分布式在线 Frank-Wolfe 凸优化的动态遗憾*

Wentao Zhang[a], Yang Shi[b], Baoyong Zhang [a,*], Deming Yuan[a]

[a]School of Automation, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, P. R. China
[b]Department of Mechanical Engineering, University of Victoria, Victoria, BC, Canada, V8W 3P6

arXiv:2302.00663v1 [math.OC] 1 Feb 2023

摘要

本文研究分布式在线凸约束优化问题，其中，多智能体系统中的各个智能体通过时变网络与邻居通信，相互协作以最小化全局代价函数。当优化问题的约束集维度高且结构复杂时，投影操作的计算成本往往高得难以承受。为解决这一问题，我们提出一种结合梯度跟踪技术的分布式在线 Frank-Wolfe 优化算法。我们严格证明了所提出优化算法的动态遗憾界为 $\mathcal{O}(\sqrt{T(1+H_T)}+D_T)$，该界显式依赖于迭代轮数 T、函数变化量H_T和梯度变化量 D_T。最后，我们通过分布式在线岭回归问题验证并比较了理论结果。

关键词：分布式在线凸优化；Frank-Wolfe 算法；动态遗憾；梯度跟踪方法。

1. 引言

近年来，分布式优化在多智能体网络中受到广泛关注，这得益于其在电力系统、传感器网络、机器学习等领域的广泛应用（例如，Nedi´c& Liu (2018),Yang 等 (2019), Li 等 (2022), Liu 等 (2020),Li 等 (2021), Xu &Soh (2016), Yuan 等 (2022),Xu 等 (2022)）。在这类优化问题中，分布式在线优化可描述为如下重复博弈：

1. 在每一轮 t，智能体 i 首先生成一个决策$\boldsymbol{x}_{i,t}\in\boldsymbol{X}$。2. 智能体 i 遭受损失 $f_{i,t}(\boldsymbol{x}_{i,t})$，对手揭示有关损失函数 $f_{i,t}$的信息。3. 随后，智能体 i 利用有关损失函数$f_{i,t}$ 的信息构造下一个决策 $\boldsymbol{x}_{i,t+1}$。

网络的主要任务是通过所有智能体之间的信息交换和局部计算，最小化问题（1）中局部目标函数之和，其中每个智能体在每一轮仅了解自身及其邻居的信息。

$$\min_{\boldsymbol{x}\in\boldsymbol{X}}\sum_{t=1}^{T}F_t(\boldsymbol{x}) \tag{1}$$

其中，$F_t(\boldsymbol{x})=\sum_{i=1}^{n}f_{i,t}(\boldsymbol{x})$，$\boldsymbol{X}$ 是 $\mathbb{R}^d$中的凸紧集，函数$f_{i,t}$ 关于 $\boldsymbol{X}$是凸的。为衡量所提出的算法的性能，定义动态遗憾 $\mathbf{Regret}_d^j(T)$，其表示在时间 T 上的总和，即智能体 j 的累积代价 $F_t(\boldsymbol{x}_{j,t})$与最优解处的累积代价 $\boldsymbol{x}_t^*$之差。

$$\mathbf{Regret}_d^j(T)=\sum_{t=1}^{T}F_t(\boldsymbol{x}_{j,t})-\sum_{t=1}^{T}F_t(\boldsymbol{x}_t^*) \tag{2}$$

其中 $\boldsymbol{x}_t^*\in\arg\min_{\boldsymbol{x}\in\boldsymbol{X}}F_t(\boldsymbol{x})$。

在这类带约束集的分布式在线（离线）优化问题中，通常使用投影操作作为处理约束的基本技术，例如 Sundhar Ram 等人 (2010年提出的分布式在线梯度下降。一般而言，投影步骤等价于求解一个凸二次问题 Hazan& Kale (2012年)。然而，在高维且复杂的约束优化问题中，例如Zhang 等人 (2017b中的多类分类、Wu &Teo (1983年)中的最优控制、Hazan& Kale (2012年)；Wai 等人 (2017年)中的矩阵补全、Zhang 等人 (2017a)中的电动汽车充电，以及Hazan(2008年)中的半正定规划，投影操作会带来沉重的计算负担。相比之下，Frank-Wolfe（FW）方法通过求解线性最小化预言机，避免了这类计算成本高昂的操作。由于计算成本低，近年来 FW 方法已广泛用于分布式在线（离线）优化。在Zhang 等人 (2017b)中，作者首次通过扩展

*本文未在任何会议上发表。

*通讯作者。*Email addresses:* iswt.zhang@gmail.com (Wentao Zhang),yshi@uvic.ca (Yang Shi), baoyongzhang@njust.edu.cn (Baoyong Zhang), dmyuan1012@gmail.com (Deming Yuan)

表 1：在线 Frank-Wolfe 凸优化相关工作的比较。

参考文献	损失函数	分布式	动态遗憾	动态网络	线性预言机	遗憾界
Zhang 等（2017b）	凸且连续	✓	✗	✗	$\mathcal{O}(T)$	$\mathcal{O}(T^{3/4})$
Wan 等人 (2020)	凸且连续	✓	✗	✗	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(T^{3/4})$
Wan 等人(2021a)	强凸且连续	✓	✗	✗	$\mathcal{O}(T^{1/3})$	$\mathcal{O}(T^{2/3}$ 对数 $T)$
Thang 等人(2022)	凸且光滑	✓	✗	✗	$\mathcal{O}(T^{3/2})$	$\mathcal{O}(\sqrt{T})$
Kalhan 等人 (2021)	凸且光滑	✗	✓	/	$\mathcal{O}(T)$	$\mathcal{O}\left(\sqrt{T}\left(1+H_T+\sqrt{D_T}\right)\right)$
					$\mathcal{O}(T^{3/2})$	$\mathcal{O}\left(1+H_T+\sqrt{T}\right)$
Wan 等人(2021b)	凸且连续	✗	✓	/	$\mathcal{O}(T$ 对数$_2$ $T)$	$\mathcal{O}\left(\text{最大值}\left\{\sqrt{T}, T^{2/3}H_T^{1/3}\right\}\right)$
	强凸且连续	✗	✓	/	$\mathcal{O}(T$ 对数$_2$ $T)$	$\mathcal{O}\left(\text{最大值}\left\{\sqrt{TH_T \log T},\ \log T\right\}\right)$
本文工作	凸且光滑	✓	✓	✓	$\mathcal{O}(T)$	$\mathcal{O}\left(\sqrt{T(1+H_T)}+D_T\right)$

Hazan & Kale（2012）提出的在线集中式FW算法，给出了一种在线分布式FW算法，并获得了静态遗憾界$\mathcal{O}(T^{3/4})$。Wan 等人（2020）考虑了在全信息反馈和多臂老虎机反馈下的一种改进变体。基于将时间 T 划分为 $\sqrt{T}$ 个大小相等的数据块这一思路，智能体之间的通信频率得以降低，并分别建立了相关的遗憾上界 $\mathcal{O}(T^{3/4})$和 $\tilde{\mathcal{O}}(T^{3/4})$。Wan 等人（2021a）在 Wan 等人（2020）的基础上，进一步利用了强凸性条件下改进的收敛结果。Thang 等人（2022）分析了光滑损失函数下精确梯度和随机梯度的两个算法版本，并给出了遗憾上界 $\mathcal{O}(\sqrt{T})$。然而，这两种算法中每个智能体额外的步长循环都会拖慢时间 t的计算。

