerant Constrained-OCO (DTC-OCO), for OCO with multi-slot feedback delay and with both short-term and long-term constraints. All existing OCO algorithms update current decisions using either no regularization or a single-regularization approach based on either the one-slot ahead decision or τ-slot ahead decision, where τ is the feedback delay. In contrast, in DTC-OCO, we propose a novel double-regularization approach to update current decisions based on both one-slot ahead decision and τ-slot ahead decision, to capture useful information from both decisions to further minimize the accumulated loss and constraint violation. The double-regularization approach together with a new penalty mechanism on the long-term constraint violation improves the performance tolerance to multi-slot delay and facilitates the bounding of the performance by DTC-OCO.
- We analyze the special structure of DTC-OCO under the double regularization approach and show that the algorithm achieves $\mathcal{O}(\max\{\tau^{\frac{1}{2}}T^{\frac{1+\delta}{2}}, T^{\nu}\})$ dynamic regret and $\mathcal{O}(\max\{\tau^{\frac{1}{2}}T^{\frac{1}{2}}, T^{\nu}\})$ static regret, where δ represents the growth rate of the accumulated variation of the per-slot optimizer, and ν measures the accumulated squared variation of the constraint functions. Furthermore, we show $\mathcal{O}(\max\{T^{\frac{1-\delta}{2}}, \tau T^{\kappa}\})$ constraint violation bound for DTC-OCO, where κ measures the accumulated variation of the constraint functions. In the special case of time-invariant constraints, DTC-OCO achieves $\mathcal{O}(\tau^{\frac{1}{2}}T^{\frac{1+\delta}{2}})$ dynamic regret, $\mathcal{O}(\tau^{\frac{1}{2}}T^{\frac{1}{2}})$ static regret, and $\mathcal{O}(\tau)$ constraint violation. To the best of our knowledge, this is the first work to simultaneously provide dynamic regret bound and constraint violation bound for OCO with both multi-slot delay and long-term constraints.
- To further improve the performance of DTC-OCO, we propose a variant with multi-step gradient descent. Our analysis shows that, for strongly convex loss functions, when the number of gradient descent steps at each slot is large enough, DTC-OCO provides improved $\mathcal{O}(\max\{\tau^2T^{\delta}, T^{\nu}\})$ dynamic regret bound and $\mathcal{O}(\tau T^{\kappa})$ constraint violation bound. For the special case of time-invariant constraints, the algorithm achieves $\mathcal{O}(\tau^2T^{\delta})$ dynamic regret and $\mathcal{O}(\tau)$ constraint violation. We note that, in prior works, even under the standard setting with one-slot feedback delay, it was unknown whether or not strong convexity helps improve the dynamic regret of OCO with long-term constraints.
- As an application example, we apply DTC-OCO to a general network resource allocation problem. For a specific cloud computing system, our simulation demonstrates that DTC-OCO is more tolerant to feedback delay and achieves much smaller accumulated loss compared with the current best alternative from [22]. Our simulation also shows that enabling multi-step gradient descent in DTC-OCO further reduces the accumulated loss.

Organizations: 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 and its performance analysis in Section IV. Then, we consider multi-step gradient descent for DTC-OCO and study its performance in Section V. The application of DTC-OCO to network resource allocation is presented in Section VI, followed by concluding remarks in Section VII.

Notations: The transpose, Euclidean norm, L_∞ norm, and L_1 norm of a vector $\mathbf{a}$ are denoted by $\mathbf{a}^T$, $\|\mathbf{a}\|$, $\|\mathbf{a}\|_\infty$, and $\|\mathbf{a}\|_1$, respectively. The notation $\mathbf{I}$ denotes an identity matrix, $\mathbf{0}$ denotes a vector of all 0's, $\mathbf{1}$ denotes a vector of all 1's, and $[N]$ denotes the set $\{1, \ldots, N\}$.

II. Related Work

In this section, we survey existing works on OCO. The differences between these works and our work are summarized in Table. I.

1) OCO with Long-Term Constraints: Among the existing works on OCO with long-term constraints, a saddle-point-typed algorithm was proposed in [12], which achieves $\mathcal{O}(T^{\frac{1}{2}})$ static regret and $\mathcal{O}(T^{\frac{3}{4}})$ constraint violation for time-invariant long-term constraints. A follow-up work [13] provided $\mathcal{O}(T^{\max\{\xi,1-\xi\}})$ static regret and $\mathcal{O}(T^{1-\frac{\xi}{2}})$ constraint violation, where $\xi \in (0,1)$ is a trade-off parameter. A virtual-queue-based algorithm was proposed in [14], which provides $\mathcal{O}(1)$ constraint violation that is currently the best result for OCO with time-invariant long-term constraints. For independent and identically distributed (i.i.d.) long-term constraints, virtual-queue-based algorithms were proposed in [15] and [16] with the standard gradient descent update and general mirror descent update, respectively. The analyses in [12]-[16] focus on the static regret. Dynamic regret bounds were provided in [17] and [18], by modifying the the saddle-point-typed

TABLE I
SUMMARY OF RELATED WORKS ON OCO

Reference	Type of benchmark	Long-term constraint	Multi-slot delay
[2], [3], [5]	Static	No	No
[4]	Static and dynamic	No	No
[6]-[11]	Dynamic	No	No
[12]-[14]	Static	Invariant	No
[15], [16]	Static	Varying	No
[17], [18]	Dynamic	Varying	No
[19]-[21]	Static	No	Yes
[22]	Static	Varying	Yes
DTC-OCO	Static and dynamic	Invariant and varying	Yes

and virtual-queue-based algorithms respectively to deal with general time-varying long-term constraints. The above works all consider the standard OCO setting where the feedback information is delayed for only one slot, while our work considers the more challenging problem of multi-slot feedback delay.

2) OCO with Multi-Slot Feedback Delay: Most existing works on OCO with multi-slot feedback delay focus on online decision design problems with only short-term constraints [19]-[21]. In [19], the standard online gradient descent algorithm [4] was extended to provide $\mathcal{O}(\tau^{\frac{1}{2}}T^{\frac{1}{2}})$ static regret. Delay-adaptive online gradient descent algorithms were proposed in [20] and [21] to accommodate adversarial feedback delay.

The impact of long-term constraints on OCO with multi-slot feedback delay was considered in [22]. However, [22] only studied the static regret. Furthermore, the constraint violation bound provided in [22] was no less than $\mathcal{O}(T^{\frac{3}{4}})$ even for time-invariant long-term constraints. This constraint violation performance is inherited from the saddle-point-typed algorithm [12]. Thus, even for time-invariant long-term constraints, whether an online algorithm can achieve a tighter constraint violation bound in the presence of multi-slot delay is unknown. Different from [22], in this work, we propose a novel double-regularization approach together with a new penalty mechanism on the long-term constraint violation. Furthermore, we provide both dynamic and static regret bounds, as well as a stronger constraint violation bound for the proposed algorithm.

3) Lyapunov Optimization: Constrained OCO is related to Lyapunov optimization [24]. The latter uses the system state and queueing information to implicitly learn the system variations and adapt the online decisions accordingly without knowing the system statistics. However, under the Lyapunov optimization framework, the system states are commonly assumed to be i.i.d. or Markovian, while OCO frameworks do not have such restriction. Furthermore, the 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 using historical information to predict the current system state [25]. However, this way of dealing with feedback delay is equivalent to extending the standard Lyapunov optimization to inaccurate system states in [26] and [27]. As a result, with inaccurate system state information, the performance gap between the online decisions and the optimal decisions over time is $\mathcal{O}(\sigma T)$, which grows linearly with T with σ being some measure of system inaccuracy. Therefore, such an approach cannot lead to the sublinear dynamic regret bound that we seek.

III. CONSTRAINED OCO WITH MULTI-SLOT DELAY

A. Problem Formulation

We consider a time-slotted system with time slots indexed by t. Let $f_t(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ be a convex loss function at time slot t. Let $\mathbf{g}_t(\mathbf{x})$ be a vector of C convex long-term constraint functions at time slot t, where $\mathbf{g}_t(\mathbf{x}) = [g_t^1(\mathbf{x}), \ldots, g_t^C(\mathbf{x})]^T : \mathbb{R}^n \to \mathbb{R}^C$. The loss function $f_t(\mathbf{x})$ and the constraint function $\mathbf{g}_t(\mathbf{x})$ both may change over t. 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 $\mathcal{X}_0$ to minimize the accumulated loss while also meeting the long-term constraints, which expressed as the following dynamic optimization problem:

$$\mathbf{P1}: \quad \min_{\{\mathbf{x}_t\}} \quad \sum_{t=1}^{T} f_t(\mathbf{x}_t)$$

$$\text{s.t.} \quad \sum_{t=1}^{T} \mathbf{g}_t(\mathbf{x}_t) \preceq \mathbf{0}, \tag{1}$$

$$\mathbf{x}_t \in \mathcal{X}_0, \quad \forall t \tag{2}$$

where T is the time horizon. 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 [12]-[14].

Under the standard constrained OCO setting [12]-[18], feedback information on the loss function $f_t(\mathbf{x})$ and the long-term constraint function $\mathbf{g}_t(\mathbf{x})$ is assumed to only have one time slot delay, and the feedback can be used to make the new decision $\mathbf{x}_{t+1}$ for the next time slot.[1] However, in many practical applications, such as wireless transmission and mobile computing mentioned in Section I, this one-slot delay assumption is unrealistic as the feedback information typically may experience a severe delay.

Therefore, in this work, we consider a general scenario where the feedback information on $f_t(\mathbf{x})$ and $\mathbf{g}_t(\mathbf{x})$ is delayed by $\tau \geq 1$ slots to arrive at the decision maker at the end

[1]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 [12]-[14].

of slot $t+\tau-1$. The multi-slot feedback delay has also been considered in [19]-[22]. Different from [19]-[21], where only short-term constraints are considered, the additional long-term constraints in (1) lead to a more complicated online optimization problem as the decisions $\{\mathbf{x}_t\}$ are correlated over time. The problem is especially more challenging as the underlying system state varies over time while decisions need to be made based on the delayed feedback.

B. Performance Metrics

Due to the lack of in-time information of the current loss and constraint functions under the OCO setting, it is very difficult, if not impossible, to obtain an optimal solution to **P1**.[2] Instead, a time-varying constrained OCO algorithm aims at selecting a sequence of online decisions $\{\mathbf{x}_t\}$ that is asymptotically no worse than some performance benchmarks.

One common static benchmark is given by

$$\mathbf{x}^\star \in \arg\min_{\mathbf{x}\in\mathcal{X}_0}\left\{\sum_{t=1}^{T} f_t(\mathbf{x}) | \mathbf{g}_t(\mathbf{x}) \preceq \mathbf{0}, \forall t\right\}, \tag{3}$$

where the decision $\mathbf{x}^\star$ is computed assuming all information of $\{f_t(\mathbf{x})\}$ and $\{\mathbf{g}_t(\mathbf{x})\}$ within T time slots is known in advance.[3] The performance gap between $\{\mathbf{x}_t\}$ and $\mathbf{x}^\star$ is referred to as the *static regret*, given by

$$\mathrm{RE_s}(T) \triangleq \sum_{t=1}^{T}\left(f_t(\mathbf{x}_t) - f_t(\mathbf{x}^\star)\right). \tag{4}$$

This static regret was adopted in [22], while [12]-[14] used a special case of it when time-invariant constraints are assumed. However, as a rather coarse performance metric, the static regret may not be a strong indicator of the actual performance of an algorithm, especially when the underlying system is inherently time-varying.

A more attractive performance benchmark for time-varying constrained OCO is the dynamic benchmark $\{\mathbf{x}_t^\star\}$, given by[4]

$$\mathbf{x}_t^\star \in \arg\min_{\mathbf{x}\in\mathcal{X}_0}\{f_t(\mathbf{x}) | \mathbf{g}_t(\mathbf{x}) \preceq \mathbf{0}\}. \tag{5}$$

In this case, the decision $\mathbf{x}_t^\star$ 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 [4] and later was modified in [17] and [18] to incorporate long-term constraints. The corresponding performance gap, referred to as *dynamic regret*, is defined by

$$\mathrm{RE_d}(T) \triangleq \sum_{t=1}^{T}\left(f_t(\mathbf{x}_t) - f_t(\mathbf{x}_t^\star)\right). \tag{6}$$

[2]In fact, even for the most basic OCO problem [4], *i.e.,* without long-term constraints (1), an optimal solution cannot be found [5].

[3]The static benchmark $\mathbf{x}^\star$ in (3) satisfies $\mathbf{g}_t(\mathbf{x}^\star) \preceq \mathbf{0}$ at each time slot t. One may define the static benchmark as $\mathbf{x}^\circ \in \arg\min_{\mathbf{x}\in\mathcal{X}_0}\{\sum_{t=1}^{T} f_t(\mathbf{x}) | \sum_{t=1}^{T}\mathbf{g}_t(\mathbf{x}) \preceq \mathbf{0}\}$, which satisfies the long-term constraints. However, even with one-slot feedback delay, [28] showed via a counterexample that it is impossible to achieve sublinear static regret $\sum_{t=1}^{T}\left(f_t(\mathbf{x}_t) - f_t(\mathbf{x}^\circ)\right)$ and sublinear constraint violations in (7) simultaneously.

[4]Similar to the discussion of static benchmark in Footnote 3, for dynamic benchmark $\mathbf{x}_t^\circ \in \arg\min_{\mathbf{x}\in\mathcal{X}_0}\{f_t(\mathbf{x}) | \sum_{t=1}^{T}\mathbf{g}_t(\mathbf{x}) \preceq \mathbf{0}\}$, it is also impossible to achieve sublinear dynamic regret $\sum_{t=1}^{T}\left(f_t(\mathbf{x}_t) - f_t(\mathbf{x}_t^\circ)\right)$ and sublinear constraint violations simultaneously.

The dynamic regret provides a more accurate measure of performance. In some cases, the gap between $\mathrm{RE_s}(T)$ and $\mathrm{RE_d}(T)$ can be as large as $\mathcal{O}(T)$ [23]. In this work, for a comprehensive performance study, we provide upper bounds on both $\mathrm{RE_s}(T)$ and $\mathrm{RE_d}(T)$.

To measure the accumulated violation of the long-term constraints, the *constraint violation*,[5] for any $c \in [C]$, is defined as in [18], [22]:

$$\mathrm{VO}^c(T) \triangleq \sum_{t=1}^{T} g_t^c(\mathbf{x}_t). \tag{7}$$

Note that the constraint violation for time-invariant constraint function $\mathbf{g}(\mathbf{x})$ defined in [12]-[14] is a special case of (7). With (7), our study accommodates both time-varying and time-invariant constraints.

It is desirable to design a constrained OCO algorithm that can provide both sublinear regrets, *i.e.,* $\mathrm{RE_d}(T) = \mathbf{o}(T)$ and $\mathrm{RE_s}(T) = \mathbf{o}(T)$, and sublinear constraint violation, *i.e.,* $\mathrm{VO}^c(T) = \mathbf{o}(T)$. Sublinearity in regret and constraint violation is important; it implies that the online decision is asymptotically no worse than the corresponding benchmark in terms of its time-averaged performance, and at the same time, the long-term constraints are satisfied.

IV. Delay-Tolerant Constrained OCO

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

A. DTC-OCO Algorithm

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

$$Q_t^c = \max\left\{-\gamma g_{t-\tau}^c(\mathbf{x}_t), Q_{t-1}^c + \gamma g_{t-\tau}^c(\mathbf{x}_t)\right\} \tag{8}$$

where $\gamma > 0$ is a weighting factor on the constraint violation that controls how fast the virtual queue varies over time. The role of $\mathbf{Q}_t$ is similar to a Lagrange multiplier vector associated with constraints in (1) of **P1** or a backlog queue for the constraint violation that is used in [14]-[16], [18]. However, unique to our proposed approach, $g_{t-\tau}^c(\mathbf{x}_t)$ is the *τ-slot delayed* constraint violation caused by the *current* decision; also, it needs to be scaled by an appropriate γ factor.

