2025 IEEE 64th Conference on Decision and Control (CDC)
December 10-12, 2025. Rio de Janeiro, Brazil

2025 IEEE 64th Conference on Decision and Control (CDC) | 979-8-3315-2627-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/CDC57313.2025.11312327

Distributed Constrained Online Nonconvex Optimization with Compressed Communication

Kunpeng Zhang[1], Lei Xu[1], Xinlei Yi[2], Ming Cao[3], Karl H. Johansson[4], Tianyou Chai[1], and Tao Yang[1]

Abstract— **This paper considers distributed online nonconvex optimization with time-varying inequality constraints over a network of agents. For time-varying graphs, we propose a distributed online primal–dual algorithm with compressed communication to efficiently utilize communication resources. We show that the proposed algorithm establishes an $\mathcal{O}(T^{\max\{1-\theta_1,\theta_1\}})$ network regret bound and an $\mathcal{O}(T^{1-\theta_1/2})$ network cumulative constraint violation bound, where T is the number of iterations and $\theta_1 \in (0,1)$ is a user-defined trade-off parameter. When Slater's condition holds, the network cumulative constraint violation bound is reduced to $\mathcal{O}(T^{1-\theta_1})$. These bounds are comparable to the state-of-the-art results established by existing distributed online algorithms with perfect communication for distributed online convex optimization with (time-varying) inequality constraints. Finally, a simulation example is presented to validate the theoretical results.**

I. INTRODUCTION

Distributed online convex optimization offers a promising framework for modeling a variety of problems in dynamic, uncertain, and adversarial environments, with wide-ranging applications such as real-time routing in data networks and online advertisement placement in web search [1]. Various projection-based distributed online algorithms with sublinear regret have been proposed to solve the distributed online convex optimization problem, see, e.g., [2]–[10], and recent survey paper [11]. For example, for the fixed communication topology, the authors of [2] propose a projection-based distributed online subgradient descent algorithm, and establish an $\mathcal{O}(\sqrt{T})$ regret bound for general convex local loss functions. For strongly convex local loss functions, the authors of [3], [4] establish an $\mathcal{O}\big(\log(T)\big)$ regret bound. For time-varying communication topology, the authors of [5] propose a projection-based distributed online weighted dual averaging algorithm, and establish an $\mathcal{O}(\sqrt{T})$ regret bound for general convex local loss functions. By utilizing proportional-integral distributed feedback on the disagreement among neighboring agents, the authors of [6] propose a projection-based distributed online proportional-integral subgradient descent algorithm, and establish an $\mathcal{O}\big(\log(T)\big)$ regret bound for strongly convex local loss functions.

The aforementioned distributed online algorithms rely on agents exchanging their local data with perfect communication. Consequently, these algorithms encounter significant limitations arising from communication bottlenecks. To overcome the limitations, distributed online algorithms with compressed communication are studied for the fixed communication topology in the literature, see [12]–[14], and recent survey paper [15]. For example, the authors of [12] propose a decentralized online gradient descent algorithm with compressed communication by introducing an auxiliary variable to estimate the neighbors' decisions at each iteration. They establish an $\mathcal{O}(\sqrt{T})$ network regret bound for general convex local loss functions. Unlike the compression strategy employed in [12], the authors of [13], [14] introduce two auxiliary variables: the first serves the same purpose as the auxiliary variable in [12], while the second ensures that the first variable does not need to be exchanged. The compression strategy is effective when the communication topology is fixed. However, it becomes ineffective for a time-varying communication topology.

Note that inequality constraints are common in practical applications. However, performing projection operations onto such constraints can result in substantial computational and storage burdens. To deal with this challenge, the authors of [16] consider the idea of long term constraints proposed in [17], where inequality constraints are allowed to be violated temporarily, with the requirement that they are ultimately satisfied over the long term. This violation is measured by a performance metric named constraint violation where the projection onto the non-negative orthant is performed after summing the constraint functions over time. Accordingly, they propose a distributed online primal–dual algorithm and establish an $\mathcal{O}(T^{1/2+c})$ regret bound and an $\mathcal{O}(T^{1-c/2})$ constraint violation bound for general convex local loss and constraint functions, where $c \in (0,1/2)$ is a user-defined parameter. The regret bound is further reduced to $\mathcal{O}(T^{c})$ for strongly convex local loss functions. The authors of [18] use performance metric named cumulative constraint violation where the projection onto the non-negative orthant is performed before summing the constraint functions over time, which is proposed in [19]. Moreover, they establish an $\mathcal{O}\big(T^{\max\{c,1-c\}}\big)$ regret bound and an $\mathcal{O}(T^{1-c/2})$ cumulative constraint violation bound with $c \in (0,1)$ for quadratic

---

*This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB3305904, the National Natural Science Foundation of China under Grant 62133003.

