Distributed Bandit Online Convex Optimization With Time-Varying Coupled Inequality Constraints

Xinlei Yi, Xiuxian Li, *Member, IEEE*, Tao Yang, *Senior Member, IEEE*, Lihua Xie, *Fellow, IEEE*, Tianyou Chai, *Fellow, IEEE*, and Karl Henrik Johansson, *Fellow, IEEE*

***Abstract*—Distributed bandit online convex optimization with time-varying coupled inequality constraints is considered, motivated by a repeated game between a group of learners and an adversary. The learners attempt to minimize a sequence of global loss functions and at the same time satisfy a sequence of coupled constraint functions, where the constraints are coupled across the distributed learners at each round. The global loss and the coupled constraint functions are the sum of local convex loss and constraint functions, respectively, which are adaptively generated by the adversary. The local loss and constraint functions are revealed in a bandit manner, i.e., only the values of loss and constraint functions are revealed to the learners at the sampling instance, and the revealed function values are held privately by each learner. Both one- and two-point bandit feedback are studied with the two corresponding distributed bandit online algorithms used by the learners. We show that sublinear expected regret and constraint violation are achieved by these two algorithms, if the accumulated variation of the comparator sequence also grows sublinearly. In particular, we show that $\mathcal{O}(T^{\theta})$ expected static regret and $\mathcal{O}(T^{7/4-\theta})$ constraint violation are achieved in the one-point bandit feedback setting, and $\mathcal{O}(T^{\max\{\kappa,1-\kappa\}})$ expected static regret and $\mathcal{O}(T^{1-\kappa/2})$ constraint violation in the two-point bandit feedback setting, where $\theta\in(3/4,5/6]$ and $\kappa\in(0,1)$ are user-defined tradeoff parameters. Finally, the tightness of the theoretical results is illustrated by numerical simulations of a simple power grid example, which also compares the proposed algorithms to algorithms existing in the literature.**

***Index Terms*—Bandit convex optimization, distributed optimization, gradient approximation, online optimization, time-varying constraints.**

Manuscript received May 26, 2020; accepted October 4, 2020. Date of publication October 13, 2020; date of current version September 27, 2021. This work was supported in part by the Knut and Alice Wallenberg Foundation the Swedish Foundation for Strategic Research, the Swedish Research Council, Ministry of Education of Republic of Singapore under Grant MoE Tier 1 RG72/19 and in part by the National Natural Science Foundation of China under Grant 61991403, Grant 61991404, and Grant 61991400, and 2020 Science and Technology Major Project of Liaoning Province under Grant 2020JH1/10100008. This article was presented in part at the 2020 American Control Conference, Sheraton Denver Downtown Hotel, Denver, CO, USA, July 2020. Recommended by Associate Editor J. Lavaei. *(Corresponding author: Tao Yang.)*

Xinlei Yi and Karl Henrik Johansson are with the Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden (e-mail: xinleiy@kth.se; kallej@kth.se).

Xiuxian Li and Lihua Xie are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore (e-mail: xxli@ieee.org; elhxie@ntu.edu.sg).

Tao Yang and Tianyou Chai are with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China (e-mail: yangtao@mail.neu.edu.cn; tychai@mail.neu.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2020.3030883

I. Introduction

Online convex optimization is a promising methodology for modeling sequential tasks and has important applications in machine learning [1], smart grids [2], sensor networks [3], [4], etc. It can be traced back to the 1990s [5]–[8]. Online convex optimization can be understood as a repeated game between a learner and an adversary [1]. At round t of the game, the learner chooses a point x_t from a known convex set $\mathbb{X}\subseteq\mathbb{R}^p$, where p is the dimension of the space. Then, the adversary observes x_t and chooses a convex loss function $f_t:\mathbb{R}^p\rightarrow\mathbb{R}$. After that, the loss function f_t is revealed to the learner who suffers a loss $f_t(x_t)$. Note that at each round, the loss function can be arbitrarily chosen by the adversary, especially with no probabilistic model imposed on the choices, which is the key difference between online and stochastic convex optimization. Such an adversary with the power to arbitrarily choose the loss functions is said to be a completely adaptive adversary [9]. The goal of the learner is to choose a sequence $\boldsymbol{x}_T=(x_1,\dots,x_T)$ such that his/her regret $\mathrm{Reg}(\boldsymbol{x}_T,\boldsymbol{y}_T)=\sum_{t=1}^{T}(f_t(x_t)-f_t(y_t))$ is minimized, where T is the total number of rounds and $\boldsymbol{y}_T=(y_1,\dots,y_T)$ is a comparator sequence. Over the past two decades, online convex optimization has been extensively studied, e.g., [1], [3], [4], [8], [10]–[19]. It has also been extended to distributed setting, e.g., [20]–[22], and nonconvex setting, e.g., [23]–[25]. All existing online algorithms require the knowledge of the entire loss function or the gradient of the loss function. In particular, it is known that the projection-based online gradient descent algorithm achieves an $\mathcal{O}(\sqrt{T})$ static regret bound for convex loss functions with bounded subgradients and that this is a tight bound up to constant factors [10].

Bandit online convex optimization is online convex optimization with bandit feedback, i.e., at each round, only the values of the loss functions are revealed, rather than the entire loss function, the gradient of the loss function, or some other information. Bandit feedback is suitable to model various applications, where the entire function or gradient information is not available, such as online source localization, online routing in data networks, and online advertisement placement in web search [26]. For such applications, existing online algorithms are inapplicable but gradient-free (zeroth-order) optimization methods are needed. Gradient-free optimization methods have a long history [27] and have an evident advantage since computing a function value is much simpler than computing its gradient. Gradient-free optimization methods have gained renewed interests in recent years, e.g., [28]–[31]. Essentially, bandit online convex optimization is a gradient-free method to solve convex optimization problems. In a bandit setting, a sublinear static regret bound may not be guaranteed if the adversary still can arbitrarily choose the loss function. Under completely adaptive adversary, Agarwal *et al.* [9] gave an example to show that any algorithm suffer at least linear regret. Therefore, the power of the adversary should be limited to achieve a sublinear regret bound. For a so-called adaptive adversary [9], the adversary chooses f_t based only on the learner's past decisions $x_1, \dots, x_{t-1}$, but not on his/her current decision x_t. In other words, the adversary chooses f_t at the beginning of round t, before the learner chooses his/her decision.

A key step in bandit online convex optimization is to estimate the gradient of the loss function by sampling the loss function. Various algorithms have been developed and can be divided into two categories depending on the number of samplings. Algorithms with one sampling at each round have been proposed in [32]–[41]. Specifically, in [32], $\mathcal{O}(T^{3/4})$ expected static regret was achieved for Lipschitz-continuous functions. In [33]–[37], smaller regret bounds were established under additional assumptions. Bubeck *et al.* [38] and Bubeck and Eldan[39] showed that $\mathcal{O}(\sqrt{T}\log(T))$ expected static regret can be achieved for Lipschitz-continuous loss functions, but they did not develop any explicit algorithm. An algorithm to achieve this bound was proposed in [40] based on the application of the ellipsoid method to online learning. Algorithms with two or more samplings at each round have been proposed in [9], [42]–[46]. The expected static regret bounds can then be reduced compared to the one-sample case. For example, Shamir [43] proposed a simple algorithm with two samplings at each round and obtained $\mathcal{O}(\sqrt{T})$ expected static regret for Lipschitz-continuous loss functions.

Aforementioned studies did not consider equality or inequality constraints. In the literature, there are few papers considering bandit online convex optimization with such constraints, although such constraints are common in applications. Mahdavi *et al.* [47] studied online convex optimization with static inequality constraints and bandit feedback for constraints, whereas Chen and Giannakis [48] studied online convex optimization with time-varying inequality constraints and bandit feedback for loss functions. Cao and Liu [49] studied online convex optimization with time-varying inequality constraints and bandit feedback for both loss and constraint functions. Moreover, most existing bandit online convex optimization studies are in a centralized setting and only few papers considered distributed bandit online convex optimization. The consensus-based distributed bandit online algorithms were proposed in [50]–[52].

This article considers the problem of distributed bandit online convex optimization with time-varying coupled inequality constraints. This problem can be interpreted as a repeated game between a group of learners and an adversary. The learners attempt to minimize a sequence of global loss functions and at the same time satisfy a sequence of coupled constraint functions. The global loss and the coupled constraint functions are the sum of local convex loss and constraint functions, respectively. They are generated adaptively by the adversary. The local loss and constraint functions are revealed in a bandit manner and the revealed information is held privately by each learner. Specifically, at each round, each learner can sample his/her local loss and constraint function at one point (i.e., one-point bandit feedback) or two points (i.e., two-point bandit feedback). Compared to existing studies, the contributions of this article are summarized as follows.

In the one-point bandit feedback setting, we propose a distributed bandit online algorithm with a one-point sampling gradient estimator to solve the considered optimization problem. To the best of our knowledge, this is the first algorithm to solve the online convex optimization problem with time-varying inequality constraints in the one-point bandit feedback setting. An advantage of our algorithm is that the total number of rounds is not used in the algorithm and, thus, does not need to be known *a priori*, which is an improvement compared to the one-point sampling algorithms in [32]–[37], [48], [50], [52]. Moreover, note that these papers did not consider bandit feedback for time-varying inequality constraints or did not even consider time-varying inequality constraints at all. Sublinear expected regret and constraint violation bounds are achieved by the proposed algorithm if $V(\boldsymbol{x}_T^*)$, the path-length of the optimal dynamic decision sequence, grows sublinearly with a known order. In particular, $\mathcal{O}(T^{\theta_1})$ expected static regret and $\mathcal{O}(T^{7/4-\theta_1})$ constraint violation are achieved, where $\theta_1 \in (3/4, 5/6]$ is a user-defined tradeoff parameter. Specifically, when there are no inequality constraints, the proposed algorithm achieves $\mathcal{O}(T^{3/4})$ expected static regret, which is the same expected static regret bound that has been achieved by the one-point sampling algorithm in [32]. However, in [32], the total number of iterations T as well as the Lipschitz constant and upper bound of the loss functions are needed for the algorithm.

In the two-point bandit feedback setting, we propose a distributed bandit online algorithm with a two-point sampling gradient estimator. This algorithm does not require the total number of rounds or any other parameters related to the loss or constraint functions, which is different from the two-point sampling algorithms in [9], [42]–[44], [46]–[49], and [51]. In an average sense, this algorithm is as efficient as the algorithms proposed in [11], [12], [47], and [53], although Jenatton *et al.* [11], Sun *et al.*[12], and Yi *et al.*[53] are in a full-information feedback setting and Mahdavi *et al.*[47] consider the bandit setting only for the constraint functions. Sublinear expected regret and constraint violation bounds are achieved by the proposed algorithm if the path-length of the optimal dynamic decision sequence grows sublinearly with a known order $\nu \in [0, 1)$. For example, $\mathcal{O}(T^{(1+\nu)/2})$ expected dynamic regret and $\mathcal{O}(T^{(3+\nu)/4})$ constraint violation are achieved by our algorithm. Thus, the bounds achieved by the centralized two-point sampling bandit algorithms in [44] and [49] are recovered by our algorithm. Moreover, $\mathcal{O}(T^{\max\{\kappa,1-\kappa\}})$ expected static regret and $\mathcal{O}(T^{1-\kappa/2})$ constraint violation are also achieved, where $\kappa \in (0, 1)$ is a user-defined parameter. Thus, the bounds achieved by the centralized two-point sampling bandit algorithm in [43] and [47] are also recovered with $\kappa = 1/2$. However, in [43] and [44], static set constraints rather than time-varying inequality constraints are considered; in [47], static inequality constraints and full-information feedback for the cost function are studied; and in [43], [44], [47], and [49], the total number of

TABLE I
COMPARISON OF THE TWO ALGORITHMS PROPOSED IN THIS ARTICLE TO RELATED WORKS ON BANDIT ONLINE CONVEX OPTIMIZATION

Reference	Problem type	Constraint type	Information feedback	Regret and constraint violation bounds
[32]	Centralized	$g_t(x) \equiv \mathbf{0}_m$	One-point sampling	$\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}_T^*)] = \mathcal{O}(T^{3/4})$
[40]	Centralized	$g_t(x) \equiv \mathbf{0}_m$	One-point sampling	$\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}_T^*)] = \mathcal{O}(T^{1/2}\log(T))$
[43]	Centralized	$g_t(x) \equiv \mathbf{0}_m$	Two-point sampling	$\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}_T^*)] = \mathcal{O}(T^{1/2})$
[44]	Centralized	$g_t(x) \equiv \mathbf{0}_m$	Two-point sampling	$\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}_T^*)] = \mathcal{O}(\max\{(TV(\boldsymbol{x}_T^*))^{1/2},\ T^{1/2}\})$
[47]	Centralized	$g(x) \leq \mathbf{0}_m$	∇f_t and two-point sampling for g	$\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}_T^*)] = \mathcal{O}(T^{1/2})$, $\mathbf{E}[\lVert[\sum_{t=1}^{T} g(x_t)]_+\rVert] = \mathcal{O}(T^{3/4})$
[48]	Centralized	$g_t(x) \leq \mathbf{0}_m$ and Slater's condition	∇g_t and one-point sampling for f_t	$\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}_T^*)] = \mathcal{O}(\max\{T^{3/4}V(\boldsymbol{x}_T^*),\ T^{3/4}\})$, $\lVert[\sum_{t=1}^{T} g(x_t)]_+\rVert = \mathcal{O}(T^{3/4})$
			∇g_t and two-point sampling for f_t	$\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}_T^*)] = \mathcal{O}(\max\{T^{1/2}V(\boldsymbol{x}_T^*),\ T^{1/2}\})$, $\lVert[\sum_{t=1}^{T} g(x_t)]_+\rVert = \mathcal{O}(T^{1/2})$
[49]	Centralized	$g_t(x) \leq \mathbf{0}_m$	Two-point sampling	$\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}_T^*)] = \mathcal{O}((TV(\boldsymbol{x}_T^*))^{1/2})$, $\mathbf{E}[\lVert[\sum_{t=1}^{T} g(x_t)]_+\rVert] = \mathcal{O}((T^3V(\boldsymbol{x}_T^*))^{1/4})$, if $V(\boldsymbol{x}_T^*) > 0$
This paper	Distributed	$g_t(x) = \sum_{i=1}^{n} g_{i,t}(x_i) \leq \mathbf{0}_m$	One-point sampling	$\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}_T^*)] = \mathcal{O}(\max\{T^{\theta_1}V(\boldsymbol{x}_T^*),\ T^{\max\{\theta_1, 1-\theta_1+2\theta_3, 1-\theta_3+\theta_2\}}\})$, $\lVert[\sum_{t=1}^{T} g(x_t)]_+\rVert = \mathcal{O}(T^{1-\theta_2/2})$, where $\theta_1 \in (0,1)$, $\theta_2 \in (0, \theta_1/3)$, and $\theta_3 \in (\theta_2, (\theta_1-\theta_2)/2]$
			Two-point sampling	$\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}_T^*)] = \mathcal{O}(\max\{T^{\kappa}V(\boldsymbol{x}_T^*),\ T^{\max\{\kappa, 1-\kappa\}}\})$, $\lVert[\sum_{t=1}^{T} g(x_t)]_+\rVert = \mathcal{O}(T^{1-\kappa/2})$, where $\kappa \in (0,1)$

rounds as well as the Lipschitz constant of the loss function are needed.

The comparison of the two algorithms proposed in this article to related studies in the literature is summarized in Table I.

The rest of this article is organized as follows. Section II introduces the preliminaries. Section III gives the problem formulation and a motivating example. Sections IV and V provide the distributed bandit online algorithms for one- and two-point bandit feedback, respectively, and present their expected regret and constraint violation bounds. Section VI gives numerical simulations for the motivating example and compares the performance of the proposed algorithms and the existing algorithms in the literature. Finally, Section VII concludes this article. Proofs are given in the Appendix.