迄今为止，关于在线 Frank-Wolfe 算法动态遗憾的研究工作很少，尤其是在分布式场景下。动态遗憾由于采用动态而非固定基准，因此比静态遗憾更严格、更有效，是一种性能指标；其动态遗憾界通常与优化问题的正则性有关。根据 Besbes 等人 （2015），众所周知，除非变化预算满足次线性条件，否则动态遗憾无法实现次线性收敛，其中变化预算为 T。基于此，与 函数变化量 H_T 和梯度变化量D_T 相关的 遗憾$_d^j$ （T）界定义为

$$H_T = \sum_{t=1}^{T} \max_{i\in\mathcal{V}} \max_{\boldsymbol{x}\in\boldsymbol{X}} |f_{i,t+1}(\boldsymbol{x}) - f_{i,t}(\boldsymbol{x})|, \tag{3}$$

$$D_T = \sum_{t=1}^{T} \max_{i\in\mathcal{V}} \max_{\boldsymbol{x}\in\boldsymbol{X}} \|\nabla f_{i,t+1}(\boldsymbol{x}) - \nabla f_{i,t}(\boldsymbol{x})\|. \tag{4}$$

在 Kalhan 等人 （2021）中，作者分析了几种集中式在线 Frank-Wolfe 算法的动态遗憾界，其中 H_T 在适用范围方面存在局限性。本文旨在在相同条件下进一步扩大该范围，因此所提出的算法具有更强的适用性。 Wan 等人（2021b）采用重启策略研究了一种新颖的集中式在线 Frank-Wolfe 算法。具体而言，在线 Frank-Wolfe 凸优化相关工作的比较总结见表 1。

不难发现，Kalhan等人(2021）和Wan等人 (2021b）仅分析了集中式在线优化的动态遗憾。众所周知，由于单机的计算瓶颈，集中式在线算法难以处理大规模优化问题。根据表 1 中的比较和分析，有必要研究分布式场景下在线 FW 算法的动态遗憾。因此，本文设计了分布式在线 Frank-Wolfe 凸优化算法（DOFW-CO），以弥补这一空白。本文的贡献如下。

首先，我们提出了一种分布式在线 Frank-Wolfe 凸优化算法，可高效处理高维问题和复杂约束集，从而减轻投影算子带来的高计算负担。同时，DOFW-CO 算法采用梯度跟踪技术，利用历史信息更新损失函数梯度的变化。此外，不同于 Zhang 等人(2017b)、Wan 等人(2021a)、Wan 等人(2020)、Thang 等人(2022中的通信拓扑，本文提出了一种适用于时变网络拓扑的分布式优化算法，相较于静态网络更具实用性和通用性。

其次，受 Kalhan 等人(2021的启发，我们为所提出的算法建立了动态遗憾界 $\mathcal{O}(\sqrt{T(1+H_T)}+D_T)$，在相同的线性预言机复杂度下，该界既能恢复集中式结果，又进一步拓宽了 H_T 相较于 Kalhan 等人(2021的适用范围。本文首次对分布式场景下 FW 方法的动态遗憾进行了分析。最后，我们通过模拟分布式在线岭回归问题，验证了所提出算法的性能。

记号： $\mathbb{R}^n$ 表示 n维欧几里得空间。 $\mathbb{Z}$ ($\mathbb{Z}_+$)表示（正）整数集。向量 $\boldsymbol{z}$ 的欧几里得范数记为 $\|\boldsymbol{z}\|$。 $[A_t]_{ij}$ 表示矩阵 A_t 第 i行第 j列的元素， $[\boldsymbol{w}]_i$ 表示向量 $\boldsymbol{w}$的第 i 个元素。

2. 问题表述

2.1. 优化问题

令 $\mathcal{G}_t = \{\mathcal{V}, \mathcal{E}_t, A_t\}$ 表示一个有向时变网络，其智能体集合为 $\mathcal{V}:=\{1,\ldots,n\}$，边集为

$\mathcal{E}_t \subseteq \mathcal{V}\times\mathcal{V}$，加权邻接矩阵为 $A_t \in \mathbb{R}^{n\times n}$。在该网络中，智能体 i 有权与智能体 i 的内部邻居集合 $\mathcal{N}_i^{\text{in}}(t) = \{j \mid (j, i) \in \mathcal{E}_t\} \cup \{i\}$ 中的智能体进行通信。此外，当 $j \in \mathcal{N}_i^{\text{in}}(t)$ 时，$[A_t]_{ij} > 0$ 成立；否则，$[A_t]_{ij} = 0$ 成立。

本文旨在为问题（1）设计一种分布式在线算法，以确保每个智能体的动态遗憾 $j \in \mathcal{V}$增长为次线性，即 $\lim_{T\to\infty}(\text{遗憾}_d^j(T)/T) = 0$，$\forall j \in \mathcal{V}$。对于网络 $\mathcal{G}_t$，问题1中的约束集和损失函数满足以下假设。

假设 1. (a) 存在一个正标量 ζ，使得当 $[A_t]_{ij} > 0$ 时，$[A_t]_{ij} > \zeta, t \in \{1,\dots,T\}$。

(b) 对任意 $t \in \{1,\dots,T\}$ 和所有 $i, j \in \mathcal{V}$，A_t 满足 $\sum_{j=1}^n [A_t]_{ij} = \sum_{i=1}^n [A_t]_{ij} = 1$。

(c) 对于某个 $Q \in \mathbb{Z}_+$，图的并集 $\bigcup_{i=kQ+1}^{(k+1)Q} \mathcal{G}_i$ 对每个整数 $k \geq 0$ 都是强连通的。

假设2。约束集 $\boldsymbol{X}$ 的直径有限，为 M，即，对于 $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{X}$，$\max_{\boldsymbol{x}_1,\boldsymbol{x}_2\in\boldsymbol{X}} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \leq M$.

假设3。（Lipschitz函数）函数 $f_{i,t}$ 是L_X-Lipschitz的，即，对于 $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{X}$, $|f_{i,t}(\boldsymbol{x}_1) - f_{i,t}(\boldsymbol{x}_2)| \leq L_X\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$，其中 L_X 是已知的正常数。

假设4。（Lipschitz梯度）梯度$\nabla f_{i,t}(\boldsymbol{x})$ 是 G_X-Lipschitz的，即，$\|\nabla f_{i,t}(\boldsymbol{x}_1) - \nabla f_{i,t}(\boldsymbol{x}_2)\| \leq G_X\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|, \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{X}$.

备注1。在集中式和分布式优化中，假设1、2、3是标准的，类似的设定可见于 Nedi´c 等（2008年）；Yi等（2021年）、Besbes等（2015年）；Wai等（2017年）。根据 Shalev-Shwartz等（2011年）中的引理2.6，假设3蕴含 $\|\nabla f_{i,t}(\boldsymbol{x})\| \leq L_X$。假设4等价于以下事实

$$f_{i,t}(\boldsymbol{x}_1) - f_{i,t}(\boldsymbol{x}_2) \leq \langle \nabla f_{i,t}(\boldsymbol{x}_2), \boldsymbol{x}_1 - \boldsymbol{x}_2\rangle + \frac{G_X}{2}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2, \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{X}. \quad (5)$$

3. 算法设计与收敛性分析

3.1. DOFW-CO 算法

本节首先提出算法 DOFW-CO。算法描述见算法1。具体而言，其关键要素包括：1）采用梯度跟踪和 Frank-Wolfe 方法；2)^在梯度跟踪步骤之后，梯度 $\nabla f_{i,t}$ 取代 Frank-Wolfe 步骤的线性预言机中的传统梯度 $\nabla f_{i,t}(\hat{\boldsymbol{x}}_{i,t})$。

算法1（DOFW-CO）分布式在线 Frank-Wolfe 凸优化

初始化：初始变量 $\boldsymbol{x}_{i,1} \in \boldsymbol{X}$ 和参数 $0 < \alpha \leq 1$.

对于 $t = 1, 2, \cdots, T$ **do**

  对于每个智能体 $i \in \mathcal{V}$ 执行

    智能体 i 接收 $\boldsymbol{x}_{j,t}$ 来自 $j \in \mathcal{N}_i^{\text{in}}(t)$，并更新日期 $\hat{\boldsymbol{x}}_{i,t} = \sum_{j\in\mathcal{N}_i^{in}(t)} [A_t]_{ij}\boldsymbol{x}_{j,t}$.