K. Zhang, L. Xu, T. Chai and T. Yang are with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China `{2110343, 2010345}@stu.neu.edu.cn`, `{tychai, yangtao}@mail.neu.edu.cn`

[2] X. Yi is with the Department of Control Science and Engineering, College of Electronics and Information Engineering, Tongji University, Shanghai, 201804, China `xinleiyi@tongji.edu.cn`

[3] M. Cao is with the Engineering and Technology Institute Groningen, Faculty of Science and Engineering, University of Groningen, AG 9747 Groningen, The Netherlands `m.cao@rug.nl`

[4] K. H. Johansson is with Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, and he is also affiliated with Digital Futures, 10044, Stockholm, Sweden `kallej@kth.se`

Authorized licensed use limited to: KTH Royal Institute of Technology. Downloaded on April 22,2026 at 14:40:24 UTC from IEEE Xplore. Restrictions apply.

local loss functions and linear constraint functions. However, [16], [18] only consider static inequality constraints. The authors of [20] extend distributed online convex optimization with long-term constraints into the time-varying constraints setting. Moreover, the same network regret and cumulative constraint violation bounds as in [18] are established. However, the distributed online algorithms proposed in [18], [20] are unable to achieve reduced network cumulative constraint violation under Slater's condition. Slater's condition is a sufficient condition for strong duality to hold in convex optimization problems [21], and can be leveraged to achieve reduced constraint violation, e.g., [22], [23]. Recently, the authors of [24] propose a novel distributed online primal–dual algorithm, and establish reduced network cumulative constraint violation bounds under Slater's condition.

Unlike the aforementioned studies that focus on distributed online convex optimization, the authors of [25] investigate distributed online nonconvex optimization where local loss functions are nonconvex. To evaluate algorithm performance, they propose a novel regret metric based on the first-order optimality condition associated with the variational inequality. For this metric, the offline benchmark seeks a stationary point of the cumulative global loss functions across all iterations. Moreover, they establish an $\mathcal{O}(\sqrt{T})$ regret bound for general nonconvex local loss functions. However, [25] does not account for inequality constraints and uses perfect communication among agents.

Motivated by the above observations, this paper considers the distributed online nonconvex optimization problem with time-varying constraints. For time-varying communication topologies, we propose a distributed online primal–dual algorithm with compressed communication to efficiently utilize communication resources. Furthermore, base on several classes of appropriately chosen parameter sequences, we analyze how compressed communication influences network regret and cumulative constraint violation. The contributions are as follows.

- To the best of our knowledge, this paper is among the first to consider (time-varying) inequality constraints for distributed online nonconvex optimization. Compared to [2]–[10], [12]–[14], [16], [18], [20], [24] which focus on convex problems, we consider nonconvex problems where the absence of the convexity assumption on local loss functions makes the analysis more challenging. Compared to [25] which investigates distributed online nonconvex optimization, we additionally consider time-varying inequality constraints, which complicate both algorithm design and performance analysis. Moreover, similar to [12]–[14], we use compressed communication instead of perfect communication used in [25]. Different from [12]–[14] which consider a fixed communication topology, we consider a time-varying communication topology.
- We show in Theorem 1 that the proposed algorithm establishes an $\mathcal{O}(T^{\max\{1-\theta_1,1+\theta_1-\theta_2\}})$ network regret bound under $\theta_1 < \theta_2 < 1$ and an $\mathcal{O}(T^{\max\{1-\theta_1,\theta_1\}})$ network regret bound under $\theta_2 \geq 1$, and establishes an $\mathcal{O}(T^{1-\theta_1/2})$ network cumulative constraint violation bound. These bounds are the same as the results established in [20], [24] where the local loss functions are convex and perfect communication is used. When Slater's condition holds, we further show in Theorem 1 that the proposed algorithm establishes an reduced $\mathcal{O}(T^{1-\theta_1})$ network cumulative constraint violation bound. This bound is the same as the results established in [24].
- Under a different setting on the algorithm parameters, we show in Theorem 2 that the proposed algorithm establishes an $\mathcal{O}(\sqrt{T})$ network regret bound and an $\mathcal{O}(T^{3/4})$ network cumulative constraint violation bound. These bounds are the same as the results established in Theorem 1 when $\theta_1 = 1/2$ and $\theta_2 \geq 1$. Moreover, the network regret bound is the same as the results established in [25] where compressed communication and inequality constraints are not considered, as well as the results established in [12], [13] where inequality constraints and nonconvex local loss functions are not considered. When Slater's condition holds, we further show that in Theorem 2 that the proposed algorithm establishes an reduced $\mathcal{O}(\sqrt{T})$ network cumulative constraint violation bound. This bound is the same as the results established in Theorem 1 when $\theta_1 = 1/2$.

The remainder of this paper is organised as follows. Section II presents the problem formulation. Section III proposes the distributed online primal–dual algorithm with compressed communication, and analyzes its network regret and cumulative constraint violation bounds without and with Slater's condition, respectively. Section IV provides a simulation example to verify the theoretical results. Finally, Section V concludes this paper. Due to the space limitations, we omit all proofs, which can be found in the arXiv version [26].

Notations: All inequalities and equalities throughout this paper are understood componentwise. $\mathbb{N}_+$, $\mathbb{R}$, $\mathbb{R}^p$ and $\mathbb{R}^p_+$ denote the sets of all positive integers, real numbers, p-dimensional and nonnegative vectors, respectively. Given m and $n \in \mathbb{N}_+$, $[m]$ denotes the set $\{1, \cdots, m\}$, and $[m, n]$ denotes the set $\{m, \cdots, n\}$ for $m < n$. Given vectors x and y, x^T denotes the transpose of the vector x, and $\langle x, y \rangle$ denotes the standard inner. $\mathbf{0}_p$ and $\mathbf{1}_p$ denote the p-dimensional column vector whose components are all 0 and 1, respectively. $\mathrm{col}(q_1, \cdots, q_n)$ denotes the concatenated column vector of $q_i \in \mathbb{R}^{m_i}$ for $i \in [n]$. For a set $\mathbb{K} \in \mathbb{R}^p$ and a vector $x \in \mathbb{R}^p$, $\mathcal{P}_{\mathbb{K}}(x)$ denotes the projection of the vector x onto the set $\mathbb{K}$, i.e., $\mathcal{P}_{\mathbb{K}}(x) = \arg\min_{y \in \mathbb{K}} \|x - y\|^2$, and $[x]_+$ denotes $\mathcal{P}_{\mathbb{R}^p_+}(x)$. For a function f and a vector x, $\nabla f(x)$ denotes the subgradient of f at x.