Notations: All inequalities and equalities are understood componentwise. $\mathbb{R}^p$ and $\mathbb{R}^p_+$ denote the set of p-dimensional vectors and nonnegative vectors, respectively. $\mathbb{N}_+$ stands for the set of positive integers. $[n]$ represents the set $\{1, \dots, n\}$ for any $n \in \mathbb{N}_+$. $[x]_j$ is the jth element of a vector $x \in \mathbb{R}^p$. $\langle x, y\rangle$ denotes the standard inner product of two vectors x and y. $x^\top$ stands for the transpose of the vector or matrix x. $\lVert \cdot \rVert$ ($\lVert \cdot \rVert_1$) represents the Euclidean norm (1-norm) for vectors and the induced 2-norm (1-norm) for matrices. $\mathbb{B}^p$ and $\mathbb{S}^p$ are the unit ball and sphere centered around the origin in $\mathbb{R}^p$ under Euclidean norm, respectively. $\mathbf{I}_n$ denotes the n-dimensional identity matrix. $\mathbf{1}_n$ ($\mathbf{0}_n$) stands for the column one (zero) vector of dimension n. $\mathrm{col}(z_1, \dots, z_k)$ represents the concatenated column vector of vectors $z_i \in \mathbb{R}^{n_i}, i \in [k]$. $\log(\cdot)$ is the natural logarithm. Given two scalar sequences $\{\alpha_t, t \in \mathbb{N}_+\}$ and $\{\beta_t > 0, t \in \mathbb{N}_+\}$, $\alpha_t = \mathcal{O}(\beta_t)$ means that $\limsup_{t\to\infty}(\alpha_t/\beta_t)$ is bounded, whereas $\alpha_t = \mathbf{o}(\beta_t)$ means that $\lim_{t\to\infty}(\alpha_t/\beta_t) = 0$. For a set $\mathbb{K} \subseteq \mathbb{R}^p$, $\mathcal{P}_{\mathbb{K}}(\cdot)$ denotes the projection operator, i.e., $\mathcal{P}_{\mathbb{K}}(x) = \arg\min_{y\in\mathbb{K}} \lVert x - y\rVert^2 \quad \forall x \in \mathbb{R}^p$. For simplicity, $[\cdot]_+$ is used to denote $\mathcal{P}_{\mathbb{K}}(\cdot)$ when $\mathbb{K} = \mathbb{R}^p_+$.

II. PRELIMINARIES

In this section, we present some definitions and properties related to graph theory and gradient approximation.

A. Graph Theory

Let $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$ denote a time-varying directed graph, where $\mathcal{V} = [n]$ is the agent set and $\mathcal{E}_t \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set. A directed edge $(j, i) \in \mathcal{E}_t$ means that agent i can receive data from agent j at time t. Let $\mathcal{N}_i^{\mathrm{in}}(\mathcal{G}_t) = \{j \in [n] \mid (j, i) \in \mathcal{E}_t\}$ and $\mathcal{N}_i^{\mathrm{out}}(\mathcal{G}_t) = \{j \in [n] \mid (i, j) \in \mathcal{E}_t\}$ be the sets of in- and out-neighbors, respectively, of agent i at time t. A directed path is a sequence of consecutive directed edges. A directed graph is said to be strongly connected if there is at least one directed path from any agent to any other agent in the graph. The mixing matrix $W_t \in \mathbb{R}^{n\times n}$ at time t fulfills $[W_t]_{ij} > 0$ if $(j, i) \in \mathcal{E}_t$ or $i = j$, and $[W_t]_{ij} = 0$ otherwise.

B. Gradient Approximation

In this section, we introduce one- and two-point sampling gradient estimators.

Let $f : \mathbb{K} \to \mathbb{R}$ be a function with $\mathbb{K} \subset \mathbb{R}^p$. We assume that $\mathbb{K}$ is convex and bounded and has a nonempty interior. Specifically, we assume that $\mathbb{K}$ contains the ball of radius $r(\mathbb{K})$ centered at the origin and is contained in the ball of radius $R(\mathbb{K})$, i.e., $r(\mathbb{K})\mathbb{B}^p \subseteq \mathbb{K} \subseteq R(\mathbb{K})\mathbb{B}^p$. Flaxman *et al.* [32] proposed the following gradient estimator:

$$\hat{\nabla}_1 f(x) = \frac{p}{\delta} f(x+\delta u) u \quad \forall x \in (1-\xi)\mathbb{K} \tag{1}$$

where $u \in \mathbb{S}^p$ is a uniformly distributed random vector, $\delta \in (0, r(\mathbb{K})\xi]$ is an exploration parameter, and $\xi \in (0,1)$ is a shrinkage coefficient. The estimator $\hat{\nabla}_1 f$ only requires to sample the function at one point, so it is a one-point sampling gradient estimator. Some intuition for this estimator can be found in [32]. Different from Nesterov and Spokoiny [28], uniform distribution rather than Gaussian distribution is used to generate u in (1) since the later may generate unbounded u. The estimator $\hat{\nabla}_1 f$ is defined over the set $(1-\xi)\mathbb{K}$ instead of $\mathbb{K}$, since otherwise the perturbations may move points outside $\mathbb{K}$. The feasibility of the perturbations is guaranteed by the following lemma.

Lemma 1 (see Observation 2 in [32]): For any $x \in (1-\xi)\mathbb{K}$ and $u \in \mathbb{S}^p$, it holds that $x + \delta u \in \mathbb{K}$ for any $\delta \in (0, r(\mathbb{K})\xi]$.

Our two-point sampling gradient estimator is defined as

$$\hat{\nabla}_2 f(x) = \frac{p}{\delta}\left(f(x+\delta u) - f(x)\right) u \quad \forall x \in (1-\xi)\mathbb{K}. \tag{2}$$

The intuition follows from directional derivatives [42].

Both estimators $\hat{\nabla}_1 f$ and $\hat{\nabla}_2 f$ are unbiased gradient estimators of $\hat{f}$, where $\hat{f}$ is the uniformly smoothed version of f defined as

$$\hat{f}(x) = \mathbf{E}_{v\in\mathbb{B}^p}\left[f(x+\delta v)\right] \quad \forall x \in (1-\xi)\mathbb{K}$$

with the expectation is taken with respect to uniform distribution. Some properties of $\hat{f}$, $\hat{\nabla}_1 f$, and $\hat{\nabla}_2 f$ are presented in the following lemma.

Lemma 2:

1) The uniform smoothing $\hat{f}$ is differentiable on $(1-\xi)\mathbb{K}$ even when f is not, and for all $x \in (1-\xi)\mathbb{K}$

$$\nabla \hat{f}(x) = \mathbf{E}_{u\in\mathbb{S}^p}\left[\hat{\nabla}_1 f(x)\right] = \mathbf{E}_{u\in\mathbb{S}^p}\left[\hat{\nabla}_2 f(x)\right].$$

2) If f is convex on $\mathbb{K}$, then $\hat{f}$ is convex on $(1-\xi)\mathbb{K}$ and

$$f(x) \le \hat{f}(x) \quad \forall x \in (1-\xi)\mathbb{K}.$$

3) If f is Lipschitz-continuous on $\mathbb{K}$ with constant $L_0(f) > 0$, then $\hat{f}$ and $\nabla \hat{f}$ are Lipschitz-continuous on $(1-\xi)\mathbb{K}$ with constants $L_0(f)$ and $pL_0(f)/\delta$, respectively. Moreover

$$\left|\hat{f}(x) - f(x)\right| \le \delta L_0(f) \quad \forall x \in (1-\xi)\mathbb{K}.$$

4) If f is bounded on $\mathbb{K}$, i.e., there exists $F_0(f) > 0$ such that $|f(x)| \le F_0(f) \quad \forall x \in \mathbb{K}$, then

$$\left|\hat{f}(x)\right| \le F_0(f)$$

$$\left\|\hat{\nabla}_1 f(x)\right\| \le \frac{pF_0(f)}{\delta} \quad \forall x \in (1-\xi)\mathbb{K}.$$

5) If f is Lipschitz-continuous on $\mathbb{K}$ with constant $L_0(f) > 0$, then

$$\left\|\hat{\nabla}_2 f(x)\right\| \le pL_0(f) \quad \forall x \in (1-\xi)\mathbb{K}.$$

Proof: See Appendix B. ■

Intuitively, the key idea of gradient-free optimization methods is using the smoothed function $\hat{f}$ to replace the original function f since they are close when δ is small, as shown in 3) of Lemma 2. Moreover, the gradient of $\hat{f}$ can be estimated by the gradient estimators $\hat{\nabla}_1 f$ or $\hat{\nabla}_2 f$, as shown in 1). The main difference between these two gradient estimators is that the norm of $\hat{\nabla}_1 f$ is large when δ is small, whereas $\hat{\nabla}_2 f$ has a bounded norm, as shown in 4) and 5), respectively. This difference leads to improved results for the two-point bandit feedback algorithm, as will be seen in the later sections.

III. Problem Formulation

We consider the problem of distributed bandit online convex optimization with time-varying coupled inequality constraints. This problem can be defined as a repeated game between a group of n learners indexed by $i \in [n]$ and an adversary. At round t of the game, the adversary first arbitrarily chooses n local convex loss functions $\{f_{i,t} : \mathbb{R}^{p_i} \to \mathbb{R}, i \in [n]\}$ and n local convex constraint functions $\{g_{i,t} : \mathbb{R}^{p_i} \to \mathbb{R}^m, i \in [n]\}$, where p_i and m are positive integers. Then, without knowing $\{f_{i,t}, i \in [n]\}$ and $\{g_{i,t}, i \in [n]\}$, all learners simultaneously choose their decisions $\{x_{i,t} \in \mathbb{X}_i, i \in [n]\}$, where $\mathbb{X}_i \subseteq \mathbb{R}^{p_i}$ are known convex sets. Each learner i samples the values of $f_{i,t}$ and $g_{i,t}$ at the point $x_{i,t}$ as well as at other potential points, i.e., the learners receive bandit feedback from the adversary. These values are held privately by each learner. At the same moment, the learners exchange data with their neighbors over a time-varying directed graph $\mathcal{G}_t$. The goal of the learners is to cooperatively choose a global decision sequence $\boldsymbol{x}_T = (x_1, \dots, x_T)$, where T is the total number of rounds and $x_t = \mathrm{col}(x_{1,t}, \dots, x_{n,t})$ is the decision vector, such that the accumulated global loss $\sum_{t=1}^T f_t(x_t)$, where $f_t(x_t) = \sum_{i=1}^n f_{i,t}(x_{i,t})$ is the global loss function, is competitive with the loss of any comparator sequence $\boldsymbol{y}_T = (y_1, \dots, y_T)$ with $y_t = \mathrm{col}(y_{1,t}, \dots, y_{n,t})$ (i.e., the regret is as small as possible) and at the same time the constraint violation is as small as possible.

Specifically, the regret of a global decision sequence $\boldsymbol{x}_T$ with respect to a comparator sequence $\boldsymbol{y}_T$ is defined as

$$\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{y}_T) = \sum_{t=1}^T f_t(x_t) - \sum_{t=1}^T f_t(y_t).$$

In the literature, there are two commonly used comparator sequences. One is the optimal dynamic decision sequence in hindsight $\boldsymbol{y}_T = \boldsymbol{x}_T^* = (x_1^*, \dots, x_T^*)$ solving the constrained convex optimization problem

$$\min \sum_{t=1}^T f_t(x_t)$$
$$\text{s.t. } x_t \in \mathbb{X}, g_t(x_t) \le \mathbf{0}_m \quad \forall t \in [T] \tag{3}$$

where $\mathbb{X} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_n \subseteq \mathbb{R}^p$ is the global decision set, $p = \sum_{i=1}^n p_i$, and $g_t(x_t) = \sum_{i=1}^n g_{i,t}(x_{i,t})$ is the coupled constraint function. In order to guarantee that problem (3) is feasible, we assume that for any $T \in \mathbb{N}_+$, the set of all feasible decision sequences $\mathcal{X}_T = \{(x_1, \dots, x_T) : x_t \in \mathbb{X}, g_t(x_t) \le \mathbf{0}_m, t \in [T]\}$ is nonempty. With this standing assumption, an optimal dynamic decision sequence to (3) always exists. In this case, $\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}_T^*)$ is called the dynamic regret for $\boldsymbol{x}_T$. The other comparator sequence is $\boldsymbol{y}_T = \check{\boldsymbol{x}}_T^* = (\check{x}_T^*, \dots, \check{x}_T^*)$, where $\check{x}_T^*$ is the optimal static decision in hindsight solving

$$\min \sum_{t=1}^T f_t(x)$$
$$\text{s.t. } x \in \mathbb{X}, g_t(x) \le \mathbf{0}_m \quad \forall t \in [T]. \tag{4}$$

Similar to above, in order to guarantee that problem (4) is feasible, we assume that for any $T \in \mathbb{N}_+$, the set of all feasible static decision sequences $\check{\mathcal{X}}_T = \{(x, \ldots, x) : x \in \mathbb{X}, g_t(x) \le \mathbf{0}_m, t \in [T]\} \subseteq \mathcal{X}_T$ is nonempty. In this case, $\mathrm{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}_T^*)$ is called the static regret. It is straightforward to see that $\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{y}_T) \le \mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}_T^*) \quad \forall \boldsymbol{y}_T \in \mathcal{X}_T$, and that $\mathrm{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}_T^*) \le \mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}_T^*)$.

For a decision sequence $\boldsymbol{x}_T$, the constraint violation is defined as

$$\left\| \left[\sum_{t=1}^{T} g_t(x_t) \right]_+ \right\|.$$

Note that this definition implicitly allows constraint violations at some times to be compensated by strictly feasible decisions at other times. This is appropriate for constraints that have a cumulative nature, such as in applications with energy budgets enforced through average power constraints.

The considered problem can be viewed as an extension of the problem studied in [53], from full information feedback to bandit feedback. As discussed in Section I, two main motivations of considering bandit feedback are that gradient information is not available in many applications [26] and computing a function value is much simpler than computing its gradient [28].

We make the following assumptions on the time-varying directed graph $\mathcal{G}_t$ as well as the loss and constraint functions.

Assumption 1: For any $t \in \mathbb{N}_+$, the directed graph $\mathcal{G}_t$ satisfies the following conditions.

1) There exists a constant $w \in (0, 1)$, such that $[W_t]_{ij} \ge w$ if $[W_t]_{ij} > 0$.
2) The mixing matrix W_t is doubly stochastic, i.e., $\sum_{i=1}^n [W_t]_{ij} = \sum_{j=1}^n [W_t]_{ij} = 1 \quad \forall i, j \in [n]$.
3) There exists an integer $\iota > 0$ such that the directed graph $(\mathcal{V}, \cup_{l=0,\ldots,\iota-1} \mathcal{E}_{t+l})$ is strongly connected.

Assumption 2:

1) For each $i \in [n]$, the set $\mathbb{X}_i$ is convex and closed. Moreover, there exist $r_i > 0$ and $R_i > 0$ such that
$$r_i \mathbb{B}^{p_i} \subseteq \mathbb{X}_i \subseteq R_i \mathbb{B}^{p_i} \tag{5}$$
and r_i is known a priori.
2) For each $i \in [n]$, $\{f_{i,t}(x)\}$ and $\{[g_{i,t}(x)]_j, j \in [m]\}$ are convex and uniformly bounded on $\mathbb{X}_i$, i.e., there exist constants $F_{f_i} > 0$ and $F_{g_i} > 0$ such that for all $t \in \mathbb{N}_+, j \in [m], x \in \mathbb{X}_i$
$$|f_{i,t}(x)| \le F_{f_i}, \text{and } |[g_{i,t}(x)]_j| \le F_{g_i}. \tag{6}$$
3) For each $i \in [n]$, $f_{i,t}$ and $g_{i,t}$ are differentiable on $\mathbb{X}_i$. Moreover, $\{\nabla f_{i,t}\}$ and $\{\nabla [g_{i,t}(x)]_j, j \in [m]\}$ are uniformly bounded on $\mathbb{X}_i$, i.e., there exist constants $G_{f_i} > 0$ and $G_{g_i} > 0$ such that for all $t \in \mathbb{N}_+, j \in [m], x \in \mathbb{X}_i$
$$\|\nabla f_{i,t}(x)\| \le G_{f_i}, \text{and } \|\nabla [g_{i,t}(x)]_j\| \le G_{g_i}. \tag{7}$$

Assumption 1 is common in the literature on distributed optimization. Assumption 2 appears often in the literature of bandit online convex optimization. From Assumption 2 and [1, Lemma 2.6], it follows that for all $t \in \mathbb{N}_+, i \in [n], j \in [m], x, y \in \mathbb{X}_i$

$$|f_{i,t}(x) - f_{i,t}(y)| \le G_{f_i} \|x - y\| \tag{8a}$$

$$|[g_{i,t}(x)]_j - [g_{i,t}(y)]_j| \le G_{g_i} \|x - y\| \tag{8b}$$

i.e., $\{f_{i,t}(x)\}$ and $\{[g_{i,t}(x)]_j\}$ are Lipschitz-continuous on $\mathbb{X}_i$ with constants G_{f_i} and G_{g_i}, respectively.

Algorithm 1: Distributed Bandit Online Descent With One-Point Sampling Gradient Estimator.

