

带事件触发机制的分布式量化随机无梯度投影算法

谢奕彬, 高文华[†]

(华南理工大学 数学学院, 广东 广州 510640)

摘要: 本文研究多智能体系统的分布式约束优化问题, 系统中的每个智能体仅知道自身的局部目标函数和全局非空约束集, 通过与邻居节点进行信息交互, 最终协同求出优化问题的最优解. 本文所提出的算法针对通信网络为时变不平衡有向图, 且每个智能体不知道它的出度的情况. 同时考虑到现实中通信带宽有限和通讯成本的限制, 应用基于编译码方案的量化技术对节点之间的通讯信息进行预处理, 再利用事件触发广播技术降低网络的通信次数. 同时引入高斯光滑函数和随机无梯度方法替代传统的次梯度方法. 本文提出了基于事件触发的分布式量化随机无梯度算法, 在目标函数为凸且 Lipschitz 连续的条件下, 证明了所提算法能收敛到最优值的邻域, 同时给出了使量化器不饱和和量化水平更新规则. 最后通过数值仿真验证了算法的有效性和可行性.

关键词: 分布式优化; 无梯度; 量化; 事件触发机制; 时变不平衡有向图

引用格式: 谢奕彬, 高文华. 带事件触发机制的分布式量化随机无梯度投影算法. 控制理论与应用, 2021, 38(8): 1175 – 1187

DOI: 10.7641/CTA.2020.00353

Distributed quantized random gradient-free algorithm with event triggered communication

XIE Yi-bin, GAO Wen-hua[†]

(School of Mathematics, South China University of Technology, Guangzhou Guangdong 510640, China)

Abstract: We study the distributed constraint optimization problem of multi-agent systems, where each agent only knows its own local objective function and a global non-empty constraint set and the optimal solution of the optimization problem is finally obtained by communicating with neighbor nodes. The proposed algorithm is for the case that the communication network is time-varying unbalanced digraphs and each agent does not know its out-degree. Considering the limited bandwidth and communication cost in reality, the quantization technology based on coding and decoding scheme is used to preprocess the communication information between nodes and the event triggered broadcasting technology is also used to reduce the communication times of the networks. Gaussian smooth function and gradient-free oracle are introduced to replace the traditional subgradient method in this paper. We propose a distributed quantized random gradient-free algorithm based on event triggering communication, and under the condition that the objective function is convex and Lipschitz continuous, it is proved that the proposed algorithm can converge to the neighborhood of the optimal value. Furthermore, the update rule of quantization level that makes the quantizer unsaturated is given. Finally, numerical simulations are provided to illustrate the validity and feasibility of the algorithm.

Key words: distributed optimization; gradient-free; quantization; event triggering; time-varying unbalanced digraph

Citation: XIE Yibin, GAO Wenhua. Distributed quantized random gradient-free algorithm with event triggered communication. *Control Theory & Applications*, 2021, 38(8): 1175 – 1187

1 引言

有线和无线技术的最新进展导致了大规模网络的出现, 带来了多智能体网络中的许多新应用, 其中分布式优化问题得到了广泛关注, 其目标函数为网络中所有智能体的目标函数之和, 每个智能体在仅知道自

身目标函数的情况下, 利用网络通信对状态信息进行更新, 最终所有的智能体的状态在最优解处达到一致. 分布式优化理论和应用已经成为当代系统和控制科学的重要发展方向之一, 并在许多领域取得了重大突破^[1-7]. 例如近年来逐渐取得广泛关注的智能电网调

收稿日期: 2020-06-14; 录用日期: 2020-12-30.

[†]通信作者. E-mail: whgao@scut.edu.cn; Tel.: +86 15302201933.

本文责任编辑: 刘志新.

国家自然科学基金项目(61803108), 广州市科技计划项目(202002030158)资助.

Supported by the National Natural Science Foundation of China (61803108) and the Guangzhou Science and Technology Planning Project (202002030158).

度策略、分布式机器学习、基于分布式优化的信息—物理系统的优化控制、自主式水下机器人的队形控制、基于算术均值融合的分布式随机集多目标跟踪等^[8-10].

分布式优化问题的特点是算法设计依赖于节点组成的通信网络拓扑结构和节点间的信息传输方式,而传输的信息依赖于目标函数模型,不同性质的目标函数假设意味着个体能获取的信息不同.下面就从网络结构、信息传输和目标函数3方面阐述已有的与本文相关的研究结果.

早期的研究大多考虑在平衡网络中分布式优化算法,如Nedić(2009)提出了分布式次梯度算法^[11-12],Srivastava(2010)研究了带白噪声的分布式次梯度算法^[13],Lin(2016)研究了通信带延迟的分布式次梯度算法^[14].

现实中智能体很可能分布在时变不平衡网络中,此时通信网络图的权重矩阵不再是双随机.为了解决不平衡网络所带来的影响,Nedić(2015)提出了subgradient-push算法^[15],只要求权重矩阵是列随机矩阵,随后(2017)提出了Push-DIGing算法,要求权重矩阵是列随机且采用了固定步长.而Xie(2018)借助优化问题的上境图形式抵消了不平衡图给分布式次梯度算法带来的影响^[16].

上述分布式算法要求智能体准确地接收到邻居的状态信息.但随着模型规模变大,例如在深度学习算法中,每次迭代的通信开销成为主要瓶颈,量化通信是解决此问题的主要方法,但这导致了许多的结果无法直接应用.Rabbat(2005)研究了分布式量化增量优化算法^[17],Nedić(2008)研究了分布式量化次梯度算法^[18],Taheri(2020)研究了量化gradient-push算法^[19].此外,Li(2011)提出了一个基于编译码方案的量化器^[20],Yi(2014)应用这个编译码的量化方案设计了分布式量化次梯度算法^[21].Zhang(2018)将节点间的量化信息减少到只用符号信息^[22].