II. PROBLEM FORMULATION

Consider the distributed online nonconvex optimization problem with time-varying constraints. At iteration t, a network of n agents is modeled by a time-varying directed graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$ with the agent set $\mathcal{V} = [n]$ and the edge set

Authorized licensed use limited to: KTH Royal Institute of Technology. Downloaded on April 22,2026 at 14:40:24 UTC from IEEE Xplore. Restrictions apply.

$\mathcal{E}_t \subseteq \mathcal{V}\times\mathcal{V}$. $(j,i)\in\mathcal{E}_t$ indicates that agent i can receive information from agent j. The sets of in- and out-neighbors of agent i are $\mathcal{N}_i^{\text{in}}(\mathcal{G}_t)=\{j\in[n]|(j,i)\in\mathcal{E}_t\}$ and $\mathcal{N}_i^{\text{out}}(\mathcal{G}_t)=\{j\in[n]|(i,j)\in\mathcal{E}_t\}$, respectively. Let $\{f_{i,t}:\mathbb{X}\to\mathbb{R}\}$ and $\{g_{i,t}:\mathbb{X}\to\mathbb{R}^{m_i}\}$ be the sequences of nonconvex local loss and convex local constraint functions, respectively, where $\mathbb{X}\subseteq\mathbb{R}^p$ is a known set, p and m_i are positive integers and $g_{i,t}\le\mathbf{0}_{m_i}$ is the local constraint. Each agent i selects a local decisions $x_{i,t}\in\mathbb{X}$ without prior access to $\{f_{i,t}\}$ and $\{g_{i,t}\}$. Upon selection, the nonconvex local loss function $\{f_{i,t}\}$ and convex local constraint function $\{g_{i,t}\}$ are privately revealed to the agent. The goal of the agent is to choose the decision sequence $\{x_{i,t}\}$ for $i\in[n]$ and $t\in[T]$ such that both network regret

$$\begin{aligned}\text{Net-Reg}(T) := \frac{1}{n}\sum_{i=1}^{n}\Big(&\sum_{t=1}^{T}\langle\nabla f_t(x_{i,t}),x_{i,t}\rangle\\ &-\inf_{x\in\mathcal{X}_T}\Big\langle\sum_{t=1}^{T}\nabla f_t(x_{i,t}),x\Big\rangle\Big),\end{aligned}\tag{1}$$

and network cumulative constraint violation

$$\text{Net-CCV}(T) := \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\|[g_t(x_{i,t})]_+\|,\tag{2}$$

increase sublinearly, where $f_t(x)=\frac{1}{n}\sum_{j=1}^{n}f_{j,t}(x)$ is the global loss function of the network at iteration t, $\mathcal{X}_T=\{x:x\in\mathbb{X},g_t(x)\le\mathbf{0}_m,\forall t\in[T]\}$ is the feasible set, and $g_t(x)=\text{col}\big(g_{1,t}(x),\cdots,g_{n,t}(x)\big)\in\mathbb{R}^m$ with $m=\sum_{i=1}^{n}m_i$ is the global constraint function of the network at iteration t. In addition, we consider the scenario where the communication between agents is compressed to efficiently utilize communication resources. Similar to existing literature on distributed online convex optimization with time-varying constraints, e.g., [20], [24], we assume that the feasible set $\mathcal{X}_T$ is nonempty for all $T\in\mathbb{N}_+$, ensuring the existence of the offline feasible static decision.

Note that the authors of [25] propose an individual regret metric $\text{Net-Reg}(T)=\max\limits_{x\in\mathbb{X}}\Big(\sum_{t=1}^{T}\langle\nabla f_{i,t}(x_{i,t}),x_{i,t}-x\rangle\Big)$ for distributed online nonconvex optimization by utilizing the first-order optimality condition associated with the variational inequality. In this paper, we consider time-varying inequality constraints, which cause the feasible set to become $\mathcal{X}_T$ instead of $\mathbb{X}$. Furthermore, the objective is to optimize the network-wide accumulated loss over all iterations, rather than the local one as considered in [25]. Therefore, we made a slight modification to the form of the regret metric in [25], transforming it into the form presented in (1).

The following commonly used assumptions are made throughout this paper.