Input: Nonincreasing sequences $\{\alpha_{i,t}\}$, $\{\beta_{i,t}\}$, $\{\gamma_{i,t}\} \subseteq (0, +\infty)$, $\{\xi_{i,t}\} \subseteq (0, 1)$, and $\{\delta_{i,t}\} \subseteq (0, r_i \xi_{i,t-1}]$, $i \in [n], t \in \mathbb{N}_+$.
Initialize: $u_{i,1} \in \mathbb{S}^{p_i}$, $z_{i,1} \in (1 - \xi_{i,1})\mathbb{X}_i$, $x_{i,1} = z_{i,1} + \delta_{i,1} u_{i,1}$, and $q_{i,1} = \mathbf{0}_m, i \in [n]$.
for $t = 2, \ldots, T$ **do**
  for $i \in [n]$ in parallel **do**
    Select vector $u_{i,t} \in \mathbb{S}^{p_i}$ independently and uniformly at random.
    Sample $f_{i,t-1}(x_{i,t-1})$ and $g_{i,t-1}(x_{i,t-1})$.
    Update
$$\tilde{q}_{i,t} = \sum_{j=1}^{n} [W_{t-1}]_{ij} q_{j,t-1} \tag{9a}$$
$$z_{i,t} = \mathcal{P}_{(1-\xi_{i,t})\mathbb{X}_i}(z_{i,t-1} - \alpha_{i,t} a_{i,t}) \tag{9b}$$
$$x_{i,t} = z_{i,t} + \delta_{i,t} u_{i,t} \tag{9c}$$
$$q_{i,t} = [(1 - \beta_{i,t}\gamma_{i,t})\tilde{q}_{i,t} + \gamma_{i,t} g_{i,t-1}(x_{i,t-1})]_+ . \tag{9d}$$
    Broadcast $q_{i,t}$ to $\mathcal{N}_i^{\mathrm{out}}(\mathcal{G}_t)$ and receive $q_{j,t}$ from $j \in \mathcal{N}_i^{\mathrm{in}}(\mathcal{G}_t)$.
  end for
end for
Output: $\boldsymbol{x}_T$.

A. Motivating Example

As a motivating example, consider a power grid with n power generation units. Each unit i has p_i conventional and renewable power generators. The units can communicate through the information infrastructure. At stage t, let $x_{i,t} \in \mathbb{X}_i$ and $\mathbb{X}_i \subset \mathbb{R}^{p_i}$ be the output and the set of feasible outputs of the generators in unit i, respectively. To generate the output, each unit i suffers a cost $f_{i,t}(x_{i,t})$. This local cost $f_{i,t}$ is usually described by a quadratic function [54], but it is unknown in advance, since fossil fuel price is fluctuating and renewable energy is uncertain and unpredictable. Except the local generator limit constraints $\mathbb{X}_i$, all units need to cooperatively take into account global constraints, such as power balance and emission constraints. The global constraints can be modeled as $\sum_{i=1}^n g_{i,t}(x_{i,t}) \le \mathbf{0}_m$, where $g_{i,t}$ is unit i's local constraint function. Again, the precise form of the constraint functions is unknown in advance either since that power demands can change from 1 h to the next, or that the emission can change due to the uncertain and unpredictable features of renewable energy. The goal of the units is to reduce the global cost while satisfying the constraints.

IV. One-Point Bandit Feedback

In this section, we propose a distributed bandit online algorithm with a one-point sampling gradient estimator to solve the considered optimization problem. We then derive expected regret and constraint violation bounds for the proposed algorithm.

A. Distributed Bandit Online Algorithm With One-Point Sampling Gradient Estimator

The proposed algorithm is given in pseudocode as Algorithm 1. In this algorithm, each agent i maintains four local sequences: the local primal decision variable sequence $\{x_{i,t}\} \subseteq \mathbb{X}_i$, the local intermediate decision variable sequence $\{z_{i,t}\} \subseteq (1-\xi_{i,t})\mathbb{X}_i$, the local dual variable sequence $\{q_{i,t}\} \subseteq \mathbb{R}^m_+$, and the estimates of the average of local dual variables $\{\tilde{q}_{i,t}\} \subseteq \mathbb{R}^m_+$. They are updated recursively by the update rules (9a)–(9d). In (9b), $a_{i,t}$ is the updating direction information for the local intermediate decision variable defined as

$$a_{i,t} = \hat{\nabla}_1 f_{i,t-1}(z_{i,t-1}) + \left(\hat{\nabla}_1 g_{i,t-1}(z_{i,t-1})\right)^\top \tilde{q}_{i,t}. \quad (10)$$

The intuition of the update rules (9a)–(9d) is as follows. The regularized Lagrangian function associated with the constrained optimization problem with cost function f and constraint function g is

$$\mathcal{A}(x,\mu) = f(x) + \mu^\top g(x) - \frac{\beta}{2}\|\mu\|^2 \quad (11)$$

where $\mu \in \mathbb{R}^m_+$ is the Lagrange multiplier and $\beta > 0$ is the regularization parameter. $\mathcal{A}(x,\mu)$ is a convex–concave function. A standard primal-dual algorithm to find its saddle point is

$$x_{k+1} = \mathcal{P}_{\mathbb{X}}\left(x_k - \alpha\left(\nabla f(x_k) + (\nabla g(x_k))^\top \mu_k\right)\right) \quad (12a)$$

$$\mu_{k+1} = [\mu_k + \gamma(g(x_k) - \beta\mu_k)]_+ \quad (12b)$$

where $\alpha > 0$ and $\gamma > 0$ are the stepsizes used in the primal and dual updates, respectively. The update rules (9a)–(9d) are the distributed, online, and gradient-free extensions of (12a) and (12b). The differences between Algorithm 1 and the centralized one-point sampling algorithm in [48] are that in [48], full-information feedback for the constraint functions is used and in the update of the dual variables in Algorithm 1, i.e., (9d), there is an additional term $-\beta_{i,t}\gamma_{i,t}\tilde{q}_{i,t}$, which comes from the regularized Lagrangian function and it plays a key role to bound the dual variables, as shown later in Lemma 5.

The sequences $\{\alpha_{i,t}\}$, $\{\beta_{i,t}\}$, $\{\gamma_{i,t}\}$, $\{\xi_{i,t}\}$, and $\{\delta_{i,t}\}$ used in Algorithm 1 are predetermined and the vector sequences $\{u_{i,t}\}$ are randomly selected. Moreover, $\{\tilde{q}_{i,t}\}$, $\{z_{i,t}\}$, $\{x_{i,t}\}$, and $\{q_{i,t}\}$ are random vector sequences generated by Algorithm 1. Let $\mathfrak{U}_t$ denote the σ-algebra generated by the independent and identically distributed random variables $u_{1,t}, \dots, u_{n,t}$ and let $\mathcal{U}_t = \bigcup_{s=1}^t \mathfrak{U}_s$. It is straightforward to see that $\tilde{q}_{t+1}, z_{i,t}, x_{i,t-1}$, and $q_{i,t}, i \in [n]$ depend on $\mathcal{U}_{t-1}$ and are independent of $\mathfrak{U}_s$ for all $s \geq t$.

B. Expected Regret and Constraint Violation Bounds

This section states the main results on the expected regret and constraint violation bounds for Algorithm 1. The following theorem characterizes these bounds based on some specially selected stepsizes, shrinkage coefficients, and exploration parameters.

Theorem 1: Suppose Assumptions 1 and 2 hold. For any $T \in \mathbb{N}_+$, let $\boldsymbol{x}_T$ be the sequence generated by Algorithm 1 with

$$\alpha_{i,t} = \frac{r_i^2}{4mp_i^2F_{g_i}^2t^{\theta_1}}, \beta_{i,t} = \frac{2}{t^{\theta_2}}, \gamma_{i,t} = \frac{1}{t^{1-\theta_2}}$$

$$\xi_{i,t} = \frac{1}{(t+1)^{\theta_3}}, \delta_{i,t} = \frac{r_i}{(t+1)^{\theta_3}}, i \in [n], t \in \mathbb{N}_+ \quad (13)$$

where $\theta_1 \in (0,1)$, $\theta_2 \in (0, \theta_1/3)$, and $\theta_3 \in (\theta_2, (\theta_1-\theta_2)/2]$ are constants. Then, for any comparator sequence $\boldsymbol{y}_T \in \mathcal{X}_T$

$$\mathbf{E}\left[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{y}_T)\right] \leq C_1 T^{\max\{\theta_1, 1-\theta_1+2\theta_3, 1-\theta_3+\theta_2\}} + C_{1,1}T^{\theta_1}V(\boldsymbol{y}_T) \quad (14a)$$

$$\mathbf{E}\left[\left\|\left[\sum_{t=1}^T g_t(x_t)\right]_+\right\|\right] \leq C_2 T^{1-\theta_2/2} \quad (14b)$$

where $C_1 = \sum_{i=1}^n\big(\frac{mF_gG_{g_i}(2r_i+R_i)}{1-\theta_3+\theta_2} + \frac{G_{f_i}(2r_i+R_i)}{1-\theta_3} + \frac{F_{f_i}^2}{4mF_{g_i}^2(1-\theta_1+2\theta_3)}\big) + C_{1,1} + \frac{C_0}{\theta_2}$, $C_2 = \sqrt{C_{2,1}(2\sum_{i=1}^n F_{f_i} + C_1)}$, $F_g = \max_{i\in[n]}\{F_{g_i}\}$, $C_{1,1} = \sum_{i=1}^n \frac{8mp_i^2F_{g_i}^2R_i^2}{r_i^2}$, $C_0 = \frac{6mn^2F_g^2\tau}{1-\lambda} + 2mnF_g^2$, $\tau = (1-\frac{w}{2n^2})^{-2} > 1$, $\lambda = (1-\frac{w}{2n^2})^{\frac{1}{\iota}}$, $C_{2,1} = 2n(1+\max_{i\in[n]}\{\frac{F_{f_i}^2}{F_{g_i}^2(1-\theta_1+2\theta_3)}\} + \frac{1}{1-\theta_2})$, w and ι are given in Assumption 1, r_i, R_i, F_{f_i}, F_{g_i}, G_{f_i}, and G_{g_i} are given in Assumption 2, and

$$V(\boldsymbol{y}_T) = \sum_{t=1}^{T-1}\sum_{i=1}^n \|y_{i,t+1} - y_{i,t}\|$$

is the accumulated variation (path-length) of the comparator sequence $\boldsymbol{y}_T$.

Proof: See Appendix C. ∎

Remark 1: From (14b), we see that Algorithm 1 achieves sublinear expected constraint violation. From (14a), we see that Algorithm 1 can achieve sublinear expected dynamic regret if $V(\boldsymbol{x}^*_T)$ grows sublinearly with a known order. In this case, there exists a known constant $\nu \in [0,1)$, such that $V(\boldsymbol{x}^*_T) = \mathcal{O}(T^\nu)$, then setting $\boldsymbol{y}_T = \boldsymbol{x}^*_T$ and $\theta_1 \in (0, 1-\nu)$ in Theorem 1 gives $\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}^*_T)] = \mathbf{o}(T)$.

Remark 2: To the best of our knowledge, Algorithm 1 is the first algorithm to solve the online convex optimization problem with time-varying inequality constraints in the one-point bandit feedback setting. In Algorithm 1, the information about the total number of rounds is not used, which is an improvement compared to the one-point sampling algorithms in [32]–[37], [48], [50], [52]. Note that these papers did not consider bandit feedback for time-varying inequality constraints or did not even consider time-varying inequality constraints at all. The potential drawback of Algorithm 1 is that in order to use the sequences defined in (13), each learner i needs to know F_{g_i}, the uniform upper bound of his/her time-varying constraint function. One way to overcome this is to let $\alpha_{i,t} = \tau_i/t^{\theta_1}$ and $\theta_3 \in (\theta_2, (\theta_1-\theta_2)/2)$, where $\tau_i > 0$ is a user-defined parameter. In this case, similar to the way we prove (14a) and (14b), we can establish similar results as (14a) and (14b) for $T \geq (4m\max_{i\in[n]}\{p_i^2F_{g_i}^2\tau_i/r_i^2\})^{1/(\theta_1-\theta_2-2\theta_3)}$ rather than any $T \in \mathbb{N}_+$.

Setting $\boldsymbol{y}_T = \check{\boldsymbol{x}}^*_T$ in Theorem 1 gives following results, which characterize the expected static regret and constraint violation bounds.

Corollary 1: Under the same conditions as in Theorem 1 with $\theta_1 \in (3/4, 5/6]$, $\theta_2 = 2\theta_1 - 3/2$, and $\theta_3 = \theta_1 - 1/2$, it holds that

$$\mathbf{E}\left[\mathrm{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}^*_T)\right] \leq C_1 T^{\theta_1} \quad (15a)$$

$$\mathbf{E}\left[\left\|\left[\sum_{t=1}^T g_t(x_t)\right]_+\right\|\right] \leq C_2 T^{7/4-\theta_1}. \quad (15b)$$

Algorithm 2: Distributed Bandit Online Descent With Two-Point Sampling Gradient Estimator.

Input: Nonincreasing sequences $\{\alpha_{i,t}\}$, $\{\beta_{i,t}\}$, $\{\gamma_{i,t}\} \subseteq (0, +\infty)$, $\{\xi_{i,t}\} \subseteq (0, 1)$, and $\{\delta_{i,t}\} \subseteq (0, r_i\xi_{i,t-1}]$, $i \in [n]$, $t \in \mathbb{N}_+$.
Initialize: $x_{i,1} \in (1 - \xi_{i,1})\mathbb{X}_i$ and $q_{i,1} = \mathbf{0}_m$, $i \in [n]$.
for $t = 2, \ldots, T$ **do**
  for $i \in [n]$ in parallel **do**
    Select vector $u_{i,t-1} \in \mathbb{S}^{p_i}$ independently and uniformly at random.
    Sample $f_{i,t-1}(x_{i,t-1} + \delta_{i,t-1}u_{i,t-1})$, $f_{i,t-1}(x_{i,t-1})$, $g_{i,t-1}(x_{i,t-1} + \delta_{i,t-1}u_{i,t-1})$, and $g_{i,t-1}(x_{i,t-1})$.
    Update

$$\tilde{q}_{i,t} = \sum_{j=1}^{n} [W_{t-1}]_{ij} q_{j,t-1} \quad (16a)$$

$$x_{i,t} = \mathcal{P}_{(1-\xi_{i,t})\mathbb{X}_i}(x_{i,t-1} - \alpha_{i,t} b_{i,t}) \quad (16b)$$

$$q_{i,t} = [(1 - \gamma_{i,t}\beta_{i,t})\tilde{q}_{i,t} + \gamma_{i,t} c_{i,t}]_+ . \quad (16c)$$

    Broadcast $q_{i,t}$ to $\mathcal{N}_i^{\text{out}}(\mathcal{G}_t)$ and receive $q_{j,t}$ from $j \in \mathcal{N}_i^{\text{in}}(\mathcal{G}_t)$.
  end for
end for
Output: $\boldsymbol{x}_T$.

Remark 3: The parameter θ_1 in Corollary 1 is a user-defined parameter influencing the step length in (13). It enables the tradeoff between the expected static regret bound and the expected constraint violation bound. Same as in [32], if there are no inequality constraints, i.e., $g_{i,t} \equiv \mathbf{0}_m \quad \forall i \in [n] \quad \forall t \in \mathbb{N}_+$, then by setting $\alpha_{i,t} = 1/t^{3/4}, \beta_{i,t} = \gamma_{i,t} = 0, \xi_{i,t} = 1/(t+1)^{1/4}$, and$\delta_{i,t} = r_i/(t+1)^{1/4}$ in (13), we have that (15a) can be replaced by $\mathbf{E}[\text{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}_T^*)] \le \hat{C}_1 T^{3/4}$, where $\hat{C}_1 = \sum_{i=1}^n (4G_{f_i}(2r_i + R_i)/3 + 6R_i^2 + 4p_i^2 F_{f_i}^2/(3r_i^2))$. Hence, Algorithm 1 achieves the same expected static regret bound as the bandit algorithm in [32]. However, in [32], the total number of rounds, the Lipschitz constant, and upper bound of the loss functions need to be known in advance to run the algorithm.

V. Two-Point Bandit Feedback

In this section, we consider a novel two-point bandit feedback algorithm.