In the basic form of DTC-OCO, we convert **P1** into solving a per-slot problem at each slot $t > \tau$, with short-term constraints only, given by

$$\mathbf{P2}: \quad \min_{\mathbf{x}\in\mathcal{X}_0} \quad [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T(\mathbf{x} - \mathbf{x}_{t-\tau})$$

[5]The constraint violation is referred to as dynamic fit $\mathrm{Fit}(T) \triangleq \|[\sum_{t=1}^{T}\mathbf{g}_t(\mathbf{x}_t)]^+\|$ in [17], where $[\mathbf{x}]^+ \triangleq \max\{\mathbf{x}, \mathbf{0}\}$ is the entry-wise positive projection operator. One can easily verify that the sublinearity of $\mathrm{VO}^c(T), \forall c \in [C]$ implies $\mathrm{Fit}(T)$ being sublinear, and vice versa.

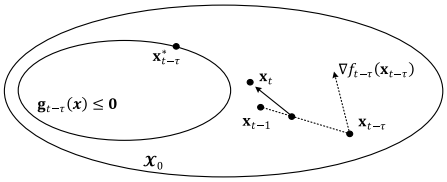


Fig. 1. An illustration of the update of $\mathbf{x}_t$ by DTC-OCO using the proposed double regularization. At each time t, as the result of solving **P2** based on $\mathbf{x}_{t-\tau}$ and $\mathbf{x}_{t-1}$, $\mathbf{x}_t$ moves from a point on the line segment between $\mathbf{x}_{t-\tau}$ and $\mathbf{x}_{t-1}$ towards $\mathbf{x}^\star_{t-\tau}$ to track the dynamic benchmark $\{\mathbf{x}^\star_t\}$.

$$+ [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T [\gamma \mathbf{g}_{t-\tau}(\mathbf{x})] + \alpha \|\mathbf{x} - \mathbf{x}_{t-\tau}\|^2 + \eta \|\mathbf{x} - \mathbf{x}_{t-1}\|^2$$

where $\alpha, \eta > 0$ are two step-size parameters that control the weights on the two regularization terms. Note that **P2** is a convex optimization problem and therefore can be solved efficiently using existing optimization tools.

As seen from the first term of the objective in **P2**, DTC-OCO uses the τ-slot delayed gradient $\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})$ for controlling $f_{t-\tau}(\mathbf{x}_t)$ to minimize the accumulated loss objective in **P1**. Compared with the original **P1**, the long-term constraints in (1) are converted into a penalty term for controlling $\mathbf{g}_{t-\tau}(\mathbf{x}_t)$ to maintain queue stability as shown in the second term of the objective in **P2**. Minimizing these two terms is equivalent to letting the new decision $\mathbf{x}_t$ minimize $f_{t-\tau}(\mathbf{x})$ while satisfying $\mathbf{g}_{t-\tau}(\mathbf{x}) \preceq \mathbf{0}$. In other words, the new decision $\mathbf{x}_t$ tries to track the dynamic benchmark $\mathbf{x}^\star_{t-\tau}$. Note that due to the intractability of original **P1** with τ-slot delay, our goal is to track the dynamic benchmark $\{\mathbf{x}^\star_t\}$ for dynamic regret minimization over time. However, due to τ-slot delay, we can only track $\mathbf{x}^\star_{t-\tau}$ instead at each time t. Furthermore, for the third and fourth terms of the objective in **P2**, we use a novel constraint penalty term containing double regularization $\alpha\|\mathbf{x} - \mathbf{x}_{t-\tau}\|^2$ and $\eta\|\mathbf{x} - \mathbf{x}_{t-1}\|^2$ to handle the asynchrony between information feedback and decision updates. Compared with the single regularization $\alpha\|\mathbf{x} - \mathbf{x}_{t-\tau}\|^2$ (or $\eta\|\mathbf{x} - \mathbf{x}_{t-1}\|^2$), this double regularization shifts the starting point of the decision update from $\mathbf{x}_{t-\tau}$ (or $\mathbf{x}_{t-1}$) to a point between $\mathbf{x}_{t-\tau}$ and $\mathbf{x}_{-1}$. The intuition behind the double regularization is that both $\mathbf{x}_{t-\tau}$ and $\mathbf{x}_{t-1}$ provide useful information in minimizing the accumulated loss and constraint violation. Thus, it is desirable for the new decision $\mathbf{x}_t$ to be not too far away from either $\mathbf{x}_{t-\tau}$ or $\mathbf{x}_{t-1}$.

An illustrative example is shown in Fig. 1: With the proposed double regularization, $\mathbf{x}_t$ moves from a point on the line segment between $\mathbf{x}_{t-\tau}$ and $\mathbf{x}_{t-1}$ towards $\mathbf{x}^\star_{t-\tau}$. This is in contrast to existing works that update $\mathbf{x}_t$ from either $\mathbf{x}_{t-\tau}$ [19]-[21] or $\mathbf{x}_{t-1}$ [22]. We will show both analytically in Sections IV-B and IV-C and numerically in Section VI, that the double regularization provides DTC-OCO a substantial performance advantage over existing algorithms in terms of regret bounds and average performance.

Thus, DTC-OCO consists of three major steps: 1) Initialize $\mathbf{x}_t \in \mathcal{X}_0, \forall t \in [\tau]$, and set $\mathbf{Q}_t = \mathbf{0}, \forall t \in [\tau]$ 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 **P2**; 3) At the end of each slot $t > \tau$, update the virtual queue $\mathbf{Q}_t$ via (8).[6] Note that when τ is unknown, in addition to $\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})$, $\mathbf{g}_{t-\tau}(\mathbf{x})$, the system may need to feedback $\mathbf{x}_{t-\tau}$. The pseudo code of DTC-OCO is given in Algorithm 1. Recall that DTC-OCO has three algorithm parameters α, η, and γ, whose choice depends on our knowledge of the system. This 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 algorithm parameters on those bounds.

Algorithm 1 The DTC-OCO Algorithm

1: Initialize $\alpha, \eta, \gamma > 0$ and $\mathbf{x}_t \in \mathcal{X}_0, \mathbf{Q}_t = \mathbf{0}, \forall t \in [\tau]$.
2: At the beginning of each slot $t > \tau$, do:
3: Update the decision $\mathbf{x}_t$ by solving **P2**.
4: Update the virtual queue $\mathbf{Q}_t$ via (8).

Remark 1. The main difference between DTC-OCO and the saddle-point-typed OCO algorithms in [12], [13], [17], [22] is that DTC-OCO uses a virtual queue to track the constraint violation. The virtual queue was also used in Lyapunov optimization [24], and was later extended to OCO in [14]-[16], [18]. Although we have borrowed some technique from Lyapunov drift analysis in a small part of our performance bound analysis, DTC-OCO is structurally different from Lyapunov optimization as explained in Section II.

Remark 2. We point out that the virtual-queue-based OCO algorithms in [14]-[16], [18] are limited to one-slot feedback delay. In addition, in [14], only time-invariant constraints are considered, and in [15], [16], time-varying constraints are considered but required to be i.i.d. over time. In contrast, DTC-OCO allows the constraints to vary arbitrarily over time. Furthermore, [14]-[16] only provides static regret bounds, while we will provide both static and dynamic regret bounds for DTC-OCO in Section IV-B. Compared with [18], 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. In summary, the virtual queue construction, algorithm design, and performance bound analysis for DTC-OCO are all substantially different from those in [14]-[16], [18].

Remark 3. In this work, we focus on analyzing the impact of feedback delay on OCO with both long-term and short-term constraints. For this purpose, we study centralized OCO, similar to the study in [22], which provide the current best algorithms. We note that when the constraint functions $\mathbf{g}_t(\mathbf{x})$ are separable among different components or blocks of $\mathbf{x}$, **P2** can be equivalently decomposed into separate subproblems, each corresponds to a component or block of $\mathbf{x}$. In this case, Algorithm 1 leads to fully distributed implementation. The design of distributed OCO of general constraint functions is beyond the scope of this work, and we refer the interested readers to [29]-[33].

[6] If the information feedback of $f_t(\mathbf{x})$ and $\mathbf{g}_t(\mathbf{x})$ are respectively delayed by τ_1 and τ_2 slots where $\tau_1 \neq \tau_2$, we can still apply DTC-OCO by setting $\tau = \max\{\tau_1, \tau_2\}$.

B. Regret and Constraint Violation Bounds

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

We make the following assumptions that are common in the literature for constrained OCO [12]-[18], [22].

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

$$\|\nabla f_t(\mathbf{x})\| \leq D, \quad \forall \mathbf{x} \in \mathcal{X}_0, \quad \forall t. \tag{9}$$

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

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

$$\|\mathbf{g}_t(\mathbf{x}) - \mathbf{g}_t(\mathbf{y})\| \leq B\|\mathbf{x}-\mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}_0, \quad \forall t. \tag{10}$$

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

$$\|\mathbf{g}_t(\mathbf{x})\| \leq G, \quad \forall \mathbf{x} \in \mathcal{X}_0, \quad \forall t. \tag{11}$$

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

$$\mathbf{g}_t(\tilde{\mathbf{x}}_t) \preceq -\epsilon \mathbf{1}, \quad \forall t. \tag{12}$$

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

$$\|\mathbf{x}-\mathbf{y}\| \leq R, \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}_0. \tag{13}$$

We first provide bounds on the virtual queue vector in the following lemma.

Lemma 1. The virtual queue vector produced by DTC-OCO is bounded for any $t > \tau$ as follows:

$$\mathbf{Q}_t \succeq \mathbf{0}, \tag{14}$$

$$\mathbf{Q}_t + \gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t) \succeq \mathbf{0}, \tag{15}$$

$$\|\mathbf{Q}_t\| \geq \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|, \tag{16}$$

$$\|\mathbf{Q}_t\| \leq \|\mathbf{Q}_{t-1}\| + \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|. \tag{17}$$

Proof: The proofs of (14)-(17) mainly follow from the virtual queue dynamics in (8). Note that the virtual queue is initialized as $\mathbf{Q}_t = \mathbf{0}, \forall t \in [\tau]$. By induction, we first assume $Q^c_{t-1} \geq 0, \forall c \in [C], \forall t > \tau$. Form the virtual queue dynamics in (8), we have $Q^c_t \geq -\gamma g^c_{t-\tau}(\mathbf{x}_t)$ if $g^c_{t-\tau}(\mathbf{x}_t) < 0$; otherwise, we have $Q^c_t \geq Q^c_{t-1} + \gamma g^c_{t-\tau}(\mathbf{x}_t)$. Combining the two cases, we have (14).

From (8), we have $Q^c_t \geq -\gamma g^c_{t-\tau}(\mathbf{x}_t), \forall c \in [C], \forall t > \tau$, which is (15).

From (8) and (14), for any $c \in [C]$ and $t > \tau$, we have $Q^c_t \geq Q^c_{t-1} + \gamma g^c_{t-\tau}(\mathbf{x}_t) \geq \gamma g^c_{t-\tau}(\mathbf{x}_t)$ if $\gamma g^c_{t-\tau}(\mathbf{x}_t) \geq 0$; otherwise, we have $Q^c_t \geq -\gamma g^c_{t-\tau}(\mathbf{x}_t)$. Combining the two cases yields $(Q^c_t)^2 \geq (\gamma g^c_{t-\tau}(\mathbf{x}_t))^2$. Summing over $c \in [C]$, we have (16).

From (8), we have $Q^c_t \leq Q^c_{t-1} + |\gamma g_{t-\tau}(\mathbf{x}_t)|, \forall c \in [C], \forall t > \tau$. By the triangle inequality, we have $\|\mathbf{Q}_t\| \leq \sqrt{\sum_{c\in[C]}(Q^c_{t-1} + |\gamma g_{t-\tau}(\mathbf{x}_t)|)^2} \leq \|\mathbf{Q}_{t-1}\| + \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|$, which gives (17). ∎

Define $L_t \triangleq \frac{1}{2}\|\mathbf{Q}_t\|^2$ as a quadratic Lyapunov function and $\Delta_t \triangleq L_{t+1} - L_t$ as the corresponding Lyapunov drift [24]. Using Lemma 1, we provide an upper bound on Δ_t in the following lemma.

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

$$\Delta_{t-1} \leq \gamma \mathbf{Q}^T_{t-1} \mathbf{g}_{t-\tau}(\mathbf{x}_t) + \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|^2. \tag{18}$$

Proof: For any $c \in [C]$ and $t > \tau$, we first prove

$$\frac{1}{2}(Q^c_t)^2 - \frac{1}{2}(Q^c_{t-1})^2 \leq \gamma Q^c_{t-1} g^c_{t-\tau}(\mathbf{x}_t) + [\gamma g^c_{t-\tau}(\mathbf{x}_t)]^2 \tag{19}$$

by considering the following two cases from (8).

1) $Q^c_{t-1} + \gamma g^c_{t-\tau}(\mathbf{x}_t) \geq -\gamma g^c_{t-\tau}(\mathbf{x}_t)$: We have $Q^c_t = Q^c_{t-1} + \gamma g^c_{t-\tau}(\mathbf{x}_t)$ from (8). It then follows that

$$\begin{aligned}\frac{1}{2}(Q^c_t)^2 &= \frac{1}{2}[Q^c_{t-1} + \gamma g^c_{t-\tau}(\mathbf{x}_t)]^2 \\ &\leq \frac{1}{2}(Q^c_{t-1})^2 + \gamma Q^c_{t-1} g^c_{t-\tau}(\mathbf{x}_t) + [\gamma g^c_{t-\tau}(\mathbf{x}_t)]^2.\end{aligned}$$

2) $-\gamma g^c_{t-\tau}(\mathbf{x}_t) > Q^c_{t-1} + \gamma g^c_{t-\tau}(\mathbf{x}_t)$: We have $Q^c_t = -\gamma g^c_{t-\tau}(\mathbf{x}_t)$ from (8). It then follows that

$$\begin{aligned}\frac{1}{2}(Q^c_t)^2 &\leq \frac{1}{2}[\gamma g^c_{t-\tau}(\mathbf{x}_t)]^2 + \frac{1}{2}[Q^c_{t-1} + \gamma g^c_{t-\tau}(\mathbf{x}_t)]^2 \\ &= \frac{1}{2}(Q^c_{t-1})^2 + \gamma Q^c_{t-1} g^c_{t-\tau}(\mathbf{x}_t) + [\gamma g^c_{t-\tau}(\mathbf{x}_t)]^2.\end{aligned}$$

Combining the above two cases, we have (19). Then, summing (19) over $c \in [C]$ yields (18). ∎

We also require the following lemma, which is borrowed from [2, Lemma 2.8].

Lemma 3. Let $\mathcal{S} \in \mathbb{R}^n$ be a nonempty convex set. Let $h(\mathbf{s}) : \mathbb{R}^n \to \mathbb{R}$ be a 2ϱ-strongly-convex function over $\mathcal{S}$ with respect to (w.r.t.) a norm $\|\cdot\|$. Let $\mathbf{s}^\star = \arg\min_{\mathbf{s}\in\mathcal{S}} h(\mathbf{s})$. Then, for any $\mathbf{u} \in \mathcal{S}$, we have $h(\mathbf{s}^\star) \leq h(\mathbf{u}) - \varrho\|\mathbf{u} - \mathbf{s}^\star\|^2$.