Assumption 1. *The set $\mathbb{X}$ is convex and closed. Moreover, it is bounded by $R(\mathbb{X})$, i.e., for any $x\in\mathbb{X}$*

$$\|x\|\le R(\mathbb{X}).\tag{3}$$

Assumption 2. *For all $i\in[n]$, $t\in\mathbb{N}_+$, the subgradients $\nabla f_{i,t}(x)$ and $\nabla g_{i,t}(x)$ exist. Moreover, there exist constants G_1 and G_2 such that*

$$\|\nabla f_{i,t}(x)\|\le G_1,\tag{4a}$$

$$\|\nabla g_{i,t}(x)\|\le G_2, x\in\mathbb{X}.\tag{4b}$$

Due to the convexity of the local constraint function $g_{i,t}$, from Assumption 2 and Lemma 2.6 in [27], for all $i\in[n]$, $t\in\mathbb{N}_+$, we have

$$\|g_{i,t}(x)-g_{i,t}(y)\|\le G_2\|x-y\|, x,y\in\mathbb{X}.\tag{5}$$

Assumption 3. *For all $i\in[n]$, $t\in\mathbb{N}_+$, there exists a constant L such that*

$$\|\nabla f_{i,t}(x)-\nabla f_{i,t}(y)\|\le L\|x-y\|, x,y\in\mathbb{X}.\tag{6}$$

Assumption 4. *For $t\in\mathbb{N}_+$, the time-varying directed graph $\mathcal{G}_t$ satisfies that*
(i) There exists a constant $w\in(0,1)$ such that $[W_t]_{ij}\ge w$ if $(j,i)\in\mathcal{E}_t$ or $i=j$, and $[W_t]_{ij}=0$ otherwise.
(ii) The mixing matrix W_t is doubly stochastic, i.e., $\sum_{i=1}^{n}[W_t]_{ij}=\sum_{j=1}^{n}[W_t]_{ij}=1$, $\forall i,j\in[n]$.
(iii) There exists an integer $B>0$ such that the time-varying directed graph $(\mathcal{V},\cup_{l=0}^{B-1}\mathcal{E}_{t+l})$ is strongly connected.

Assumption 5. *The compressor $\mathcal{C}:\mathbb{R}^p\to\mathbb{R}^p$ satisfies*

$$\mathbf{E}_{\mathcal{C}}[\|\mathcal{C}(x)-x\|_d^2]\le C,\forall x\in\mathbb{R}^p,\tag{7}$$

for some real norm parameter $d\ge1$ and constant $C\ge0$. Here $\mathbf{E}_{\mathcal{C}}$ denotes the expectation over the internal randomness of the stochastic compression operator $\mathcal{C}$.

Assumption 6. *There exists a point $x_s\in\mathbb{X}$ and a positive constant ς_s such that*

$$g_t(x_s)\le-\varsigma_s\mathbf{1}_m, t\in\mathbb{N}_+.\tag{8}$$

III. DISTRIBUTED ONLINE PRIMAL–DUAL ALGORITHM WITH COMPRESSED COMMUNICATION

A. Algorithm Description

To achieve reduced network cumulative constraint violation, the authors of [24] propose a distributed online primal–dual algorithm for distributed online convex optimization with time-varying constraints. Here, we first give a subgradient descent variant of this algorithm in the following:

$$x_{i,t}=\sum_{j=1}^{n}[W_t]_{ij}z_{j,t},\tag{9a}$$

$$v_{i,t+1}=\gamma_t[g_{i,t}(x_{i,t})]_+,\tag{9b}$$

$$z_{i,t+1}=\mathcal{P}_{\mathbb{X}}(x_{i,t}-\alpha_t\omega_{i,t+1}),\tag{9c}$$

$$\omega_{i,t+1}=\nabla f_{i,t}(x_{i,t})+\big(\nabla g_{i,t}(x_{i,t})\big)^T v_{i,t+1},\tag{9d}$$

where γ_t is the regularization parameter, and α_t is the step-size. To improve communication efficiency of the algorithm (9), the distributed online primal–dual algorithm with compressed communication is proposed by using the class of compressors satisfying Assumption 5, which is presented in pseudo-code as Algorithm 1.

Authorized licensed use limited to: KTH Royal Institute of Technology. Downloaded on April 22,2026 at 14:40:24 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 Distributed Online Primal–Dual Algorithm with Compressed Communication

```
Input: non-increasing stepsize sequence $\{\alpha_t\} \subseteq (0, +\infty)$,
    non-increasing scaling parameter sequence $\{s_t\} \subseteq (0, +\infty)$, and
    non-decreasing regularization parameter sequence $\{\gamma_t\} \subseteq (0, +\infty)$.
Initialize: $\hat{z}_{j,0} = \mathbf{0}_p$ for $j \in [n]$, and $z_{i,1} \in \mathbb{X}$, $\forall i \in [n]$.
  for $t = 1, \cdots$ do
    for $i = 1, \cdots, n$ in parallel do
      Broadcast $\mathcal{C}\big((z_{i,t} - \hat{z}_{i,t-1})/s_t\big)$ to $\mathcal{N}_i^{\text{out}}(\mathcal{G}_t)$ and
      receive $\mathcal{C}\big((z_{j,t} - \hat{z}_{j,t-1})/s_t\big)$ from $j \in \mathcal{N}_i^{\text{in}}(\mathcal{G}_t)$.
      Update
        $\hat{z}_{j,t} = \mathcal{P}_{\mathbb{X}}\big(\hat{z}_{j,t-1} + s_t\mathcal{C}((z_{j,t} - \hat{z}_{j,t-1})/s_t)\big)$,   (10)
          where $j \in \big\{\mathcal{N}_i^{\text{in}}(\mathcal{G}_t) \cup \{i\}\big\}$.
      Select
        $x_{i,t} = \sum_{j=1}^{n} [W_t]_{ij} \hat{z}_{j,t}$.   (11)
      Observe $\nabla f_{i,t}(x_{i,t})$, $\nabla g_{i,t}(x_{i,t})$, and $g_{i,t}(x_{i,t})$.
      Update
        $v_{i,t+1} = \gamma_t [g_{i,t}(x_{i,t})]_+$,   (12a)
        $\omega_{i,t+1} = \nabla f_{i,t}(x_{i,t}) + \big(\nabla g_{i,t}(x_{i,t})\big)^T v_{i,t+1}$,   (12b)
        $z_{i,t+1} = \mathcal{P}_{\mathbb{X}}(x_{i,t} - \alpha_t \omega_{i,t+1})$.   (12c)
    end for
  end for
Output: $\{x_{i,t}\}$.
```

B. Performance Analysis

In this section, we establish network regret and cumulative constraint violation bounds for Algorithm 1 in the following lemma and theorems without and with Slater's condition, respectively.