    梯度值 $\nabla f_{i,t}(\hat{\boldsymbol{x}}_{i,t})$ 已被揭示，并且智能体 i 执行梯度跟踪步骤：

    如果 $t = 1$ 则

      $\widehat{\nabla} f_{i,1} = \nabla f_{i,1}(\hat{\boldsymbol{x}}_{i,1})$,

    else

      $\widehat{\nabla} f_{i,t} = \widehat{\nabla} f_{i,t-1} + \nabla f_{i,t}(\hat{\boldsymbol{x}}_{i,t}) - \nabla f_{i,t-1}(\hat{\boldsymbol{x}}_{i,t-1})$.

    结束如果 $\widehat{\nabla} f_{i,t} = \sum_{j\in\mathcal{N}_i^{in}(t)} [A_t]_{ij}\widehat{\nabla} f_{j,t}$,

    Frank-Wolfe 步骤：更新 $\boldsymbol{v}_{i,t} = \arg\min_{\boldsymbol{x}\in\boldsymbol{X}} \left\langle \boldsymbol{x}, \widehat{\nabla} f_{i,t}\right\rangle$,

    $\boldsymbol{x}_{i,t+1} = \hat{\boldsymbol{x}}_{i,t} + \alpha(\boldsymbol{v}_{i,t} - \hat{\boldsymbol{x}}_{i,t})$。

  结束循环

结束循环

3.2. 主要收敛结果

本节将详细分析算法1 中定义于（2）的动态遗憾上界。为便于证明和分析，我们定义运行平均向量 $\boldsymbol{x}_{avg,t}$ 和 $\boldsymbol{v}_{avg,t}$、时间 t 的最大函数变化量 $f_{t,sup}$、时间 t 的最大梯度变化量 $g_{t,sup}$，以及智能体 i 的梯度差 $\boldsymbol{\delta}_{i,t}$，如下所示：

$$\begin{cases} \boldsymbol{x}_{avg,t} &= \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i,t}, \boldsymbol{v}_{avg,t} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{v}_{i,t} \\ f_{t,sup} &= \max_{i\in\mathcal{V}} \max_{\boldsymbol{x}\in\boldsymbol{X}} |f_{i,t+1}(\boldsymbol{x}) - f_{i,t}(\boldsymbol{x})| \\ g_{t,sup} &= \max_{i\in\mathcal{V}} \max_{\boldsymbol{x}\in\boldsymbol{X}} \|\nabla f_{i,t+1}(\boldsymbol{x}) - \nabla f_{i,t}(\boldsymbol{x})\| \\ \boldsymbol{\delta}_{i,t} &= \nabla f_{i,t}(\hat{\boldsymbol{x}}_{i,t}) - \nabla f_{i,t-1}(\hat{\boldsymbol{x}}_{i,t-1}) \end{cases} \quad (6)$$

定理1. 设决策序列 $\{\boldsymbol{x}_{i,t}\}$ 由算法1 生成，并假设1-4 成立。则对于 $T \geq 2$ 和 $j \in \mathcal{V}$，遗憾有如下界：

$$\textbf{Regret}_d^j(T) \leq C_1 + C_2\alpha T + \frac{2n}{\alpha}H_T + \frac{C_3}{\alpha} + C_4 D_T \quad (7)$$

其中

$$C_1 = nL_X\sum_{i=1}^n \|\boldsymbol{x}_{i,1} - \boldsymbol{x}_{avg,1}\| + \frac{2Mn\Gamma}{1-\sigma}\sum_{i=1}^n \|\nabla f_{i,1}(\hat{\boldsymbol{x}}_{i,1})\| + \left(nL_X + 2MG_X + \frac{4Mn\Gamma G_X}{1-\sigma}\right)\frac{n\Gamma}{1-\sigma}\sum_{i=1}^n \|\boldsymbol{x}_{i,1}\|,$$

$$C_2 = 2n^2L_XM + \left(4MG_X + nL_X + \frac{4Mn\Gamma G_X}{1-\sigma}\right)\frac{n^2M\Gamma}{1-\sigma} + \frac{nG_XM^2}{2}, C_3 = nL_XM, C_4 = \frac{2Mn^2\Gamma}{1-\sigma} + nM,$$

$\sigma = \left(1-\zeta/4n^2\right)^{1/Q}$, $\Gamma = \left(1-\zeta//4n^2\right)^{(1-2Q)/Q}$.

证明：根据假设 3，可得

$$\begin{aligned} & F_t(\boldsymbol{x}_{j,t}) - F_t(\boldsymbol{x}_t^*) \\ & = F_t(\boldsymbol{x}_{j,t}) - F_t(\boldsymbol{x}_{avg,t}) + F_t(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_t^*) \\ & \le nL_X\|\boldsymbol{x}_{j,t} - \boldsymbol{x}_{avg,t}\| + F_t(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_t^*) \\ & \le nL_X\sum_{i=1}^{n}\|\boldsymbol{x}_{i,t} - \boldsymbol{x}_{avg,t}\| + F_t(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_t^*). \end{aligned} \tag{8}$$

根据算法1 以及A_{t-1}的双随机性，我们得到 $\boldsymbol{x}_{avg,t} = \frac{1}{n}\sum_{i=1}^{n}[\hat{}\boldsymbol{x}_{i,t-1} + \alpha(\boldsymbol{v}_{i,t-1} - \hat{}\boldsymbol{x}_{i,t-1})] = \boldsymbol{x}_{avg,t-1} + \alpha(\boldsymbol{v}_{avg,t-1} - \boldsymbol{x}_{avg,t-1})$。因此，考虑假设2，对于任意 $t \ge 2$，有

$$\begin{aligned} \|\boldsymbol{x}_{i,t} - \boldsymbol{x}_{avg,t}\| &= \|\hat{\boldsymbol{x}}_{i,t-1} - \boldsymbol{x}_{avg,t-1} + \alpha(\boldsymbol{v}_{i,t-1} - \hat{\boldsymbol{x}}_{i,t-1}) \\ &\quad - \alpha(\boldsymbol{v}_{avg,t-1} - \boldsymbol{x}_{avg,t-1})\| \\ &\le \|\hat{\boldsymbol{x}}_{i,t-1} - \boldsymbol{x}_{avg,t-1}\| + 2\alpha M. \end{aligned} \tag{9}$$

结合该条件并回顾式（2）中定义的遗憾概念，由式（8）可得

$$\begin{aligned} & \text{遗憾}_d^j(T) \\ & \le nL_X\sum_{i=1}^{n}\|\boldsymbol{x}_{i,1} - \boldsymbol{x}_{avg,1}\| + nL_X\sum_{t=1}^{T-1}\sum_{i=1}^{n}\|\hat{\boldsymbol{x}}_{i,t} - \boldsymbol{x}_{avg,t}\| \\ & \quad + 2\alpha T n^2 L_X M + \sum_{t=1}^{T}\left[F_t(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_t^*)\right]. \end{aligned} \tag{10}$$

结合该不等式以及附录中的引理1、引理3 和引理4，即可得出定理7）中的条件1。证明完毕。□

定理1 给出了动态遗憾的主要结果。容易看出，算法1的遗憾界取决于 α的选择。因此，通过选择合适的步长，可得以下推论。

推论1。 假设定理 1中的条件成立。那么，若 $H_T = o(T)$ $D_T = o(T)$ 成立，取 $\alpha = \gamma\sqrt{\frac{H_T+1}{T}}$，则有

$$\mathbf{Regret}_d^j(T) \le \mathcal{O}\left(\sqrt{T(1+H_T)} + D_T\right) \tag{11}$$

其中， γ 是一个正的调整常数，可确保 $\alpha \le 1$。

证明。 根据（7），可得 遗憾$_d^j(T)$ $\le \mathcal{O}\left(\alpha T + \frac{1}{\alpha}(1+H_T) + D_T\right)$. 由此，取 11），即可容易得到式（ $\alpha T = \frac{1}{\alpha}(1+H_T)$ 。证明完毕。□