A main goal of this paper is to examine the impact of multi-slot feedback delay on the dynamic regret bound for OCO with long-term constraints, 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 $\{\mathbf{x}^\star_t\}$ (commonly referred to as the path length [4]) as

$$\Delta_{\mathbf{x}^\star} \triangleq \sum_{t=1}^{T} \|\mathbf{x}^\star_t - \mathbf{x}^\star_{t-1}\|. \tag{20}$$

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

$$\Delta_{\mathbf{g}} \triangleq \sum_{t=1}^{T} \max_{\mathbf{x}\in\mathcal{X}_0} \|\mathbf{g}_t(\mathbf{x}) - \mathbf{g}_{t-1}(\mathbf{x})\|. \tag{21}$$

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

$$\Delta_{2,\mathbf{g}} \triangleq \sum_{t=1}^{T} \max_{\mathbf{x}\in\mathcal{X}_0} \|\mathbf{g}_t(\mathbf{x}) - \mathbf{g}_{t-1}(\mathbf{x})\|^2. \tag{22}$$

Note that, in terms of the growth order, $\Delta_{2,\mathbf{g}}$ is usually smaller than $\Delta_{\mathbf{g}}$ for a constraint function sequence $\{\mathbf{g}_t(\mathbf{x})\}$ that varies sublinearly [18].[7]

[7]For instance $\max_{\mathbf{x}\in\mathcal{X}_0}\{\|\mathbf{g}_t(\mathbf{x}) - \mathbf{g}_{t-1}(\mathbf{x})\|\} \propto T^\xi$ for any t, then $\Delta_{\mathbf{g}} = \mathcal{O}(T^{1+\xi})$ and $\Delta_{2,\mathbf{g}} = \mathcal{O}(T^{1+2\xi})$. For sublinear $\Delta_{\mathbf{g}}$ or $\Delta_{2,\mathbf{g}}$, we have $\xi < 0$ and thus $\Delta_{2,\mathbf{g}}$ grows slower than $\Delta_{\mathbf{g}}$. In particular, if $\xi = -\frac{1}{2}$, we have $\Delta_{\mathbf{g}} = \mathcal{O}(T^{\frac{1}{2}})$ and $\Delta_{2,\mathbf{g}} = \mathcal{O}(1)$.

Using results in Lemmas 1-3, and with the tuning freedom brought by the double regularization, we provide an upper bound on the dynamic regret $\mathrm{RE_d}(T)$ for DTC-OCO with τ-slot feedback delay in the following theorem.

Theorem 1. Under Assumptions 1-3, if we choose $\eta \geq \gamma^2 B^2$, the dynamic regret of DTC-OCO is upper bounded by

$$\mathrm{RE_d}(T) \leq \frac{D^2}{4\alpha}T + \frac{\gamma^2 G^2}{2} + \gamma^2 \Delta_{2,\mathbf{g}} + (\alpha\tau+\eta)(R^2 + 2R\Delta_{\mathbf{x}^\star}) + DR\tau. \quad (23)$$

Proof: The objective function of **P2** is $2(\alpha+\eta)$-strongly-convex over $\mathcal{X}_0$ w.r.t. Euclidean norm $\|\cdot\|$ due to the double regularization. Since $\mathbf{x}_t$ minimizes **P2** over $\mathcal{X}_0$ for any $t > \tau$, we have

$$\begin{aligned}
&[\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T(\mathbf{x}_t - \mathbf{x}_{t-\tau}) + \alpha\|\mathbf{x}_t - \mathbf{x}_{t-\tau}\|^2 \\
&\quad + [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T[\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)] + \eta\|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 \\
&\overset{(a)}{\leq} [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T(\mathbf{x}^\star_{t-\tau} - \mathbf{x}_{t-\tau}) + \alpha\|\mathbf{x}^\star_{t-\tau} - \mathbf{x}_{t-\tau}\|^2 \\
&\quad + [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T[\gamma \mathbf{g}_{t-\tau}(\mathbf{x}^\star_{t-\tau})] \\
&\quad + \eta\|\mathbf{x}^\star_{t-\tau} - \mathbf{x}_{t-1}\|^2 - (\alpha+\eta)\|\mathbf{x}_t - \mathbf{x}^\star_{t-\tau}\|^2 \quad (24) \\
&\overset{(b)}{\leq} [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T(\mathbf{x}^\star_{t-\tau} - \mathbf{x}_{t-\tau}) \\
&\quad + \alpha(\|\mathbf{x}^\star_{t-\tau} - \mathbf{x}_{t-\tau}\|^2 - \|\mathbf{x}_t - \mathbf{x}^\star_{t-\tau}\|^2) \\
&\quad + \eta(\|\mathbf{x}^\star_{t-\tau} - \mathbf{x}_{t-1}\|^2 - \|\mathbf{x}_t - \mathbf{x}^\star_{t-\tau}\|^2), \quad (25)
\end{aligned}$$

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 (15), $\gamma > 0$, and $\mathbf{g}_{t-\tau}(\mathbf{x}^\star_{t-\tau}) \preceq \mathbf{0}, \forall t > \tau$, in (5), such that $[\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T[\gamma \mathbf{g}_{t-\tau}(\mathbf{x}^\star_{t-\tau})] \leq 0, \forall t > \tau$.

Now, we bound the second and third terms in (25). From $\|\mathbf{a}+\mathbf{b}\|^2 \geq \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\|\mathbf{a}\|\|\mathbf{b}\|$, we have

$$\begin{aligned}
&\|\mathbf{x}^\star_{t-\tau} - \mathbf{x}_{t-\tau}\|^2 - \|\mathbf{x}_t - \mathbf{x}^\star_{t-\tau}\|^2 \\
&\leq \|\mathbf{x}^\star_{t-\tau} - \mathbf{x}_{t-\tau}\|^2 - \|\mathbf{x}^\star_t - \mathbf{x}_t\|^2 - \|\mathbf{x}^\star_{t-\tau} - \mathbf{x}^\star_t\|^2 \\
&\quad + 2\|\mathbf{x}^\star_t - \mathbf{x}_t\|\|\mathbf{x}^\star_{t-\tau} - \mathbf{x}^\star_t\| \leq \Phi_{t-\tau} + 2R\phi_{t-\tau}, \quad (26)
\end{aligned}$$

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

$$\begin{aligned}
&\|\mathbf{x}^\star_{t-\tau} - \mathbf{x}_{t-1}\|^2 - \|\mathbf{x}_t - \mathbf{x}^\star_{t-\tau}\|^2 \\
&\leq \Psi_{t-\tau} - \|\mathbf{x}^\star_{t-\tau} - \mathbf{x}_{t-\tau+1}\|^2 + 2R\psi_{t-\tau}, \quad (27)
\end{aligned}$$

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

Substituting (26) and (27) into (25) and adding $f_{t-\tau}(\mathbf{x}_{t-\tau})$ on both sides, we have

$$\begin{aligned}
&f_{t-\tau}(\mathbf{x}_{t-\tau}) + [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T(\mathbf{x}_t - \mathbf{x}_{t-\tau}) + \alpha\|\mathbf{x}_t - \mathbf{x}_{t-\tau}\|^2 \\
&\quad + [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T[\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)] + \eta\|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 \\
&\leq f_{t-\tau}(\mathbf{x}_{t-\tau}) + [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T(\mathbf{x}^\star_{t-\tau} - \mathbf{x}_{t-\tau}) \\
&\quad + \alpha(\Phi_{t-\tau} + 2R\phi_{t-\tau}) + \eta(\Psi_{t-\tau} + 2R\psi_{t-\tau}). \quad (28)
\end{aligned}$$

Applying the first-order condition of convexity

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

to the right-hand side (RHS) of (28), and rearranging terms, we have

$$\begin{aligned}
&f_{t-\tau}(\mathbf{x}_{t-\tau}) - f_{t-\tau}(\mathbf{x}^\star_{t-\tau}) \\
&\leq -[\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T(\mathbf{x}_t - \mathbf{x}_{t-\tau}) - \alpha\|\mathbf{x}_t - \mathbf{x}_{t-\tau}\|^2 \\
&\quad - [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T[\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)] - \eta\|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 \\
&\quad + \alpha(\Phi_{t-\tau} + 2R\phi_{t-\tau}) + \eta(\Psi_{t-\tau} + 2R\psi_{t-\tau}). \quad (29)
\end{aligned}$$

We now bound the right-hand side of (29). Note that

$$\begin{aligned}
&- [\mathbf{Q}_{t-1} + \gamma \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T[\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)] \\
&\overset{(a)}{\leq} -\Delta_{t-1} + \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|^2 - \gamma^2 \mathbf{g}^T_{t-\tau-1}(\mathbf{x}_{t-1})\mathbf{g}_{t-\tau}(\mathbf{x}_t) \\
&\overset{(b)}{=} -\Delta_{t-1} + \frac{\gamma^2}{2}(\|\mathbf{g}_{t-\tau}(\mathbf{x}_t)\|^2 - \|\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|^2) \\
&\quad + \frac{\gamma^2}{2}\|\mathbf{g}_{t-\tau}(\mathbf{x}_t) - \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|^2 \\
&\overset{(c)}{\leq} -\Delta_{t-1} + \gamma^2\left(\frac{1}{2}\varphi_{t-\tau} + B^2\|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 + \varpi_{t-\tau}\right), \quad (30)
\end{aligned}$$

where $\varphi_{t-\tau} \triangleq \|\mathbf{g}_{t-\tau}(\mathbf{x}_t)\|^2 - \|\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|^2$ and $\varpi_{t-\tau} \triangleq \|\mathbf{g}_{t-\tau}(\mathbf{x}_{t-1}) - \mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|^2$. Here, (a) follows from rearranging terms of (18) in Lemma 2 such that $-\gamma \mathbf{Q}^T_{t-1}\mathbf{g}_{t-\tau}(\mathbf{x}_t) \leq -\Delta_{t-1} + \|\gamma \mathbf{g}_{t-\tau}(\mathbf{x}_t)\|^2$, (b) is because $\mathbf{a}^T\mathbf{b} = \frac{1}{2}(\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - \|\mathbf{a}-\mathbf{b}\|^2)$, and (c) follows from $\mathbf{g}_t(\mathbf{x})$ being Lipschitz continuous in (10) and the fact that $\frac{1}{2}\|\mathbf{a}+\mathbf{b}\|^2 \leq \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2$.

Substituting (30) into (29), we have

$$\begin{aligned}
&f_{t-\tau}(\mathbf{x}_{t-\tau}) - f_{t-\tau}(\mathbf{x}^\star_{t-\tau}) \\
&\leq -[\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T(\mathbf{x}_t - \mathbf{x}_{t-\tau}) - \alpha\|\mathbf{x}_t - \mathbf{x}_{t-\tau}\|^2 \\
&\quad + (\gamma^2 B^2 - \eta)\|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 - \Delta_{t-1} + \frac{\gamma^2}{2}\varphi_{t-\tau} + \gamma^2\varpi_{t-\tau} \\
&\quad + \alpha(\Phi_{t-\tau} + 2R\phi_{t-\tau}) + \eta(\Psi_{t-\tau} + 2R\psi_{t-\tau}) \\
&\overset{(a)}{\leq} \frac{D^2}{4\alpha} - \Delta_{t-1} + \frac{\gamma^2}{2}\varphi_{t-\tau} + \gamma^2\varpi_{t-\tau} \\
&\quad + \alpha(\Phi_{t-\tau} + 2R\phi_{t-\tau}) + \eta(\Psi_{t-\tau} + 2R\psi_{t-\tau}), \quad (31)
\end{aligned}$$

where (a) follows from $\eta \geq \gamma^2 B^2$, the bound on $\nabla f_t(\mathbf{x})$ in (9), and completing the square such that

$$\begin{aligned}
&- [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T(\mathbf{x}_t - \mathbf{x}_{t-\tau}) - \alpha\|\mathbf{x}_t - \mathbf{x}_{t-\tau}\|^2 \\
&= -\left\|\frac{\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})}{2\sqrt{\alpha}} + \sqrt{\alpha}(\mathbf{x}_t - \mathbf{x}_{t-\tau})\right\|^2 + \frac{1}{4\alpha}\|\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})\|^2 \\
&\leq \frac{1}{4\alpha}\|\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})\|^2 \leq \frac{D^2}{4\alpha}. \quad (32)
\end{aligned}$$

Summing (31) over $t \in [\tau+1, T]$, we have

$$\begin{aligned}
&\sum_{t=\tau+1}^{T} f_{t-\tau}(\mathbf{x}_{t-\tau}) - f_{t-\tau}(\mathbf{x}^\star_{t-\tau}) = \sum_{t=1}^{T-\tau} f_t(\mathbf{x}_t) - f_t(\mathbf{x}^\star_t) \\
&\overset{(a)}{\leq} \frac{D^2}{4\alpha}T + \frac{\gamma^2 G^2}{2} + \gamma^2\Delta_{\mathbf{g}} + (\alpha\tau+\eta)(R^2 + 2R\Delta_{\mathbf{x}^\star}), \quad (33)
\end{aligned}$$

where (a) follows from Δ_{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 $t \in \{\tau+1, \ldots, T\}$ are upper bounded by 0, G^2, $\Delta_{2,\mathbf{g}}$, τR^2, $\tau\Delta_{\mathbf{x}^\star}$, R^2, and $\Delta_{\mathbf{x}^\star}$, respectively.

Finally, adding $\sum_{t=T-\tau+1}^{T} f_t(\mathbf{x}_t) - f_t(\mathbf{x}_t^\star)$ on both sides of (33), and noting that the convexity of $f_t(\mathbf{x})$ implies

$$f_t(\mathbf{x}_t) - f_t(\mathbf{x}_t^\star) \leq \|\nabla f_t(\mathbf{x}_t)\| \|\mathbf{x}_t^\star - \mathbf{x}_t\| \leq DR, \quad (34)$$

we complete the proof. ∎

Next, we provide an upper bound on the static regret $\mathrm{RE_s}(T)$ yielded by DTC-OCO.

Theorem 2. Under Assumptions 1-3, if we choose $\eta \geq \gamma^2 B^2$, the static regret of DTC-OCO is upper bounded by

$$\mathrm{RE_s}(T) \leq \frac{D^2}{4\alpha}T + \frac{\gamma^2 G^2}{2} + \gamma^2 \Delta_{2,\mathbf{g}} + (\alpha\tau + \eta)R^2 + DR\tau. \quad (35)$$

Proof: To show (35), we use the techniques in the proof for the dynamic regret $\mathrm{RE_d}(T)$ in Theorem 1. Replacing all the per-slot optimizers with the static benchmark $\mathbf{x}^\star$ in the proof of Theorem 1, we can show that (31) still holds by redefining $\Phi_{t-\tau} \triangleq \|\mathbf{x}^\star - \mathbf{x}_{t-\tau}\|^2 - \|\mathbf{x}^\star - \mathbf{x}_t\|^2$, $\phi_{t-\tau} \triangleq 0$, $\Psi_{t-\tau} \triangleq \|\mathbf{x}^\star - \mathbf{x}_{t-1}\|^2 - \|\mathbf{x}^\star - \mathbf{x}_t\|^2$, and $\psi_{t-\tau} \triangleq 0$. Summing the above version of (31) over $t \in \{\tau+1, \ldots, T\}$, noting that $\Phi_{t-\tau}$ and $\Psi_{t-\tau}$ are still telescoping, and leveraging (34), we complete the proof. ∎

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

Lemma 4. The virtual queue vector produced by DTC-OCO satisfies the following inequality for any $c \in [C]$:

$$\mathrm{VO}^c(T) \leq \frac{1}{\gamma}\|\mathbf{Q}_T\| + \tau\Delta_{\mathbf{g}} + G\tau. \quad (36)$$

Proof: From (8), we have $Q_t^c \geq Q_{t-1}^c + \gamma g_{t-\tau}^c(\mathbf{x}_t), \forall t > \tau$. Summing it over $t \in \{\tau+1, \ldots, T\}$ 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 $\mathrm{VO}^c(T)$ in (7), we have

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

Noting $\|\mathbf{a}\|_\infty \leq \|\mathbf{a}\|$, the bound on $\mathbf{g}_t(\mathbf{x})$ in (11), and the definition of $\Delta_{\mathbf{g}}$ in (21), we complete the proof. ∎

From Lemma 4, we see that one can bound the constraint violation $\mathrm{VO}^c(T)$ by bounding the virtual queue vector $\mathbf{Q}_T$. Following this, we obtain an upper bound on the constraint violation for DTC-OCO in the following theorem.