Lemma 1. *Suppose Assumptions 1–5 hold. For all $i \in [n]$, let $\{x_{i,t}\}$ be the sequences generated by Algorithm 1 with $\gamma_t = \gamma_0/\alpha_t$, where $\gamma_0 \in (0, 1/(4G_2^2)]$ is a constant. Then, for any $T \in \mathbb{N}_+$, and any $y \in \mathcal{X}_T$, it holds that*

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{E}_{\mathcal{C}}[\langle \nabla f_t(x_{i,t}), x_{i,t} - y\rangle] \\
&\le \hat{\vartheta} + \vartheta_2 \sum_{t=1}^{T} \alpha_t + \tilde{p}\sqrt{C}\tilde{\vartheta}_1 \sum_{t=1}^{T} s_t + 2\tilde{p}\sqrt{C}R(\mathbb{X}) \sum_{t=1}^{T} \frac{s_t}{\alpha_t} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{E}_{\mathcal{C}}[\Delta_{i,t}(y)], \qquad (13a)
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{E}_{\mathcal{C}}[\|[g_t(x_{i,t})]_+\|] \\
&\le \sqrt{\vartheta_3 T + \vartheta_4 T \tilde{\Lambda}_T(y) + 4n\tilde{p}^2 C G_2^2 \tilde{\vartheta}_4 T \sum_{t=1}^{T} s_t^2}, \qquad (13b)
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{E}_{\mathcal{C}}[\|[g_t(x_{i,t})]_+\|] \\
&\le nG_2\vartheta_1 + \vartheta_5 \sum_{t=1}^{T} \alpha_t + \vartheta_6 \sum_{i=1}^{n}\sum_{t=1}^{T}\mathbf{E}_{\mathcal{C}}[\|[g_{i,t}(x_{i,t})]_+\|] \\
&\quad + n\tilde{p}\sqrt{C}G_2\tilde{\vartheta}_2 \sum_{t=1}^{T} s_t, \qquad (13c)
\end{aligned}$$

where

$$\vartheta_1 = \frac{2\tau}{\lambda(1-\lambda)} \sum_{i=1}^{n} \|\hat{z}_{i,1}\|, \hat{\vartheta} = 2R(\mathbb{X})L\vartheta_1 + G_1\vartheta_1,$$

$$\tilde{\vartheta}_2 = \frac{4 - 4\lambda + 2n\tau}{1-\lambda}, \tilde{\vartheta}_1 = 2R(\mathbb{X})L\tilde{\vartheta}_2 + G_1\tilde{\vartheta}_2,$$

$$\vartheta_2 = 2G_1^2 + \tilde{\vartheta}_1^2, \tilde{\vartheta}_3 = \frac{16n\tau^2}{\lambda^2(1-\lambda^2)}\Big(\sum_{i=1}^{n} \|\hat{z}_{i,1}\|\Big)^2,$$

$$\vartheta_3 = 2G_2^2\tilde{\vartheta}_3, \tilde{\vartheta}_4 = \frac{16n^2\tau^2}{(1-\lambda)^2} + 32, \vartheta_4 = \frac{4\max\{1, 2G_2^2\tilde{\vartheta}_4\}}{\min\{1, \frac{1}{2\gamma_0}\}},$$

$$\vartheta_5 = nG_1G_2\tilde{\vartheta}_2, \vartheta_6 = 1 + G_2^2\gamma_0\tilde{\vartheta}_2.$$

Lemma 1 provides network regret and cumulative constraint violation bounds for Algorithm 1 under arbitrary parameter sequences. To obtain sublinear network regret and cumulative constraint violation bounds, we specifically design the parameter sequences for Algorithm 1.

Firstly, we choose the scaling parameter sequence $\{s_t\}$ produced by $\{1/t^{\theta_2}\}$ in the following theorem.

Theorem 1. *Suppose Assumptions 1–5 hold. For all $i \in [n]$, let $\{x_{i,t}\}$ be the sequences generated by Algorithm 1 with*

$$\alpha_t = \frac{\alpha_0}{t^{\theta_1}}, \gamma_t = \frac{\gamma_0}{\alpha_t}, s_t = \frac{s_0}{t^{\theta_2}}, \qquad (14)$$

where $\theta_1 \in (0, 1)$, $\alpha_0 > 0$, $\gamma_0 \in (0, 1/(4G_2^2)]$, $s_0 > 0$, and $\theta_2 > \theta_1$ are constants. Then, for any $T \in \mathbb{N}_+$,

$$\begin{aligned}
&\mathbf{E}_{\mathcal{C}}[\text{Net-Reg}(T)] \\
&= \begin{cases} \mathcal{O}(T^{\max\{1-\theta_1, 1+\theta_1-\theta_2\}}), \text{if } \theta_1 < \theta_2 < 1, \\ \mathcal{O}(T^{\max\{1-\theta_1, \theta_1\}}), \text{if } \theta_2 \ge 1, \end{cases} \qquad (15) \\
&\mathbf{E}_{\mathcal{C}}[\text{Net-CCV}(T)] = \mathcal{O}(T^{1-\theta_1/2}). \qquad (16)
\end{aligned}$$