A. Distributed Bandit Online Algorithm With Two-Point Sampling Gradient Estimator

With two-point bandit feedback at each round, each learner samples the values of his/her local loss and constraint at two points. This gives the freedom to design a more efficient algorithm, which at the same time avoids the potential drawback of Algorithm 1 stated in Remark 2 on knowing the upper bounds of the time-varying constraint functions. The proposed algorithm is given in pseudocode as Algorithm 2. In (16b), $b_{i,t}$ is the updating direction information for the local primal decision variable defined as

$$b_{i,t} = \hat{\nabla}_2 f_{i,t-1}(x_{i,t-1}) + \left(\hat{\nabla}_2 g_{i,t-1}(x_{i,t-1})\right)^\top \tilde{q}_{i,t}. \quad (17)$$

Similarly, in (16c), $c_{i,t}$ is the updating direction information for the local dual variable defined as

$$c_{i,t} = \hat{\nabla}_2 g_{i,t-1}(x_{i,t-1})(x_{i,t} - x_{i,t-1}) + g_{i,t-1}(x_{i,t-1}). \quad (18)$$

In addition to that Algorithm 2 uses a two-point sampling gradient estimator, another difference between Algorithms 1 and 2 is that when updating the local dual variable, in Algorithm 2, $c_{i,t}$ is used to replace $g_{i,t-1}(x_{i,t-1})$, which is a key difference between Algorithm 2 and the centralized two-point sampling algorithm in [49]. This modification is inspired by the algorithms proposed in [13] and [53] and helps to avoid using the uniform upper bound of each learner's time-varying constraint function, i.e., to remove the potential drawback stated in Remark 2.

B. Expected Regret and Constraint Violation Bounds

This section states the main results on the expected regret and constraint violation bounds for Algorithm 2.

Theorem 2: Suppose Assumptions 1 and 2 hold. For any $T \in \mathbb{N}_+$, let $\boldsymbol{x}_T$ be the sequence generated by Algorithm 2 with

$$\alpha_t = \frac{1}{t^\kappa}, \beta_t = \frac{1}{t^\kappa}, \gamma_t = \frac{1}{t^{1-\kappa}}$$

$$\xi_{i,t} = \frac{1}{t+1}, \delta_{i,t} = \frac{r_i}{t+1}, i \in [n], t \in \mathbb{N}_+ \quad (19)$$

where $\kappa \in (0, 1)$ is a constant. Then, for any comparator sequence $\boldsymbol{y}_T \in \mathcal{X}_T$

$$\mathbf{E}\left[\text{Reg}(\boldsymbol{x}_T, \boldsymbol{y}_T)\right] \le C_3 T^{\max\{\kappa, 1-\kappa\}} + 2R_{\max} T^\kappa V(\boldsymbol{y}_T) \quad (20a)$$

$$\mathbf{E}\left[\left\| \left[\sum_{t=1}^T g_t(x_t)\right]_+ \right\|\right] \le C_4 T^{1-\kappa/2} \quad (20b)$$

where $C_3 = \sum_{i=1}^n (2G_{f_i}(r_i + R_i) + 8R_i^2 + \frac{2\sqrt{m}B_1 G_{g_i} R_i}{\kappa} + \frac{p_i^2 G_{f_i}^2}{1-\kappa}) + \frac{\hat{C}_0}{\kappa}$, $C_4 = \sqrt{C_{4,1}(2\sum_{i=1}^n F_{f_i} + C_3)}$, $C_{4,1} = \sum_{i=1}^n 2(\frac{2mp_i^2 G_{g_i}^2 + 1}{1-\kappa} + 1)$, $\hat{C}_0 = \frac{6n^2\sqrt{m}\tau B_1 F_g}{1-\lambda} + 2nB_1^2$, $B_1 = \sqrt{m}F_g + \sqrt{mp}G_g R_{\max}$, and $R_{\max} = \max_{i\in[n]}\{R_i\}$.

Proof: See Appendix D. ■

Remark 4: The bounds obtained in (20a) and (20b) are the same as the bounds achieved in [53] under the same assumptions, although Yi *et al.* [53] considered a full-information feedback setting. In other words, in an average sense, Algorithm 2, which only uses two-point bandit feedback, is as efficient as the algorithm proposed in [53], which uses full-information feedback. By comparing (13), (14a), and (14b) with (19), (20a), and (20b), respectively, we see that if a two-point sampling gradient estimator is used, then not only the uses of F_{g_i}, the uniform upper bound of the time-varying constraint functions, is avoided, but also the upper bounds of the expected regret and constraint violation are both reduced. An advantage of Algorithm 2 is that the total number of rounds or any other parameters related to loss or constraint functions are not used, which is different from the two-point sampling algorithms in [9], [42]–[44], [46]–[49], [51].

Remark 5: Similar to the analysis in Remark 1, from (20b), we know that Algorithm 2 achieves sublinear expected constraint violation. Algorithm 2 can also achieve sublinear expected dynamic regret if $V(\boldsymbol{x}_T^*)$ grows sublinearly with a known order. In this case, there exists a known constant $\nu \in [0, 1)$, such that $V(\boldsymbol{x}_T^*) = \mathcal{O}(T^\nu)$. Then, setting $\boldsymbol{y}_T = \boldsymbol{x}_T^*$ and $\kappa \in$

$(0, 1-\nu)$ in Theorem 2 gives $\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}_T^*)] = \mathbf{o}(T)$. One special case is to set $\kappa = (1-\nu)/2$ in (20a) and (20b). It gives $\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}_T^*)] = \mathcal{O}(T^{(1+\nu)/2})$ and $\mathbf{E}[\|[\sum_{t=1}^T g_t(x_t)]_+\|] = \mathcal{O}(T^{(3+\nu)/4})$, which recovers the bounds achieved by the centralized two-point sampling bandit algorithms in [44] and [49].

Setting $\boldsymbol{y}_T = \check{\boldsymbol{x}}_T^*$ in Theorem 2 gives the following results.

Corollary 2: Under the same conditions as stated in Theorem 2, it holds that

$$\mathbf{E}\left[\mathrm{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}_T^*)\right] \le C_3 T^{\max\{\kappa, 1-\kappa\}} \tag{21a}$$

$$\mathbf{E}\left[\left\|\left[\sum_{t=1}^T g_t(x_t)\right]_+\right\|\right] \le C_4 T^{1-\kappa/2}. \tag{21b}$$

Remark 6: The parameter κ for the sequences $\{\alpha_{i,t}\}$, $\{\beta_{i,t}\}$, and $\{\gamma_{i,t}\}$ in Corollary 2 enables the user to tradeoff the expected static regret bound for the expected constraint violation bound. For example, setting $\kappa = 1/2$ in Corollary 2 gives $\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \check{\boldsymbol{x}}_T^*)] = \mathcal{O}(\sqrt{T})$ and $\mathbf{E}[\|[\sum_{t=1}^T g_t(x_t)]_+\|] = \mathcal{O}(T^{3/4})$. These two bounds are the same as the bounds achieved in [11], [12], and [47]. In other words, Algorithm 2 is as efficient as the algorithms proposed in [11], [12], and [47]. However, Jenatton *et al.* [11] and Sun *et al.* [12] use full-information feedback and Mahdavi *et al.* [47] consider bandit setting only for the constraint functions. The algorithms proposed in [11], [12], and [47] are centralized and the constraint functions considered in [11] and [47] are time-invariant. Moreover, in [12] and [47], the total number of rounds and in [11], [12], [47], the upper bounds of the loss and constraint functions and their subgradients need to be known in advance to execute the algorithms. Also, an $\mathcal{O}(\sqrt{T})$ expected static regret bound was achieved by the bandit algorithm in [43]. However, in [43], static set constraints (rather than time-varying inequality constraints) are considered and the proposed algorithm is centralized (rather than distributed). Moreover, in [43], the total number of rounds and the Lipschitz constant need to be known in advance.

Remark 7: If the learners exchange data with their neighbors over a static complete graph rather than the time-varying directed graph, then with some modifications to the proposed algorithms and proofs, we can show that all the results on constraint violation still hold if we replace the constraint violation metric $\|[\sum_{t=1}^T g_t(x_t)]_+\|$ by the more stricter metric $\sum_{t=1}^T \|[g(x_t)]_+\|^2$. It is unclear how to extend this over general time-varying directed graphs. We leave this for future work.

VI. Numerical Simulations

This section evaluates the performance of Algorithms 1 and 2 in solving the power generation example introduced in Section III-A. The local cost and constraint functions are denoted

$$f_{i,t}(x_{i,t}) = x_{i,t}^\top \Pi_{i,t}^\top \Pi_{i,t} x_{i,t} + \langle \pi_{i,t}, x_{i,t}\rangle$$

$$g_{i,t}(x_{i,t}) = x_{i,t}^\top \Phi_{i,t}^\top \Phi_{i,t} x_{i,t} + \langle \phi_{i,t}, x_{i,t}\rangle + c_{i,t}$$

where $\Pi_{i,t} \in \mathbb{R}^{p_i \times p_i}$, $\pi_{i,t} \in \mathbb{R}_+^{p_i}$, $\Phi_{i,t} \in \mathbb{R}^{p_i \times p_i}$, $\phi_{i,t} \in \mathbb{R}^{p_i}$, and $c_{i,t} \in \mathbb{R}$. At each time t, an undirected graph is used as the communication graph. Specifically, connections between vertices are random and the probability of two vertices being connected is $\rho > 0$. Moreover, edges $(i, i+1), i \in [n-1]$ are added and $[W_t]_{ij} = 1/n$ if $(j,i) \in \mathcal{E}_t$ and $[W_t]_{ii} = 1 - \sum_{j \in \mathcal{N}_i^{\mathrm{in}}(\mathcal{G}_t)} [W_t]_{ij}$. The parameters are set as: $n = 50$, $m = 1$, $p_i = 6$, $\mathbb{X}_i = [-10, 10]^{p_i}$, and $\rho = 0.2$. Each element of $\Pi_{i,t}$, $\pi_{i,t}$, $\Phi_{i,t}$, $\phi_{i,t}$, and $c_{i,t}$ are drawn from the discrete uniform distribution in $[-5, 5]$, [0,10], $[-5, 5]$, $[-5, 5]$, and $[-5, -1]$, respectively. Under aforementioned settings, Assumptions 1 and 2 hold.

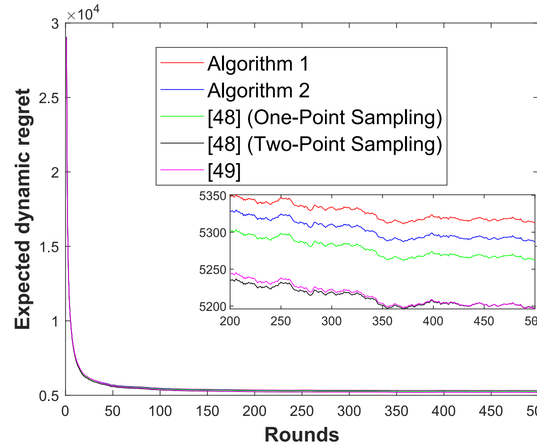


Fig. 1. Comparison of evolutions of the expected dynamic regret $\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}_T^*)]/T$.

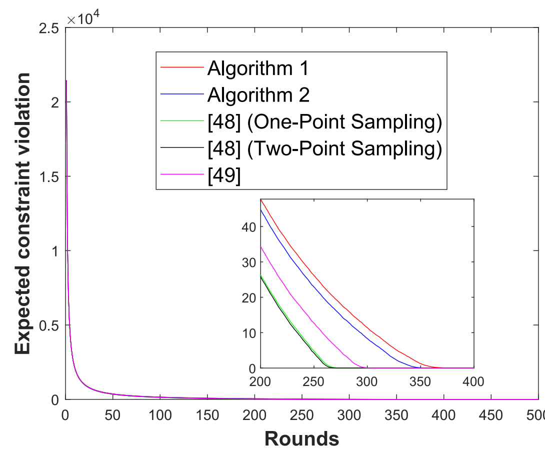


Fig. 2. Comparison of evolutions of the expected constraint violation $\mathbf{E}[\|[\sum_{t=1}^T g_t(x_t)]_+\|]/T$.

Since there are no other distributed bandit online algorithms to solve the problem of online optimization with time-varying coupled inequality constraints, we compare our Algorithms 1 and 2 with the centralized one- and two-point sampling algorithms in [48], which use full-information feedback for the constraint functions, and the centralized two-point sampling algorithm in [49]. Figs. 1 and 2 show the evolutions of $\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{x}_T^*)]/T$ and $\mathbf{E}[\|[\sum_{t=1}^T g_t(x_t)]_+\|]/T$, respectively. The average is taken over 100 realizations. Note that $\mathbf{E}[\|[\sum_{t=1}^T g_t(x_t)]_+\|]/T \to 0$. This is in agreement with (14b), (20b), and the theoretical results shown in [48] and [49]. From the zoomed figures, we see that the centralized algorithms in [48] and [49] achieve smaller expected dynamic regret and constraint violation than our distributed algorithms, which is reasonable. We also see that Algorithm 2 achieves smaller expected dynamic regret and constraint violation than Algorithm 1, which is consistent with our theoretical results.

VII. Conclusion

In this article, we considered the distributed bandit online convex optimization problem with time-varying coupled inequality constraints. We proposed distributed bandit online algorithms with one- and two-point bandit feedback. We showed that sublinear expected regret and constraint violation can be achieved by both proposed algorithms. We showed that the results can be cast as nontrivial extensions of existing literature on online optimization and bandit feedback. Future research directions include considering an adaptive choice of the number of samplings at each round by different learners, relaxing the doubly stochastic assumption, studying sampling noise, achieving a smaller regret bound under stronger assumptions for the cost functions, and trying to establish sublinear constraint violation under a stricter constraint violation metric.

Appendix

A. Useful Lemmas

The following two lemmas are used in the proofs.

Lemma 3: Let $\mathbb{K}$ be a nonempty closed convex subset of $\mathbb{R}^p$ and let $a, b,$ and$c$ be three vectors in $\mathbb{R}^p$. The following statements hold.

1) For each $x \in \mathbb{R}^p$, $\mathcal{P}_{\mathbb{K}}(x)$ exists and is unique.
2) $\mathcal{P}_{\mathbb{K}}(x)$ is nonexpansive, i.e.,
$$\|\mathcal{P}_{\mathbb{K}}(x) - \mathcal{P}_{\mathbb{K}}(y)\| \le \|x - y\| \quad \forall x, y \in \mathbb{R}^p. \tag{22}$$
3) If $a \le b$, then
$$\|[a]_+\| \le \|b\| \text{and} [a]_+ \le [b]_+. \tag{23}$$
4) If $x_1 = \mathcal{P}_{\mathbb{K}}(c - a)$, then
$$2\langle x_1 - y, a\rangle \le \|y - c\|^2 - \|y - x_1\|^2 - \|x_1 - c\|^2 \quad \forall y \in \mathbb{K}. \tag{24}$$

Proof: The first two parts are from [55, Th. 1.5.5].

Substituting $x = a$ and $y = a - b$ into (22) with $\mathbb{K} = \mathbb{R}^p_+$ gives (23). If $a \le b$, then it is straightforward to see $[a]_+ \le [b]_+$ since all inequalities are understood componentwise.