Theorem 3. Under Assumptions 1-3, the constraint violation of DTC-OCO for any $c \in [C]$ is upper bounded by

$$\mathrm{VO}^c(T) \leq 2G + \frac{2\gamma^2 G^2 + DR + (\alpha+\eta)R^2}{\epsilon\gamma^2} + \tau\Delta_{\mathbf{g}} + G\tau. \quad (37)$$

Proof: From Lemma 3, we can show that inequality (24) still holds for any $t > \tau$ after replacing the per-slot optimizer $\mathbf{x}_{t-\tau}^\star$ with the interior point $\tilde{\mathbf{x}}_{t-\tau}$. We have

$$\begin{aligned}
&[\mathbf{Q}_{t-1} + \gamma\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T[\gamma\mathbf{g}_{t-\tau}(\tilde{\mathbf{x}}_{t-\tau})] \\
&\overset{(a)}{\leq} -\epsilon\gamma[\mathbf{Q}_{t-1} + \gamma\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T\mathbf{1} \\
&\overset{(b)}{\leq} -\epsilon\gamma\|\mathbf{Q}_{t-1} + \gamma\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\| \\
&\overset{(c)}{\leq} -\epsilon\gamma(\|\mathbf{Q}_{t-1}\| - \|\gamma\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|), \quad (38)
\end{aligned}$$

where (a) follows from the existence of interior point in (12) and the virtual queue bound in (15), (b) is because $\|\mathbf{a}\| \leq \|\mathbf{a}\|_1$, and (c) follows from $|\|\mathbf{a}\| - \|\mathbf{b}\|| \leq \|\mathbf{a} - \mathbf{b}\|$. Applying (38) to the aforementioned version of (24) with $\tilde{\mathbf{x}}_{t-\tau}$ and rearranging terms, we have

$$\begin{aligned}
&\gamma\mathbf{Q}_{t-1}^T\mathbf{g}_{t-\tau}(\mathbf{x}_t) \\
&\leq -\epsilon\gamma(\|\mathbf{Q}_{t-1}\| - \|\gamma\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})\|) - \alpha\|\mathbf{x}_t - \mathbf{x}_{t-\tau}\|^2 \\
&\quad - [\gamma\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T[\gamma\mathbf{g}_{t-\tau}(\mathbf{x}_t)] - \eta\|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2 \\
&\quad + [\nabla f_{t-\tau}(\mathbf{x}_{t-\tau})]^T(\tilde{\mathbf{x}}_{t-\tau} - \mathbf{x}_t) + \alpha\|\tilde{\mathbf{x}}_{t-\tau} - \mathbf{x}_{t-\tau}\|^2 \\
&\quad + \eta\|\tilde{\mathbf{x}}_{t-\tau} - \mathbf{x}_{t-1}\|^2 - (\alpha+\eta)\|\mathbf{x}_t - \tilde{\mathbf{x}}_{t-\tau}\|^2 \\
&\overset{(a)}{\leq} -\epsilon\gamma\|\mathbf{Q}_{t-1}\| + \epsilon\gamma^2 G + \gamma^2 G^2 + DR + (\alpha+\eta)R^2, \quad (39)
\end{aligned}$$

where (a) follows from the Cauchy-Schwartz inequality $|\mathbf{a}^T\mathbf{b}| \leq \|\mathbf{a}\|\|\mathbf{b}\|$, the bound on $\nabla f_t(\mathbf{x})$ in (9), the bound on $\mathbf{g}_t(\mathbf{x})$ in (11), and the bound on $\mathcal{X}_0$ in (13). Substituting (39) into (18) in Lemma 2 and noting that $\|\mathbf{g}_{t-\tau}(\mathbf{x}_t)\|^2 \leq G^2$ from (11) yields

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

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

$$\|\mathbf{Q}_{t-1}\| > \gamma G + \frac{2\gamma^2 G^2 + DR + (\alpha+\eta)R^2}{\epsilon\gamma}. \quad (40)$$

If (40) holds, we have $\|\mathbf{Q}_t\| < \|\mathbf{Q}_{t-1}\|$, *i.e.,* the virtual queue decreases; otherwise, from the virtual queue bound in (17), there is a maximum increase from $\|\mathbf{Q}_{t-1}\|$ to $\|\mathbf{Q}_t\|$ since $\|\mathbf{Q}_t\| - \|\mathbf{Q}_{t-1}\| \leq \|\gamma\mathbf{g}_{t-\tau}(\mathbf{x}_t)\| \leq \gamma G$. Therefore, the virtual queue is upper bounded for any $t > \tau$ by

$$\|\mathbf{Q}_t\| \leq 2\gamma G + \frac{2\gamma^2 G^2 + DR + (\alpha+\eta)R^2}{\epsilon\gamma}. \quad (41)$$

Substituting (41) into (36), we complete the proof. ∎

TABLE II
DYNAMIC REGRET AND CONSTRAINT VIOLATION BOUNDS OF DTC-OCO

Constraint	Know δ?	$\mathrm{RE_d}(T)$	$\mathrm{VO}^c(T)$
Varying	Yes	$\mathcal{O}(\max\{\tau^{\frac{1}{2}}T^{\frac{1+\delta}{2}}, T^\nu\})$	$\mathcal{O}(\max\{T^{\frac{1-\delta}{2}}, \tau T^\kappa\})$
Varying	No	$\mathcal{O}(\max\{\tau^{\frac{1}{2}}T^{\frac{1}{2}+\delta}, T^\nu\})$	$\mathcal{O}(\max\{T^{\frac{1}{2}}, \tau T^\kappa\})$
Invariant	Yes	$\mathcal{O}(\max\{\tau^{\frac{1}{2}}T^{\frac{1+\delta}{2}}\})$	$\mathcal{O}(\tau)$
Invariant	No	$\mathcal{O}(\max\{\tau^{\frac{1}{2}}T^{\frac{1}{2}+\delta}\})$	$\mathcal{O}(\tau)$

C. Discussion on the Regret and Constraint Violation Bounds

With the regret and constraint violation bounds obtained above, we now discuss the sufficient conditions for DTC-OCO to yield sublinear regret and constraint violation. We also highlight several prominent advantages of DTC-OCO over existing constrained OCO algorithms. For clarity, we summarize the performance bounds of DTC-OCO in terms of the growth rate over T for general convex loss functions under different conditions in Table II.

1) Sublinear Regret and Constraint Violation: From Theorems 1-3, we derive the following corollaries regarding the growth rates of the regret and constraint violation over T. We assume the time variabilities of the dynamic benchmark $\{\mathbf{x}_t^\star\}$ and constraint functions $\{\mathbf{g}_t(\mathbf{x})\}$ satisfy $\Delta_{\mathbf{x}^\star}=\mathcal{O}(T^\delta)$, $\Delta_{\mathbf{g}}=\mathcal{O}(T^\kappa)$, and $\Delta_{2,\mathbf{g}}=\mathcal{O}(T^\nu)$, for some constant parameters $\delta,\nu,\kappa\geq 0$ [17], [18]. Corollaries 1 and 2 provide two sets of performance bounds depending on whether the value of δ is known to set the step-size parameter α in DTC-OCO. The proofs of these two corollaries can be obtained from substituting the corresponding algorithm parameters α,η,γ, specified in each corollary into the bounds in (23), (35), (37); Thus, they are omitted for brevity.

Corollary 1. Suppose the value of δ is known. If feedback delay τ is known, let $\alpha=\tau^{-\frac{1}{2}}T^{\frac{1-\delta}{2}}$, $\eta=B^2\gamma^2$, and $\gamma=1$ in DTC-OCO. Then,

$$\mathrm{RE_d}(T)=\mathcal{O}\left(\max\left\{\tau^{\frac{1}{2}}T^{\frac{1+\delta}{2}},T^\nu\right\}\right), \tag{42}$$

$$\mathrm{RE_s}(T)=\mathcal{O}\left(\max\left\{\tau^{\frac{1}{2}}T^{\frac{1}{2}},T^\nu\right\}\right), \tag{43}$$

$$\mathrm{VO}^c(T)=\mathcal{O}\left(\max\left\{T^{\frac{1-\delta}{2}},\tau T^\kappa\right\}\right). \tag{44}$$

If feedback delay τ is unknown, let $\alpha=T^{\frac{1-\delta}{2}}$, $\eta=B^2\gamma^2$, and $\gamma=1$ in DTC-OCO. Then,

$$\mathrm{RE_d}(T)=\mathcal{O}\left(\max\left\{\tau T^{\frac{1+\delta}{2}},T^\nu\right\}\right), \tag{45}$$

$$\mathrm{RE_s}(T)=\mathcal{O}\left(\max\left\{\tau T^{\frac{1}{2}},T^\nu\right\}\right), \tag{46}$$

$$\mathrm{VO}^c(T)=\mathcal{O}\left(\max\left\{T^{\frac{1-\delta}{2}},\tau T^\kappa\right\}\right) \tag{47}$$

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

In Corollary 1, we set the step-size parameter α in DTC-OCO with the knowledge of T and δ. When T is unknown, the standard doubling trick [2], [7] can applied to adjust T over time. The value of δ may be estimated over time through some expert-tracking algorithm [34].

Corollary 2. Suppose the value of δ is unknown. If feedback delay τ is known, let $\alpha=\tau^{-\frac{1}{2}}T^{\frac{1}{2}}$, $\eta=B^2\gamma^2$, and $\gamma=1$ in DTC-OCO. Then,

$$\mathrm{RE_d}(T)=\mathcal{O}\left(\max\left\{\tau^{\frac{1}{2}}T^{\frac{1}{2}+\delta},T^\nu\right\}\right), \tag{48}$$

$$\mathrm{RE_s}(T)=\mathcal{O}\left(\max\left\{\tau^{\frac{1}{2}}T^{\frac{1}{2}},T^\nu\right\}\right), \tag{49}$$

$$\mathrm{VO}^c(T)=\mathcal{O}\left(\max\left\{T^{\frac{1}{2}},\tau T^\kappa\right\}\right). \tag{50}$$

If feedback delay τ is unknown, let $\alpha=T^{\frac{1}{2}}$, $\eta=B^2\gamma^2$, and $\gamma=1$ in DTC-OCO. Then,

$$\mathrm{RE_d}(T)=\mathcal{O}\left(\max\left\{\tau T^{\frac{1}{2}+\delta},T^\nu\right\}\right), \tag{51}$$

$$\mathrm{RE_s}(T)=\mathcal{O}\left(\max\left\{\tau T^{\frac{1}{2}},T^\nu\right\}\right), \tag{52}$$

$$\mathrm{VO}^c(T)=\mathcal{O}\left(\max\left\{T^{\frac{1}{2}},\tau T^\kappa\right\}\right). \tag{53}$$

In particular, if $\tau=\mathcal{O}(1)$, $\delta<\frac{1}{2}$, $\nu<1$, and $\kappa<1$, both the dynamic and static regrets are sublinear in T, and the constraint violation is sublinear in T.

From Corollaries 1 and 2, a sufficient condition for DTC-OCO to yield sublinear dynamic and static regrets and sublinear constraint violation is that the accumulated variation $\Delta_{\mathbf{x}^\star}$ of the dynamic benchmark $\{\mathbf{x}_t^\star\}$ and the accumulated variations $\Delta_{\mathbf{g}}$ and $\Delta_{2,\mathbf{g}}$ of the constraints $\{\mathbf{g}_t\}$ evolve sufficiently slowly. This is the case for many online applications, where the system tends to stabilize over time. Otherwise, if the system varies too drastically, it has been shown via a counter example in [23] and stated in [17], [18] that, no online algorithm can track the system due to the lack of in-time information.

We now highlight some advantages of DTC-OCO over the online algorithm in [22]. The performance analysis in [22] focuses on the static regret. In contrast, we provide both dynamic and static regret bounds for DTC-OCO. Furthermore, to show sublinear static regret and constraint violation bounds, [22] requires that T is sufficiently large to satisfy $\frac{(1+C)\max\{D,B\}^2+2}{\sqrt{\tau T}}+[(5C+1)\max\{D,B\}^2+2]\sqrt{\frac{\tau}{T}}\leq\sqrt{\frac{1}{3}}$. In comparison, our performance bounds for DTC-OCO hold for any T. Finally, to compute the optimal step-sizes in [22] require knowledge of the values of C and D, which we do not need for DTC-OCO.

2) Special Case of One-Slot Feedback Delay: No existing algorithm provides a dynamic regret bound for constrained OCO with multi-slot feedback delay. Thus, to compare our dynamic regret bound with existing ones, we consider the special case of one-slot feedback delay, and compare DTC-OCO with [17] and [18] under this setting.

We point out a few differences of the dynamic regrets obtained by DTC-OCO and that in [17], and highlight the advantages of DTC-OCO. The dynamic regret and constraint violation bounds achieved by [17] rely on a key assumption that the slack constant ϵ is larger than the maximum variation of the constraints, *i.e.,* $\epsilon>\max_{t\in[T]}\max_{\mathbf{x}\in\mathcal{X}_0}\|\mathbf{g}_t(\mathbf{x})-\mathbf{g}_{t-1}(\mathbf{x})\|$, which may be difficult to satisfy in general. In contrast, DTC-OCO only assumes $\epsilon>0$ as indicated in (12). Furthermore, the optimal step-sizes used in [17] require the knowledge of the accumulated variation measure κ on the constraint function sequence $\{\mathbf{g}_t(\mathbf{x})\}$. In comparison, DTC-OCO only needs an upper bound B on the gradient $\nabla\mathbf{g}_t(\mathbf{x})$, which is much easier to acquire than κ. Finally, when δ is unknown, [17] 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.

We now compare the condition to achieve sublinear dynamic regret in DTC-OCO and that in the online algorithm in [18]. To achieve sublinear dynamic regret and constraint violation, [18] relies on two additional assumptions: the accumulated variation of the convex loss functions $\{f_t(\mathbf{x})\}$ is sublinear regardless of the trajectory of the online decision sequence, *i.e.,* $\sum_{t=1}^T\max_{\mathbf{x}\in\mathcal{X}_0}\|f_t(\mathbf{x})-f_{t-1}(\mathbf{x})\|=\mathbf{o}(T)$, and the accumulated variation of the optimal dual points $\{\lambda_t^\star\}$ of the optimization problem **P1** is sublinear, *i.e.,* $\sum_{t=1}^T\|\lambda_{t+1}^\star-$

$\lambda_t^\star\| = \mathbf{o}(T)$. The two assumptions are not required for DTC-OCO. In terms of information needed for the online update, DTC-OCO requires only the gradient information $\nabla f_t(\mathbf{x}_t)$ of the loss function $f_t(\mathbf{x}_t)$ at the online decision point $\mathbf{x}_t$. In contrast, instead of a gradient-based algorithm, the online algorithm in [18] directly minimizes the loss function $f_t(\mathbf{x})$, which requires complete information feedback of $f_t(\mathbf{x})$.

3) Special Case of Time-invariant Constraints: When the constraints are time-invariant, the following corollary suggests that the static regret of DTC-OCO is the same as the current best $\mathcal{O}(\tau^{\frac{1}{2}}T^{\frac{1}{2}})$ static regret for unconstrained OCO with multi-slot delay [19]; Furthermore, in the special case of one-slot feedback delay, the static regret and constraint violation of DTC-OCO are the same as the current best $\mathcal{O}(T^{\frac{1}{2}})$ static regret and $\mathcal{O}(1)$ constraint violation [14], respectively.