Moreover, if Assumption 6 also holds, then

$$\mathbf{E}_{\mathcal{C}}[\text{Net-CCV}(T)] = \mathcal{O}(T^{1-\theta_1}). \qquad (17)$$

Remark 1. *We show in Theorem 1 that Algorithm 1 establishes sublinear network regret and cumulative constraint violation bounds as in* (15)–(16). *These bounds characterize the impact of compressed communication on the network regret and cumulative constraint violation bounds, which is captured by θ_2. When $\theta_2 \ge 1$, they are the same as the state-of-the-art results established by the distributed online algorithms without compressed communication in [20], [24]. When $\theta_1 < \theta_2 < 1$, compressed communication may enable the network regret bound to become larger due to $1 - \theta_2 > 0$. In addition, when Slater's condition holds, the network cumulative constraint violation bound is further reduced as in* (17). *The bound remains the same as the results*

Authorized licensed use limited to: KTH Royal Institute of Technology. Downloaded on April 22,2026 at 14:40:24 UTC from IEEE Xplore. Restrictions apply.

established by the distributed online algorithm with perfect communication in [24].

We then choose the scaling parameter sequence $\{s_t\}$ produced by $\{\mu^t\}$, which is also adopted by the distributed algorithms in [14], [28].

Theorem 2. *Suppose Assumptions 1–5 hold. For all $i \in [n]$, let $\{x_{i,t}\}$ be the sequences generated by Algorithm 1 with*

$$\alpha_t = \alpha_0\sqrt{\frac{\Psi_t}{t}}, \gamma_t = \frac{\gamma_0}{\alpha_t}, s_t = s_0\mu^t, \tag{18}$$

where $\Psi_t = \sum_{k=1}^{t}\mu^k$, $\alpha_0 > 0$, $\gamma_0 \in (0, 1/(4G_2^2)]$, $s_0 > 0$, and $\mu \in (0,1)$ are constants. Then, for any $T \in \mathbb{N}_+$,

$$\mathbf{E}_{\mathcal{C}}[\text{Net-Reg}(T)] = \mathcal{O}(\sqrt{T}), \tag{19}$$

$$\mathbf{E}_{\mathcal{C}}[\text{Net-CCV}(T)] = \mathcal{O}(T^{3/4}). \tag{20}$$

Moreover, if Assumption 6 also holds, then

$$\mathbf{E}_{\mathcal{C}}[\text{Net-CCV}(T)] = \mathcal{O}(\sqrt{T}). \tag{21}$$

Remark 2. *We show in Theorem 2 that Algorithm 1 establishes an $\mathcal{O}(\sqrt{T})$ network regret bound as in* (19) *and an $\mathcal{O}(T^{3/4})$ cumulative constraint violation bound as in* (20)*. These bounds are the same as the results established in* (15)–(16) *with $\theta_1 = 1/2$ and $\theta_2 \geq 1$. Moreover, the network regret bound is the same as the results established in [25] where compressed communication and inequality constraints are not considered, and the results established in [12], [13] where inequality constraints and nonconvex local loss functions are not considered. In addition, when Slater's condition holds, the network cumulative constraint violation bound is further reduced as in* (21)*, which is the same as the results established in* (17) *with $\theta_1 = 1/2$.*

IV. SIMULATION EXAMPLE

To evaluate the performance of Algorithm 1, we consider a distributed online localization problem with long-term constraints over a network of 100 sensors as follows:

$$\min_{x} \quad \sum_{t=1}^{T}\sum_{i=1}^{n}\frac{1}{4}\big|\|S_i - x\|^2 - D_{i,t}\big|^2, \tag{22a}$$

$$\text{s.t.} \quad x \in \mathbb{X}, B_{i,t}x - b_{i,t} \leq \mathbf{0}_{m_i}, \forall i \in [n], \forall t \in [T], \tag{22b}$$

where $S_i \in \mathbb{R}^p$ denotes the position of sensor i, $D_{i,t} = \|S_i - X_{0,t}\|^2 + \tau_{i,t}$ denotes the distance between the positions of the target and sensor i, $X_{0,t} \in \mathbb{R}^p$ and $\tau_{i,t} \in \mathbb{R}$ denote the position of the target and the measurement noise at iteration t, respectively, and $B_{i,t} \in \mathbb{R}^{m_i \times p}$ and $b_{i,t} \in \mathbb{R}^{m_i}$ denote thecoefficientmatrixandcoefficientvectorofthelocallinearconstraints,respectively. The communication topology is modeled by a time-varying undirected graph. Specifically, at each iteration t, the graph is first randomly generated where the probability of any two sensors being connected is ρ. Then, to make sure that Assumption 4 is satisfied, we add edges $(i, i+1)$ for $i \in [24]$ when $t \in \{4c+1\}$, edges $(i, i+1)$ for $i \in [25, 49]$ when $t \in \{4c+2\}$, edges $(i, i+1)$ for $i \in [50, 74]$ when $t \in \{4c+3\}$, edges $(i, i+1)$ for $i \in [75, 99]$ when $t \in \{4c+4\}$ for $c = \{0, 1, \cdots\}$. Moreover, let $[W_t]_{ij} = \frac{1}{n}$ if $(j, i) \in \mathcal{E}_t$ and $[W_t]_{ii} = 1 - \sum_{j=1}^{n}[W_t]_{ij}$.