Denote $h(y) = \|c - y\|^2 + 2\langle a, y\rangle$. Then, $x_1 = \arg\min_{y\in\mathbb{K}} h(y)$. This optimality condition implies that
$$\langle x_1 - y, \nabla h(x_1)\rangle \le 0 \quad \forall y \in \mathbb{K}.$$
Substituting $\nabla h(x_1) = 2x_1 - 2c + 2a$ into aforementioned inequality yields (24). ■

Lemma 4: For any constants $\theta \in [0, 1]$, $\kappa \in [0, 1)$, and $s \le T \in \mathbb{N}_+$, it holds that
$$(t + 1)^\kappa \left(\frac{1}{t^\theta} - \frac{1}{(t + 1)^\theta}\right) \le \frac{1}{t} \quad \forall t \in \mathbb{N}_+ \tag{25a}$$
$$\sum_{t=s}^{T} \frac{1}{t^\kappa} \le \frac{T^{1-\kappa}}{1 - \kappa} \tag{25b}$$
$$\sum_{t=s}^{T} \frac{1}{t} \le 2\log(T), \text{if} T \ge 3. \tag{25c}$$

Proof:

1) Denote $h_t(\theta) = \frac{1}{t^\theta} - \frac{1}{(t+1)^\theta}$. Then, for any fixed $t \in \mathbb{N}_+$, $\max_{\theta\in[0,1]}\{h_t(\theta)\} = h_t(1)$ since $\frac{dh_t(\theta)}{d\theta} \ge 0 \quad \forall\theta \in [0, 1]$. Hence, $(t + 1)^\kappa h_t(\theta) \le (t + 1)^\kappa h_t(1) = \frac{(t+1)^\kappa}{t(t+1)} \le \frac{1}{t}$, i.e., (25a) holds.
2) (25b) holds since
$$\sum_{t=s}^{T} \frac{1}{t^\kappa} \le \int_{s-1}^{T} \frac{1}{t^\kappa} dt = \frac{T^{1-\kappa} - (s - 1)^{1-\kappa}}{1 - \kappa} \le \frac{T^{1-\kappa}}{1 - \kappa}.$$
3) (25c) holds since
$$\sum_{t=s}^{T} \frac{1}{t} \le \frac{1}{s} + \int_{s}^{T} \frac{1}{t} dt = \frac{1}{s} + \log(T) - \log(s) \le 2\log(T).$$
■

B. Proof of Lemma 2

1) $\nabla \hat{f}(x) = \mathbf{E}_{u\in\mathbb{S}^p}[\hat{\nabla}_1 f(x)]$ is the result of [32, Lemma 1]. $\nabla \hat{f}(x) = \mathbf{E}_{u\in\mathbb{S}^p}[\hat{\nabla}_2 f(x)]$ since $\mathbf{E}_{u\in\mathbb{S}^p}[f(x)u] = f(x)\mathbf{E}_{u\in\mathbb{S}^p}[u] = \mathbf{0}_p$.
2) $(1 - \xi)\mathbb{K}$ is convex since $\mathbb{K}$ is convex.
For any $x, y \in (1 - \xi)\mathbb{K}$ and $\alpha \in [0, 1]$, then $\alpha x + (1 - \alpha)y \in (1 - \xi)\mathbb{K}$ since $(1 - \xi)\mathbb{K}$ is convex and $\alpha x + (1 - \alpha)y + \delta v \in \mathbb{K}$ due to Lemma 1. Moreover
$$\hat{f}(\alpha x + (1 - \alpha)y) = \mathbf{E}_{v\in\mathbb{B}^p}[f(\alpha x + (1 - \alpha)y + \delta v)]$$
$$\le \mathbf{E}_{v\in\mathbb{B}^p}[\alpha f(x + \delta v) + (1 - \alpha)f(y + \delta v)]$$
$$= \alpha\hat{f}(x) + (1 - \alpha)\hat{f}(y).$$
Hence, $\hat{f}$ is convex on $(1 - \xi)\mathbb{K}$.
From Lemma 1, we know that $(1 - \xi)\mathbb{K}$ is a subset of the interior of $\mathbb{K}$. Then, for any $x \in (1 - \xi)\mathbb{K}$, from [56, Th. 3.1.15], we know that $\nabla f(x)$ exists. Moreover
$$\hat{f}(x) = \mathbf{E}_{v\in\mathbb{B}^p}[f(x + \delta v)]$$
$$\ge \mathbf{E}_{v\in\mathbb{B}^p}[f(x) + \delta\langle\nabla f(x), v\rangle] = f(x).$$
3) For any $x, y \in (1 - \xi)\mathbb{K}$
$$\left|\hat{f}(x) - \hat{f}(y)\right| = |\mathbf{E}_{v\in\mathbb{B}^p}[f(x + \delta v) - f(y + \delta v)]|$$
$$\le \mathbf{E}_{v\in\mathbb{B}^p}[|f(x + \delta v) - f(y + \delta v)|]$$
$$\le \mathbf{E}_{v\in\mathbb{B}^p}[L_0(f)\|x - y\|] = L_0(f)\|x - y\|.$$
Hence, $\hat{f}$ is Lipschitz-continuous on $(1 - \xi)\mathbb{K}$ with constant $L_0(f)$.
Similarly
$$\left\|\nabla\hat{f}(x) - \nabla\hat{f}(y)\right\|$$
$$= \frac{p}{\delta}\|\mathbf{E}_{u\in\mathbb{S}^p}[f(x + \delta u)u - f(y + \delta u)u]\|$$
$$\le \frac{p}{\delta}\mathbf{E}_{u\in\mathbb{S}^p}[|f(x + \delta u) - f(y + \delta u)|\,\|u\|]$$
$$\le \frac{p}{\delta}\mathbf{E}_{u\in\mathbb{S}^p}[L_0(f)\|x - y\|] = \frac{pL_0(f)}{\delta}\|x - y\|.$$
Hence, $\nabla\hat{f}$ is Lipschitz-continuous on $(1 - \xi)\mathbb{K}$ with constant $pL_0(f)/\delta$.
For any $x \in (1 - \xi)\mathbb{K}$
$$\left|\hat{f}(x) - f(x)\right| = |\mathbf{E}_{v\in\mathbb{B}^p}[f(x + \delta v)] - \mathbf{E}_{v\in\mathbb{B}^p}[f(x)]|$$
$$\le \mathbf{E}_{v\in\mathbb{B}^p}[|f(x + \delta v) - f(x)|]$$
$$\le \mathbf{E}_{v\in\mathbb{B}^p}[\delta L_0(f)\|v\|] \le \mathbf{E}_{v\in\mathbb{B}^p}[\delta L_0(f)] = \delta L_0(f).$$
4) For any $x \in (1 - \xi)\mathbb{K}$ and $u \in \mathbb{S}^p$
$$\left|\hat{f}(x)\right| = |\mathbf{E}_{v\in\mathbb{B}^p}[f(x + \delta v)]|$$

$$\leq \mathbf{E}_{v\in\mathbb{B}^p}\left[|f(x+\delta v)|\right] \leq F_0(f)$$

and

$$\left\|\hat{\nabla}_1 f(x)\right\| = \left\|\frac{p}{\delta}f(x+\delta u)u\right\|$$
$$\leq \frac{p}{\delta}|f(x+\delta u)|\|u\| \leq \frac{pF_0(f)}{\delta}.$$

5) For any $x \in (1-\xi)\mathbb{K}$ and $u \in \mathbb{S}^p$

$$\left\|\hat{\nabla}_2 f(x)\right\| = \left\|\frac{p}{\delta}(f(x+\delta u)-f(x))u\right\|$$
$$\leq \frac{pL_0(f)}{\delta}\|x+\delta u - x\|\|u\| = pL_0(f).$$

C. Proof of Theorem 1

To prove Theorem 1, the following three lemmas are used. Lemma 5 presents the results on the local dual variables, whereas Lemma 6 provides an upper bound for the regret of one round. Lemma 7 provides the expected regret constraint violation bounds for Algorithm 1 for the general case.

To simplify notation, we denote $\beta_t = \beta_{i,t}$, $\gamma_t = \gamma_{i,t}$, and $\xi_t = \xi_{i,t}$.

Lemma 5: Suppose Assumptions 1 and 2 hold. For all $i \in [n]$ and $t \in \mathbb{N}_+$, $\tilde{q}_{i,t}$ and $q_{i,t}$ generated by Algorithm 1 satisfy

$$\|\tilde{q}_{i,t+1}\| \leq \frac{\sqrt{m}F_g}{\beta_t}, \|q_{i,t}\| \leq \frac{\sqrt{m}F_g}{\beta_t} \quad (26a)$$

$$\|\tilde{q}_{i,t+1} - \bar{q}_t\| \leq 2\sqrt{m}nF_g\tau\sum_{s=1}^{t-1}\gamma_{s+1}\lambda^{t-1-s} \quad (26b)$$

$$\frac{\Delta_{t+1}}{2\gamma_{t+1}} \leq (\bar{q}_t - q)^\top g_t(x_t) + 2mnF_g^2\gamma_{t+1} + \frac{n\beta_{t+1}}{2}\|q\|^2 + d_1(t) \quad (26c)$$

where $\bar{q}_t = \frac{1}{n}\sum_{i=1}^n q_{i,t}$

$$\Delta_t = \sum_{i=1}^n \|q_{i,t} - q\|^2 - (1-\beta_t\gamma_t)\sum_{i=1}^n\|q_{i,t-1} - q\|^2 \quad (27)$$

q is an arbitrary vector in $\mathbb{R}^m_+$, and $d_1(t) = 2mn^2F_g^2\tau\sum_{s=1}^t \gamma_{s+1}\lambda^{t-s}$.

Proof:

1) From (6), we have

$$\|g_{i,t}(x_{i,t})\| \leq \sqrt{m}F_g, \forall i \in [n] \quad \forall t \in \mathbb{N}_+. \quad (28)$$

We prove (26a) by induction.

It is straightforward to see that $q_{i,1} = \tilde{q}_{i,2} = \mathbf{0}_m \quad \forall i \in [n]$, thus $\|\tilde{q}_{i,2}\| \leq \frac{\sqrt{m}F_g}{\beta_1}, \|q_{i,1}\| \leq \frac{\sqrt{m}F_g}{\beta_1} \quad \forall i \in [n]$. Assume that (26a) is true at time t for all $i \in [n]$. We show that it remains true at time $t+1$. First, from (23), (9d), (28), $1-\gamma_{t+1}\beta_{t+1} \geq 0$, and $\beta_t \geq \beta_{t+1}$, we know that for all $i \in [n]$

$$\|q_{i,t+1}\| \leq (1-\gamma_{t+1}\beta_{t+1})\|\tilde{q}_{i,t+1}\| + \gamma_{t+1}\|g_{i,t}(x_{i,t})\|$$
$$\leq (1-\gamma_{t+1}\beta_{t+1})\frac{\sqrt{m}F_g}{\beta_t} + \gamma_{t+1}\sqrt{m}F_g$$
$$\leq (1-\gamma_{t+1}\beta_{t+1})\frac{\sqrt{m}F_g}{\beta_{t+1}} + \gamma_{t+1}\sqrt{m}F_g \leq \frac{\sqrt{m}F_g}{\beta_{t+1}}.$$

Then, the convexity of norms and $\sum_{j=1}^n[W_t]_{ij} = 1$ yield

$$\|\tilde{q}_{i,t+2}\| \leq \sum_{j=1}^n[W_{t+1}]_{ij}\|q_{j,t+1}\| \leq \sum_{j=1}^n[W_t]_{ij}\frac{\sqrt{m}F_g}{\beta_{t+1}}$$
$$= \frac{\sqrt{m}F_g}{\beta_{t+1}} \quad \forall i \in [n].$$

Thus, (26a) follows.

2) Note that (9d) can be rewritten as

$$q_{i,t+1} = \sum_{j=1}^n [W_t]_{ij}q_{j,t} + \epsilon^q_{i,t} \quad (29)$$

where $\epsilon^q_{i,t} = [(1-\gamma_{t+1}\beta_{t+1})\tilde{q}_{i,t+1} + \gamma_{t+1}g_{i,t}(x_{i,t})]_+ - \tilde{q}_{i,t+1}$. Then, (22), (26a), and (28) give

$$\|\epsilon^q_{i,t}\| \leq \|-\gamma_{t+1}\beta_{t+1}\tilde{q}_{i,t+1} + \gamma_{t+1}g_{i,t}(x_{i,t})\|$$
$$\leq 2\sqrt{m}F_g\gamma_{t+1} \quad \forall i \in [n]. \quad (30)$$

Then, from Assumption 1, [20, Lemma 2], $q_{i,1} = \mathbf{0}_m \quad \forall i \in [n]$, and (30), we know that for any $i \in [n]$ and $t \in \mathbb{N}_+$

$$\|q_{i,t+1} - \bar{q}_{t+1}\| \leq 2\sqrt{m}nF_g\tau\sum_{s=1}^t\gamma_{s+1}\lambda^{t-s}. \quad (31)$$

Thus, (26b) follows since $\sum_{j=1}^n[W_t]_{ij} = 1$ and $\|\tilde{q}_{i,t+1} - \bar{q}_t\| = \|\sum_{j=1}^n[W_t]_{ij}q_{j,t} - \bar{q}_t\| \leq \sum_{j=1}^n[W_t]_{ij}\|q_{j,t} - \bar{q}_t\|$.

3) Applying (22) to (9d) yields

$$\|q_{i,t} - q\|^2 \leq \|(1-\beta_t\gamma_t)\tilde{q}_{i,t} + \gamma_t g_{i,t-1}(x_{i,t-1}) - q\|^2$$
$$= \|\tilde{q}_{i,t} - q\|^2 + \gamma_t^2\|g_{i,t-1}(x_{i,t-1}) - \beta_t\tilde{q}_{i,t}\|^2$$
$$+ 2\gamma_t[\tilde{q}_{i,t} - q]^\top g_{i,t-1}(x_{i,t-1}) - 2\beta_t\gamma_t[\tilde{q}_{i,t} - q]^\top\tilde{q}_{i,t}. \quad (32)$$

For the first term of the right-hand side of (32), by convexity of norms and $\sum_{j=1}^n[W_{t-1}]_{ij} = 1$, it can be concluded that

$$\|\tilde{q}_{i,t} - q\|^2 = \left\|\sum_{j=1}^n[W_{t-1}]_{ij}q_{j,t-1} - \sum_{j=1}^n[W_{t-1}]_{ij}q\right\|^2$$
$$\leq \sum_{j=1}^n[W_{t-1}]_{ij}\|q_{j,t-1} - q\|^2. \quad (33)$$

For the second term of the right-hand side of (32), (26a), and (28) yield

$$\gamma_t^2\|g_{i,t-1}(x_{i,t-1}) - \beta_t\tilde{q}_{i,t}\|^2 \leq \left(2\sqrt{m}F_g\gamma_t\right)^2. \quad (34)$$

For the fourth term of the right-hand side of (32), we have

$$2\gamma_t[\tilde{q}_{i,t} - q]^\top g_{i,t-1}(x_{i,t-1})$$
$$= 2\gamma_t[\bar{q}_{t-1} - q]^\top g_{i,t-1}(x_{i,t-1})$$
$$+ 2\gamma_t[\tilde{q}_{i,t} - \bar{q}_{t-1}]^\top g_{i,t-1}(x_{i,t-1}). \quad (35)$$

Moreover, from (28) and (26b), we have

$$2\gamma_t[\tilde{q}_{i,t} - \bar{q}_{t-1}]^\top g_{i,t-1}(x_{i,t-1})$$
$$\leq 2\gamma_t\|\tilde{q}_{i,t} - \bar{q}_{t-1}\|\|g_{i,t-1}(x_{i,t-1})\| \leq \frac{2\gamma_t d_1(t-1)}{n}. \quad (36)$$

For the last term of the right-hand side of (32), neglecting the nonnegative term $\beta_t\gamma_t\|\tilde{q}_{i,t}\|^2$ gives

$$-2\beta_t\gamma_t[\tilde{q}_{i,t}-q]^\top\tilde{q}_{i,t} \le \beta_t\gamma_t\left(\|q\|^2-\|\tilde{q}_{i,t}-q\|^2\right). \tag{37}$$

Combining (32)–(37), summing over $i\in[n]$, dividing by $2\gamma_t$, using $\sum_{i=1}^n[W_{t-1}]_{ij}=1 \quad \forall t\in\mathbb{N}_+$, setting $t=t+1$, and rearranging the terms yields (26c). ■

Lemma 6: Suppose Assumptions 1 and 2 hold. For all $i\in[n]$, let $\{x_t\}$ be the sequence generated by Algorithm 1 and $\{y_t\}$ be an arbitrary sequence in $\mathbb{X}$, then

$$\begin{aligned}
&f_t(x_t)-f_t(y_t)\\
&\le(\bar{q}_t)^\top(g_t(y_t)-g_t(x_t))+2d_1(t)+d_2(t)\\
&\quad+\sum_{i=1}^n\frac{p_i^2F_{f_i}^2\alpha_{i,t+1}}{\delta_{i,t}^2}+\sum_{i=1}^n\frac{2R_i\|y_{i,t+1}-y_{i,t}\|}{\alpha_{i,t+1}}\\
&\quad+d_3(t)+\mathbf{E}_{\mathfrak{U}_t}[d_4(t)] \quad \forall t\in\mathbb{N}_+
\end{aligned}\tag{38}$$

where $d_1(t)$ is given in Lemma 5, $d_2(t)=\sum_{i=1}^n\{(2\delta_{i,t}+R_i\xi_t)(\sqrt{m}G_{g_i}\|q_{i,t}\|+G_{f_i})+\frac{2R_i^2(\xi_t-\xi_{t+1})}{\alpha_{i,t+1}}\}$, $d_3(t)=2m\max_{i\in[n]}\{\frac{p_i^2F_{g_i}^2\alpha_{i,t+1}}{\delta_{i,t}^2}\}(n\|q\|^2+\sum_{i=1}^n\|q_{i,t}-q\|^2)$, $d_4(t)=\sum_{i=1}^n\frac{\|\check{y}_{i,t}-z_{i,t}\|^2-\|\check{y}_{i,t+1}-z_{i,t+1}\|^2}{2\alpha_{i,t+1}}$, and $\check{y}_{i,t}=(1-\xi_t)y_{i,t}$.