Corollary 3. If δ is known, let $\alpha = \gamma^2 = \tau^{-\frac{1}{2}}T^{\frac{1-\delta}{2}}$ and $\eta = B^2\gamma^2$ in DTC-OCO. Then, $\mathrm{RE_d}(T) = \mathcal{O}(\tau^{\frac{1}{2}}T^{\frac{1+\delta}{2}})$, $\mathrm{RE_s}(T) = \mathcal{O}(\tau^{\frac{1}{2}}T^{\frac{1}{2}})$, and $\mathrm{VO}^c(T) = \mathcal{O}(\tau)$. In particular, if $\tau = \mathcal{O}(1)$ and $\delta < 1$, both the dynamic regret $\mathrm{RE_d}(T)$ and static regret $\mathrm{RE_s}(T)$ are sublinear, and the constraint violation $\mathrm{VO}^c(T)$ is upper bounded by a constant. If δ is unknown, let $\alpha = \gamma^2 = \tau^{-\frac{1}{2}}T^{\frac{1}{2}}$ and $\eta = B^2\gamma^2$ in DTC-OCO. Then, $\mathrm{RE_d}(T) = \mathcal{O}(\tau^{\frac{1}{2}}T^{\frac{1}{2}+\delta})$, $\mathrm{RE_s}(T) = \mathcal{O}(\tau^{\frac{1}{2}}T^{\frac{1}{2}})$, and $\mathrm{VO}^c(T) = \mathcal{O}(\tau)$.

We further compare the constraint violation bound of DTC-OCO and those in [17], [22] under time-invariant constraints. The constraint violation bound in [17] is no less than $\mathcal{O}(T^{\frac{2}{3}})$ under one-slot feedback delay. In [22], the constraint violation is no less than $\mathcal{O}(\tau^{\frac{1}{4}}T^{\frac{3}{4}})$ under τ-slot delay. In contrast, for DTC-OCO, the constraint violation $\mathrm{VO}^c(T)$ is $\mathcal{O}(\tau)$ under τ-slot delay, which is smaller than the ones in [17], [22] as T is usually much greater than τ.

V. Delay-Tolerant Constrained OCO with Multi-Step Gradient Descent

In the previous section, we have proposed DTC-OCO and derived its performance bounds for general convex loss functions. In this section, we propose a variation of DTC-OCO to enable multi-step gradient descent in the objective of the per-slot optimization problem **P2**. With this algorithm, we obtain improved bounds on both the dynamic regret and constraint violation for strongly convex loss functions.

A. DTC-OCO with Multi-Step Gradient Descent

It has been shown in [11] that, for strongly-convex loss functions, multi-step gradient descent provides stronger bounding performance for OCO with short-term constraints, under the standard one-slot feedback delay setting. In this work, we will further show that, when the loss functions are strongly convex, performing multi-step gradient descent can achieve smaller dynamic regret bound and constraint violation bound for OCO with long-term constraints and multi-slot feedback delay.

Below, we show how to configure DTC-OCO to incorporate multi-step gradient descent. At the beginning of each slot $t > \tau$, we first initialize an intermediate decision $\hat{\mathbf{x}}_{t-\tau}^0 = \mathbf{x}_{t-\tau}$. Then, we perform M-step gradient descent to generate $\hat{\mathbf{x}}_{t-\tau}^M$

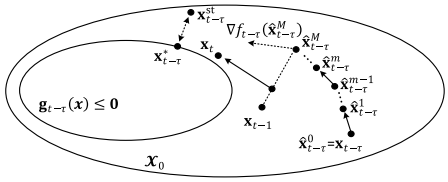


Fig. 2. An illustration of updating $\mathbf{x}_t$ by DTC-OCO under the double regularization and multi-step gradient descent. At each time t, under the M-step gradient descent, $\hat{\mathbf{x}}_{t-\tau}^m$ moves from $\hat{\mathbf{x}}_{t-\tau}^{m-1}$ towards $\mathbf{x}_{t-\tau}^{\mathrm{st}}$ to track the dynamic benchmark $\{\mathbf{x}_t^{\mathrm{st}}\}$. As the result of solving $\hat{\mathbf{P2}}$ based on $\hat{\mathbf{x}}_{t-\tau}^M$ and $\mathbf{x}_{t-1}$, $\mathbf{x}_t$ moves from a point on the line segment between $\hat{\mathbf{x}}_{t-\tau}^M$ and $\mathbf{x}_{t-1}$ towards $\mathbf{x}_{t-\tau}^\star$ to track the dynamic benchmark $\{\mathbf{x}_t^\star\}$.

for any $M > 0$. Note that if $M = 0$, we readily have $\hat{\mathbf{x}}_{t-\tau}^M = \mathbf{x}_{t-\tau}$. For each additional gradient descent step $m \in [M]$, we update $\hat{\mathbf{x}}_{t-\tau}^m$ by solving the following optimization problem:

$$\min_{\mathbf{x}\in\mathcal{X}_0} \quad [\nabla f_{t-\tau}(\hat{\mathbf{x}}_{t-\tau}^{m-1})]^T(\mathbf{x} - \hat{\mathbf{x}}_{t-\tau}^{m-1}) + \alpha\|\mathbf{x} - \hat{\mathbf{x}}_{t-\tau}^{m-1}\|^2.$$

The solution to the above optimization problem is the standard projected gradient descent, given by

$$\hat{\mathbf{x}}_{t-\tau}^m = \mathcal{P}_{\mathcal{X}_0}\left\{\hat{\mathbf{x}}_{t-\tau}^{m-1} - \frac{1}{2\alpha}\nabla f_{t-\tau}(\hat{\mathbf{x}}_{t-\tau}^{m-1})\right\} \tag{54}$$

where $\mathcal{P}_{\mathcal{X}_0}\{\mathbf{x}\} \triangleq \arg\min_{\mathbf{y}\in\mathcal{X}_0}\|\mathbf{y} - \mathbf{x}\|^2$ is the projection operator to project $\mathbf{x}$ onto the set $\mathcal{X}_0$, and $\alpha > 0$ can be seen as a step-size parameter.

To quantify the impact of multi-step gradient descent on the dynamic regret, we define the dynamic benchmark $\{\mathbf{x}_t^{\mathrm{st}}\}$ that is obtained assuming the short-term constraints only, given by

$$\mathbf{x}_t^{\mathrm{st}} \in \arg\min_{\mathbf{x}\in\mathcal{X}_0} f_t(\mathbf{x}). \tag{55}$$

The benefit of using the multi-step gradient descent is that each gradient descent step helps track the dynamic benchmark $\{\mathbf{x}_t^{\mathrm{st}}\}$. We will show analytically in Section V-B that at each gradient descent step m, the distance between $\hat{\mathbf{x}}_{t-\tau}^m$ and $\mathbf{x}_{t-\tau}^{\mathrm{st}}$ is strictly less than the distance between $\hat{\mathbf{x}}_{t-\tau}^{m-1}$ and $\mathbf{x}_{t-\tau}^{\mathrm{st}}$ in the previous step.

Note that the update of $\hat{\mathbf{x}}_{t-\tau}^m$ in (54) is obtained by only considering the short-term constraints in $\mathcal{X}_0$. For the long-term constraints in (1) of **P1**, we use the same virtual queue dynamics in (8) as the basic form of DTC-OCO. We then replace $\mathbf{x}_{t-\tau}$ with $\hat{\mathbf{x}}_{t-\tau}^M$ in **P2** to obtain the following optimization problem to find the decision $\mathbf{x}_t$ for time slot t:

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

where $\alpha, \eta, \gamma, M > 0$ are four algorithm parameters. Different from the basic form of DTC-OCO, the double regularization in $\hat{\mathbf{P2}}$ is on $\hat{\mathbf{x}}_{t-\tau}^M$ and $\mathbf{x}_{t-1}$, instead of on $\mathbf{x}_{t-\tau}$ and $\mathbf{x}_{t-1}$ as in **P2**. An illustrative example for updating $\mathbf{x}_t$ is shown in Fig. 2: With the M-step gradient descent, $\hat{\mathbf{x}}_{t-\tau}^m$ moves from $\hat{\mathbf{x}}_{t-\tau}^{m-1}$ towards $\mathbf{x}_{t-\tau}^{\mathrm{st}}$ to track the dynamic benchmark $\{\mathbf{x}_t^{\mathrm{st}}\}$. Note that the difference between the two dynamic benchmarks $\{\mathbf{x}_t^{\mathrm{st}}\}$ in (55) and $\{\mathbf{x}_t^\star\}$ in (5) is that $\{\mathbf{x}_t^\star\}$ satisfies both the long-term and short-term constraints while $\{\mathbf{x}_t^{\mathrm{st}}\}$ only satisfies the short-term constraints. When the long-term

Algorithm 2 DTC-OCO with Multi-Step Gradient Descent

1: Initialize $\alpha, \eta, \gamma, M > 0$ and $\mathbf{x}_t \in \mathcal{X}_0, \mathbf{Q}_t = \mathbf{0}, \forall t \in [\tau]$.
2: At the beginning of each slot $t > \tau$, do:
3: Initialize intermediate decision $\hat{\mathbf{x}}_{t-\tau}^0 = \mathbf{x}_{t-\tau}$.
4: **for** $m = 1$ **to** M
5: Update $\hat{\mathbf{x}}_{t-\tau}^m$ via (54).
6: **end for**
7: Update the decision $\mathbf{x}_t$ by solving $\hat{\mathbf{P2}}$.
8: Update the virtual queue $\mathbf{Q}_t$ via (8).

constraints are relatively loose as compared with the short-term constraints, $\{\mathbf{x}_t^{\rm st}\}$ is close to $\{\mathbf{x}_t^\star\}$. Thus, tracking $\{\mathbf{x}_t^{\rm st}\}$ also helps minimize the dynamic regret. With the double regularization, $\mathbf{x}_t$ moves from a point on the line segment between $\mathbf{x}_{t-1}$ and $\hat{\mathbf{x}}_{t-\tau}^M$ towards $\mathbf{x}_{t-\tau}^\star$ to track the dynamic benchmark $\{\mathbf{x}_t^\star\}$ for dynamic regret minimization. We will show analytically in the next two subsections and numerically in Section VI that this method of utilizing multi-step gradient descent together with double regularization will improve the performance bounds and averaged performance of DTC-OCO for strongly convex loss functions.

DTC-OCO with multi-step gradient descent is summarized in Algorithm 2. The algorithm consists of four major steps: 1) Initialize $\mathbf{x}_t \in \mathcal{X}_0, \forall t \in [\tau]$, and set $\mathbf{Q}_t = \mathbf{0}, \forall t \in [\tau]$ and $\mathbf{g}_0(\mathbf{x}) \equiv \mathbf{0}$; 2) At the beginning of each slot $t > \tau$, perform M-step gradient descent to generate $\hat{\mathbf{x}}_{t-\tau}^M$; 3) With both $\hat{\mathbf{x}}_{t-\tau}^M$ and $\mathbf{x}_{t-1}$, obtain the current decision $\mathbf{x}_t$ by solving $\hat{\mathbf{P2}}$; 4) At the end of each slot $t > \tau$, update the virtual queue $\mathbf{Q}_t$ via (8). Note that this algorithm has four algorithm parameters α, η, γ, M. Their choice will be discussed in Section V-C, after we derive the regret bound and constraint violation bound in the next subsection.

B. Dynamic Regret and Constraint Violation Bounds

We now derive the performance bounds of Algorithm 2 for strongly convex loss function $f_t(\mathbf{x})$. Note that strongly convex loss functions arise in many system control and signal processing applications, *e.g.,* support vector machine, Lasso regression, softmax classifier, and robust subspace tracking. Furthermore, for general applications with convex loss functions, adding a regularization term $\mu\|\mathbf{x}\|^2$ can make the overall objective function strongly convex without significantly impacting on the system performance [10].

In Sections V-B and V-C, we make the following assumptions that are common in existing works on OCO with strongly convex loss functions [8]-[11].

Assumption 4. The loss function $f_t(\mathbf{x})$ satisfies the following conditions:

4.1) $f_t(\mathbf{x})$ is 2μ-strongly convex over $\mathcal{X}_0$: $\exists\ \mu > 0$, s.t., for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}_0$ and t

$$f_t(\mathbf{y}) \geq f_t(\mathbf{x}) + [\nabla f_t(\mathbf{x})]^T(\mathbf{y} - \mathbf{x}) + \mu\|\mathbf{y} - \mathbf{x}\|^2. \quad (56)$$

4.2) $f_t(\mathbf{x})$ is $2L$-smooth over $\mathcal{X}_0$: $\exists\ L > 0$, s.t., for any $\mathbf{x}, \mathbf{y} \in \mathcal{X}_0$ and t

$$f_t(\mathbf{y}) \leq f_t(\mathbf{x}) + [\nabla f_t(\mathbf{x})]^T(\mathbf{y} - \mathbf{x}) + L\|\mathbf{y} - \mathbf{x}\|^2. \quad (57)$$

For our analysis, we need the following lemma, which is borrowed from [11, Lemma 5].

Lemma 5. Let $\mathcal{S} \in \mathbb{R}^n$ be a nonempty convex set. Let $h(\mathbf{s}) : \mathbb{R}^n \to \mathbb{R}$ be a 2ϱ-strongly convex and 2ζ-smooth function over $\mathcal{S}$ w.r.t. a norm $\|\cdot\|$. Let $\mathbf{v} = \arg\min_{\mathbf{s}\in\mathcal{S}}\{[\nabla h(\mathbf{u})]^T(\mathbf{s} - \mathbf{u}) + \upsilon\|\mathbf{s} - \mathbf{u}\|^2\}$ and $\mathbf{s}^\star = \arg\min_{\mathbf{s}\in\mathcal{S}} h(\mathbf{s})$. Then, for any $\upsilon \geq \zeta$, we have $\|\mathbf{v} - \mathbf{s}^\star\|^2 \leq \frac{\upsilon - \varrho}{\upsilon + \varrho}\|\mathbf{u} - \mathbf{s}^\star\|^2$.

Similar to Proposition 2 in [8], Lemma 5 shows that for strongly convex and smooth loss function $h(\mathbf{s})$, the distance between the new decision $\mathbf{v}$ and the optimum $\mathbf{s}^\star$ is strictly smaller than the distance between the previous decision $\mathbf{u}$ and $\mathbf{s}^\star$, *i.e.,* $\|\mathbf{v} - \mathbf{s}^\star\| < \|\mathbf{u} - \mathbf{s}^\star\|$. This indicates that performing multi-step gradient descent help the updated decision approach the optimal decision faster than the decision yielded by one-step gradient descent. Such property is utilized in [8]-[11] to improve the dynamic regret for OCO with short-term constraints under one-slot feedback delay.

We now examine the effect of multi-step gradient descent on the dynamic regret of OCO with long-term constraints and multi-slot delay. We define the accumulated squared difference between the dynamic benchmarks $\{\mathbf{x}_t^\star\}$ in (5) and $\{\mathbf{x}_t^{\rm st}\}$ in (55) as

$$\Pi_{\mathbf{x}} \triangleq \sum_{t=1}^{T} \|\mathbf{x}_t^{\rm st} - \mathbf{x}_t^\star\|^2. \quad (58)$$

Note that $\mathbf{x}_t^\star$ satisfies $\mathbf{g}_t(\mathbf{x}_t^\star) \preceq \mathbf{0}$, while $\mathbf{x}_t^{\rm st}$ is only subject to the short term constraints $\mathbf{x}_t^{\rm st} \in \mathcal{X}_0$. Therefore, $\Pi_{\mathbf{x}}$ naturally quantifies the impact of the long-term constraint functions $\{\mathbf{g}_t(\mathbf{x})\}$ on the dynamic benchmark $\{\mathbf{x}_t^\star\}$.

In the following theorem, we provide a dynamic regret bound for DTC-OCO with multi-step gradient descent. To prove the theorem, we have used the techniques in the proof of Theorem 1, the results in Lemma 5, as well as the properties of strong convexity and smoothness.