备注 2。特别是，推论1 中给出的这一结果与Kalhan 等人(2021的集中式结果一致，且在$\mathcal{O}(\sqrt{T}(1 + H_T + \sqrt{D_T}))$线性预言机复杂度相同的情况下，比其上界更不保守、更紧。此外，当期望得到次线性遗憾界时， H_T 在Kalhan 等人(2021中的取值范围从 $\mathcal{O}(\sqrt{T})$扩展至 $\mathcal{O}(T)$，从而有效拓宽了优化问题的应用领域。

备注 3。请注意，在实际应用中，很难准确获得 H_T 这一先验知识。以下给出两点讨论：

i）若可以提前获知满足 $H_T \le \hat{}H_T = \mathcal{O}(T^\theta)$ $0 < \theta < 1$ 的估计上界，则通过设置步长$\alpha = \gamma\sqrt{(T^\theta+1)/T}$，可得到次线性动态遗憾。

ii）当损失函数 $F_t(x) = F(x)$ 为时不变时，$H_T = 0$, $D_T = 0$ 成立，并得到动态遗憾界$\mathcal{O}(\sqrt{T})$，这与分布式离线场景中的结果相符，例如 Hou 等人 (2022)。

4. 仿真

本节针对岭回归问题进行了若干数值仿真，以验证我们提出的算法。岭回归问题可表述为

$$\min_{\boldsymbol{x}\in\boldsymbol{X}}\sum_{t=1}^{T}\sum_{i=1}^{n}\left[\frac{1}{2}\left(\boldsymbol{a}_{i,t}^\top\boldsymbol{x} - l_{i,t}\right)^2 + \lambda_1\|\boldsymbol{x}\|_2^2\right] \tag{12}$$

其中， $\boldsymbol{X} := \{\boldsymbol{x} |\ \mathbf{1}^T\boldsymbol{x} = 1\}$, $\lambda_1 = 5\times10^{-6}$ 是一个正则化参数，数对 $(\boldsymbol{a}_{i,t}, l_{i,t}) \in \mathbb{R}^d \times \mathbb{R}$ 表示智能体 i 在时间 t 才知晓的特征和标签信息。特征向量$\boldsymbol{a}_{i,t}$ 是随机生成的，其各分量在 -5 到5之间服从均匀分布；标签$l_{i,t}$ 满足$l_{i,t} = \boldsymbol{a}_{i,t}^\top\boldsymbol{x}_0 + \frac{2\xi_{i,t}}{\sqrt{d}}$，其中 $[\boldsymbol{x}_0]_i = 1/d$ 和$\xi_{i,t}$ 在区间 $[0,1]$ 内均匀生成。本次仿真中，我们在由 $n = 20$ 和 $d = 8$ 构成的网络上执行该算法。在以下各情形中，全局平均动态遗憾、遗憾$_d^j(T)/T$的上包络和下包络分别记为$\frac{1}{n}\sum_{j=1}^{n}[$遗憾$_d^j(T)/T]$、$\sup_j\{$遗憾$_d^j(T)/T\}$和$\inf_j\{$遗憾$_d^j(T)/T\}$，用于衡量算法的性能。从图1可以清楚地看出，在条件$\alpha = 1/(4T^{0.4})$下，算法1的三个平均动态遗憾均收敛，这与理论结果相符。

为了研究算法1的计算成本节省程度，我们比较了算法1与分布式在线梯度下降（DOGD）算法 Yan 等人（2013）在两个维度 $d = 8$ 和160下的全局平均动态遗憾和计算时间，其中，算法 DOGD 的动态遗憾分析可视为 Shahrampour 与 Jadbabaie（2017）中满足距离度量函数 $\mathcal{R}(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$ 且映射 $A = I$的特殊情形。算法和 DOGD 的步长 $\alpha_1 = 1/(4T^{0.4})$ $\alpha_2 = 1/(2T^{0.4})$ 以及 $\alpha_3 = 1/(T^{0.4}_{\alpha_4})$ $1/\ 4T^{0.4}$ 1 ( ) 分别设定为相应取值。从图2和图3可以看出，算法1能够达到与算法 DOGD 相近的收敛性能，同时耗费更少的计算时间，这有力地体现了线性预言机的优势。具体而言，当维度 d 从8增至160时，与算法计算成本的显著增加相比，算法1的计算成本仅略有增加

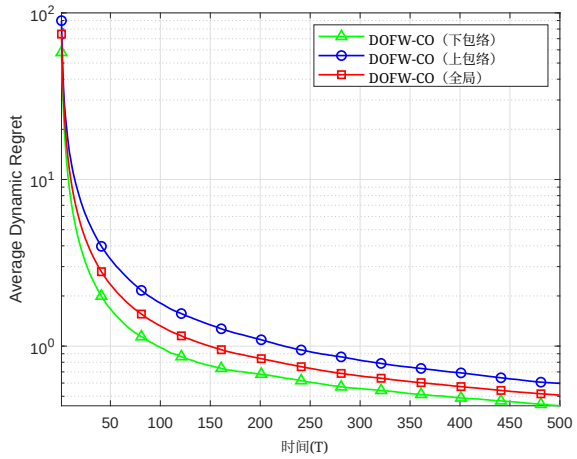

图 1：算法 1的收敛性能。

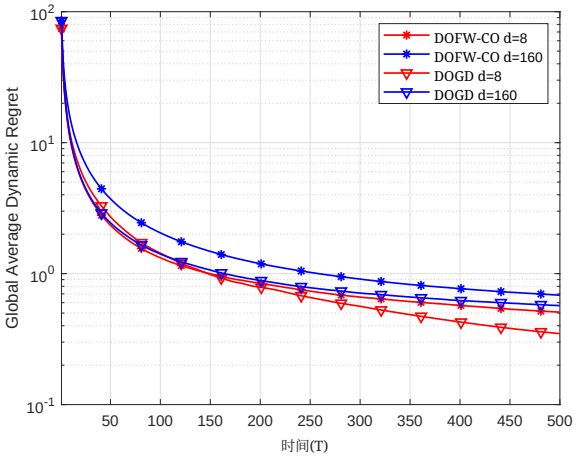

图 2：算法之间遗憾的比较结果1 和 DOGD，在 $d=8$ 和 160 下。

DOGD。就这一点而言，约束优化问题的维度越高，算法 1 在节省计算成本方面的优势就越突出、越重要。

5. 结论

针对分布式在线凸优化问题，本文提出了 DOFW-CO 算法，以降低高维且复杂约束集所带来的高昂投影步骤计算成本。在该问题中，各智能体通过局部计算以及在时变有向网络中与邻居交换信息，相互协作以最小化全局代价函数。我们严格证明了所提出优化算法的动态遗憾上界为 $\mathcal{O}(\sqrt{T}(1+H_T)+D_T)$。最后，通过分布式在线岭回归问题的仿真，对该算法的性能进行了验证和比较。未来，~~一个值得探索~~的方向是改进

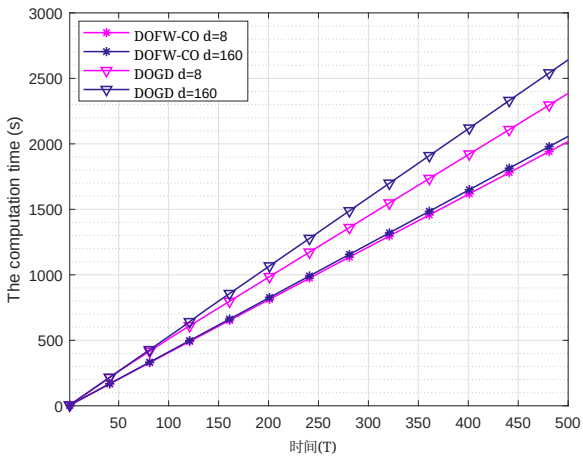

图 3：算法的计算时间比较 1 与 DOGD 在 $d=8$ 和 160 下。

版 DOFW-CO 算法，以提升收敛速度或节省计算资源。

附录 A. 一些关键引理

对于算法 1 的收敛性分析，以下引理至关重要。定义 $\Phi(t,\ s)=A_tA_{t-1}\ldots A_s$为 A_t 对所有 t，s 满足 $t\geq s\geq 1$时的转移矩阵。