In this paper, we show in Theorem 1 that, both without and with Slater's condition, Algorithm 1 establishes the same network regret and cumulative constraint violation bounds as the state-of-the-art results on distributed online convex optimization with long-term constraints, established by the distributed online algorithms with perfect communication in [24]. To verify the theoretical results, we compare Algorithm 1 with the algorithm in [24]. We set $\rho = 0.1$, $\mathbb{X} = [-5, 5]^p$, $p = 2$, $m_i = 2$, and randomly choose each component of S_i from the uniform distribution in the interval $[-10, 10]$. We assume that the position of the target evolves by

$$X_{0,t+1} = X_{0,t} + \begin{bmatrix} \frac{(-1)^{Q_t}\sin(t/50)}{10t} \\ \frac{-Q_t\cos(t/70)}{40t} \end{bmatrix},$$

where Q_t is randomly generated from Bernoulli distribution with a success probability of 0.5, and $X_{0,0} = [0.8, 0.95]^T$. Moreover, $\tau_{i,t}$ is randomly generated from the uniform distribution from in the interval $[0, 0.001]$. Furthermore, each component of $B_{i,t}$ is randomly generated from the uniform distribution in the interval $[0, 2]$, and each component of $b_{i,t}$ is randomly generated from the uniform distribution in the interval $[b, b+1]$ with $b > 0$. Note that $b > 0$ guarantees Slater's condition holds. Here we choose $b = 0.01$. In addition, we select the following compressor for Algorithm 1:

$$\mathcal{C}(x) = \Delta\Big\lfloor\frac{x}{\Delta} + \frac{\mathbf{1}_p}{2}\Big\rfloor,$$

where Δ is a positive integer. This compressor satisfies Assumption 5 with $d = \infty$ and $C = \Delta^2/4$, which is also used in [14], [28]–[30]. Transmitting $\mathcal{C}(x)$ requires pq bits if each integer is encoded using q bits. Here we set $\Delta = 1$ and $q = 8$.

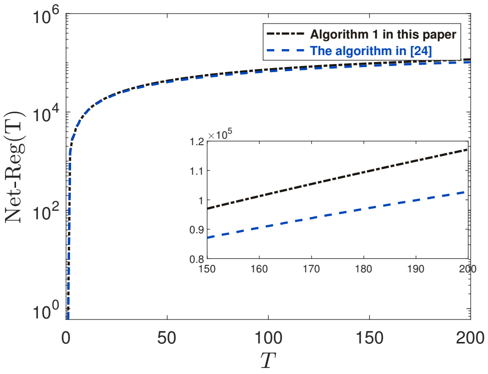

Fig. 1: Evolutions of network regret.

Figs. 1 and 2 illustrate the evolutions of network regret and cumulative constraint violation, respectively. As shown in Fig.1, our Algorithm 1 exhibits almost the same network regret as that of the algorithm in [24]. Similarly, Fig. 2

Authorized licensed use limited to: KTH Royal Institute of Technology. Downloaded on April 22,2026 at 14:40:24 UTC from IEEE Xplore. Restrictions apply.

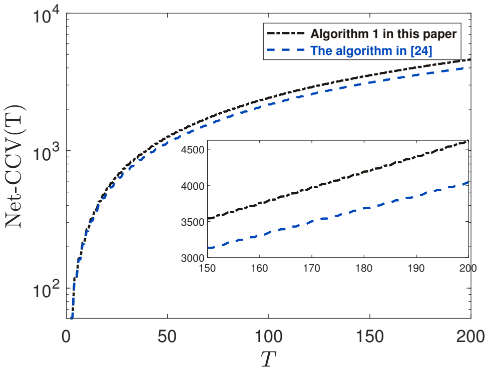

Fig. 2: Evolutions of network cumulative constraint violation.

demonstrates that our Algorithm 1 also has almost the same network cumulative constraint violation as that of the algorithm in [24]. However, due to compressed communication, our Algorithm 1 requires significantly fewer bits than those required by the algorithm in [24]. These simulation results are consistent with the results in Theorem 1.

V. CONCLUSIONS

This paper studied the distributed online nonconvex optimization problem with time-varying constraints. To better utilize communication resources, we proposed a distributed online primal–dual algorithm with compressed communication. More importantly, the algorithm was able to handle time-varying communication topologies. We showed that the algorithm established sublinear network regret and cumulative constraint violation bounds. Moreover, the network cumulative constraint violation bounds were further reduced when Slater's condition held. In the future, we plan to investigate distributed bandit nonconvex optimization with time-varying constraints since gradient information is unavailable in many real-world applications.

REFERENCES

[1] E. Hazan, "Introduction to online convex optimization," *Foundations and Trends in Optimization*, vol. 2, no. 3-4, pp. 157–325, 2016.

[2] S. Sundhar Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, pp. 516–545, 2010.

[3] F. Yan, S. Sundaram, S. Vishwanathan, and Y. Qi, "Distributed autonomous online learning: Regrets and intrinsic privacy-preserving properties," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 11, pp. 2483–2493, 2012.