Proof: For any $i\in[n]$, $t\in\mathbb{N}_+$, and $x\in(1-\xi_t)\mathbb{X}_i$, denote

$$\hat{f}_{i,t}(x)=\mathbf{E}_{v\in\mathbb{B}^p}[f_{i,t}(x+\delta_{i,t}v)]$$
$$\hat{g}_{i,t}(x)=\mathbf{E}_{v\in\mathbb{B}^p}[g_{i,t}(x+\delta_{i,t}v)].$$

From Lemma 2, (6), (28), (8a), and (8b), we know that $\hat{f}_{i,t}(x)$ and $\hat{g}_{i,t}(x)$ are convex on $(1-\xi_t)\mathbb{X}_i$, and for any $i\in[n]$, $t\in\mathbb{N}_+$, and $x\in(1-\xi_t)\mathbb{X}_i$

$$\nabla\hat{f}_{i,t}(x)=\mathbf{E}_{\mathfrak{U}_t}\left[\hat{\nabla}_1f_{i,t}(x)\right] \tag{39a}$$
$$f_{i,t}(x)\le\hat{f}_{i,t}(x)\le f_{i,t}(x)+G_{f_i}\delta_{i,t} \tag{39b}$$
$$\left\|\hat{\nabla}_1f_{i,t}(x)\right\|\le\frac{p_iF_{f_i}}{\delta_{i,t}} \tag{39c}$$
$$\nabla\hat{g}_{i,t}(x)=\mathbf{E}_{\mathfrak{U}_t}\left[\hat{\nabla}_1g_{i,t}(x)\right] \tag{39d}$$
$$g_{i,t}(x)\le\hat{g}_{i,t}(x)\le g_{i,t}(x)+G_{g_i}\delta_{i,t}\mathbf{1}_m \tag{39e}$$
$$\left\|\hat{\nabla}_1g_{i,t}(x)\right\|\le\frac{\sqrt{m}p_iF_{g_i}}{\delta_{i,t}} \tag{39f}$$
$$\|\hat{g}_{i,t}(x)\|\le\sqrt{m}F_{g_i}. \tag{39g}$$

Then, (8a), (8b), (5), and (39b) yield

$$|f_{i,t}(x_{i,t})-f_{i,t}(z_{i,t})|\le G_{f_i}\|x_{i,t}-z_{i,t}\|\le G_{f_i}\delta_{i,t} \tag{40a}$$

$$\begin{aligned}
&\|g_{i,t}(x_{i,t})-g_{i,t}(z_{i,t})\|\\
&\le\sqrt{m}G_{g_i}\|x_{i,t}-z_{i,t}\|\le\sqrt{m}G_{g_i}\delta_{i,t}
\end{aligned}\tag{40b}$$

$$\begin{aligned}
&\hat{f}_{i,t}(\check{y}_{i,t})-f_{i,t}(y_{i,t})\\
&=f_{i,t}(\check{y}_{i,t})-f_{i,t}(y_{i,t})+\hat{f}_{i,t}(\check{y}_{i,t})-f_{i,t}(\check{y}_{i,t})\\
&\le G_{f_i}\|\check{y}_{i,t}-y_{i,t}\|+\hat{f}_{i,t}(\check{y}_{i,t})-f_{i,t}(\check{y}_{i,t})\\
&\le G_{f_i}R_i\xi_t+G_{f_i}\delta_{i,t}
\end{aligned}\tag{40c}$$

$$f_{i,t}(z_{i,t})-\hat{f}_{i,t}(z_{i,t})\le 0 \tag{40d}$$
$$\|g_{i,t}(\check{y}_{i,t})-g_{i,t}(y_{i,t})\|\le\sqrt{m}G_{g_i}R_i\xi_t. \tag{40e}$$

From that $\hat{f}_{i,t}(x)$ is convex on $(1-\xi_t)\mathbb{X}_i$, we have that

$$\begin{aligned}
&\hat{f}_{i,t}(z_{i,t})-\hat{f}_{i,t}(\check{y}_{i,t})\le\left\langle\nabla\hat{f}_{i,t}(z_{i,t}),z_{i,t}-\check{y}_{i,t}\right\rangle\\
&=\left\langle\mathbf{E}_{\mathfrak{U}_t}\left[\hat{\nabla}_1f_{i,t}(z_{i,t})\right],z_{i,t}-\check{y}_{i,t}\right\rangle\\
&=\mathbf{E}_{\mathfrak{U}_t}\left[\left\langle\hat{\nabla}_1f_{i,t}(z_{i,t}),z_{i,t}-\check{y}_{i,t}\right\rangle\right]
\end{aligned}\tag{41}$$

where the first equality holds from (39a) and the last equality holds since $z_{i,t}$ is independent of $\mathfrak{U}_t$.

Next, we rewrite the right-hand side of (41) into two terms and bound them individually.

$$\begin{aligned}
&\mathbf{E}_{\mathfrak{U}_t}\left[\left\langle\hat{\nabla}_1f_{i,t}(z_{i,t}),z_{i,t}-\check{y}_{i,t}\right\rangle\right]\\
&=\mathbf{E}_{\mathfrak{U}_t}\left[\left\langle\hat{\nabla}_1f_{i,t}(z_{i,t}),z_{i,t}-z_{i,t+1}\right\rangle\right]\\
&\quad+\mathbf{E}_{\mathfrak{U}_t}\left[\left\langle\hat{\nabla}_1f_{i,t}(z_{i,t}),z_{i,t+1}-\check{y}_{i,t}\right\rangle\right].
\end{aligned}\tag{42}$$

For the first term of the right-hand side of (42), the Cauchy–Schwarz inequality and (39c) give

$$\begin{aligned}
&\left\langle\hat{\nabla}_1f_{i,t}(z_{i,t}),z_{i,t}-z_{i,t+1}\right\rangle\\
&\le\left\|\hat{\nabla}_1f_{i,t}(z_{i,t})\right\|\|z_{i,t}-z_{i,t+1}\|\le\frac{p_iF_{f_i}}{\delta_{i,t}}\|z_{i,t}-z_{i,t+1}\|\\
&\le\frac{p_i^2F_{f_i}^2\alpha_{i,t+1}}{\delta_{i,t}^2}+\frac{1}{4\alpha_{i,t+1}}\|z_{i,t}-z_{i,t+1}\|^2.
\end{aligned}\tag{43}$$

For the second term of the right-hand side of (42), it follows from (10) that

$$\begin{aligned}
&\mathbf{E}_{\mathfrak{U}_t}\left[\left\langle\hat{\nabla}_1f_{i,t}(z_{i,t}),z_{i,t+1}-\check{y}_{i,t}\right\rangle\right]\\
&=\mathbf{E}_{\mathfrak{U}_t}\left[\left\langle\left(\hat{\nabla}_1g_{i,t}(z_{i,t})\right)^\top\tilde{q}_{i,t+1},\check{y}_{i,t}-z_{i,t+1}\right\rangle\right]\\
&\quad+\mathbf{E}_{\mathfrak{U}_t}\left[\langle a_{i,t+1},z_{i,t+1}-\check{y}_{i,t}\rangle\right]\\
&=\mathbf{E}_{\mathfrak{U}_t}\left[\left\langle\left(\hat{\nabla}_1g_{i,t}(z_{i,t})\right)^\top\tilde{q}_{i,t+1},\check{y}_{i,t}-z_{i,t}\right\rangle\right]\\
&\quad+\mathbf{E}_{\mathfrak{U}_t}\left[\left\langle\left(\hat{\nabla}_1g_{i,t}(z_{i,t})\right)^\top\tilde{q}_{i,t+1},z_{i,t}-z_{i,t+1}\right\rangle\right]\\
&\quad+\mathbf{E}_{\mathfrak{U}_t}\left[\langle a_{i,t+1},z_{i,t+1}-\check{y}_{i,t}\rangle\right].
\end{aligned}\tag{44}$$

For the first term of the right-hand side of (44), noting that $x_{i,t}$ and $\tilde{q}_{i,t+1}$ are dependent of $\mathfrak{U}_t$, from (39d), $\tilde{q}_{i,t+1}\ge\mathbf{0}_m$, $\bar{q}_t\ge\mathbf{0}_m$, (39e), and that $\hat{g}_{i,t}$ is convex, we have

$$\begin{aligned}
&\mathbf{E}_{\mathfrak{U}_t}\left[\left\langle\left(\hat{\nabla}_1g_{i,t}(z_{i,t})\right)^\top\tilde{q}_{i,t+1},\check{y}_{i,t}-z_{i,t}\right\rangle\right]\\
&=\left\langle\left(\mathbf{E}_{\mathfrak{U}_t}\left[\hat{\nabla}_1g_{i,t}(z_{i,t})\right]\right)^\top\tilde{q}_{i,t+1},\check{y}_{i,t}-z_{i,t}\right\rangle\\
&=\left\langle(\nabla\hat{g}_{i,t}(z_{i,t}))^\top\tilde{q}_{i,t+1},\check{y}_{i,t}-z_{i,t}\right\rangle
\end{aligned}$$

$$
\begin{aligned}
&\le [\tilde{q}_{i,t+1}]^\top \hat{g}_{i,t}(\check{y}_{i,t}) - [\tilde{q}_{i,t+1}]^\top \hat{g}_{i,t}(z_{i,t}) \\
&= [\bar{q}_t]^\top [\hat{g}_{i,t}(\check{y}_{i,t}) - \hat{g}_{i,t}(z_{i,t})] \\
&\quad + [\tilde{q}_{i,t+1} - \bar{q}_t]^\top [\hat{g}_{i,t}(\check{y}_{i,t}) - \hat{g}_{i,t}(z_{i,t})] \\
&\le [\bar{q}_t]^\top [g_{i,t}(\check{y}_{i,t}) + \delta_{i,t} G_{g_i} \mathbf{1}_m - g_{i,t}(z_{i,t})] \\
&\quad + [\tilde{q}_{i,t+1} - \bar{q}_t]^\top [\hat{g}_{i,t}(\check{y}_{i,t}) - \hat{g}_{i,t}(z_{i,t})]. \qquad (45)
\end{aligned}
$$

From (26b) and (39g), we have

$$
[\tilde{q}_{i,t+1} - \bar{q}_t]^\top [\hat{g}_{i,t}(\check{y}_{i,t}) - \hat{g}_{i,t}(z_{i,t})] \le \frac{2d_1(t)}{n}. \qquad (46)
$$

For the second term of the right-hand side of (44), from the Cauchy–Schwarz inequality, (39f), and (33), we have

$$
\begin{aligned}
&\left\langle \left(\hat{\nabla}_1 g_{i,t}(z_{i,t})\right)^\top \tilde{q}_{i,t+1}, z_{i,t} - z_{i,t+1} \right\rangle \\
&= q^\top \hat{\nabla}_1 g_{i,t}(z_{i,t})(z_{i,t} - z_{i,t+1}) \\
&\quad + (\tilde{q}_{i,t+1} - q)^\top \hat{\nabla}_1 g_{i,t}(z_{i,t})(z_{i,t} - z_{i,t+1}) \\
&\le \frac{2mp_i^2 F_{g_i}^2 \alpha_{i,t+1}}{\delta_{i,t}^2} \|q\|^2 + \frac{1}{8\alpha_{i,t+1}} \|z_{i,t+1} - z_{i,t}\|^2 \\
&\quad + \frac{2mp_i^2 F_{g_i}^2 \alpha_{i,t+1}}{\delta_{i,t}^2} \|\tilde{q}_{i,t+1} - q\|^2 + \frac{1}{8\alpha_{i,t+1}} \|z_{i,t+1} - z_{i,t}\|^2 \\
&\le 2m \max_{i\in[n]} \left\{ \frac{p_i^2 F_{g_i}^2 \alpha_{i,t+1}}{\delta_{i,t}^2} \right\} \|q\|^2 + \frac{1}{4\alpha_{i,t+1}} \|z_{i,t+1} - z_{i,t}\|^2 \\
&\quad + 2m \max_{i\in[n]} \left\{ \frac{p_i^2 F_{g_i}^2 \alpha_{i,t+1}}{\delta_{i,t}^2} \right\} \sum_{j=1}^n [W_t]_{ij} \|q_{j,t} - q\|^2. \qquad (47)
\end{aligned}
$$

For the last term of the right-hand side of (44), noting that $\check{y}_{i,t} \in (1-\xi_t)\mathbb{X}_i \subseteq (1-\xi_{t+1})\mathbb{X}_i$ since $\xi_t \ge \xi_{t+1}$ and applying (24) to the update rule (9b) yields

$$
\begin{aligned}
&2\alpha_{i,t+1}\langle a_{i,t+1}, z_{i,t+1} - \check{y}_{i,t}\rangle \\
&\le \|\check{y}_{i,t} - z_{i,t}\|^2 - \|\check{y}_{i,t} - z_{i,t+1}\|^2 - \|z_{i,t+1} - z_{i,t}\|^2 \\
&= \|\check{y}_{i,t+1} - z_{i,t+1}\|^2 - \|\check{y}_{i,t} - z_{i,t+1}\|^2 + \|\check{y}_{i,t} - z_{i,t}\|^2 \\
&\quad - \|\check{y}_{i,t+1} - z_{i,t+1}\|^2 - \|z_{i,t+1} - z_{i,t}\|^2. \qquad (48)
\end{aligned}
$$

The first two terms of the right-hand side of (48) can be bounded by

$$
\begin{aligned}
&\|\check{y}_{i,t+1} - z_{i,t+1}\|^2 - \|\check{y}_{i,t} - z_{i,t+1}\|^2 \\
&\le \|\check{y}_{i,t+1} - \check{y}_{i,t}\| \|\check{y}_{i,t+1} + \check{y}_{i,t} - 2z_{i,t+1}\| \\
&\le 4R_i \|(1-\xi_{t+1}) y_{i,t+1} - (1-\xi_t) y_{i,t}\| \\
&= 4R_i \|(1-\xi_{t+1})(y_{i,t+1} - y_{i,t}) + (\xi_t - \xi_{t+1}) y_{i,t}\| \\
&\le 4R_i \|y_{i,t+1} - y_{i,t}\| + 4R_i^2(\xi_t - \xi_{t+1}) \qquad (49)
\end{aligned}
$$

where the last inequality holds since $\{\xi_t\} \subseteq (0,1)$ is nonincreasing.

Combining (40c)–(49), taking expectation in $\mathfrak{U}_t$, summing over $i \in [n]$, and rearranging the terms yields (38). ■

Lemma 7: Suppose Assumptions 1 and 2 hold. For any $T \in \mathbb{N}_+$, let $\boldsymbol{x}_T$ be the sequence generated by Algorithm 1. Then, for any comparator sequence $\boldsymbol{y}_T \in \mathcal{X}_T$

$$
\begin{aligned}
&\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{y}_T)] \\
&\le \sum_{t=1}^T \mathbf{E}[d_2(t)] + C_0 \sum_{t=1}^T \gamma_{t+1} + \sum_{t=1}^T \sum_{i=1}^n \frac{p_i^2 F_{f_i}^2 \alpha_{i,t+1}}{\delta_{i,t}^2} \\
&\quad + \sum_{i=1}^n \frac{2R_i^2}{\alpha_{i,T+1}} + \sum_{t=1}^{T-1} \sum_{i=1}^n \frac{2R_i \|y_{i,t+1} - y_{i,t}\|}{\alpha_{i,t+1}} \\
&\quad + \frac{1}{2} \sum_{t=1}^T \tilde{\alpha}_t \mathbf{E}\left[\|q_{i,t}\|^2\right] \qquad (50a)
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{E}\left[\left\| \left[\sum_{t=1}^T g_t(x_t)\right]_+ \right\|^2\right] \\
&\le d_5(T) \Bigg\{ \sum_{t=1}^T \mathbf{E}[d_2(t)] + C_0 \sum_{t=1}^T \gamma_{t+1} \\
&\quad + \sum_{t=1}^T \sum_{i=1}^n \frac{p_i^2 F_{f_i}^2 \alpha_{i,t+1}}{\delta_{i,t}^2} + \sum_{i=1}^n \frac{2R_i^2}{\alpha_{i,T+1}} + 2T \sum_{i=1}^n F_{f_i} \\
&\quad + \frac{1}{2} \sum_{t=1}^T \tilde{\alpha}_t \mathbf{E}\left[\|q_{i,t} - q_c\|^2\right] \Bigg\} \qquad (50b)
\end{aligned}
$$

where $\tilde{\alpha}_t = \sum_{i=1}^n (4m \max_{i\in[n]}\{\frac{p_i^2 F_{g_i}^2 \alpha_{i,t+1}}{\delta_{i,t}^2}\} + \frac{1}{\gamma_{t+1}} - \frac{1}{\gamma_t} - \beta_{t+1})$, $d_5(T) = 2n(\frac{1}{\gamma_1} + \sum_{t=1}^T (4m \max_{i\in[n]}\{\frac{p_i^2 F_{g_i}^2 \alpha_{i,t+1}}{\delta_{i,t}^2}\} + \beta_{t+1}))$, and $q_c = 2[\sum_{t=1}^T g_t(x_t)]_+ / d_5(T) \in \mathbb{R}_+^m$.