引理 1. 设决策序列 $\{\boldsymbol{x}_{i,t}\}$ 由算法1生成。在假设 1 和 2 下，对于 $T\geq 2$，有

$$\sum_{t=1}^{T}\sum_{i=1}^{n}\|\hat{\boldsymbol{x}}_{i,t}-\boldsymbol{x}_{avg,t}\|\leq\frac{n\Gamma}{1-\sigma}\sum_{j=1}^{n}\|\boldsymbol{x}_{j,1}\|+\alpha T\frac{n^2M\Gamma}{1-\sigma}.$$

证明。 根据算法 1，可得

$$\begin{aligned}\hat{\boldsymbol{x}}_{i,t}&=\sum_{j=1}^{n}[A_t]_{ij}\boldsymbol{x}_{j,t} \qquad (A.1)\\&=\sum_{j=1}^{n}[A_t]_{ij}\hat{\boldsymbol{x}}_{j,t-1}+\alpha\sum_{j=1}^{n}[A_t]_{ij}(\boldsymbol{v}_{j,t-1}-\hat{\boldsymbol{x}}_{j,t-1})\\&=\sum_{j=1}^{n}[\Phi(t,1)]_{ij}\boldsymbol{x}_{j,1}+\alpha\sum_{l=1}^{t-1}\sum_{j=1}^{n}[\Phi(t,l+1)]_{ij}(\boldsymbol{v}_{j,l}-\hat{\boldsymbol{x}}_{j,l}).\end{aligned}$$

根据算法 1，项$\boldsymbol{x}_{avg,\ t}$ 可进一步简化如下：

$$\begin{aligned}\boldsymbol{x}_{avg,t}&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}[A_{t-1}]_{ij}\boldsymbol{x}_{i,t-1}+\frac{\alpha}{n}\sum_{i=1}^{n}(\boldsymbol{v}_{i,t-1}-\hat{\boldsymbol{x}}_{j,t-1})\\&=\boldsymbol{x}_{avg,t-1}+\frac{\alpha}{n}\sum_{i=1}^{n}(\boldsymbol{v}_{i,t-1}-\hat{\boldsymbol{x}}_{j,t-1})\\&=\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_{j,1}+\frac{\alpha}{n}\sum_{l=1}^{t-1}\sum_{j=1}^{n}(\boldsymbol{v}_{j,l}-\hat{\boldsymbol{x}}_{j,l}) \qquad (A.2)\end{aligned}$$

其中，第三个等式结合了相邻权重矩阵 A_{t-1} 的双随机性。

结合（A.1）和（A.2），对于 $t \geq 2$，我们得到

$$
\begin{aligned}
&\|\hat{\boldsymbol{x}}_{i,t} - \boldsymbol{x}_{avg,t}\| \\
&\leq \left\| \sum_{j=1}^{n} [\Phi(t,1)]_{ij} \boldsymbol{x}_{j,1} - \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{x}_{j,1} \right\| + \\
&\left\| \alpha \sum_{l=1}^{t-1} \sum_{j=1}^{n} [\Phi(t,l+1)]_{ij} (\boldsymbol{v}_{j,l} - \hat{\boldsymbol{x}}_{j,l}) - \frac{\alpha}{n} \sum_{l=1}^{t-1} \sum_{j=1}^{n} (\boldsymbol{v}_{j,l} - \hat{\boldsymbol{x}}_{j,l}) \right\| \\
&\leq \sum_{j=1}^{n} \left| [\Phi(t,1)]_{ij} - \frac{1}{n} \right| \|\boldsymbol{x}_{j,1}\| \\
&+ \alpha \sum_{l=1}^{t-1} \sum_{j=1}^{n} \left| [\Phi(t,l+1)]_{ij} - \frac{1}{n} \right| \|\boldsymbol{v}_{j,l} - \hat{\boldsymbol{x}}_{j,l}\| \\
&\leq \Gamma \sigma^{t-1} \sum_{j=1}^{n} \|\boldsymbol{x}_{j,1}\| + \alpha n M \Gamma \sum_{l=1}^{t-1} \sigma^{t-l-1} \\
&\leq \Gamma \sigma^{t-1} \sum_{j=1}^{n} \|\boldsymbol{x}_{j,1}\| + \alpha \frac{nM\Gamma}{1-\sigma} \qquad \text{(A.3)}
\end{aligned}
$$

其中第三个不等式由 $\Phi(t,s)$ [1]的性质以及 $\boldsymbol{v}_{j,l}, \hat{\boldsymbol{x}}_{j,l} \in \boldsymbol{X}$ 这一事实得出。从 $i=1$ 到 n 以及从$t=1$ 到 T，对（A.3）两边求和，得到

$$
\begin{aligned}
&\sum_{t=1}^{T} \sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_{i,t} - \boldsymbol{x}_{avg,t}\| \\
&\leq \sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_{i,1} - \boldsymbol{x}_{avg,1}\| + n\Gamma \sum_{t=2}^{T} \sigma^{t-1} \sum_{j=1}^{n} \|\boldsymbol{x}_{j,1}\| + \alpha T \frac{n^2 M \Gamma}{1-\sigma} \\
&\leq \frac{n\Gamma}{1-\sigma} \sum_{j=1}^{n} \|\boldsymbol{x}_{j,1}\| + \alpha T \frac{n^2 M \Gamma}{1-\sigma} \qquad \text{(A.4)}
\end{aligned}
$$

其中第二个不等式由以下事实得出：$\sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_{i,1} - \boldsymbol{x}_{avg,1}\| \leq \sum_{i=1}^{n} \|\sum_{j=1}^{n} [A_1]_{ij} \boldsymbol{x}_{j,1} - \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{x}_{j,1}\| \leq \sum_{i=1}^{n} \sum_{j=1}^{n} |[A_1]_{ij} - \frac{1}{n}| \|\boldsymbol{x}_{j,1}\| \leq n\Gamma \sum_{j=1}^{n} \|\boldsymbol{x}_{j,1}\|$。证明完毕。□

引理 2. 设决策序列 $\{\boldsymbol{x}_{i,t}\}$ 由算法1生成。在假设 2 和 4成立的条件下，对于任意 $T \geq 2$，我们有

$$
\begin{aligned}
&\textbf{(a)} \quad \sum_{i=1}^{n} \overline{\nabla} f_{i,t} = \sum_{i=1}^{n} \nabla f_{i,t}(\hat{\boldsymbol{x}}_{i,t}), \ \forall t \geq 1. \qquad \text{(A.5)} \\
&\textbf{(b)} \quad \sum_{t=2}^{T} \sum_{i=1}^{n} \|\boldsymbol{\delta}_{i,t}\| \leq 2G_X \sum_{t=2}^{T} \sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_{i,t-1} - \boldsymbol{x}_{avg,t-1}\| \\
&\qquad + nD_T + nMG_X \alpha T. \qquad \text{(A.6)}
\end{aligned}
$$

[1]引理（Nedi'等人（2008年））设假设 1 成立。那么，对于所有 $i\ j \in \mathcal{V}$ $|[\Phi\ t\ s\ ]_{ij} - \frac{1}{n}| \leq \Gamma\sigma^{(t-s)}$ $\sigma = 1-$，，有（，），其中（）且 $\zeta/4n^{2\ 1/Q}$ $\Gamma = 1-\zeta/4n^{2\ (1-2Q)/Q}$（）。