为了进一步解决通信成本问题,可以让智能体选择在必要时才与邻居通信.Yu(2019)研究了基于事件触发机制和传输延迟的分布式优化算法^[23],Hayashi和Naoki(2020)提出了基于事件触发机制的定步长次梯度分布式算法^[24],Li(2018)结合了事件触发机制和量化的技术,提出了基于事件触发机制的分布式量化次梯度算法^[25].

上述文献为解决目标函数不可微都采用了次梯度的方法,当目标函数的次梯度求解太过于繁琐时,Nesterov(2017)提出了一种随机无梯度方法^[26].随机无梯度方法仅要求目标函数是凸的且Lipschitz连续,其求解过程不需用到梯度信息,具有广泛的应用前景.文献[27-33]将随机无梯度方法应用到分布式算法上.特别地,Yuan(2015)基于gradient-push提出了不平衡

图上的随机无梯度算法^[32],但该算法不易拓展到约束优化问题.Ding(2017)提出了分布式量化随机无梯度投影算法^[33],但该算法仅适用于时变平衡图的情况且在文章中没有给出具体的量化方案.

目前,分布式随机无梯度方法有待深入研究,本文在文献[33]的基础上,应用文献[34]中对梯度加权的方法,将随机无梯度投影算法拓展到时变不平衡图上,该算法仅要求权重矩阵是行随机的,这在基于广播的通信机制中具有更好的实用性.同时为减少通信成本,本文结合了编译码动态量化规则和事件触发机制提出了基于事件触发机制的分布式量化随机无梯度投影算法,降低了对通信网络的带宽要求且更大效率地利用通信资源.

本文的主要结果是如下3个方面:

1) 在分布式优化算法的基础上,考虑到了通信带宽和通信成本的限制,采用了基于事件触发的动态编译码方案进行信息传输,有效地降低了对网络通信带宽的要求和通信次数.此外,采用了随机无梯度方法替代传统的次梯度方法,仅要求目标函数是凸的且Lipschitz连续.

2) 在所有的量化器不饱和的条件下,第3.2节给出一种时变不平衡有向图上的分布式算法求解优化问题的近似解.

3) 在所有的量化器不饱和的条件下,第4节对所提算法进行收敛性分析,给出算法的收敛速率,并求出了量化器不饱和的量化水平下界.

本文余下部分的结构如下:第2节介绍了分布式优化问题并给出预备知识;第3节陈述了编译码量化规则和事件触发设置,设计了基于事件触发的分布式量化随机无梯度算法;第4节研究了所提出的算法的收敛性并给出使量化器不饱和的量化水平下界;第5节提供了具体的仿真来验证分布式优化算法的有效性;最后,第6节给出结论.

记号:所有向量均视为列向量.智能体 i 涉及的变量(可能是标量、集合和函数)均标有下标 i ,例如 $x_i(k)$ 智能体 i 在迭代时刻 k 的状态值.用 $\hat{\cdot}$ 代表涉及量化的相关变量估计,例如 $\hat{x}_{ij}(k)$ 代表智能体 j 得到对 $x_i(k)$ 的量化估计值.用 $\tilde{\cdot}$ 代表涉及事件触发机制的相关变量估计,例如 $\tilde{x}_{ij}(k)$ 为考虑事件触发机制后智能体 j 得到对 $x_i(k)$ 的估计值.设 $\mathbf{1}_m$ 为所有元素均为1的 m 维列向量, I 为 m 维单位矩阵, \mathbb{R} 和 \mathbb{R}^n 分别为全体实数集合和 n 维实数空间.向量 x 的欧式范数和无穷范数分别记为 $\|x\|$ 和 $\|x\|_\infty$,

$$\|x\| = (x^T x)^{\frac{1}{2}}, \|x\|_\infty = \max_{1 \leq i \leq m} |x_i|.$$

向量 x 在非空闭凸集 X 上的投影记为 $P_X(x)$,

$$P_X(x) = \arg \min_{y \in X} \|x - y\|.$$

$\nabla f(x)$ 代表 $f(x)$ 的梯度, 问题(1)–(2)的最优解分别记为 x^* 和 x_μ^* , 即

$$x^* = \arg \min_{x \in X} f(x), \quad x_\mu^* = \arg \min_{x \in X} f_\mu(x).$$

2 问题描述与预备知识

本文考虑一个由 $V = \{1, 2, \dots, m\}$ 标记的智能体网络. 智能体之间可以局部的交换信息. 多智能体系统的共同目标是协同求解一个凸优化问题, 该问题的目标函数是所有智能体局部目标函数的和,

$$\min f(x) = \sum_{i=1}^m f_i(x), \quad (1)$$

$$\text{s.t. } x \in X,$$

其中: $f_i: \mathbb{R}^n \rightarrow \mathbb{R}$ 是智能体 i 的局部目标函数, 全局约束集 $X \subseteq \mathbb{R}^n$.

假设 1 a) X 是一个有限的非空的闭凸集, 半径为 C_X . 即 $\|x\| \leq C_X, \forall x \in X$.

b) 优化问题(1)的最优解集非空.

c) 局部目标函数 $f_i(x)$ 为凸的且 $L_0(f_i)$ -Lipschitz连续, 即存在 $L_0(f_i) > 0$ 使得

$$\|f_i(x) - f_i(y)\| \leq L_0(f_i)\|x - y\|, \quad \forall x, y \in X.$$

注 1 文献[34]采用次梯度算法, 假设中不要求目标函数满足Lipschitz连续的条件, 只要求次梯度存在且有界, 而根据文献[35], 次梯度有界其本质等价于目标函数Lipschitz连续. 在非光滑优化算法中, 存在目标函数非光滑非Lipschitz连续的优化算法, 如文献[35], 但该算法依然需要计算次梯度. 随机无梯度方法如文献[27–33]皆要求目标函数Lipschitz连续, 以保证光滑后的函数能逼近原函数. 本文假设1是随机无梯度方法的通常假设.