Theorem 4. Under Assumptions 1-4, if we choose $\alpha \geq L$, $\eta \geq \max\{\gamma^2 B^2, \alpha\tau^2\}$, and $M > \log_\rho(\frac{1}{8})$ with $\rho \triangleq \frac{\alpha - \mu}{\alpha + \mu} < 1$, the dynamic regret yielded by DTC-OCO with multi-step gradient descent is upper bounded by

$$\begin{aligned}\mathrm{RE_d}(T) \leq \frac{1}{4\xi}\Pi_\nabla + \frac{L + \xi}{1 - 8\rho^M}\Bigg[\frac{D^2}{2\alpha^2}T + \frac{\gamma^2}{\alpha}G^2 + \frac{2\gamma^2}{\alpha}\Delta_{2,\mathbf{g}} + 7\Pi_{\mathbf{x}} \\ + \frac{2\eta}{\alpha}(R^2 + 2R\Delta_{\mathbf{x}^\star}) + \left(\frac{2\eta}{\alpha} + 1\right)\tau R^2\Bigg],\ \forall \xi > 0 \quad (59)\end{aligned}$$

where $\Pi_\nabla \triangleq \sum_{t=1}^T \|\nabla f_t(\mathbf{x}_t^\star)\|^2$ is the accumulated squared gradients at the dynamic benchmark $\mathbf{x}_t^\star, t \in [T]$.

Proof: Under Assumption 4, we have

$$\begin{aligned}\mathrm{RE_d}(T) &\overset{(a)}{\leq} \sum_{t=1}^{T}\left([\nabla f_t(\mathbf{x}_t^\star)]^T(\mathbf{x}_t - \mathbf{x}_t^\star) + L\|\mathbf{x}_t - \mathbf{x}_t^\star\|^2\right) \\ &\overset{(b)}{\leq} \frac{1}{4\xi}\Pi_\nabla + (L + \xi)\sum_{t=1}^{T}\|\mathbf{x}_t - \mathbf{x}_t^\star\|^2 \quad (60)\end{aligned}$$

where (a) is because $f_t(\mathbf{x})$ is $2L$-smooth in (57), and (b) follows from $\mathbf{a}^T\mathbf{b} \leq \frac{1}{4\xi}\|\mathbf{a}\|^2 + \xi\|\mathbf{b}\|^2$ for any $\xi > 0$.

We now bound the RHS of (60). Note that

$$\sum_{t=1}^{T}\|\mathbf{x}_t-\mathbf{x}_t^\star\|^2 \le \sum_{t=1}^{\tau}\|\mathbf{x}_t-\mathbf{x}_t^\star\|^2+\sum_{t=\tau+1}^{T}\|\mathbf{x}_t-\mathbf{x}_t^\star\|^2$$
$$\overset{(a)}{\le} \tau R^2+\sum_{t=\tau+1}^{T}\|\mathbf{x}_t-\mathbf{x}_{t-\tau}^\star+\mathbf{x}_{t-\tau}^\star-\mathbf{x}_t^\star\|^2$$
$$\overset{(b)}{\le} \tau R^2+2\sum_{t=\tau+1}^{T}\left(\|\mathbf{x}_t-\mathbf{x}_{t-\tau}^\star\|^2+\|\mathbf{x}_t^\star-\mathbf{x}_{t-\tau}^\star\|^2\right)$$
$$\overset{(c)}{\le} \tau R^2+2\sum_{t=\tau+1}^{T}\|\mathbf{x}_t-\mathbf{x}_{t-\tau}^\star\|^2+2\tau^2\sum_{t=1}^{T}\|\mathbf{x}_t^\star-\mathbf{x}_{t-1}^\star\|^2 \quad (61)$$

where (a) follows from (13) that $\mathcal{X}_0$ is assumed to be bounded, (b) is because $\|\mathbf{a}-\mathbf{b}\|^2\le 2(\|\mathbf{a}\|^2+\|\mathbf{b}\|^2)$, and (c) follows from $\|\sum_{i=1}^n \mathbf{a}_i\|^2 \le n\sum_{i=1}^n\|\mathbf{a}_i\|^2$, which leads to $\|\mathbf{x}_t^\star-\mathbf{x}_{t-\tau}^\star\|^2\le\tau\sum_{i=1}^{\tau}\|\mathbf{x}_{t-\tau+i}^\star-\mathbf{x}_{t-\tau+i-1}^\star\|^2$.

We then bound the second term at the RHS of (61). The objective function of $\hat{\mathbf{P2}}$ is $2(\alpha+\eta)$-strongly convex over $\mathcal{X}_0$ for any $t>\tau$ due to the double regularization. We now show that (25) in the proof of Theorem 1 still holds by replacing $\mathbf{x}_{t-\tau}$ with $\hat{\mathbf{x}}_{t-\tau}^M$:

$$\begin{aligned}&[\nabla f_{t-\tau}(\hat{\mathbf{x}}_{t-\tau}^M)]^T(\mathbf{x}_t-\hat{\mathbf{x}}_{t-\tau}^M)+\alpha\|\mathbf{x}_t-\hat{\mathbf{x}}_{t-\tau}^M\|^2\\ &\quad+[\mathbf{Q}_{t-1}+\gamma\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T[\gamma\mathbf{g}_{t-\tau}(\mathbf{x}_t)]+\eta\|\mathbf{x}_t-\mathbf{x}_{t-1}\|^2\\ &\le[\nabla f_{t-\tau}(\hat{\mathbf{x}}_{t-\tau}^M)]^T(\mathbf{x}_{t-\tau}^\star-\hat{\mathbf{x}}_{t-\tau}^M)\\ &\quad+\alpha(\|\mathbf{x}_{t-\tau}^\star-\hat{\mathbf{x}}_{t-\tau}^M\|^2-\|\mathbf{x}_t-\mathbf{x}_{t-\tau}^\star\|^2)\\ &\quad+\eta(\|\mathbf{x}_{t-\tau}^\star-\mathbf{x}_{t-1}\|^2-\|\mathbf{x}_t-\mathbf{x}_{t-\tau}^\star\|^2). \quad (62)\end{aligned}$$

Note from (57) that $f_{t-\tau}(\mathbf{x})$ is $2L$-smooth over $\mathcal{X}_0$, and we set $\alpha\ge L$. Thus, we have

$$\begin{aligned}f_{t-\tau}(\mathbf{x}_t)-f_{t-\tau}(\hat{\mathbf{x}}_{t-\tau}^M)-\alpha\|\mathbf{x}_t-\hat{\mathbf{x}}_{t-\tau}^M\|^2\\ \le[\nabla f_{t-\tau}(\hat{\mathbf{x}}_{t-\tau}^M)]^T(\mathbf{x}_t-\hat{\mathbf{x}}_{t-\tau}^M). \quad (63)\end{aligned}$$

From the convexity of $f_{t-\tau}(\mathbf{x})$, we have

$$[\nabla f_{t-\tau}(\hat{\mathbf{x}}_{t-\tau}^M)]^T(\mathbf{x}_{t-\tau}^\star-\hat{\mathbf{x}}_{t-\tau}^M)\le f_{t-\tau}(\mathbf{x}_{t-\tau}^\star)-f_{t-\tau}(\hat{\mathbf{x}}_{t-\tau}^M). \quad (64)$$

Applying (63) and (64) to the LHS and RHS of (62), respectively, we have

$$\begin{aligned}&f_{t-\tau}(\mathbf{x}_t)-f_{t-\tau}(\hat{\mathbf{x}}_{t-\tau}^M)-\alpha\|\mathbf{x}_t-\hat{\mathbf{x}}_{t-\tau}^M\|^2+\alpha\|\mathbf{x}_t-\hat{\mathbf{x}}_{t-\tau}^M\|^2\\ &\quad+[\mathbf{Q}_{t-1}+\gamma\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T[\gamma\mathbf{g}_{t-\tau}(\mathbf{x}_t)]+\eta\|\mathbf{x}_t-\mathbf{x}_{t-1}\|^2\\ &\le f_{t-\tau}(\mathbf{x}_{t-\tau}^\star)-f_{t-\tau}(\hat{\mathbf{x}}_{t-\tau}^M)\\ &\quad+\alpha(\|\mathbf{x}_{t-\tau}^\star-\hat{\mathbf{x}}_{t-\tau}^M\|^2-\|\mathbf{x}_t-\mathbf{x}_{t-\tau}^\star\|^2)\\ &\quad+\eta(\|\mathbf{x}_{t-\tau}^\star-\mathbf{x}_{t-1}\|^2-\|\mathbf{x}_t-\mathbf{x}_{t-\tau}^\star\|^2). \quad (65)\end{aligned}$$

Rearranging terms of (65), we have

$$\begin{aligned}&\alpha\|\mathbf{x}_t-\mathbf{x}_{t-\tau}^\star\|^2\\ &\le\alpha\|\hat{\mathbf{x}}_{t-\tau}^M-\mathbf{x}_{t-\tau}^\star\|^2+f_{t-\tau}(\mathbf{x}_{t-\tau}^\star)-f_{t-\tau}(\mathbf{x}_t)\\ &\quad-[\mathbf{Q}_{t-1}+\gamma\mathbf{g}_{t-\tau-1}(\mathbf{x}_{t-1})]^T[\gamma\mathbf{g}_{t-\tau}(\mathbf{x}_t)]-\eta\|\mathbf{x}_t-\mathbf{x}_{t-1}\|^2\\ &\quad+\eta(\|\mathbf{x}_{t-\tau}^\star-\mathbf{x}_{t-1}\|^2-\|\mathbf{x}_t-\mathbf{x}_{t-\tau}^\star\|^2). \quad (66)\end{aligned}$$

We next bound the first term on RHS of (66). Applying Lemma 5 to the update of $\hat{\mathbf{x}}_{t-\tau}^m$ in (54), for any $\alpha\ge L$, we have

$$\|\hat{\mathbf{x}}_{t-\tau}^m-\mathbf{x}_{t-\tau}^{\mathrm{st}}\|^2\le\rho\|\hat{\mathbf{x}}_{t-\tau}^{m-1}-\mathbf{x}_{t-\tau}^{\mathrm{st}}\|^2,\quad\forall m\in[M]$$

where $\rho=\frac{\alpha-\mu}{\alpha+\mu}<1$. Combining the above M inequalities and noting that $\hat{\mathbf{x}}_{t-\tau}^0=\mathbf{x}_{t-\tau}$ by initialization, we have

$$\|\hat{\mathbf{x}}_{t-\tau}^M-\mathbf{x}_{t-\tau}^{\mathrm{st}}\|^2\le\rho^M\|\mathbf{x}_{t-\tau}-\mathbf{x}_{t-\tau}^{\mathrm{st}}\|^2.$$

From the above inequality and using $\|\mathbf{a}+\mathbf{b}\|^2\le 2(\|\mathbf{a}\|^2+\|\mathbf{b}\|^2)$, we have

$$\begin{aligned}&\|\hat{\mathbf{x}}_{t-\tau}^M-\mathbf{x}_{t-\tau}^\star\|^2\le 2\|\hat{\mathbf{x}}_{t-\tau}^M-\mathbf{x}_{t-\tau}^{\mathrm{st}}\|^2+2\|\mathbf{x}_{t-\tau}^{\mathrm{st}}-\mathbf{x}_{t-\tau}^\star\|^2\\ &\le 2\rho^M\|\mathbf{x}_{t-\tau}-\mathbf{x}_{t-\tau}^{\mathrm{st}}\|^2+2\|\mathbf{x}_{t-\tau}^{\mathrm{st}}-\mathbf{x}_{t-\tau}^\star\|^2\\ &\le 4\rho^M\|\mathbf{x}_{t-\tau}-\mathbf{x}_{t-\tau}^\star\|^2+(4\rho^M+2)\|\mathbf{x}_{t-\tau}^{\mathrm{st}}-\mathbf{x}_{t-\tau}^\star\|^2. \quad (67)\end{aligned}$$

For the second and third terms at the RHS of (66), we have

$$\begin{aligned}&f_{t-\tau}(\mathbf{x}_{t-\tau}^\star)-f_{t-\tau}(\mathbf{x}_t)\overset{(a)}{\le}f_{t-\tau}(\mathbf{x}_{t-\tau}^\star)-f_{t-\tau}(\mathbf{x}_{t-\tau}^{\mathrm{st}})\\ &\overset{(b)}{\le}[\nabla f_{t-\tau}(\mathbf{x}_{t-\tau}^\star)]^T(\mathbf{x}_{t-\tau}^{\mathrm{st}}-\mathbf{x}_{t-\tau}^\star)\overset{(c)}{\le}\frac{D^2}{4\alpha}+\alpha\|\mathbf{x}_{t-\tau}^{\mathrm{st}}-\mathbf{x}_{t-\tau}^\star\|^2 \quad (68)\end{aligned}$$

where (a) follows from the definition of $\mathbf{x}_t^{\mathrm{st}}$ in (55), (b) is because $f_t(\mathbf{x})$ is convex over $\mathcal{X}_0$, and (c) follows from $\mathbf{a}^T\mathbf{b}\le\frac{1}{4\alpha}\|\mathbf{a}\|^2+\alpha\|\mathbf{b}\|^2$ for any $\alpha>0$ and Assumption 1 that $\nabla f_t(\mathbf{x})$ is bounded as in (9).

From the proof of Theorem 1, we can bound the last two terms at the RHS of (66) by (30) and (27), respectively. Substituting (27), (30), (67), (68) into (66), we have

$$\begin{aligned}&\alpha\|\mathbf{x}_t-\mathbf{x}_{t-\tau}^\star\|^2\\ &\overset{(a)}{\le}4\alpha\rho^M\|\mathbf{x}_{t-\tau}-\mathbf{x}_{t-\tau}^\star\|^2+\alpha(4\rho^M+3)\|\mathbf{x}_{t-\tau}^{\mathrm{st}}-\mathbf{x}_{t-\tau}^\star\|^2\\ &\quad-\eta\|\mathbf{x}_{t-\tau+1}^\star-\mathbf{x}_{t-\tau}^\star\|^2+\frac{D^2}{4\alpha}-\Delta_{t-1}+\frac{\gamma^2}{2}\varphi_{t-\tau}+\gamma^2\varpi_{t-\tau}\\ &\quad+\eta(\Psi_{t-\tau}+2R\psi_{t-\tau}) \quad (69)\end{aligned}$$

where (a) follows from $\eta\ge B^2\gamma^2$, which leads to $(\gamma^2B^2-\eta)\|\mathbf{x}_t-\mathbf{x}_{t-1}\|^2\le 0$. Dividing both sides of (69) by α and summing them over $t\in\{\tau+1,\ldots,T\}$, we have

$$\begin{aligned}&\sum_{t=\tau+1}^{T}\|\mathbf{x}_t-\mathbf{x}_{t-\tau}^\star\|^2\\ &\overset{(a)}{\le}4\rho^M\sum_{t=1}^{T}\|\mathbf{x}_t-\mathbf{x}_t^\star\|^2+(4\rho^M+3)\sum_{t=1}^{T}\|\mathbf{x}_t^{\mathrm{st}}-\mathbf{x}_t^\star\|^2\\ &\quad-\frac{\eta}{\alpha}\sum_{t=2}^{T-\tau+1}\|\mathbf{x}_t^\star-\mathbf{x}_{t-1}^\star\|^2+\frac{D^2}{4\alpha^2}T+\frac{\gamma^2}{2\alpha}G^2+\frac{\gamma^2}{\alpha}\Delta_{2,\mathbf{g}}\\ &\quad+\frac{\eta}{\alpha}(R^2+2R\Delta_{\mathbf{x}^\star}) \quad (70)\end{aligned}$$

where (a) follows from the fact that the sums of Δ_{t-1}, $\varphi_{t-\tau}$, $\varpi_{t-\tau}$, $\Psi_{t-\tau}$ and $\psi_{t-\tau}$ over $t\in\{\tau+1,\ldots,T\}$ are upper bounded by 0, G^2, $\Delta_{2,\mathbf{g}}$, R^2, and $\Delta_{\mathbf{x}^\star}$, respectively, similar to (33).