证明。(a) 我们利用数学归纳法证明这一部分。注意，根据算法1，$\overline{\nabla} f_{i,\ 1} = \nabla f_{i,\ 1}(\hat{\boldsymbol{x}}_{i,\ 1})$。因此，当$t=1$时，(A.5)中的等式成立。现在假设在某个 t，$\sum_{i=1}^{n} \overline{\nabla} f_{i,\ t} = \sum_{i=1}^{n} \nabla f_{i,\ t}(\hat{\boldsymbol{x}}_{i,\ t})$ 成立，我们将证明该等式在 $t+1$ 时也成立。事实上，

$$
\begin{aligned}
&\sum_{i=1}^{n} \overline{\nabla} f_{i,t+1} \\
&= \sum_{i=1}^{n} \widehat{\nabla} f_{i,t} + \sum_{i=1}^{n} \nabla f_{i,t+1}(\hat{\boldsymbol{x}}_{i,t+1}) - \sum_{i=1}^{n} \nabla f_{i,t}(\hat{\boldsymbol{x}}_{i,t}) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} [A_t]_{ij} \overline{\nabla} f_{j,t} + \sum_{i=1}^{n} \nabla f_{i,t+1}(\hat{\boldsymbol{x}}_{i,t+1}) - \sum_{i=1}^{n} \overline{\nabla} f_{i,t} \\
&= \sum_{i=1}^{n} \nabla f_{i,t+1}(\hat{\boldsymbol{x}}_{i,t+1}) \qquad \text{(A.7)}
\end{aligned}
$$

其中，最后一个等号由 A_t的双随机性得到。

(b) 利用（6），对于任意 $t \geq 2$，我们有

$$
\begin{aligned}
\sum_{i=1}^{n} \|\boldsymbol{\delta}_{i,t}\| &\leq \sum_{i=1}^{n} \|\nabla f_{i,t}(\hat{\boldsymbol{x}}_{i,t}) - \nabla f_{i,t-1}(\hat{\boldsymbol{x}}_{i,t})\| \\
&\quad + \sum_{i=1}^{n} \|\nabla f_{i,t-1}(\hat{\boldsymbol{x}}_{i,t}) - \nabla f_{i,t-1}(\hat{\boldsymbol{x}}_{i,t-1})\| \\
&\leq n g_{t-1,sup} + \sum_{i=1}^{n} (G_X \|\hat{\boldsymbol{x}}_{i,t} - \hat{\boldsymbol{x}}_{i,t-1}\|) \\
&\leq n g_{t-1,sup} + G_X \sum_{i=1}^{n} (\|\hat{\boldsymbol{x}}_{i,t} - \boldsymbol{x}_{avg,t-1}\| \\
&\quad + \|\hat{\boldsymbol{x}}_{i,t-1} - \boldsymbol{x}_{avt,t-1}\|) \\
&\leq n g_{t-1,sup} + 2G_X \sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_{i,t-1} - \boldsymbol{x}_{avg,t-1}\| \\
&\quad + nMG_X \alpha \qquad \text{(A.8)}
\end{aligned}
$$

其中，最后一个不等式基于以下事实得出：$\sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_{i,t} - \boldsymbol{x}_{avg,t-1}\| \leq \sum_{i=1}^{n} \sum_{j=1}^{n} [A_t]_{ij} \|\boldsymbol{x}_{j,t} - \boldsymbol{x}_{avg,t-1}\| \leq \sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_{i,t-1} - \boldsymbol{x}_{avg,t-1} + \alpha(\boldsymbol{v}_{i,t-1} - \hat{\boldsymbol{x}}_{i,t-1})\| \leq \sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_{i,t-1} - \boldsymbol{x}_{avg,t-1}\| + \alpha nM.$

然后，将（A.8）两边从 $t=2$ 到 T求和，即可得到（A.6）中的不等式。证明完毕。□

引理 3. 设序列 $\{\nabla f_{i,t}, \nabla f_{i,t}(\hat{\boldsymbol{x}}_{i,t})\}$ 由算法 1生成。则在假设 1 和4下，对于任意 $T \geq 2$，有

$$
\begin{aligned}
&\sum_{t=1}^{T} \sum_{i=1}^{n} \left\| \widehat{\nabla} f_{i,t} - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j,t}(\hat{\boldsymbol{x}}_{j,t}) \right\| \\
&\leq \frac{n\Gamma}{1-\sigma} \sum_{j=1}^{n} \|\nabla f_{j,1}(\hat{\boldsymbol{x}}_{j,1})\| + \frac{n\Gamma}{1-\sigma} \sum_{t=2}^{T} \sum_{i=1}^{n} \|\boldsymbol{\delta}_{i,t}\|. \qquad \text{(A.9)}
\end{aligned}
$$

证明。类似于引理 1 的证明，并结合算法 1，（A.5），对于任意 $t \geq 2$，可以验证

$$\widehat{\nabla} f_{i,t} = \sum_{j=1}^{n} [\Phi(t,2)]_{ij} \widehat{\nabla} f_{j,1} + \sum_{l=2}^{t} \sum_{j=1}^{n} [\Phi(t,l)]_{ij} \boldsymbol{\delta}_{j,l}, \quad \text{(A.10)}$$

$$\frac{1}{n} \sum_{i=1}^{n} \nabla f_{i,t}(\hat{\boldsymbol{x}}_{i,t}) = \frac{1}{n} \sum_{i=1}^{n} \overline{\nabla} f_{i,1} + \frac{1}{n} \sum_{l=2}^{t} \sum_{i=1}^{n} \boldsymbol{\delta}_{i,l}. \quad \text{(A.11)}$$

^与（A.3）类似，结合事实$\nabla f_{i,\ 1} = \sum_{j=1}^{n} [A_1]_{ij} \overline{\nabla} f_{j,\ 1}$，$\overline{\nabla} f_{i,\ 1} = \nabla f_{i,\ 1}(\hat{\boldsymbol{x}}_{i,\ 1})$，由（A.10）和（A.11）可得$\| \widehat{\nabla} f_{i,\ t} - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j,\ t}(\hat{\boldsymbol{x}}_{j,\ t}) \| \leq \sum_{j=1}^{n} \Gamma \sigma^{t-1} \| \nabla f_{j,\ 1}(\hat{\boldsymbol{x}}_{j,\ 1}) \| + \sum_{l=2}^{t} \sum_{j=1}^{n} \Gamma \sigma^{t-l} \| \boldsymbol{\delta}_{j,\ l} \|$。这意味着

$$\begin{aligned}
&\sum_{t=1}^{T} \sum_{i=1}^{n} \left\| \widehat{\nabla} f_{i,t} - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j,t}(\hat{\boldsymbol{x}}_{j,t}) \right\| \\
&\leq \sum_{i=1}^{n} \left\| \widehat{\nabla} f_{i,1} - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j,1}(\hat{\boldsymbol{x}}_{j,1}) \right\| \\
&+ \sum_{t=2}^{T} \sum_{j=1}^{n} n\Gamma \sigma^{t-1} \| \nabla f_{j,1}(\hat{\boldsymbol{x}}_{j,1}) \| + \sum_{t=2}^{T} \sum_{l=2}^{t} \sum_{j=1}^{n} n\Gamma \sigma^{t-l} \| \boldsymbol{\delta}_{j,l} \| \\
&\leq \sum_{i=1}^{n} \sum_{j=1}^{n} \left| [A_1]_{ij} - \frac{1}{n} \right| \| \nabla f_{j,1}(\hat{\boldsymbol{x}}_{j,1}) \| \\
&+ \frac{\sigma n \Gamma}{1-\sigma} \sum_{j=1}^{n} \| \nabla f_{j,1}(\hat{\boldsymbol{x}}_{j,1}) \| + n\Gamma \left(\sum_{l=0}^{T-2} \sigma^{l} \right) \left(\sum_{t=2}^{T} \sum_{i=1}^{n} \| \boldsymbol{\delta}_{i,t} \| \right) \\
&\leq \frac{n\Gamma}{1-\sigma} \sum_{j=1}^{n} \| \nabla f_{j,1}(\hat{\boldsymbol{x}}_{j,1}) \| + \frac{n\Gamma}{1-\sigma} \sum_{t=2}^{T} \sum_{i=1}^{n} \| \boldsymbol{\delta}_{i,t} \| \quad \text{(A.12)}
\end{aligned}$$