由于每个局部目标函数 $f_i(x)$ 都是非光滑的, 则 $f_i(x)$ 的梯度不存在, 而次梯度信息在有些实际问题中计算相对复杂, 此时基于梯度或次梯度的方法不能有效的解决优化问题(1). 为此, 本文通过高斯光滑函数 $f_{i,\mu_i}(x)$ 逼近原目标函数进而得到随机无梯度, 用于代替传统的梯度信息. 由文献[26]知高斯光滑函数 $f_{i,\mu_i}(x)$ 和随机无梯度 $\tilde{\nabla} f_{i,\mu_i}(x)$ 定义如下:

$$f_{i,\mu_i}(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\mathbb{R}^n} f_i(x + \mu_i u) e^{-\frac{1}{2}\|u\|^2} du,$$

其中 μ_i 是 $f_{i,\mu_i}(x)$ 的非负光滑参数.

$$\tilde{\nabla} f_{i,\mu_i}(x) = \frac{f_i(x + \mu_i u) - f_i(x - \mu_i u)}{2\mu_i} u,$$

其中 u 是单位球体上均匀分布的随机向量. 由文[26]可知 $f_{i,\mu_i}(x)$ 是可微的, 且有

$$\mathbb{E}[\tilde{\nabla} f_{i,\mu_i}(x(k)) | F(k)] = \nabla f_{i,\mu_i}(x(k)),$$

其中 $F(k)$ 是由迭代算法从0步到 $k-1$ 步产生的随机变量构成的 σ -代数, 即

$$\begin{cases} F(k) = \{x_i(\tau), u(\tau); 0 \leq \tau \leq k-1, i \in V\}, \\ F(0) = \{x_i(0), i \in V\}. \end{cases}$$

注 2 1) 次梯度的计算需要一定的技巧与时间, 且在次梯度的计算过程中, 有时需要额外的内存来储存中间变量, 这对于有记忆限制的模型不再适用. 2) 当系统的确切表达式不可知, 仅知道系统的输入与输出时, 基于次梯度的算法不再适用但可以采用随机无梯度方法^[36–37].

引理 1^[26] 若 $f_i(x)$ 满足假设1(c)则

a) $f_{i,\mu_i}(x)$ 是凸函数, 且满足

$$f_i(x) \leq f_{i,\mu_i}(x) \leq f_i(x) + \sqrt{n}\mu_i L_0(f_i).$$

b) $f_{i,\mu_i}(x)$ 是 $L(f_{i,\mu_i})$ -Lipschitz连续的, 且 $L(f_{i,\mu_i}) \leq L_0(f_i)$.

c) 存在常数 $C_{\nabla f} \leq nL_0(f_i)$ 使得

$$\mathbb{E}[\|\tilde{\nabla} f_{i,\mu_i}(x)\| | F(k)] \leq C_{\nabla f}.$$

将问题(1)的目标函数用所对应的高斯光滑函数替代, 得到问题(2).

多智能体之间的信息通信可以建模成时变有向

$$\begin{aligned} \min f_\mu(x) &= \sum_{i=1}^m f_{i,\mu_i}(x), \\ \text{s.t. } x &\in X. \end{aligned} \quad (2)$$

图 $G(k) = (V, E(k))$, 其中节点 $i \in V$ 代表第 i 个智能体, 有向边 $(i, j) \in E(k)$ 表示在 k 时刻信息可以从智能体 i 传送到智能体 j . $A(k) = [a_{ij}(k)]$ 为时变有向图 $G(k)$ 的非负权重矩阵, 当 $(i, j) \in E(k)$ 则有 $a_{ji}(k) > 0$, 否则 $a_{ji}(k) = 0$. 如果存在一系列的有向边 $(i_1, i_2), (i_2, i_3), \dots, (i_{t-1}, i_t)$, 使得 $(i_{l-1}, i_l) \in E(k), l = 2, \dots, t$, 称 k 时刻 i_1 是可达 i_t . 如果任意节点都是可达其他节点的, 称时变有向图是强连通的. 本文对时变有向图的连通性有如下常见的假设, 见文献[15, 34, 38].

假设 2 时变有向图是 B -强连通的, 即对任意 $k \geq 0$, 存在一个正整数 B 使得

$$G(k) \cup G(k+1) \cdots \cup G(k+B-1)$$

是强连通的.

在假设2的条件下, 时变有向图 $G(k)$ 不需要在任意时刻 k 都保持连通, 但 $G(k)$ 是 B -强连通的, 以保证在 B 连续时间段内信息能从一个节点传输到另一个节点.

注 3 对于时变不平衡有向图, 权重矩阵的双随机条件不再满足, 许多平衡图下的算法不再适用, 因此有必要设计新的分布式优化算法. 文献[15]提出gradient-push算法, 该算法要求权重矩阵是列随机的, 即 $\sum_{i=1}^m a_{ij}(k) = 1$, 为此每个节点要求知道它的出度. 而本文仅要求权重矩阵是行随机的, 相比于双随机矩阵和列随机矩阵, 行随机矩阵在通信机制中有更好的实用性.

定义转移矩阵 $A(k:s)$:

$$\begin{cases} A(k:s) = A(k)A(k-1)\cdots A(s+1)A(s), & k \geq s, \\ A(k:k+1) = I. \end{cases}$$

引理 2^[34] 在假设2成立的条件下, 对于所有 $s \geq 0$, 存在一随机向量 $\pi(k)$, 即 $\pi^T(k)1_m = 1$,

a) 对任意 $i, j \in V$ 和 $k \geq s$, 有

$$\| [A(k:s)]_{ij} - \pi_j(k) \| \leq C\xi^{k-s},$$

其中常数 $C > 0, 0 < \xi < 1$.

b) 对任意 $i \in V$ 和 $k \geq 0$, 有 $\pi_i(k) \geq \eta$ 成立, 其中 $\eta > 0$.

c) $\pi^T(k) = \pi^T(k+1)A(k)$.