[4] K. I. Tsianos and M. G. Rabbat, "Distributed strongly convex optimization," in *Annual Allerton Conference on Communication, Control, and Computing*, 2012, pp. 593–600.

[5] S. Hosseini, A. Chapman, and M. Mesbahi, "Online distributed convex optimization on dynamic networks," *IEEE Transactions on Automatic Control*, vol. 61, no. 11, pp. 3545–3550, 2016.

[6] D. Mateos-Núnez and J. Cortés, "Distributed online convex optimization over jointly connected digraphs," *IEEE Transactions on Network Science and Engineering*, vol. 1, no. 1, pp. 23–37, 2014.

[7] A. Koppel, F. Y. Jakubiec, and A. Ribeiro, "A saddle point algorithm for networked online convex optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5149–5164, 2015.

[8] S. Shahrampour and A. Jadbabaie, "Distributed online optimization in dynamic environments using mirror descent," *IEEE Transactions on Automatic Control*, vol. 63, no. 3, pp. 714–725, 2017.

[9] D. Yuan, Y. Hong, D. W. C. Ho, and S. Xu, "Distributed mirror descent for online composite optimization," *IEEE Transactions on Automatic Control*, vol. 66, no. 2, pp. 714–729, 2021.

[10] X. Li, X. Yi, and L. Xie, "Distributed online convex optimization with an aggregative variable," *IEEE Transactions on Control of Network Systems*, vol. 9, no. 1, pp. 438–449, 2022.

[11] X. Li, L. Xie, and N. Li, "A survey on distributed online optimization and online games," *Annual Reviews in Control*, vol. 56, p. 100904, 2023.

[12] X. Cao and T. Başar, "Decentralized online convex optimization with compressed communications," *Automatica*, vol. 156, p. 111186, 2023.

[13] Z. Tu, X. Wang, Y. Hong, L. Wang, D. Yuan, and G. Shi, "Distributed online convex optimization with compressed communication," in *Advances in Neural Information Processing Systems*, 2022, pp. 34 492–34 504.

[14] X. Ge, H. Zhang, W. Xu, and H. Bao, "Distributed online bandit optimization with communication compression," in *International Conference on Information Science and Technology*, 2023, pp. 678–686.

[15] X. Cao, T. Başar, S. Diggavi, Y. C. Eldar, K. B. Letaief, H. V. Poor, and J. Zhang, "Communication-efficient distributed learning: An overview," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 4, pp. 851–873, 2023.

[16] D. Yuan, D. W. Ho, and G. Jiang, "An adaptive primal–dual subgradient algorithm for online distributed constrained optimization," *IEEE Transactions on Cybernetics*, vol. 48, no. 11, pp. 3045–3055, 2017.

[17] M. Mahdavi, R. Jin, and T. Yang, "Trading regret for efficiency: Online convex optimization with long term constraints," *Journal of Machine Learning Research*, vol. 13, no. 1, pp. 2503–2528, 2012.

[18] D. Yuan, A. Proutiere, and G. Shi, "Distributed online linear regressions," *IEEE Transactions on Information Theory*, vol. 67, no. 1, pp. 616–639, 2021.

[19] J. Yuan and A. Lamperski, "Online convex optimization for cumulative constraints," in *Advances in Neural Information Processing Systems*, 2018, pp. 6140–6149.

[20] X. Yi, X. Li, T. Yang, L. Xie, T. Chai, and K. H. Johansson, "Regret and cumulative constraint violation analysis for distributed online constrained convex optimization," *IEEE Transactions on Automatic Control*, vol. 68, no. 5, pp. 2875–2890, 2023.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[22] H. Yu, M. Neely, and X. Wei, "Online convex optimization with stochastic constraints," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

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

[24] X. Yi, X. Li, T. Yang, L. Xie, Y. Hong, T. Chai, and K. H. Johansson, "Distributed online convex optimization with time-varying constraints: Tighter cumulative constraint violation bounds under slater's condition," *IEEE Transactions on Automatic Control*, 2025, DOI 10.1109/TAC.2025.3547606.

[25] K. Lu and L. Wang, "Online distributed optimization with nonconvex objective functions: Sublinearity of first-order optimality condition-based regret," *IEEE Transactions on Automatic Control*, vol. 67, no. 6, pp. 3029–3035, 2021.

[26] K. Zhang, X. Lei, X. Yi, M. Cao, K. H. Johansson, T. Chai, and T. Yang, "Distributed constrained online nonconvex optimization with compressed communication," *arXiv:2503.22410v2*, 2025.

[27] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2012.

[28] X. Yi, S. Zhang, T. Yang, T. Chai, and K. H. Johansson, "Communication compression for distributed nonconvex optimization," *IEEE Transactions on Automatic Control*, vol. 68, no. 9, pp. 5477–5492, 2022.

[29] H. Li, S. De, Z. Xu, C. Studer, H. Samet, and T. Goldstein, "Training quantized nets: A deeper understanding," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[30] S. Khirirat, S. Magnússon, and M. Johansson, "Compressed gradient methods with hessian-aided error compensation," *IEEE Transactions on Signal Processing*, vol. 69, pp. 998–1011, 2021.

Authorized licensed use limited to: KTH Royal Institute of Technology. Downloaded on April 22,2026 at 14:40:24 UTC from IEEE Xplore. Restrictions apply.