Proof:

1) For any $\lambda \in (0,1)$ and nonnegative sequence $\zeta_1, \zeta_2, \ldots$, it holds that

$$
\sum_{t=1}^T \sum_{s=1}^t \zeta_{s+1} \lambda^{t-s} = \sum_{t=1}^T \zeta_{t+1} \sum_{s=0}^{T-t} \lambda^s \le \frac{1}{1-\lambda} \sum_{t=1}^T \zeta_{t+1}. \qquad (51)
$$

Thus

$$
\sum_{t=1}^T d_1(t) \le \frac{2\sqrt{m} n^2 \tau B_1 F_g}{1-\lambda} \sum_{t=1}^T \gamma_{t+1}. \qquad (52)
$$

The definition of Δ_t given by (27) yields

$$
\begin{aligned}
&-\sum_{t=1}^T \frac{\Delta_{t+1}}{2\gamma_{t+1}} \\
&= \frac{1}{2} \sum_{i=1}^n \sum_{t=1}^T \left[\frac{1}{\gamma_t} \|q_{i,t} - q\|^2 - \frac{1}{\gamma_{t+1}} \|q_{i,t+1} - q\|^2 \right] \\
&\quad + \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^n \left(\frac{1}{\gamma_{t+1}} - \frac{1}{\gamma_t} - \beta_{t+1} \right) \|q_{i,t} - q\|^2 \\
&= \frac{1}{2} \sum_{i=1}^n \left[\frac{1}{\gamma_1} \|q_{i,1} - q\|^2 - \frac{1}{\gamma_{T+1}} \|q_{i,T+1} - q\|^2 \right] \\
&\quad + \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^n \left(\frac{1}{\gamma_{t+1}} - \frac{1}{\gamma_t} - \beta_{t+1} \right) \|q_{i,t} - q\|^2
\end{aligned}
$$

$$\leq \frac{n}{2\gamma_1}\|q\|^2+\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}\left(\frac{1}{\gamma_{t+1}}-\frac{1}{\gamma_t}-\beta_{t+1}\right)\|q_{i,t}-q\|^2 \tag{53}$$

where the last inequality holds since $q_{i,1} = \mathbf{0}_m$ and $\|q_{i,T+1} - q\|^2 \geq 0$.

From the properties of conditional expectation, we know that

$$\mathbf{E}_{\mathcal{U}_T}\left[\mathbf{E}_{\mathfrak{U}_t}\left[d_4(t)\right]\right] = \mathbf{E}\left[d_4(t)\right] \quad \forall t \in [T] \tag{54}$$

where we recall the definition $\mathcal{U}_T = \bigcup_{s=1}^T \mathfrak{U}_s$.

Noting that $\{\alpha_t\}$ is nonincreasing and (5), for any $s \in [T]$, we have

$$\sum_{t=s}^{T} d_4(t) = \frac{1}{2}\sum_{t=s}^{T}\sum_{i=1}^{n}\left(\frac{1}{\alpha_{i,t}}\|\check{y}_{i,t} - z_{i,t}\|^2 - \frac{1}{\alpha_{i,t+1}}\|\check{y}_{i,t+1} - z_{i,t+1}\|^2\right) + \frac{1}{2}\sum_{t=s}^{T}\sum_{i=1}^{n}\left(\frac{1}{\alpha_{i,t+1}} - \frac{1}{\alpha_{i,t}}\right)\|\check{y}_{i,t} - z_{i,t}\|^2$$
$$\leq \frac{1}{2\alpha_{i,s}}\sum_{i=1}^{n}\|\check{y}_{i,s} - z_{i,s}\|^2 - \frac{1}{2\alpha_{i,T+1}}\sum_{i=1}^{n}\|\check{y}_{i,T+1} - z_{i,T+1}\|^2 + 2\sum_{i=1}^{n}\left(\frac{1}{\alpha_{i,T+1}} - \frac{1}{\alpha_{i,s}}\right)R_i^2 \leq \sum_{i=1}^{n}\frac{2R_i^2}{\alpha_{i,T+1}}. \tag{55}$$

Let $g_c : \mathbb{R}_+^m \to \mathbb{R}$ be a function defined as

$$g_c(q) = \left(\sum_{t=1}^{T} g_t(x_t)\right)^\top q - \frac{d_5(T)}{4}\|q\|^2. \tag{56}$$

Combining (26c) and (38), summing over $t \in [T]$, using (52)–(56) and $g_t(y_t) \leq \mathbf{0}_m, \boldsymbol{y}_T \in \mathcal{X}_T$, and taking expectation in $\mathcal{U}_T$ yields

$$\mathbf{E}\left[g_c(q)\right] + \mathbf{E}\left[\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{y}_T)\right] \leq \sum_{t=1}^{T}\mathbf{E}\left[d_2(t)\right] + C_0\sum_{t=1}^{T}\gamma_{t+1} + \sum_{t=1}^{T}\sum_{i=1}^{n}\frac{p_i^2F_{f_i}^2\alpha_{i,t+1}}{\delta_{i,t}^2} + \sum_{i=1}^{n}\frac{2R_i^2}{\alpha_{i,T+1}} + \sum_{t=1}^{T}\sum_{i=1}^{n}\frac{2R_i\|y_{i,t+1} - y_{i,t}\|}{\alpha_{i,t+1}} + \frac{1}{2}\sum_{t=1}^{T}\tilde{\alpha}_t\mathbf{E}\left[\|q_{i,t} - q\|^2\right] \quad \forall q \in \mathbb{R}_+^m. \tag{57}$$

Then, substituting $q = \mathbf{0}_m$ into (57), setting $y_{i,T+1} = y_{i,T}$, and noting that $\{\alpha_t\}$ is nonincreasing yields (50a).

2) Substituting $q = q_c$ into $g_c(q)$ gives

$$g_c(q_c) = \frac{\left\|\left[\sum_{t=1}^{T} g_t(x_t)\right]_+\right\|^2}{d_5(T)}. \tag{58}$$

Moreover, (6) gives

$$|\,\mathrm{Reg}(\boldsymbol{x}_T, \boldsymbol{y}_T)| \leq 2T\sum_{i=1}^{n}F_{f_i} \quad \forall \boldsymbol{y}_T \in \mathcal{X}_T. \tag{59}$$

Substituting $q = q_c$ and $y_t = \check{x}_T^*, t \in [T+1]$ into (57), combining (58)–(59), and rearranging the terms gives (50b). ■

Before proving Theorem 1, let us generally explain why choosing the sequences in (13). The intuition of the choice is to let the terms in the right-hand side of (50a) and (50b) be as small as possible. Specifically, the first four terms in the right-hand side of (50a) need to be sublinear. Moreover, $\tilde{\alpha}_t$ should be nonpositive otherwise it is unclear how to show that the last terms in the right-hand side of (50a) and (50b) are sublinear. We are now ready to prove Theorem 1.

1) Applying (25a), (25b), and (26a) to the first three terms of the right-hand side of (50a) and noting $\theta_2 < \theta_3$ gives

$$\sum_{t=1}^{T}\mathbf{E}\left[d_2(t)\right] \leq \sum_{i=1}^{n}\frac{mF_gG_{g_i}(2r_i + R_i)}{1 - \theta_3 + \theta_2}T^{1-\theta_3+\theta_2} + \sum_{i=1}^{n}\frac{G_{f_i}(2r_i + R_i)}{1 - \theta_3}T^{1-\theta_3} + C_{1,1}\log(T) \tag{60a}$$

$$C_0\sum_{t=1}^{T}\gamma_{t+1} \leq \frac{C_0}{\theta_2}T^{\theta_2} \tag{60b}$$

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{p_i^2F_{f_i}^2\alpha_{i,t+1}}{\delta_{i,t}^2} \leq \sum_{i=1}^{n}\frac{F_{f_i}^2}{4mF_{g_i}^2(1 - \theta_1 + 2\theta_3)}T^{1-\theta_1+2\theta_3}. \tag{60c}$$

From (13) and $\theta_1 - 2\theta_3 \geq \theta_2$, we know that

$$\tilde{\alpha}_t = \frac{1}{(t+1)^{\theta_1-2\theta_3}} + \frac{t+1}{(t+1)^{\theta_2}} - \frac{t}{t^{\theta_2}} - \frac{2}{(t+1)^{\theta_2}} \leq \frac{1}{(t+1)^{\theta_2}} + \frac{t+1}{(t+1)^{\theta_2}} - \frac{t}{t^{\theta_2}} - \frac{2}{(t+1)^{\theta_2}} = \frac{t}{(t+1)^{\theta_2}} - \frac{t}{t^{\theta_2}} < 0. \tag{61}$$

Combining (50a) and (60a)–(61) yields (14a).

2) Using (25b) and noting $\theta_1 - 2\theta_3 \geq \theta_2$ gives

$$d_5(T) \leq C_{2,1}T^{1-\theta_2}. \tag{62}$$

Combining (50b) and (60a)–(62) gives

$$\mathbf{E}\left[\left\|\left[\sum_{t=1}^{T} g_t(x_t)\right]_+\right\|^2\right] \leq C_2^2T^{2-\theta_2}. \tag{63}$$

Finally, combining (63) and $(\mathbf{E}[\|[\sum_{t=1}^T g_t(x_t)]_+\|])^2 \leq \mathbf{E}[\|[\sum_{t=1}^T g_t(x_t)]_+\|^2]$ (which follows from Jensen's inequality) gives (14b).

D. Proof of Theorem 2

The proof is similar to the proof of Theorem 1 with some modifications. Lemmas 5–7 are replaced by Lemmas 8–10.

To simplify notation, we denote $\alpha_t=\alpha_{i,t}$, $\beta_t=\beta_{i,t}$, $\gamma_t=\gamma_{i,t}$, and $\xi_t=\xi_{i,t}$.

Lemma 8: Suppose Assumptions 1 and 2 hold. For all $i\in[n]$ and $t\in\mathbb{N}_+$, $\tilde{q}_{i,t}$ and $q_{i,t}$ generated by Algorithm 2 satisfy

$$\|\tilde{q}_{i,t+1}\|\le\frac{B_1}{\beta_t},\|q_{i,t}\|\le\frac{B_1}{\beta_t} \tag{64a}$$

$$\|\tilde{q}_{i,t+1}-\bar{q}_t\|\le 2nB_1\tau\sum_{s=1}^{t-1}\gamma_{s+1}\lambda^{t-1-s} \tag{64b}$$

$$\begin{aligned}&\frac{\Delta_{t+1}}{2\gamma_{t+1}}\\&\le(\bar{q}_t-q)^\top g_t(x_t)+2nB_1^2\gamma_{t+1}+d_6(t)\\&\quad+\frac{1}{2}\sum_{i=1}^n\left(2mp_i^2G_{g_i}^2\alpha_{t+1}+\beta_{t+1}\right)\|q\|^2+d_7(t)\end{aligned} \tag{64c}$$

where q is an arbitrary vector in $\mathbb{R}^m_+$, $d_6(t)=2\sqrt{m}n^2B_1F_g\tau\sum_{s=1}^t\gamma_{s+1}\lambda^{t-s}$, and $d_7(t)=\frac{1}{4\alpha_{t+1}}\sum_{i=1}^n\|x_{i,t+1}-x_{i,t}\|^2+\sum_{i=1}^n[\tilde{q}_{i,t+1}]^\top\hat{\nabla}_2g_{i,t}(x_{i,t})(x_{i,t+1}-x_{i,t})$.

Proof: From the fifth part in Lemma 2 and (8b), we know that for all $i\in[n]$, $x\in(1-\xi_{i,t})\mathbb{X}_i$, and $t\in\mathbb{N}_+$

$$\left\|\hat{\nabla}_2g_{i,t}(x)\right\|\le\sqrt{m}p_iG_{g_i}. \tag{65}$$

Hence, (5), (6), (18), and (65) yield

$$\begin{aligned}\|c_{i,t+1}\|&\le\|g_{i,t}(x_{i,t})\|+\left\|\hat{\nabla}_2g_{i,t}(x_{i,t})\right\|\|x_{i,t+1}-x_{i,t}\|\\&\le\sqrt{m}F_{g_i}+2\sqrt{m}p_iG_{g_i}R_i\le B_1\quad\forall i\in[n]\quad\forall t\in\mathbb{N}_+.\end{aligned} \tag{66}$$

Replacing $z_{i,t}$ and $g_{i,t}(z_{i,t})$ by $x_{i,t}$ and $c_{i,t+1}$, respectively, and following steps similar to those used to prove (26a) and (26b) yields (64a) and (64b).

Applying (22) to (16c) yields

$$\begin{aligned}\|q_{i,t}-q\|^2&\le\|(1-\beta_t\gamma_t)\tilde{q}_{i,t}+\gamma_tc_{i,t}-q\|^2\\&=\|\tilde{q}_{i,t}-q\|^2+\gamma_t^2\|c_{i,t}-\beta_t\tilde{q}_{i,t}\|^2\\&\quad+2\gamma_t[\tilde{q}_{i,t}]^\top\hat{\nabla}_2g_{i,t-1}(x_{i,t-1})(x_{i,t}-x_{i,t-1})\\&\quad-2\gamma_tq^\top\hat{\nabla}_2g_{i,t-1}(x_{i,t-1})(x_{i,t}-x_{i,t-1})\\&\quad+2\gamma_t[\tilde{q}_{i,t}-q]^\top g_{i,t-1}(x_{i,t-1})\\&\quad-2\beta_t\gamma_t[\tilde{q}_{i,t}-q]^\top\tilde{q}_{i,t}.\end{aligned} \tag{67}$$

For the fourth term of the right-hand side of (67), (65) and the Cauchy–Schwarz inequality yield

$$\begin{aligned}&-2\gamma_tq^\top\hat{\nabla}_2g_{i,t-1}(x_{i,t-1})(x_{i,t}-x_{i,t-1})\\&\le2\gamma_t\left(mp_i^2G_{g_i}^2\alpha_t\|q\|^2+\frac{1}{4\alpha_t}\|x_{i,t}-x_{i,t-1}\|^2\right).\end{aligned} \tag{68}$$

Replacing (32) by (67), using (68), and following steps similar to those used to prove (26c) yields (64c). ■

Lemma 9: Suppose Assumptions 1 and 2 hold. For all $i\in[n]$, let $\{x_t\}$ be the sequence generated by Algorithm 2 and $\{y_t\}$ be an arbitrary sequence in $\mathbb{X}$, then

$$\begin{aligned}&f_t(x_t)-f_t(y_t)\\&\le(\bar{q}_t)^\top(g_t(y_t)-g_t(x_t))+2d_6(t)-\mathbf{E}_{\mathfrak{U}_t}[d_7(t)]\\&\quad+\sum_{i=1}^np_i^2G_{f_i}^2\alpha_{t+1}+\sum_{i=1}^n\frac{2R_i\|y_{i,t+1}-y_{i,t}\|}{\alpha_{t+1}}\\&\quad+d_8(t)+\mathbf{E}_{\mathfrak{U}_t}[d_9(t)]\quad\forall t\in\mathbb{N}_+\end{aligned} \tag{69}$$

where $d_8(t)=\sum_{i=1}^n\{(\delta_{i,t}+R_i\xi_t)(\sqrt{m}G_{g_i}\|q_{i,t}\|+G_{f_i})+\frac{2R_i^2(\xi_t-\xi_{t+1})}{\alpha_{t+1}}\}$, $d_9(t)=\frac{1}{2\alpha_{t+1}}\sum_{i=1}^n(\|\check{y}_{i,t}-x_{i,t}\|^2-\|\check{y}_{i,t+1}-x_{i,t+1}\|^2)$, and $\check{y}_{i,t}=(1-\xi_t)y_{i,t}$.