智能体在传递信息之前必须将模拟信号转换成数字信号, 而这就必须涉及到量化器的设计和使用. 最基本的量化操作就是取整, 一般称之为均匀量化. 这种量化器易于实现和分析, 但遗憾的是如果要量化的信息具有非常大的幅值, 则对量化结果进行编码就需要相当多的比特数. 这将会给通讯网络带来巨大的负载. 许多研究人员提出了新的量化器形式, 例如抖动量化器, 对数量化器等, 这些量化操作相比于基础的均匀量化来说均有不同方面性能的提升, 但这种静态的量化方案所产生的量化误差是从始至终都存在的. 文献[20]提出了一种动态量化的方案, 通过在量化编译中引入动态的伸缩函数 $\beta(k)$, 随着智能体状态的变化动态地调整量化信号, 以此逐步消除量化对信号传输的影响.

传统的分布式优化算法通常会固定采样周期, 在每个周期开始的时候, 传感器会对智能体的状态进行采样量化并发送出去, 但这种模式会造成不必要的网络资源的浪费. 一个最明智的方法是通过事件触发更新. 其主要的思想是让智能体在必要时刻才进行信息传输, 更确切地说, 只有当系统的某一项指标超过了阈值, 通信才会发生. 从这个意义上讲, 事件触发策略比传统的固定采样周期的策略在减少不必要的通讯方面更具有优势.

下面本文结合动态量化编译和事件触发机制, 提出基于事件触发机制的分布式量化随机无梯度投影算法解决不平衡图上的分布式优化问题.

3 基于事件触发的分布式量化随机无梯度算法

3.1 编译码量化规则和事件触发设置

由于通信带宽的限制, 在将信息发送之前, 智能体必须对信息进行量化. 对于向量 $x_i = (x_i^1, x_i^2, \dots, x_i^n)^T \in \mathbb{R}^n$, 定义一个 n 维均匀量化器 $Q_K(x_i)$:

$$Q_K(x_i) = [q_K(x_i^1) \ q_K(x_i^2) \ \cdots \ q_K(x_i^n)]^T.$$

对 x_i 的任意分量 $x_i^p (p=1, 2, \dots, n)$, 量化函数 $q_K(\cdot)$:

$R \rightarrow \Lambda = \{0, -K, -K+1, \dots, 0, 1, \dots, K\}$ 的定义如下:

$$q_K(x_i^p) = \begin{cases} 0, & -\frac{1}{2} \leq x_i^p < \frac{1}{2}, \\ l, & \frac{2l-1}{2} \leq x_i^p < \frac{2l+1}{2}, \\ K, & x_i^p \geq \frac{2K+1}{2}, \\ -q_K(-x_i^p), & x_i^p < -\frac{1}{2}, \end{cases}$$

其中 $l = 1, 2, \dots, K$. 其作用是将 x_i^p 转换为一个整数, 从而降低信息传输所需要的比特数. 例如 $x_i^p = \frac{1}{2}$, 则

$$Q_K(x_i^p) = 1, \quad x_i^p = \frac{1}{4}, \quad \text{则 } Q_K(x_i^p) = 0. \quad \text{如果 } \|x\|_\infty \leq$$

$K + \frac{1}{2}$, 本文称 n 维均匀量化器 $Q_K(x)$ 不饱和, 在这种情况下有量化误差 $\|x - Q_K(x)\|_\infty \leq \frac{1}{2}$. 更多的细节

可见文献[39]第5章.

上述静态均匀量化器的量化水平为 $2K+1$, 所需的比特数为 $n \log_2(2K+1)$. 可以通过降低量化水平减少信息传输所需的比特数, 从而减少信息传输所需的带宽, 但量化水平过小会导致量化器饱和从而算法不收敛. 为此, 本文采用与文献[20, 25]类似的动态编译码方案替代传统的静态均匀量化器, 降低量化水平的同时确保量化器不饱和.

在 k 时刻, 有向通信通道 (i, j) 上信息发送端智能体 i 的编码器设计如下:

$$\begin{cases} \xi_{ij}(0) = 0, \\ s_{ij}(k) = Q_K\left(\frac{1}{\beta(k-1)}(x_i(k) - \xi_{ij}(k_{ij}))\right), \\ \xi_{ij}(k) = \beta(k-1)I_{ij}(k)s_{ij}(k) + \xi_{ij}(k_{ij}), \end{cases}$$

其中: $\xi_{ij}(k)$ 是编码器的内部变量; $x_i(k)$ 和 $s_{ij}(k)$ 分别为编码器的输入和输出; k_{ij} 表示智能体 i 在 k 时刻之前的最新事件触发时刻; $\beta(k-1)$ 是伸缩函数, 用于放大缩小信号 $x_i(k) - \xi(k_{ij})$; $I_{ij}(k)$ 为示性函数,

$$I_{ij}(k) = \begin{cases} 1, & a_{ji}(k) > 0, \\ 0, & a_{ji}(k) = 0. \end{cases}$$

信息接收端智能体 j 的解码器设计如下:

$$\begin{cases} \hat{x}_{ij}(0) = 0, \\ \hat{x}_{ij}(k) = \beta(k-1)I_{ij}(k)s_{ij}(k_{ij}) + \hat{x}_{ij}(k_{ij}). \end{cases}$$

因为信息发送端的 $\xi_{ij}(k)$ 和信息接收端的 $\hat{x}_{ij}(k)$ 有相同的更新方程, 所以在事件触发时刻 k_{ij} 有

$$\xi(k_{ij}) = \hat{x}_{ij}(k_{ij}). \quad (3)$$

注 4 静态的均匀编码方式是直接对状态值 $x_i(k)$ 进行量化编码, 但当 $x_i(k)$ 具有较大的幅值, 就需要相当多的比特数. 而本文知道在 $x_i(k)$ 收敛的情况下, $x_i(k) - \hat{x}_{ij}(k_{ij})$ 会收

敛到零, 即 $x_i(k) - \xi_{ij}(k_{ij})$ 也会收敛到零, 文献[20, 25]通过在量化编译中引入动态的伸缩函数 $\beta(k)$, 使得 $s_{ij}(k)$ 的编码范围 $\frac{1}{\beta(k-1)}(x_i(k) - \xi_{ij}(k_{ij}))$ 变化不会太大, 从而有效的降低通信带宽.