将上述不等式代入（A.12）中，我们可以很容易地得到（A.9）。证明完毕。□

引理 4。设决策序列 $\{\boldsymbol{x}_{i,t}\}$ 由算法1生成。在假设2、3 和 4下，对任意 $T \geq 2$，有

$$\begin{aligned}
&\sum_{t=1}^{T} F_t(\boldsymbol{x}_{avg,t}) - \sum_{t=1}^{T} F_t(\boldsymbol{x}_t^*) \\
&\leq \frac{2n}{\alpha} H_T + \frac{n L_X M}{\alpha} + 2M \sum_{t=1}^{T-1} \sum_{i=1}^{n} \left\| \sum_{j=1}^{n} \nabla f_{j,t}(\hat{\boldsymbol{x}}_{j,t}) - \widehat{\nabla} f_{i,t} \right\| \\
&\quad + 2MG_X \sum_{t=1}^{T-1} \sum_{i=1}^{n} \| \hat{\boldsymbol{x}}_{i,t} - \boldsymbol{x}_{avg,t} \| + \frac{n G_X M^2}{2} \alpha T + nMD_T.
\end{aligned} \quad \text{(A.13)}$$

证明。利用假设4中的光滑性，可得

$$\begin{aligned}
&F_{t+1}(\boldsymbol{x}_{avg,t+1}) - F_{t+1}(\boldsymbol{x}_{avg,t}) \quad \text{(A.14)} \\
&\leq \langle \nabla F_{t+1}(\boldsymbol{x}_{avg,t}), \boldsymbol{x}_{avg,t+1} - \boldsymbol{x}_{avg,t} \rangle \\
&\quad + \frac{n G_X}{2} \| \boldsymbol{x}_{avg,t+1} - \boldsymbol{x}_{avg,t} \|^2 \\
&\leq \alpha \sum_{i=1}^{n} \left\langle \frac{1}{n} \nabla F_{t+1}(\boldsymbol{x}_{avg,t}), \boldsymbol{v}_{i,t} - \boldsymbol{x}_{avg,t} \right\rangle + \frac{n G_X M^2 \alpha^2}{2}.
\end{aligned}$$

进一步可以验证

$$\begin{aligned}
&\left\langle \frac{1}{n} \nabla F_{t+1}(\boldsymbol{x}_{avg,t}), \boldsymbol{v}_{i,t} - \boldsymbol{x}_{avg,t} \right\rangle \\
&\leq \left\langle \frac{1}{n} \nabla F_{t+1}(\boldsymbol{x}_{avg,t}) - \widehat{\nabla} f_{i,t}, \boldsymbol{v}_{i,t} - \boldsymbol{x}_{avg,t} \right\rangle \\
&\quad + \left\langle \widehat{\nabla} f_{i,t}, \boldsymbol{x}_t^* - \boldsymbol{x}_{avg,t} \right\rangle \\
&\leq \frac{M}{n} \| \nabla F_{t+1}(\boldsymbol{x}_{avg,t}) - \nabla F_t(\boldsymbol{x}_{avg,t}) \| + 2M \left\| \frac{1}{n} \nabla F_t(\boldsymbol{x}_{avg,t}) \right. \\
&\quad \left. - \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j,t}(\hat{\boldsymbol{x}}_{j,t}) + \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j,t}(\hat{\boldsymbol{x}}_{j,t}) - \widehat{\nabla} f_{i,t} \right\| \\
&\quad + \frac{1}{n} \langle \nabla F_t(\boldsymbol{x}_{avg,t}), \boldsymbol{x}_t^* - \boldsymbol{x}_{avg,t} \rangle \\
&\leq M g_{t,sup} + 2M \left\| \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j,t}(\hat{\boldsymbol{x}}_{j,t}) - \widehat{\nabla} f_{i,t} \right\| \quad \text{(A.15)} \\
&\quad + \frac{2MG_X}{n} \sum_{j=1}^{n} \| \hat{\boldsymbol{x}}_{j,t} - \boldsymbol{x}_{avg,t} \| + \frac{1}{n} [F_t(\boldsymbol{x}_t^*) - F_t(\boldsymbol{x}_{avg,t})]
\end{aligned}$$

其中，第一个不等式是利用以下最优性条件得到的：

$$\left\langle \widehat{\nabla} f_{i,t}, \boldsymbol{x}_t^* \right\rangle \geq \min_{\boldsymbol{x} \in \boldsymbol{X}} \left\langle \widehat{\nabla} f_{i,t}, \boldsymbol{x} \right\rangle = \left\langle \widehat{\nabla} f_{i,t}, \boldsymbol{v}_{i,t} \right\rangle \quad \text{(A.16)}$$

最后一个不等式则基于 $F_t(\boldsymbol{x})$ 的凸性条件以及假设4推导得出。然后，由 (A.14) 和 (A.15) 可得

$$\begin{aligned}
&F_{t+1}(\boldsymbol{x}_{avg,t+1}) - F_{t+1}(\boldsymbol{x}_{avg,t}) \\
&\leq \alpha n M g_{t,sup} + \alpha [F_t(\boldsymbol{x}_t^*) - F_t(\boldsymbol{x}_{avg,t})] + \Psi_t. \quad \text{(A.17)}
\end{aligned}$$

其中 $\Psi_t = 2\alpha M \sum_{i=1}^{n} \left\| \frac{1}{n} \sum_{j=1}^{n} \nabla f_{j,\ t}(\hat{\boldsymbol{x}}_{j,\ t}) - \widehat{\nabla} f_{i,\ t} \right\| + 2\alpha MG_X \sum_{j=1}^{n} \| \hat{\boldsymbol{x}}_{j,\ t} - \boldsymbol{x}_{avg,\ t} \| + \frac{n G_X M^2 \alpha^2}{2}$。基于该不等式，我们还可以进一步得到

$$\begin{aligned}
&F_{t+1}(\boldsymbol{x}_{avg,t+1}) - F_{t+1}(\boldsymbol{x}_{t+1}^*) \\
&\leq F_{t+1}(\boldsymbol{x}_{avg,t}) - F_{t+1}(\boldsymbol{x}_{t+1}^*) + \alpha n M g_{t,sup} + \Psi_t \\
&\quad + \alpha [F_t(\boldsymbol{x}_t^*) - F_t(\boldsymbol{x}_{avg,t})] \\
&\leq n f_{t,sup} + F_t(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_t^*) + F_t(\boldsymbol{x}_t^*) - F_{t+1}(\boldsymbol{x}_{t+1}^*) \\
&\quad + \alpha n M g_{t,sup} + \alpha [F_t(\boldsymbol{x}_t^*) - F_t(\boldsymbol{x}_{avg,t})] + \Psi_t \quad \text{(A.18)}
\end{aligned}$$

其中，最后一个不等式是利用以下事实得到的：

$$\begin{aligned}
&F_{t+1}(\boldsymbol{x}_{avg,t}) - F_{t+1}(\boldsymbol{x}_{t+1}^*) \\
&= F_{t+1}(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_{avg,t}) + F_t(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_t^*) \\
&\quad + F_t(\boldsymbol{x}_t^*) - F_{t+1}(\boldsymbol{x}_{t+1}^*)
\end{aligned}$$

$$\leq nf_{t,sup} + F_t(\boldsymbol{x}_{avg,\ t}) - F_t(\boldsymbol{x}_t^*) + F_t(\boldsymbol{x}_t^*) - F_{t+1}(\boldsymbol{x}_{t+1}^*)。$$