Substituting (70) into (61) and rearranging terms, we have

$$\begin{aligned}&(1-8\rho^M)\sum_{t=1}^{T}\|\mathbf{x}_t-\mathbf{x}_t^\star\|^2\\ &\overset{(a)}{\le}2(4\rho^M+3)\sum_{t=1}^{T}\|\mathbf{x}_t^{\mathrm{st}}-\mathbf{x}_t^\star\|^2+\frac{D^2}{2\alpha^2}T+\frac{\gamma^2}{\alpha}G^2+\frac{2\gamma^2}{\alpha}\Delta_{2,\mathbf{g}}\\ &\quad+\left(\frac{2\eta}{\alpha}+1\right)\tau R^2+\frac{2\eta}{\alpha}(R^2+2R\Delta_{\mathbf{x}^\star}) \quad (71)\end{aligned}$$

TABLE III
IMPROVED DYNAMIC REGRET AND CONSTRAINT VIOLATION OF DTC-OCO WITH MULTI-STEP GRADIENT DESCENT

Constraint	Know δ?	$\mathrm{RE_d}(T)$	$\mathrm{VO}^c(T)$
Varying	Yes/No	$\mathcal{O}(\max\{\tau^2 T^{\delta}, T^{\nu}\})$	$\mathcal{O}\left(\tau T^{\kappa}\right)$
Invariant	Yes/No	$\mathcal{O}(\tau^2 T^{\delta})$	$\mathcal{O}(\tau)$

where (a) follows from $\eta \geq \alpha\tau^2$, which leads to $\tau^2 \sum_{t=1}^{T} \|\mathbf{x}_t^\star - \mathbf{x}_{t-1}^\star\|^2 - \frac{\eta}{\alpha} \sum_{t=2}^{T-\tau+1} \|\mathbf{x}_t^\star - \mathbf{x}_{t-1}^\star\|^2 \leq \frac{\eta}{\alpha}\tau R^2$.

Note that since $M > \log_\rho(\frac{1}{8})$, we have $8\rho^M < 1$. Dividing both sides of (71) by $1 - 8\rho^M$, and applying the resulting inequality to the second term in (60), we have (59). ■

Note that the dynamic regret bound in (59) of Theorem 4 reduces as the number of gradient descent steps M increases. However, even if $M \to \infty$, the dynamic regret bound is still fundamentally limited by the system dynamics. Later in Section VI-B, we show that although the accumulated loss by DTC-OCO decreases as M increases, the decreasing amount becomes negligible as M becomes large enough. In practice, the choice of M may depend on the trade off between the computational capacity of the decision maker and the actual performance gain provided by multi-step gradient descent.

Regarding the performance on constraint violation, the following theorem shows that the constraint violation bound (37) in Theorem 3 also holds for DTC-OCO with multi-step gradient descent. The theorem can be straightforwardly proven by replacing $\mathbf{x}_{t-\tau}$ with $\hat{\mathbf{x}}_{t-\tau}^M$ in the proof of Theorem 3. Thus, we omit the proof for brevity.

Theorem 5. Under Assumptions 1-4, the constraint violation bound in (37) holds for DTC-OCO with multi-step gradient descent.

C. Improved Dynamic Regret and Constraint Violation Bounds

We now discuss the sufficient conditions for DTC-OCO with multi-step gradient descent to yield sublinear dynamic regret and constraint violation, and highlight its performance advantage. For clarity, we summarize the performance bounds in terms of the growth rate over T for strongly convex loss functions in Table III.

From Theorems 4 and 5, we have the following corollaries regarding the dynamic regret and constraint violation bounds for DTC-OCO with multi-step gradient descent. Corollaries 4 and 5 provide two sets of performance bounds depending on whether the long-term constraint functions are time-varying or not. The proofs of these two corollaries can be obtained by substituting the expression of the algorithm parameters α, η, γ, and M, specified in each corollary, into the bounds in (59) and (37). Thus, they are omitted for brevity.

Corollary 4. Suppose $\Pi_\nabla = \mathcal{O}(T^\delta)$ and $\Pi_{\mathbf{x}} = \mathcal{O}(T^\delta)$.[8] Assume the constraint functions are time-varying. Let $\alpha = T^{\frac{1}{2}} + L$, $\eta = \max\{\alpha\tau^2, B^2\gamma^2\}$, and $\gamma^2 = \tau T^{\frac{1}{2}}$ in DTC-OCO with multi-step gradient descent. Then, for any $M > \log_\rho(\frac{1}{8})$,

$$\mathrm{RE_d}(T) = \mathcal{O}\left(\max\left\{\tau^2 T^{\delta}, T^{\nu}\right\}\right), \tag{72}$$

$$\mathrm{VO}^c(T) = \mathcal{O}\left(\tau T^{\kappa}\right). \tag{73}$$

Corollary 5. Suppose $\Pi_\nabla = \mathcal{O}(T^\delta)$ and $\Pi_{\mathbf{x}} = \mathcal{O}(T^\delta)$. Assume the constraint functions are time-invariant. Let $\alpha = T^{\frac{1}{2}} + L$, $\eta = \max\{\alpha\tau^2, B^2\gamma^2\}$, and $\gamma^2 = \tau T^{\frac{1}{2}}$ in DTC-OCO with multi-step gradient descent. Then, for any $M > \log_\rho(\frac{1}{8})$,

$$\mathrm{RE_d}(T) = \mathcal{O}\left(\tau^2 T^{\delta}\right), \tag{74}$$

$$\mathrm{VO}^c(T) = \mathcal{O}\left(\tau\right). \tag{75}$$

With time-varying constraint functions, recall from Corollary 1 that for DTC-OCO with a single-step gradient descent, the dynamic regret is $\mathcal{O}(\max\{\tau^{\frac{1}{2}} T^{\frac{1+\delta}{2}}, T^{\nu}\})$ and the constraint violation is $\mathcal{O}(\max\{T^{\frac{1-\delta}{2}}, \tau T^{\kappa}\})$. Corollary 4 indicates that by increasing the number of gradient descent steps M, the dynamic regret is improved to $\mathcal{O}(\max\{\tau^2 T^{\delta}, T^{\nu}\})$ and the constraint violation is improved to $\mathcal{O}\left(\tau T^{\kappa}\right)$.

When the constraint functions are time-invariant, we have shown in Corollary 3 that DTC-OCO with a single-step gradient descent achieves $\mathcal{O}(\tau^{\frac{1}{2}} T^{\frac{1+\delta}{2}})$ dynamic regret and $\mathcal{O}(\tau)$ constraint violation. Corollary 5 indicates that performing multi-step gradient descent can improve the dynamic regret to $\mathcal{O}(\tau^2 T^{\delta})$ while maintaining $\mathcal{O}(\tau)$ constraint violation. Note that different scales of τ in the dynamic regrets are the results of different proof techniques to show the desired dynamic regret growth rates. As a constant, τ does not impact the growth rates of the dynamic regret or constraint violation.

In summary, our results above show that for a strongly convex loss function, the multi-step gradient descent with double regularization together improves the dynamic regret bound and the constraint violation bound of constrained OCO, even in the presence of multi-slot delay. To the best of our knowledge, no existing literature has considered performing multi-step gradient descent to improve the performance bounds for OCO with long-term constraints, even with one-slot delay.

Finally, we also point out that in the case of a strongly convex loss function, with multi-step gradient descent, Corollaries 4 and 5 show that the optimal algorithm parameters α, η, γ, and M do not depend on the accumulated variation measure δ on the dynamic benchmark $\{\mathbf{x}_t^\star\}$. This is in contrast to Corollaries 1 and 3 for the basic form of DTC-OCO, where the optimal α, η, γ may depend on δ.

VI. APPLICATION TO NETWORK RESOURCE ALLOCATION

DTC-OCO can be applied to many applications in the areas of wireless communications, machine learning, mobile computing, and smart grid, as mentioned in the introduction. In this section, we apply DTC-OCO to a general problem of resource allocation in a networked system [17], [18], [24], [35], [36]. We present numerical results to demonstrate the performance advantage of DTC-OCO over the current best alternative given in [22].

[8] The accumulated squared gradients Π_∇ can be very small [11]. Particularly, if $\mathbf{x}_t^\star$ is an interior point of the convex set $\mathcal{X}_0$ or there is no short-term constraint, we have $\nabla f_t(\mathbf{x}_t^\star) = \mathbf{0}, \forall t$ and thus $\Pi_\nabla = 0$. The accumulated squared difference between $\{\mathbf{x}_t^{\mathrm{st}}\}$ and $\{\mathbf{x}_t^\star\}$ can also be small. In particular, if $\|\mathbf{x}_t^{\mathrm{st}} - \mathbf{x}_t^\star\| \propto T^{\frac{1-\delta}{2}}, \forall t$, we have $\Pi_{\mathbf{x}} = \mathcal{O}(T^\delta)$.

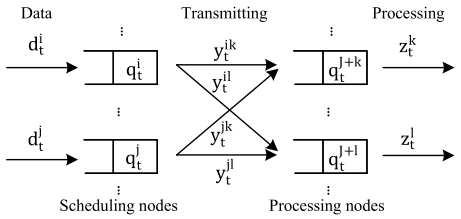


Fig. 3. An illustration of general online network resource allocation.

A. Online Network Resource Allocation

Fig. 3 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 machine learning, the scheduling nodes may be mobile devices, and the processing nodes may be parameter servers. 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 $\mathbf{d}_t = [d_t^1, \ldots, d_t^J, \mathbf{0}_{1\times K}]^T$. A central controller decides the transmission rate y_t^{jk} of the link (j,k) connecting scheduling node j and processing node k, as well as the processing rate z_t^k at processing node k. In a compact form, the decision vector at time t is $\mathbf{x}_t = [y_t^{11}, \ldots, y_t^{JK}, z_t^1, \ldots, z_t^K]^T$. 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 in the convex set as

$$\mathcal{X}_0 \triangleq \{\mathbf{x} | \mathbf{0} \preceq \mathbf{x} \preceq \mathbf{x}_{\max}\}$$

where $\mathbf{x}_{\max} = [y_{\max}^{11}, \ldots, y_{\max}^{JK}, z_{\max}^1, \ldots, z_{\max}^K]^T$ 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. Denote the queue backlog vector as $\mathbf{q}_t = [q_t^1, \ldots q_t^J, q_t^{J+1}, \ldots, q_t^{J+K}]^T$. Then, we can express the update of the queue backlog 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

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

The goal for the central controller is to minimize the network cost while controlling the long-term averaged outgoing data rate to be no less than the incoming data rate to maintain queue stability. Since the controller can only receive delayed feedback of system parameters $\mathbf{d}_t$, $\mathbf{q}_t$, and $f_t(\mathbf{x})$ from the scheduling nodes and processing nodes over time, it needs to employ an online control solution based on the feedback. This online network resource allocation problem is a special case of the OCO problem **P1**, where the convex set $\mathcal{X}_0$ for the decisions is defined above, the convex loss functions $f_t(\mathbf{x})$ are the network cost functions, and the convex constraint functions are given by

$$\mathbf{g}_t(\mathbf{x}) \triangleq \mathbf{C}\mathbf{x} + \mathbf{d}_t,$$

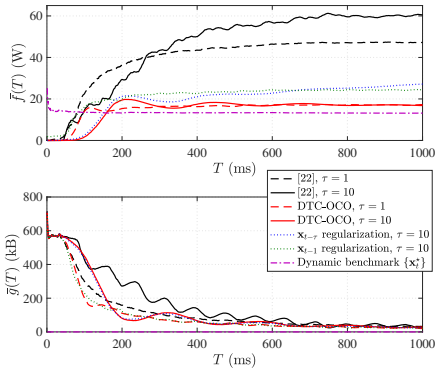
which represents the net change in the queue backlog due to incoming and outgoing data. Due to possible communication delay between the scheduling nodes and the central controller, we consider that the feedback information of $\mathbf{g}_t(\mathbf{x})$ is delayed for τ time slots at the central controller. Note that in this network resource allocation problem with data queues to achieve sublinear constraint violation, *i.e.,* $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T \mathbf{g}_t(\mathbf{x}_t) \to \mathbf{0}$, is equivalent to maintaining queue stability.

A special case of the above problem where there is no feedback delay has been considered using Lyapunov optimization techniques [24], [35], [36]. Furthermore, solutions for the standard OCO setting with one-slot feedback delay have been proposed in [17] and [18]. However, in practical systems, due to limited resources for feedback, the central controller typically experiences multi-slot feedback delay of the system parameters. The proposed DTC-OCO algorithm provides a suitable online solution to this problem.

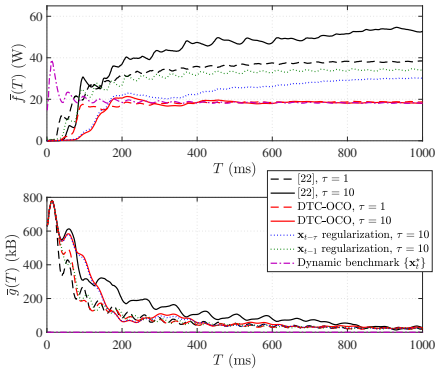
B. Numerical Performance Evaluation

We apply DTC-OCO to the above online network resource allocation problem. The analysis of sublinear regret bounds derived in Section IV implies that the algorithm can produce an efficient solution. Furthermore, the sublinear constraint violation bound guarantees queue stability. Now, we further study the numerical performance of DTC-OCO in this practical problem setting. We compare DTC-OCO with the online algorithm proposed in [22], which is the only existing algorithm to accommodate long-term constraints and multi-slot feedback delay. To further verify the benefit of the proposed double regularization in DTC-OCO, we also study two simplified versions of DTC-OCO that apply a single regularization term on either $\mathbf{x}_{t-\tau}$ or $\mathbf{x}_{t-1}$.

We consider a specific example of a mobile cloud computing system consisting of $J = 10$ scheduling nodes, and $K = 10$ processing nodes. Following the typical long-term evolution (LTE) specifications [37], 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 data arrived at each time t is $\mathbf{d}_t$ in kB. The maximum data transmission and processing rates, in MBps, are set randomly with uniform distribution as $y_{\max}^{jk} \sim \mathcal{U}(10, 100)$ and $z_{\max}^k \sim \mathcal{U}(100, 250)$, respectively. Based on the channel capacity, we can express the transmission power of each link (j,k) in terms of its transmission rate y^{jk} as $\frac{\sigma_n^2}{L_t^{jk}}(2^{\frac{y^{jk}}{B_W}} - 1)$, where $\sigma_n^2 = N_0 B_W + N_F[\text{dBm}]$ is the noise power, and L_t^{jk} captures the path-loss, instantaneous channel gain, and the gap to capacity. Note that L_t^{jk} is time-varying due to the fluctuation of wireless channels. We assume each processing node k follows a quadratic power-frequency relationship, where the power consumed for processing rate z^k is given by $\theta(\xi_t^k z^k)^2$, where $\theta = 120$ W/(GHz)^2 [38], and ξ_t^k represents the computational complexity of the computing tasks [39], [40]. Since the computing tasks change over time, ξ_t^k is time-varying [41]. We consider both the data transmission power and processing power for given decision $\mathbf{x}$ on transmission rate and processing rate at time t by defining the following



(a) I.i.d. over time.



(b) Noisy periodic variation over time.

Fig. 4. $\bar{f}(T)$ and $\bar{g}(T)$ vs. T under different feedback delay τ. The static benchmark $\mathbf{x}^\star$ is not included as its performance is much worse than the one by the dynamic benchmark $\{\mathbf{x}_t^\star\}$.

network cost function:

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

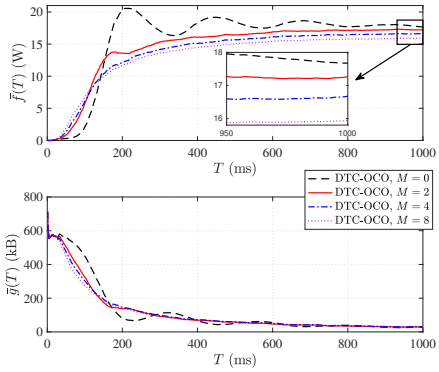
Note that information on the channel gain and task complexity at the central controller may be severely delayed due to limited resources for feedback. Therefore, we consider the feedback information of $f_t(\mathbf{x})$ is delayed for τ time slots. Furthermore, our proposed online solution for network resource allocation can be applied to other cost functions that are convex w.r.t. $\mathbf{x}$.