定义 k 时刻在通信通道 (i, j) 上量化器的量化误差 $\Delta_{ij}(k)$,

$$\Delta_{ij}(k) = Q_K\left(\frac{1}{\beta(k-1)}(x_i(k) - \xi_{ij}(k_{ij}))\right) - \frac{1}{\beta(k-1)}(x_i(k) - \xi_{ij}(k_{ij})).$$

当 $a_{ji}(k) > 0$, 即 $I_{ij}(k) = 1$ 时, 定义 $\xi_{ij}(k)$ 和 $x_i(k)$ 之间的误差 $\hat{e}_{ij}(k)$:

$$\begin{aligned} \hat{e}_{ij}(k) &= \xi_{ij}(k) - x_i(k) = \\ &= \beta(k-1)I_{ij}(k)s_{ij}(k) + \xi_{ij}(k_{ij}) - x_i(k) = \\ &= \beta(k-1)\left(s_{ij}(k) - \frac{x_i(k) - \xi_{ij}(k_{ij})}{\beta(k-1)}\right) = \\ &= \beta(k-1)\Delta_{ij}(k). \end{aligned} \quad (4)$$

下面考虑事件触发机制的设置. 智能体 i 在 k_{ij} 时刻将发送量化器的输出值 $s_{ij}(k_{ij})$ 给邻居智能体 j , 记 k_i^l 是由智能体 i 的一系列事件触发时刻组成的数列. 在不失一般性的前提下, 假设 $0 \in k_i^l, \forall i \in V$. 令 $\tilde{x}_{ij}(k)$ 为智能体 j 在最新事件触发时间接收到智能体 i 的量化值 $\hat{x}_{ij}(k_{ij})$, 由式(3)有

$$\tilde{x}_{ij}(k) = \begin{cases} \xi_{ij}(k), & k \in k_i^l, \\ \tilde{x}_{ij}(k-1), & \text{其他.} \end{cases}$$

当 $\|\xi_{ij}(k) - \xi_{ij}(k_{ij})\| > \omega_{ij}(k)$ 时, 选择 k 作为事件触发时刻, $\omega_{ij}(k)$ 是触发条件的正阈值. 由触发条件的设置有 $\|\xi_{ij}(k) - \tilde{x}_{ij}(k)\| < \omega_{ij}(k)$. 在这篇文章中, 设定 $\omega_{ij}(k) = \beta(k)\tilde{H}$, \tilde{H} 为任意常数, 其大小决定了触发频率从而影响收敛速度. 定义事件触发机制对状态值的传输所带来的误差为 $\tilde{e}_{ij}(k) = \tilde{x}_{ij}(k) - \xi_{ij}(k)$, 有

$$\|\tilde{e}_{ij}(k)\| = \|\tilde{x}_{ij}(k) - \xi_{ij}(k)\| \leq \beta(k)\tilde{H}. \quad (5)$$

假设 3 伸缩函数 $\beta(k)$ 需满足 $\sum_{k=0}^{\infty} \beta(k) < \infty$.

3.2 基于事件触发的分布式量化随机无梯度投影算法

为解决分布式优化问题(1), 本文提出如下的基于事件触发的分布式量化随机无梯度投影算法:

$$\begin{cases} \tilde{v}_i(k) = \sum_{j=1}^m a_{ij}(k)\tilde{x}_{ji}(k), \\ x_i(k+1) = P_X(\tilde{v}_i(k) - g(k)\frac{\tilde{\nabla} f_{i,\mu_i}(\tilde{v}_i(k))}{z_i(k)}), \\ z_i(k+1) = \sum_{j=1}^m a_{ij}(k)z_j(k), \end{cases} \quad (6)$$

其中: $\tilde{x}_{ji}(k)$ 是在最新触发时刻智能体 i 接收到智能体 j 的状态值, 任意选取初值 $\tilde{x}_{ji}(0), \forall j \in V; g(k) > 0$ 是单调递减的步长; P_X 是非空闭凸集 X 的投影算子; $\tilde{\nabla} f_{i,\mu_i}(\tilde{v}_i(k))$ 是 $f_{i,\mu_i}(x)$ 在 $\tilde{v}_i(k)$ 处的随机无梯度; $z_i(k)$ 是引入的辅助变量, 初值为 $z_i(0) = 1, \forall i \in V$. 为简化起见, 假设算法(6)生成的序列只有一个维度, 即 $n = 1$. 对于 $x_i(k) \in \mathbb{R}^n$ 的情况, 借助 Kronecker 乘积, 下面的分析方法将类似. 改写式(6)得

$$\begin{cases} x_i(k+1) = v_i(k) + \varepsilon_i(k), \\ \varepsilon_i(k) = P_X(v_i(k) + \tilde{E}_i(k) + \hat{E}_i(k) - \\ g(k)\frac{\tilde{\nabla} f_{i,\mu_i}(\tilde{v}_i(k))}{z_i(k)}) - v_i(k), \end{cases} \quad (7)$$

其中: $v_i(k) = \sum_{j=1}^m a_{ij}(k)x_j(k), \tilde{E}_i(k) = \sum_{j=1}^m a_{ij}(k)\tilde{e}_{ji}(k),$

$$\hat{E}_i(k) = \sum_{j=1}^m a_{ij}(k)\hat{e}_{ji}(k).$$

假设 4 对步长作如下假设:

$$\sum_{k=0}^{\infty} g(k) = \infty, \quad \sum_{k=0}^{\infty} g^2(k) < \infty, \quad (8)$$

并且存在常数 $0 < \gamma < 1$ 使得

$$\frac{g(k+1)}{g(k)} \geq \gamma > \xi. \quad (9)$$

注 5 式(8)是分布式优化算法典型的步长约束, 一般可取 $g(k) = \frac{1}{k}$. 此时, 当 $k \rightarrow \infty, \frac{g(k+1)}{g(k)} \rightarrow 1$, 可知式(9)也不难满足, 类似的步长约束出现在文献[20, 24]中.