Proof: Replacing $z_{i,t}$, $a_{i,t}$, and (39c) by $x_{i,t}$, $b_{i,t}$, and $\|\hat{\nabla}_2f_{i,t}(x)\|\le p_iG_{f_i}$, respectively, deleting (47), and following steps similar to those used to prove (38) yields (69). ■

Lemma 10: Suppose Assumptions 1 and 2 hold. For any $T\in\mathbb{N}_+$, let $\boldsymbol{x}_T$ be the sequence generated by Algorithm 2. Then, for any comparator sequence $\boldsymbol{y}_T\in\mathcal{X}_T$

$$\begin{aligned}&\mathbf{E}[\mathrm{Reg}(\boldsymbol{x}_T,\boldsymbol{y}_T)]\\&\le\sum_{t=1}^T\mathbf{E}[d_8(t)]+\hat{C}_0\sum_{t=1}^T\gamma_{t+1}+\sum_{i=1}^n\frac{2R_i^2}{\alpha_{T+1}}\\&\quad+\frac{1}{2}\sum_{t=1}^T\sum_{i=1}^n\left(\frac{1}{\gamma_{t+1}}-\frac{1}{\gamma_t}-\beta_{t+1}\right)\mathbf{E}\left[\|q_{i,t}\|^2\right]\\&\quad+\sum_{t=1}^T\sum_{i=1}^np_i^2G_{f_i}^2\alpha_{t+1}+\frac{2R_{\max}V(\boldsymbol{y}_T)}{\alpha_T}\end{aligned} \tag{70a}$$

$$\begin{aligned}&\mathbf{E}\left[\left\|\left[\sum_{t=1}^Tg_t(x_t)\right]_+\right\|^2\right]\\&\le d_{10}(T)\left\{\sum_{t=1}^T\mathbf{E}[d_8(t)]+\hat{C}_0\sum_{t=1}^T\gamma_{t+1}+\sum_{i=1}^n\frac{2R_i^2}{\alpha_{T+1}}\right.\\&\quad+\frac{1}{2}\sum_{t=1}^T\sum_{i=1}^n\left(\frac{1}{\gamma_{t+1}}-\frac{1}{\gamma_t}-\beta_{t+1}\right)\mathbf{E}\left[\|q_{i,t}-\hat{q}_c\|^2\right]\\&\quad\left.+\sum_{t=1}^T\sum_{i=1}^np_i^2G_{f_i}^2\alpha_{t+1}+2T\sum_{i=1}^nF_{f_i}\right\}\end{aligned} \tag{70b}$$

where $d_{10}(T)=2n(\frac{1}{\gamma_1}+\sum_{t=1}^T(2mp_i^2G_{g_i}^2\alpha_{t+1}+\beta_{t+1}))$ and $\hat{q}_c=2[\sum_{t=1}^Tg_t(x_t)]_+/d_{10}(T)\in\mathbb{R}^m_+$.

Proof: With Lemmas 8 and 9 at hand, the proof of Lemma 10 follows steps similar to those used to prove Lemma 7. ■

With Lemmas 8–10 at hand, the proof of (20a) and (20b) in Theorem 2 follows steps similar to those used to prove (14a) and (14b) in Theorem 1.

References

[1] S. Shalev-Shwartz, "Online learning and online convex optimization," *Found. Trends Mach. Learn.*, vol. 4, no. 2, pp. 107–194, 2012.

[2] X. Zhou, E. DallAnese, L. Chen, and A. Simonetto, "An incentive-based online optimization framework for distribution grids," *IEEE Trans. Autom. Control*, vol. 63, no. 7, pp. 2019–2031, Jul. 2018.

[3] S. Shahrampour and A. Jadbabaie, "Distributed online optimization in dynamic environments using mirror descent," *IEEE Trans. Autom. Control*, vol. 63, no. 3, pp. 714–725, Mar. 2018.

[4] D. Yuan, D. W. Ho, and G.-P. Jiang, "An adaptive primal-dual subgradient algorithm for online distributed constrained optimization," *IEEE Trans. Cybern.*, vol. 48, no. 11, pp. 3045–3055, Nov. 2018.

[5] N. Cesa-Bianchi, P. M. Long, and M. K. Warmuth, "Worst-case quadratic loss bounds for prediction using linear functions and gradient descent," *IEEE Trans. Neural Netw.*, vol. 7, no. 3, pp. 604–619, May 1996.

[6] C. Gentile and M. K. Warmuth, "Linear hinge loss and average margin," in *Proc. Adv. Neural Inf. Process. Syst.*, 1999, pp. 225–231.

[7] G. J. Gordon, "Regret bounds for prediction problems," in *Proc. Conf. Learn. Theory*, 1999, pp. 29–40.

[8] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proc. Int. Conf. Mach. Learn.*, 2003, pp. 928–936.

[9] A. Agarwal, O. Dekel, and L. Xiao, "Optimal algorithms for online convex optimization with multi-point bandit feedback," in *Proc. Conf. Learn. Theory*, 2010, pp. 28–40.

[10] E. Hazan, A. Agarwal, and S. Kale, "Logarithmic regret algorithms for online convex optimization," *Mach. Learn.*, vol. 69, no. 2–3, pp. 169–192, 2007.

[11] R. Jenatton, J. Huang, and C. Archambeau, "Adaptive algorithms for online convex optimization with long-term constraints," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 402–411.

[12] W. Sun, D. Dey, and A. Kapoor, "Safety-aware algorithms for adversarial contextual bandit," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3280–3288.

[13] M. J. Neely and H. Yu, "Online convex optimization with time-varying constraints," 2017, *arXiv:1702.04783*.

[14] H. Yu, M. Neely, and X. Wei, "Online convex optimization with stochastic constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1428–1438.

[15] H. Yu and M. J. Neely, "A low complexity algorithm with $O(\sqrt{T})$ regret and finite constraint violations for online convex optimization with long term constraints," *J. Mach. Learn. Res.*, vol. 21, no. 1, pp. 1–24, 2020.

[16] J. Yuan and A. Lamperski, "Online convex optimization for cumulative constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 6140–6149.

[17] X. Wei, H. Yu, and M. J. Neely, "Online primal-dual mirror descent under stochastic constraints," in *Proc. Abstr. SIGMETRICS/Perform. Joint Int. Conf. Meas. Model. Comput. Syst.*, 2020, pp. 3–4.

[18] N. Liakopoulos, A. Destounis, G. Paschos, T. Spyropoulos, and P. Mertikopoulos, "Cautious regret minimization: Online optimization with long-term budget constraints," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3944–3952.

[19] O. Sadeghi and M. Fazel, "Online continuous DR-submodular maximization with long-term budget constraints," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2020, pp. 4410–4419.

[20] S. Lee and M. M. Zavlanos, "On the sublinear regret of distributed primal-dual algorithms for online constrained optimization," 2017, *arXiv:1705.11128*,.

[21] X. Li, X. Yi, and L. Xie, "Distributed online optimization for multi-agent networks with coupled inequality constraints," *IEEE Trans. Autom. Control*, to be published, doi: 10.1109/TAC.2020.3021011.

[22] Y. Zhang, R. J. Ravier, V. Tarokh, and M. M. Zavlanos, "Distributed online convex optimization with improved dynamic regret," 2019, *arXiv:1911.05127*.

[23] E. Hazan, K. Singh, and C. Zhang, "Efficient regret minimization in non-convex games," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1433–1441.

[24] L. Yang, L. Deng, M. H. Hajiesmaili, C. Tan, and W. S. Wong, "An optimal algorithm for online non-convex learning," in *Proc. ACM Meas. Anal. Comput. Syst.*, 2018, pp. 1–25.

[25] S. Park, J. Mulvaney Kemp, M. Jin, and J. Lavaei, "Diminishing regret for online nonconvex optimization," 2020. [Online]. Available: https://lavaei.ieor.berkeley.edu/regret_ONO_2020_1.pdf

[26] E. Hazan, "Introduction to online convex optimization," *Found. Trends Optim.*, vol. 2, no. 3/4, pp. 157–325, 2016.

[27] J. Matyas, "Random optimization," *Autom. Remote Control*, vol. 26, no. 2, pp. 246–253, 1965.

[28] Y. Nesterov and V. Spokoiny, "Random gradient-free minimization of convex functions," *Found. Comput. Math.*, vol. 17, no. 2, pp. 527–566, 2017.

[29] D. Yuan and D. W. Ho, "Randomized gradient-free method for multiagent optimization over time-varying networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1342–1347, Jun. 2015.

[30] Y. Pang and G. Hu, "Randomized gradient-free distributed optimization methods for a multi-agent system with unknown cost function," *IEEE Trans. Autom. Control*, vol. 65, no. 1, pp. 333–340, Jan. 2020.

[31] Y. Tang, J. Zhang, and N. Li, "Distributed zero-order algorithms for nonconvex multi-agent optimization," *IEEE Trans. Control Netw. Syst.*, to be published, doi: 10.1109/TCNS.2020.3024321.

[32] A. D. Flaxman, A. T. Kalai, and H. B. McMahan, "Online convex optimization in the bandit setting: Gradient descent without a gradient," in *Proc. 17th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2005, pp. 385–394.

[33] V. Dani, S. M. Kakade, and T. P. Hayes, "The price of bandit information for online optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 345–352.

[34] J. D. Abernethy, E. Hazan, and A. Rakhlin, "Competing in the dark: An efficient algorithm for bandit linear optimization," in *Proc. Conf. Learn. Theory*, 2008, pp. 263–273.

[35] J. D. Abernethy, E. Hazan, and A. Rakhlin, "Interior-point methods for full-information and bandit online learning," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4164–4175, Jul. 2012.

[36] A. Saha and A. Tewari, "Improved regret guarantees for online smooth convex optimization with bandit feedback," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2011, pp. 636–642.

[37] E. Hazan and K. Levy, "Bandit convex optimization: Towards tight bounds," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 784–792.

[38] S. Bubeck, O. Dekel, T. Koren, and Y. Peres, "Bandit convex optimization: $\sqrt{T}$ regret in one dimension," in *Proc. Conf. Learn. Theory*, 2015, pp. 266–278.

[39] S. Bubeck and R. Eldan, "Multi-scale exploration of convex functions and bandit convex optimization," in *Proc. Conf. Learn. Theory*, 2016, pp. 583–589.

[40] E. Hazan and Y. Li, "An optimal algorithm for bandit convex optimization," 2016, *arXiv:1603.04350*.

[41] X. Hu, L. Prashanth, A. György, and C. Szepesvári, "(Bandit) convex optimization with biased noisy gradient oracles," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2016, pp. 819–828.

[42] J. C. Duchi, M. I. Jordan, M. J. Wainwright, and A. Wibisono, "Optimal rates for zero-order convex optimization: The power of two function evaluations," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2788–2806, May 2015.

[43] O. Shamir, "An optimal algorithm for bandit and zero-order convex optimization with two-point feedback," *J. Mach. Learn. Res.*, vol. 18, no. 52, pp. 1–11, 2017.

[44] T. Yang, L. Zhang, R. Jin, and J. Yi, "Tracking slowly moving clairvoyant: Optimal dynamic regret of online learning with true and noisy gradient," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 449–457.

[45] T. Tatarenko and M. Kamgarpour, "Minimizing regret in bandit online optimization in unconstrained and constrained action spaces," 2018, *arXiv:1806.05069*.

[46] I. Shames, D. Selvaratnam, and J. H. Manton, "Online optimization using zeroth order oracles," *IEEE Control Syst. Lett.*, vol. 4, no. 1, pp. 31–36, Jan. 2010.

[47] 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, 2012.

[48] T. Chen and G. B. Giannakis, "Bandit convex optimization for scalable and dynamic IoT management," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 1276–1286, Feb. 2019.

[49] X. Cao and K. R. Liu, "Online convex optimization with time-varying constraints and bandit feedback," *IEEE Trans. Autom. Control*, vol. 64, no. 7, pp. 2665–2680, Jul. 2019.

[50] D. Yuan, D. W. Ho, Y. Hong, and G. Jiang, "Online bandit convex optimization over a network," in *Proc. Chin. Control Conf.*, 2016, pp. 8090–8095.

[51] D. Yuan, A. Proutiere, and G. Shi, "Distributed online linear regression," 2019, *arXiv:1902.04774*.

[52] D. Yuan, A. Proutiere, and G. Shi, "Distributed online optimization with long-term constraints," 2019, *arXiv:1912.09705*.

[53] 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, 2020.

[54] A. Abdelaziz, E. Ali, and S. A. Elazim, "Combined economic and emission dispatch solution using flower pollination algorithm," *Int. J. Elect. Power Energy Syst.*, vol. 80, pp. 264–274, 2016.

[55] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*. New York, NY, USA: Springer-Verlag, 2007.

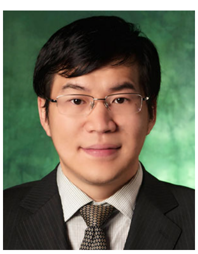
[56] Y. Nesterov, *Lectures on Convex Optimization*, 2nd ed. Berlin, Germany: Springer, 2018.

Xinlei Yi received the B.S. degree in mathematics from the China University of Geosciences, Wuhan, China, in 2011, and the M.S. degree in mathematics from Fudan University, Shanghai, China, in 2014. He is currently working toward the Ph.D. degree in electrical engineering with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden.

His current research interests include online optimization, distributed optimization, and event-triggered control.

Xiuxian Li (Member, IEEE) received the B.S. degree in mathematics and applied mathematics and the M.S. degree in pure mathematics from Shandong University, Jinan, China, in 2009 and 2012, respectively, and the Ph.D. degree in mechanical engineering from The University of Hong Kong, Hong Kong, in 2016.

Since 2016, he has been a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He was a Senior Research Associate with the Department of Biomedical Engineering, City University of Hong Kong, Hong Kong from 2018 to 2019. He held a Visiting Position with the King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, in September 2019. His research interests include distributed optimization, cooperative and distributed control, machine learning, mathematical programming, formation control, and multiagent networks.



Tao Yang (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Washington State University, Pullman, WA, USA, in 2012.

From 2012 to 2014, he was an ACCESS Postdoctoral Researcher with the ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden. He then joined the Pacific Northwest National Laboratory as a Postdoc, and was promoted to Scientist/Engineer II in 2015. From 2016 to 2019, he was an Assistant Professor with the Department of Electrical Engineering, University of North Texas, Denton, TX, USA. He is currently a Professor with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang, China. His research interests include distributed control and optimization with applications to power systems, cyber-physical systems, networked control systems, and multiagent systems.

Dr. Yang was the recipient of the Ralph E. Powe Junior Faculty Enhancement Award in 2018 and Best Student Paper Award (as an Advisor) of the 14th IEEE International Conference on Control and Automation.

Lihua Xie (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Newcastle, Callaghan, NSW, Australia, in 1992.

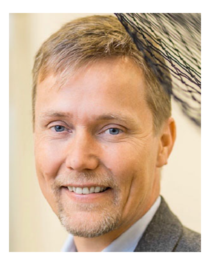
Since 1992, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a Professor and Director with the Delta-NTU Corporate Laboratory for Cyber-Physical Systems. He served as the Head of the Division of Control and Instrumentation from 2011 to 2014. He held teaching appointments with the Department of Automatic Control, Nanjing University of Science and Technology from 1986 to 1989. His research interests include robust control and estimation, networked control systems, multiagent networks, localization, and unmanned systems.

Dr. Xie is an Editor-in-Chief for the *Unmanned Systems* and an Associate Editor for the IEEE TRANSACTIONS ON NETWORK CONTROL SYSTEMS. He has served as an Editor for IET Book Series in Control and an Associate Editor for a number of journals, including IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Automatica*, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II. He was an IEEE Distinguished Lecturer from 2012 to 2014 and an Elected Member of Board of Governors, IEEE Control System Society from 2016 to 2018. He is a Fellow of IFAC.

Tianyou Chai (Fellow, IEEE) received the Ph.D. degree in control theory and engineering from Northeastern University, Shenyang, China, in 1985.

In 1988, he became a Professor with Northeastern University. He is the Founder and Director of the Center of Automation, which became a National Engineering and Technology Research Center and a State Key Laboratory. He has served as the Director of the Department of Information Science, National Natural Science Foundation of China from 2010 to 2018. He has developed control technologies with applications to various industrial processes. He has authored/coauthored 230 peer-reviewed international journal papers. His current research interests include modeling, control, optimization, and integrated automation of complex industrial processes.

Dr. Chai's paper titled *Hybrid intelligent control for optimal operation of shaft furnace roasting process* was selected as one of three best papers for the Control Engineering Practice Paper Prize for 2011–2013. He was the recipient of five prestigious awards, for his contributions, of the National Natural Science, National Science and Technology Progress and National Technological Innovation, the 2007 Industry Award for Excellence in Transitional Control Research from IEEE Multiple-conference on Systems and Control, and the 2017 Wook Hyun Kwon Education Award from Asian Control Association. He is a member of Chinese Academy of Engineering and a Fellow of IFAC.



Karl Henrik Johansson (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from Lund University, Lund, Sweden, in 1992 and 1997, respectively.

He is currently a Professor with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. He has held visiting positions with UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests include networked control systems, cyber-physical systems, and applications in transportation, energy, and automation.

Dr. Johansson has served on the IEEE Control Systems Society Board of Governors, the IFAC Executive Board, and the European Control Association Council. He was the recipient of several best paper awards and other distinctions from IEEE and ACM. He has been awarded Distinguished Professor with the Swedish Research Council and Wallenberg Scholar with the Knut and Alice Wallenberg Foundation. He was the recipient of the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from IFAC. He is Fellow of the Royal Swedish Academy of Engineering Sciences, and he is an IEEE Control Systems Society Distinguished Lecturer.