We assume both τ and δ are unknown, and thus set $\alpha = T^{\frac{1}{2}}$, $\eta = \|\mathbf{C}\|_2^2$, and $\gamma = 1$ in DTC-OCO.[9] 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 of $\bar{f}(T)$, the step-sizes for the algorithm in [22] are selected such that the algorithm has a steady-state value of $\bar{g}(T)$ similar to that of DTC-OCO. We consider two different models for the time-varying system parameters $\{d_t^j\}$, $\{L_t^{jk}\}$, $\{\xi_t^k\}$.

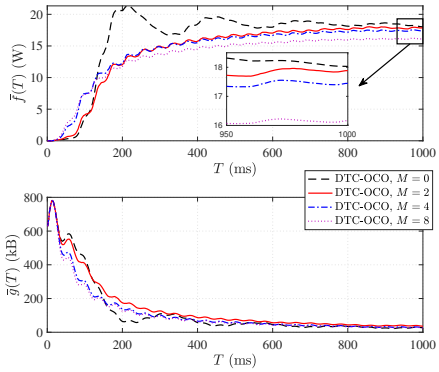
1) I.i.d. over time: All the system parameters are i.i.d. over t with uniform distribution: $d_t^j \sim \mathcal{U}(10, 100)$, $L_t^{jk}[\text{dB}] \sim \mathcal{U}(-126, -120)$, and $\xi_t^k \sim \mathcal{U}(1, 3)$.

2) Noisy periodic variation over time: We assume the variation of each system parameter over time is periodic with some additive noise that is uniformly distributed: $d_t^j = 30\sin\left(\frac{\pi t}{20}\right) + n_t^{j,d}$, $L_t^{jk}[\text{dB}] = -120 - 3\sin\left(\frac{\pi t}{20}\right) - n_t^{jk,L}$, and $\xi_t^k = 0.5\sin(\frac{\pi t}{20}) + n_t^{k,\xi}$, where $n_t^{j,d} \sim \mathcal{U}(40, 70)$, $n_t^{jk,L} \sim \mathcal{U}(6, 9)$, and $n_t^{k,\xi} \sim \mathcal{U}(1, 3)$.

[9]We consider τ and δ are unknown, since they may be difficult to obtain in practice. For our specific application to network resource allocation, following Corollary 2 to set the algorithm parameters α, η, and γ already provides DTC-OCO with relatively good performance. Therefore, we do not further tune α, η, and γ to reach the best performance. In general, the best algorithm parameters are problem dependent and may require fine tuning.



(a) I.i.d. over time.



(b) Noisy periodic variation over time.

Fig. 5. $\bar{f}(T)$ and $\bar{g}(T)$ vs. T under different gradient descent steps M.

Fig. 4 shows time-averaged cost $\bar{f}(T)$ and constraint violation $\bar{g}(T)$ over T at different values of feedback delay τ. for the above two cases. Fig. 4a is under the i.i.d. modeling of system parameters and Fig. 4b is for noisy periodic variation of system parameters. We observe that the network cost by DTC-OCO can approach that of the dynamic benchmark in (5), indicating that sublinear dynamic regret is achieved. Compared with [22], DTC-OCO achieves a much lower network cost and is much more tolerant to the feedback delay τ. The reason for this result are two fold: First, in DTC-OCO, the constraint function $\mathbf{g}_{t-\tau}(\mathbf{x})$ is penalized directly instead of its first-order approximation as in [22]. This improves the control on the constraint violation; Second, the double regularization approach in DTC-OCO prevents the online decision $\mathbf{x}_t$ from being far way from either $\mathbf{x}_{t-\tau}$ or $\mathbf{x}_{t-1}$. This results in an improved algorithm performance, since both $\mathbf{x}_{t-\tau}$ and $\mathbf{x}_{t-1}$ provide useful information in minimizing the accumulated loss and constraint violation. We also observe that the network cost by DTC-OCO is much lower than those by the two simplified versions of DTC-OCO using single regularization. This indicates that the double regularization approach is essential in the superior performance demonstrated by DTC-OCO. Our simulation result shows that the time-averaged cost $\bar{f}(T)$ by DTC-OCO can be very close to that by the dynamic benchmark; Furthermore, $\bar{g}(T)$ by DTC-OCO decreases over T. These results validate our theoretical analysis that DTC-OCO can achieve sublinear dynamic regret and constraint violation.

We further evaluate the performance of DTC-OCO with multi-step gradient descent for the two models of time-varying system parameters. We set the feedback delay $\tau = 10$. Fig. 5 shows $\bar{f}(T)$ and $\bar{g}(T)$ over T under different numbers of gradient descent steps M. The case of $M = 0$ represents the basic form of the DTC-OCO algorithm. We use the same algorithm parameters α, η, γ as in the previous experiments. We observe that, as M increases, the steady-state value of $\bar{f}(T)$ decreases for both models of time-varying system parameters. At the same time, the fluctuations of the network cost also reduces with M. The reason is that, as shown in Theorem 4, performing multi-step gradient descent improves the dynamic regret bound, which in turn improves the convergence behavior. We also observe that the impact of different values of M on the time-averaged constraint violation is small. This confirms the results in Theorem 5. These simulation results validate that enabling multi-step gradient descent in DTC-OCO can further improve system performance.

VII. Conclusions

This paper considers OCO with short-term and long-term constraints under multi-slot feedback delay. We have proposed an efficient algorithm, DTC-OCO, where we use a novel constraint penalty with double regularization to handle the asynchrony between information feedback and decision updates. Our analysis on the regret bound and the constraint violation bound takes into account the impact of multi-slot feedback delay and the double regularization structure on the performance of DTC-OCO, and we have derived conditions under which the algorithm achieves both sublinear dynamic and static regrets and sublinear constraint violation. We have further extended the DTC-OCO algorithm to enable multi-step gradient descent in each update, which is shown to improve both the dynamic regret bound and the constraint violation bound for strongly convex loss functions. We apply DTC-OCO to online network resource allocation in a mobile cloud computing system as an numerical example. Simulation results demonstrate that DTC-OCO provides substantial performance advantage over the stat-of-art alternative, yielding a much lower cost and superior capability to tolerate feedback delay.

References

[1] J. Wang, B. Liang, M. Dong, G. Boudreau, and H. Abou-zeid, "Delay-tolerant constrained OCO with application to network resource allocation," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, 2021.

[2] S. Shalev-Shwartz, "Online learning and online convex optimization," *Found. Trends Mach. Learn.*, vol. 4, pp. 107–194, Feb. 2012.

[3] E. Hazan, "Introduction to online convex optimization," *Found. Trends Optim.*, vol. 2, pp. 157–325, Aug. 2016.

[4] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proc. Intel. Conf. Mach. Learn. (ICML)*, 2003.

[5] E. Hazan, A. Agarwal, and S. Kale, "Logarithmic regret algorithms for online convex optimization," *Mach. Learn.*, vol. 69, pp. 169–192, 2007.

[6] 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.

[7] 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.

[8] A. Mokhtari, S. Shahrampour, A. Jababaie, and A. Ribeiro, "Online optimization in dynamic environments: Improved regret rates for strongly convex problems," in *Proc. IEEE Conf. Decision Control (CDC)*, 2016.

[9] A. S. Bedi, P. Sarma, and K. Rajawat, "Tracking moving agents via inexact online gradient descent algorithm," *IEEE J. Sel. Topics Signal Process.*, vol. 12, pp. 202–217, 2018.

[10] R. Dixit, A. S. Bedi, R. Tripathi, and K. Rajawat, "Online learning with inexact proximal online gradient descent algorithms," *IEEE Trans. Signal Process.*, vol. 67, pp. 1338–1352, 2019.

[11] L. Zhang, T. Yang, J. Yi, R. Jin, and Z. Zhou, "Improved dynamic regret for non-degenerate functions," in *Proc. Adv. Neural Info. Proc. Sys. (NIPS)*, 2017.

[12] 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.

[13] 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.

[14] 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.

[15] H. Yu, M. J. Neely, and X. Wei, "Online convex optimization with stochastic constraints," in *Proc. Adv. Neural Info. Proc. Sys. (NIPS)*, 2017.

[16] X. Wei, H. Yu, and M. J. Neely, "Online primal-dual mirror descent under stochastic constraints," *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 4, Jun. 2020.

[17] 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.

[18] 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.

[19] J. Langford, A. J. Smola, and M. Zinkevich, "Slow learners are fast," in *Proc. Adv. Neural Info. Proc. Sys. (NIPS)*, 2009.

[20] K. Quanrud and D. Khashabi, "Online learning with adversarial delays," in *Proc. Adv. Neural Info. Proc. Sys. (NIPS)*, 2015.

[21] H. B. McMahan and M. Streeter, "Delay-tolerant algorithms for asynchronous distributed online learning," in *Proc. Adv. Neural Info. Proc. Sys. (NIPS)*, 2014.

[22] X. Cao, J. Zhang, and H. V. Poor, "Constrained online convex optimization with feedback delays," *IEEE Trans. Automat. Contr.*, Oct. 2020.

[23] O. Besbes, Y. Gur, and A. Zeevi, "Non-stationary stochastic optimization," *Oper. Res.*, vol. 63, pp. 1227–1244, Sep. 2015.

[24] M. J. Neely, *Stochastic Network Optimization with Application on Communication and Queueing Systems*. Morgan & Claypool, 2010.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] M. Akbari, B. Gharesifard, and T. Linder, "Distributed online convex optimization on time-varying directed graphs," *IEEE Trans. Control Netw. Syst.*, vol. 4, pp. 417–428, Sep. 2017.

[31] 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.

[32] 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.

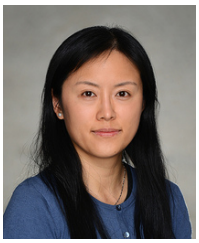
[33] X. Yi, X. Li, L. Xie, and K. H. Johansson, "Distributed online convex optimization with time-varying coupled inequality constraints," *IEEE Trans. Signal Process.*, vol. 68, pp. 731–746, Jan. 2020.

[34] L. Zhang, S. Lu, and Z. Zhou, "Adaptive online learning in dynamic environments," in *Proc. Adv. Neural Info. Proc. Sys. (NIPS)*, 2018.

[35] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Found. Trends Netw.*, pp. 1–144, 2006.

[36] 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.
[37] H. Holma and A. Toskala, *WCDMA for UMTS - HSPA evolution and LTE*. John Wiely & Sons, 2010.
[38] R. Urgaonkar, 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.
[39] 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.
[40] S. Sundar and B. Liang, "Offloading dependent tasks with communication delay and deadline constraint," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, 2018.
[41] 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.

Juncheng Wang (Student Member, IEEE) received the B.Eng. degree in Electrical Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2014, and the M.Sc. degree in Electrical and Computer Engineering from the University of Alberta, Edmonton, AB, Canada, in 2017. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. His research interests include online learning, stochastic optimization, resource allocation, and network virtualization.



Min Dong (Senior Member, IEEE) received the B.Eng. degree from Tsinghua University, Beijing, China, in 1998, and the Ph.D. degree in electrical and computer engineering with a minor in applied mathematics from Cornell University, Ithaca, NY, in 2004. From 2004 to 2008, she was with Qualcomm Research, Qualcomm Inc., San Diego, CA. Since 2008, she has been with Ontario Tech University, where she is currently a Professor with the Department of Electrical, Computer, and Software Engineering and the Associate Dean (Academic) of the Faculty of Engineering and Applied Science. She also holds a status-only Professor appointment with the Department of Electrical and Computer Engineering at the University of Toronto. Her research interests include wireless communications, statistical signal processing, learning techniques, optimization and control applications in cyber-physical systems.

Dr. Dong received the Early Researcher Award from the Ontario Ministry of Research and Innovation in 2012, the Best Paper Award at IEEE ICCC in 2012, and the 2004 IEEE Signal Processing Society Best Paper Award. She is a co-author of the Best Student Paper at IEEE SPAWC 2021 and the Best Student Paper of Signal Processing for Communications and Networking at IEEE ICASSP 2016. She is an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. She served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2010-2014) and the IEEE SIGNAL PROCESSING LETTERS (2009-2013). She served on the Steering Committee of the IEEE TRANSACTIONS ON MOBILE COMPUTING (2019-2021). She was an elected member of the Signal Processing for Communications and Networking (SP-COM) Technical Committee of IEEE Signal Processing Society (2013-2018). She was the lead Co-Chair of the Communications and Networks to Enable the Smart Grid Symposium at the IEEE International Conference on Smart Grid Communications in 2014.

Ben Liang (Fellow, IEEE) received honors-simultaneous B.Sc. (valedictorian) and M.Sc. degrees in Electrical Engineering from Polytechnic University (now the engineering school of New York University) in 1997 and the Ph.D. degree in Electrical Engineering with a minor in Computer Science from Cornell University in 2001. He was a visiting lecturer and postdoctoral research associate at Cornell University in the 2001 - 2002 academic year. He joined the Department of Electrical and Computer Engineering at the University of Toronto in 2002, where he is now Professor and L. Lau Chair in Electrical and Computer Engineering. His current research interests are in networked systems and mobile communications. He is an associate editor for the IEEE Transactions on Mobile Computing and has served on the editorial boards of the IEEE Transactions on Communications, the IEEE Transactions on Wireless Communications, and the Wiley Security and Communication Networks. He regularly serves on the organizational and technical committees of a number of conferences. He is a Fellow of IEEE and a member of ACM and Tau Beta Pi.

Gary Boudreau (Senior Member, IEEE) received a B.A.Sc. in Electrical Engineering from the University of Ottawa in 1983, an M.A.Sc. in Electrical Engineering from Queens University in 1984 and a Ph.D. in electrical engineering from Carleton University in 1989. From 1984 to 1989 he was employed as a communications systems engineer with Canadian Astronautics Limited and from 1990 to 1993 he worked as a satellite systems engineer for MPR Teltech Ltd. For the period spanning 1993 to 2009 he was employed by Nortel Networks in a variety of wireless systems and management roles within the CDMA and LTE basestation product groups. In 2010 he joined Ericsson Canada where he is currently Director of RAN Architecture and Performance in the North American CTO office. His interests include digital and wireless communications, signal processing and machine learning.

Hatem Abou-Zeid (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Queens University in 2014. From 2015 to 2017, he was at Cisco Systems designing scalable traffic engineering and routing solutions for service provider networks. Prior to joining the University of Calgary, he was at Ericsson Canada for 4 years leading 5G radio access designs and contributing to intellectual property development in the areas of RAN intelligence, low latency communications, spectrum sharing, and multi-radio access technologies. Several wireless access algorithms and traffic engineering techniques that he co-invented and co-developed are deployed in 5G mobile networks and data centers worldwide. He is also an Adjunct Professor at Queens University, Ontario Tech University, and Carleton University in Canada.

Dr. Abou-Zeids research interests are broadly in 5G/6G wireless networking, extended reality and aerial communications, and robust machine learning for networks. His work has resulted in 18 patent filings and 50 journal and conference publications in several IEEE flagship venues, and he is a co-author of a Best Paper Award at IEEE ICC 2022. He is an avid supporter of industry-university partnerships and he served on the Ericsson Government Industry Relations and Talent Development Committees where he directed multiple academic research partnerships. He also served on the TPC of several IEEE Conferences and was the Co-Chair of the IEEE ICC Workshop on Wireless Network Innovations for Mobile Edge Learning, and currently serves as Corporate Co-Chair of the IEEE LCN Conference 2022. He proposed the concepts of predictive resource allocation for video streaming in mobile networks at Queens University and was awarded the DAAD RISE Fellowship at Bell Labs and nominated for an Outstanding Thesis Medal.