算法的详细流程如下:

初始化: 生成初始状态变量 $\tilde{x}(0), \hat{x}_{ij}(0), \xi_{ij}(0)$, 常数 \tilde{H} , 选择伸缩函数 $\beta(k)$, 阈值函数 $\omega(k) = \beta(k)\tilde{H}$ 和步长 $g(k)$. 利用式(6)更新状态 $x_i(0)$.

迭代: 在第 k 步,

S1) 信息发送端智能体 i 的编码器更新:

$$\begin{cases} s_{ij}(k) = Q_K\left(\frac{1}{\beta(k-1)}(x_i(k) - \xi_{ij}(k_{ij}))\right), \\ \xi_{ij}(k) = \beta(k-1)I_{ij}(k)s_{ij}(k) + \xi_{ij}(k_{ij}). \end{cases}$$

S2) 信息接收端智能体 j 的解码器更新:

$$\hat{x}_{ij}(k) = \beta(k-1)I_{ij}(k)s_{ij}(k_{ij}) + \hat{x}_{ij}(k_{ij}).$$

S3) 判断 k 是否是事件触发时刻: 若

$$\|\xi_{ij}(k) - \xi_{ij}(k_{ij})\| > \omega_{ij}(k),$$

更新变量 $\tilde{x}_j(k) = \hat{x}(k)$, 并更新 k_{ij} . 否则

$$\tilde{x}_{ij}(k) = \tilde{x}_{ij}(k-1).$$

S4) 利用式(6)进行状态更新.

4 收敛性分析和量化水平估计

为证明所提出算法的收敛性, 本文先给出下面4个

引理.

引理 3^[40] 投影的非扩张性.

$$\|P_X(y) - P_X(x)\| \leq \|y - x\|, \forall x, y \in \mathbb{R}^m. \quad (10)$$

引理 4 卷积上界公式:

$$\sum_{l=0}^k \xi^{k-l} g(l) \leq \frac{\gamma}{\gamma - \xi} g(k). \quad (11)$$

证 该引理的证明与文献[24]中的引理4证明类

似. 记 $\omega(k) = \sum_{l=0}^k \xi^{k-l} \left(\frac{g(l)}{g(k)}\right)$, 由假设4有

$$\begin{aligned} \omega(k+1) &= \sum_{r=0}^{k+1} \xi^{k-l+1} \frac{g(r)}{g(k+1)} = \\ &= \sum_{r=0}^k \xi^{k-l+1} \frac{g(r)}{g(k+1)} + 1 \leq \\ &= \frac{\xi}{\gamma} \omega(k) + 1. \end{aligned}$$

上式进行迭代有

$$w(k+1) \leq \sum_{\tau=0}^{k+1} \left(\frac{\xi}{\gamma}\right)^\tau \leq \frac{\gamma}{\gamma - \xi},$$

由 $w(k)$ 的定义可知不等式(11)成立. 证毕.

引理 5^[41] 设 $\{\iota(k)\}, \{\chi(k)\}$ 和 $\{\psi(k)\}$ 是非负的随机序列, $\{\zeta(k)\}$ 为确定性的非负序列. 记 $\mathcal{F}(k)$ 是通过 $\iota(1), \dots, \iota(k), \chi(1), \dots, \chi(k), \psi(1), \dots, \psi(k)$ 产生的 σ -代数, 若有

$$1) \sum_{k=0}^{\infty} \zeta(k) < \infty \text{ a.s.}, \sum_{k=0}^{\infty} \psi(k) < \infty \text{ a.s.};$$

$$2) E[\iota(k+1)|\mathcal{F}(k)] \leq (1 + \zeta(k))\iota(k) - \chi(k) + \psi(k).$$

那么存在一个非负变量 ι , 使得

$$\lim_{k \rightarrow \infty} \iota(k) = \iota \text{ a.s. 且 } \sum_{k=0}^{\infty} \chi(k) < \infty.$$

引理 6 在量化器不饱和的条件下有

$$\begin{aligned} E[\|x_i(k+1) - \bar{x}(k+1)\| | F(k)] &\leq \\ C\xi^k \sum_{j=1}^m \|\delta_j(0)\| + CQ\gamma \frac{g(k)}{\gamma - \xi}, \end{aligned}$$

其中:

$$\bar{x}(k) = \sum_{i=1}^m \pi_i(k)x_i(k) = \pi^T(k)x(k),$$

$$\delta(k) = x(k) - 1_m \bar{x}(k),$$

$$Q = \max_k \frac{\beta(k)}{g(k)} (2m\tilde{H} + m) + C_z C_{\nabla f},$$

$$C_z = \sup_{l>0} \max_i \frac{1}{z_i(l)}.$$

证 记

$$x(k) = [x_1(k) \ x_2(k) \ \dots \ x_m(k)]^T,$$

$$\varepsilon(k) = [\varepsilon_1(k) \ \varepsilon_2(k) \ \dots \ \varepsilon_m(k)]^T,$$