将（A.18）两边求和，可得

$$\sum_{t=1}^{T-1}\left\{F_{t+1}(\boldsymbol{x}_{avg,t+1}) - F_{t+1}(\boldsymbol{x}_{t+1}^*) - [F_t(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_t^*)]\right\}$$
$$+\alpha\sum_{t=1}^{T-1}[F_t(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_t^*)]$$
$$\leq nH_T + F_1(\boldsymbol{x}_1^*) - F_T(\boldsymbol{x}_T^*) + \alpha nMD_T + \sum_{t=1}^{T-1}\Psi_t \quad \text{(A.19)}$$

回顾以下事实：$0<\alpha\leq 1$和 $F_T(\boldsymbol{x}_{avg,\ T})-F_T(\boldsymbol{x}_T^*)\geq 0$，可得

$$\alpha\sum_{t=1}^{T}[F_t(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_t^*)] \quad \text{(A.20)}$$
$$\leq [F_T(\boldsymbol{x}_{avg,T}) - F_T(\boldsymbol{x}_T^*)] + \alpha\sum_{t=1}^{T-1}[F_t(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_t^*)]$$
$$\leq nH_T + [F_1(\boldsymbol{x}_{avg,1}) - F_T(\boldsymbol{x}_T^*)] + \alpha nMD_T + \sum_{t=1}^{T-1}\Psi_t.$$

此外，在假设 2 和假设 3 下可得 $F_1(\boldsymbol{x}_{avg,\ 1}) - F_T(\boldsymbol{x}_T^*) = F_1(\boldsymbol{x}_{avg,\ 1}) - F_1(\boldsymbol{x}_T^*) + \sum_{t=1}^{T-1}[F_t(\boldsymbol{x}_T^*) - F_{t+1}(\boldsymbol{x}_T^*)] \leq nL_X\|\boldsymbol{x}_{avg,\ 1} - \boldsymbol{x}_T^*\| + \sum_{t=1}^{T-1}\sum_{i=1}^{n}|f_{i,\ t}(\boldsymbol{x}_T^*) - f_{i,\ t+1}(\boldsymbol{x}_T^*)| \leq nL_XM + nH_T。$ This, together with (A.20), yields that

$$\alpha\sum_{t=1}^{T}[F_t(\boldsymbol{x}_{avg,t}) - F_t(\boldsymbol{x}_t^*)]$$
$$\leq 2nH_T + nL_XM + \alpha nMD_T + \sum_{t=1}^{T-1}\Psi_t. \quad \text{(A.21)}$$

这进一步通过回顾A.13）中的定义，推出了式（ Ψ_t。证明完毕。□

参考文献

Besbes, O., Gur, Y., & Zeevi, A. (2015). 非平稳随机优化. *Operations Research*, 63 , 1227–1244. Hazan, E. (2008). 半定规划的稀疏近似解. In *Latin American Symposium on Theoretical Informatics* (pp. 306–316). Hazan, E., & Kale, S. (2012). 无投影在线学习. In *Proceedings of the 29th International Conference on International Conference on Machine Learning* 1843 1850 (pp. –).

Hou, J., Zeng, X., Wang, G., Sun, J., & Chen, J. (2022). 用于随机优化的分布式动量 Frank-Wolfe 算法. *IEEE/CAA Journal of Automatica Sinica*, . Doi:10.1109/JAS.2022.105923. Kalhan, D. S., Bedi, A. S., Koppel, A., Rajawat, K., Hassani, H., Gupta, A. K., & Banerjee, A. (2021). 基于 Frank-Wolfe 算法的动态在线学习. *IEEE Transactions on Signal Processing*, 69 932 947, – . Li, W., Zeng, X., Hong, Y., & Ji, H. (2021). 半定规划的分布式共识求解器：一种优化视角. *Automatica*, 131, 109737.

Li, X., Xie, L., & Li, N. (2022). 去中心化在线学习综述. *arXiv preprint, arXiv:2205.00473*. Liu, C., Li, H., & Shi, Y. (2020). 一种用于具有不同耦合源的多智能体优化的酉分布式次梯度方法. *Automatica*, 114 , 108834. Nedi´c, A., & Liu, J. (2018). 控制中的分布式优化. *Annual Review of Control, Robotics, and Autonomous Systems*, 1 77 103, – . Nedi´c, A., Olshevsky, A., Ozdaglar, A., & Tsitsiklis, J. N. (2008). 分布式次梯度方法与量化效应. In *2008 47th IEEE Conference on Decision and Control* 4177 4184 (pp. –).

Shahrampour, S., & Jadbabaie, A. (2017). 在动态环境中使用镜像下降进行分布式在线优化. *IEEE Transactions on Automatic Control 63* 714 725 , , – .

Shalev-Shwartz, S. et al. (2011). 在线学习与在线凸优化. *Foundations and Trends in Machine Learning*, 4 , 107 194– .

Sundhar Ram, S., Nedi´c, A., & Veeravalli, V. V. (2010). 用于凸优化的分布式随机次梯度投影算法. *Journal of Optimization Theory and Applications*, 147, 516– 545.

Thang, N. K., Srivastav, A., Trystram, D., & Youssef, P. (2022). 用于去中心化在线凸优化的随机条件梯度算法. *Journal of Parallel and Distributed Computing 169* 334 351 , , – .

Wai, H.-T., Lafond, J., Scaglione, A., & Moulines, E. (2017). 用于凸与非凸问题的去中心化 Frank-Wolfe 算法. *IEEE Transactions on Automatic Control*, 62 , 5522–5537. Wan, Y., Tu, W.-W., & Zhang, L. (2020). 具有 $o(\sqrt{T})$通信复杂度的无投影分布式在线凸优化. In *Proceedings of the 37th International Conference on Machine Learning* 9818 9828 (pp. –).

Wan, Y., Wang, G., & Zhang, L. (2021a). 具有强凸损失的无投影分布式在线学习. *arXiv preprint, arXiv:2103.11102* .

Wan, Y., Xue, B., & Zhang, L. (2021b). 动态环境中的无投影在线学习. *Proceedings of the AAAI Conference on Artificial Intelligence* 10067 10075 , (pp. –). Wu, Z., & Teo, K. (1983). 涉及一类非线性二阶双曲偏微分方程的最优控制问题的条件梯度方法. *Journal of Mathematical Analysis and Applications 91* 376 393 , , – .

Xu, J.-M., & Soh, Y. C. (2016). 一种用于大规模动态优化问题的分布式同步扰动方法. *Automatica*, 72 , 194–204. Xu, L., Yi, X., Sun, J., Shi, Y., Johansson, K. H., & Yang, T. (2022). 具有线性收敛的量化分布式非凸优化. In *2022 IEEE 61st Conference on Decision and Control (CDC)* 5837 5842 (pp. –). Yan, F., Sundaram, S., Vishwanathan, S., & Qi, Y. (2013). 分布式自主在线学习：遗憾值与内在隐私保护特性. *IEEE Transactions on Knowledge and Data Engineering 25* 2483 2493 , , – .

Yang, T., Yi, X., Wu, J., Yuan, Y., Wu, D., Meng, Z., Hong, Y., Wang, H., Lin, Z., & Johansson, K. H. (2019). 分布式优化综述. *Annual Reviews in Control*, 47 , 278–305. Yi, X., Li, X., Yang, T., Xie, L., Chai, T., & Johansson, K. H. (2021). 具有时变耦合不等式约束的分布式赌博机在线凸优化. *IEEE Transactions on Automatic Control 66* 4620 4635 , , – .

Yuan, D., Zhang, B., Ho, D. W., Zheng, W. X., & Xu, S. (2022). 随机量化下的分布式在线赌博机优化. *Automatica*, 146 , 110590. Zhang, L., Kekatos, V., & Giannakis, G. B. (2017a). 可扩展的电动汽车充电协议. *IEEE Transactions on Power Systems*, 32 1451 1462, – . Zhang, W., Zhao, P., Zhu, W., Hoi, S. C. H., & Zhang, T. (2017b). 网络中的无投影分布式在线学习. In *Proceedings of the 34th International Conference on Machine Learning* 4054 4062 (pp. –).