那么式(7)可以写成紧凑的向量形式,

$$x(k+1) = A(k)x(k) + \varepsilon(k). \quad (12)$$

令 $\bar{\varepsilon}(k) = 1_m \pi^T(k+1)\varepsilon(k)$, 由

$$\pi^T(k) = \pi^T(k+1)A(k),$$

等式(12)两边同时乘以 $1_m \pi^T(k+1)$ 得

$$1_m \bar{x}(k+1) = 1_m \bar{x}(k) + \bar{\varepsilon}(k). \quad (13)$$

由式(12)减式(13)得

$$\delta(k+1) = (A(k) - 1_m \pi^T(k))x(k) + \varepsilon(k) - \bar{\varepsilon}(k). \quad (14)$$

由

$$\begin{aligned} (A(k) - 1_m \pi^T(k))1_m \bar{x}(k) &= \\ (A(k)1_m - 1_m \pi^T(k)1_m)\bar{x}(k) &= 0, \end{aligned}$$

式(14)等价于

$$\delta(k+1) = (A(k) - 1_m \pi^T(k))\delta(k) + \varepsilon(k) - \bar{\varepsilon}(k). \quad (15)$$

由引理2(c)有

$$\prod_{l=k}^s (A(l) - 1_m \pi^T(l)) = A(s:k) - 1_m \pi^T(k),$$

通过数学归纳法从式(15)可得

$$\begin{aligned} \delta(k+1) &= (A(k:0) - 1_m \pi^T(0))\delta(0) + \\ &= \sum_{l=0}^k (A(k:l+1) - 1_m \pi^T(l+1)) \times \\ &= (\varepsilon(l) - \bar{\varepsilon}(l)). \end{aligned} \quad (16)$$

将式(16)写成标量形式,

$$\begin{aligned} \delta_i(k+1) &= \sum_{j=1}^m (A(k:0) - 1_m \pi^T(0))_{ij} \delta_j(0) + \\ &= \sum_{j=1}^m \sum_{l=0}^k (A(k:l+1) - 1_m \pi^T(l+1))_{ij} \times \\ &= (\varepsilon(l) - \bar{\varepsilon}(l))_j. \end{aligned} \quad (17)$$

由引理2(a)和式(17)可得

$$\begin{aligned} \|\delta_i(k+1)\| &\leq \\ C\xi^k \sum_{j=1}^m \|\delta_j(0)\| + mC \sum_{l=0}^k \xi^{k-l} \|\varepsilon(l) - \bar{\varepsilon}(l)\|_{\infty}. \end{aligned} \quad (18)$$

令

$$\tilde{e}(l) = \max_i \max_j \|\tilde{e}_{ji}(l)\|, \hat{e}(l) = \max_i \max_j \|\hat{e}_{ji}(l)\|,$$

由 $\|I - 1_m \pi^T(l)\|_{\infty} \leq 2$, 结合 $\varepsilon(l)$ 和 $\bar{\varepsilon}(l)$ 的定义和引理1有

$$\begin{aligned} E[\|\varepsilon(l) - \bar{\varepsilon}(l)\|_{\infty} | F(k)] &\leq \\ E[\|(I - 1_m \pi^T(l))\varepsilon(l)\|_{\infty} | F(k)] &\leq \\ \|I - 1_m \pi^T(l)\|_{\infty} \max_i E[\|\tilde{E}_i(l) + \end{aligned}$$

$$\begin{aligned} \hat{E}_i(l) - g(l) \frac{\tilde{\nabla} f_{i,\mu_i}(\tilde{v}_i(l))}{z_i(l)} \|F(k)\| \leq \\ 2(\|\tilde{e}(l)\| + \|\hat{e}(l)\|) + 2g(l)C_z C_{\nabla f}. \end{aligned} \quad (19)$$

由式(4)–(5)和假设 $\Delta(l) \leq \frac{1}{2}$, 结合式(18)–(19)可得

$$\begin{aligned} E[\|\delta_i(k+1)\| | F(k)] \leq \\ C\xi^k \sum_{j=1}^m \|\delta_j(0)\| + \\ 2mC \sum_{l=0}^k \xi^{k-l} (\beta(l)(\tilde{H} + \Delta(l)) + g(l)C_z C_{\nabla f}) \leq \\ C\xi^k \sum_{j=1}^m \|\delta_j(0)\| + C \sum_{l=0}^k \xi^{k-l} g(l)Q \leq \\ C\xi^k \sum_{j=1}^m \|\delta_j(0)\| + CQ\gamma \frac{g(k)}{\gamma - \xi}. \end{aligned} \quad (20)$$

最后一个不等号是由引理4得出. 证毕.

下面是本文的主要结果. 在量化器不饱和的条件下, 定理1和定理2证明了所提算法的收敛性, 定理3给出了算法的收敛速率, 定理4求出了使得量化器不饱和的最小的量化水平.

定理 1 在量化器不饱和及假设1–4成立的条件下, 迭代序列 $x_i(k)$ 几乎必然收敛到问题(2)的最优解. 即存在 $x_\mu^* \in X$ 使得

$$\lim_{k \rightarrow \infty} x_i(k) = x_\mu^* \text{ a.s.}$$

证 记 $\tilde{E}(k) = \max_i \tilde{E}_i(k)$, $\hat{E}(k) = \max_i \hat{E}_i(k)$, 由投影的非扩张性有

$$\begin{aligned} E[\|x_i(k+1) - x_\mu^*\|^2 | F(k)] \leq \\ E[\|v_i(k) + \tilde{E}_i(k) + \hat{E}_i(k) - \\ g(k) \frac{\tilde{\nabla} f_{i,\mu_i}(\tilde{v}_i(k))}{z_i(k)} - x_\mu^*\|^2 | F(k)] \leq \\ E[\|v_i(k) - g(k) \frac{\tilde{\nabla} f_{i,\mu_i}(\tilde{v}_i(k))}{z_i(k)} - x_\mu^*\|^2 | F(k)] + \\ \|\tilde{E}(k) + \hat{E}(k)\|^2 + 2\|\tilde{E}(k) + \hat{E}(k)\| \times \\ E[\|v_i(k) - g(k) \frac{\tilde{\nabla} f_{i,\mu_i}(\tilde{v}_i(k))}{z_i(k)} - x_\mu^*\| | F(k)] := \\ d_1 + d_2 + d_3. \end{aligned} \quad (21)$$

下面估计式(21)右端 d_1, d_2, d_3 的上界.

记 $L = \max_i L(f_{i,\mu_i})$, $\tilde{e}(k) = \max_i \max_j \|\tilde{e}_{ij}(k)\|$, $\hat{e}(k) = \max_i \max_j \|\hat{e}_{ij}(k)\|$,

$$\begin{aligned} -2g(k) \frac{E[\tilde{\nabla} f_{i,\mu_i}(\tilde{v}_i(k)) | F(k)]}{z_i(k)} (v_i(k) - x_\mu^*) = \\ -2g(k) \frac{\nabla f_{i,\mu_i}(\tilde{v}_i(k))}{z_i(k)} (\tilde{v}_i(k) - x_\mu^* + v_i(k) - \tilde{v}_i(k)) \leq \\ 2 \frac{g(k)}{z_i(k)} (|f_{i,\mu_i}(\tilde{v}_i(k)) - f_{i,\mu_i}(\bar{x}(k))| - (f_{i,\mu_i}(\bar{x}(k)) - \end{aligned}$$

$$\begin{aligned} f_{i,\mu_i}(x_\mu^*)) + \nabla f_{i,\mu_i}(\tilde{v}_i(k)) \|\tilde{E}(k) + \hat{E}(k)\| \leq \\ 2 \frac{g(k)}{z_i(k)} (L \|\tilde{v}_i(k) - \bar{x}(k)\| - (f_{i,\mu_i}(\bar{x}(k)) - \\ f_{i,\mu_i}(x_\mu^*)) + \nabla f_{i,\mu_i}(\tilde{v}_i(k)) \|\tilde{E}(k) + \hat{E}(k)\| \leq \\ 2 \frac{g(k)}{z_i(k)} (L \sum_{j=1}^m a_{ij}(k) \|x_j(k) - \bar{x}(k)\| - \\ (f_{i,\mu_i}(\bar{x}(k)) - f_{i,\mu_i}(x_\mu^*)) + (\nabla f_{i,\mu_i}(\tilde{v}_i(k)) + L) \times \\ \|\tilde{e}(k) + \hat{e}(k)\|). \end{aligned} \quad (22)$$

倒数第3个不等式是由于 $f_{i,\mu_i}(x)$ 的凸性, 最后一个不等式是由于范数的凸性和权重矩阵 $A(k)$ 是行随机的, 结合 d_1 的定义和式(22)有

$$\begin{aligned} d_1 \leq \sum_{j=1}^m a_{ij}(k) \|x_j(k) - x_\mu^*\|^2 + \frac{g^2(k)}{z_i^2(k)} C_{\nabla f}^2 + \\ 2 \frac{g(k)}{z_i(k)} (L \sum_{j=1}^m a_{ij}(k) \|x_j(k) - \bar{x}(k)\| - \\ (f_{i,\mu_i}(\bar{x}(k)) - f_{i,\mu_i}(x_\mu^*)) + \\ (\nabla f_{i,\mu_i}(\tilde{v}_i(k)) + L) \|\tilde{e}(k) + \hat{e}(k)\|). \end{aligned} \quad (23)$$

由引理1(c)和假设量化误差有界 $\Delta(k) \leq \frac{1}{2}$ 有

$$\begin{aligned} d_1 \leq \sum_{j=1}^m a_{ij}(k) \|x_j(k) - x_\mu^*\|^2 + \frac{g^2(k)}{z_i^2(k)} C_{\nabla f}^2 + \\ 2 \frac{g(k)}{z_i(k)} (LC\xi^k \sum_{j=1}^m \|\delta_j(0)\| + LCQ\gamma \frac{g(k)}{\gamma - \xi} - \\ (f_{i,\mu_i}(\bar{x}(k)) - f_{i,\mu_i}(x_\mu^*)) + \\ (C_{\nabla f} + L)\beta(k)(\tilde{H} + \frac{1}{2})). \end{aligned} \quad (24)$$

同理有

$$\begin{aligned} d_2 \leq \beta^2(k)(\tilde{H} + \frac{1}{2})^2, \\ d_3 \leq 2\beta(k)(\tilde{H} + \frac{1}{2}) \sum_{j=1}^m a_{ij}(k) \|x_j(k) - x_\mu^*\| + \\ 2 \frac{\beta(k)g(k)}{z_i(k)} (\tilde{H} + \frac{1}{2}) C_{\nabla f}. \end{aligned} \quad (26)$$

由引理2(c)有

$$\begin{aligned} \sum_{i=1}^m \pi_i(k+1) \sum_{i=1}^m a_{ij}(k) \|x_j(k) - x_\mu^*\|^2 = \\ \sum_{i=1}^m \pi_i(k) \|x_i(k) - x_\mu^*\|^2. \end{aligned} \quad (27)$$

在式(21)的两边乘以 $\pi(k+1)$ 并在 $i \in V$ 上求和, 结合式(23)–(27)有

$$\begin{aligned} \sum_{i=1}^m E[\pi_i(k+1) \|x_i(k+1) - x_\mu^*\|^2 | F(k)] \leq \\ \sum_{i=1}^m \pi_i(k) \|x_i(k) - x_\mu^*\|^2 + \frac{g^2(k)}{z_i^2(k)} \sum_{i=1}^m \pi_i(k+1) C_{\nabla f}^2 + \\ \xi^k 2LC \sum_{i=1}^m \pi_i(k+1) \frac{g(k)}{z_i(k)} \sum_{j=1}^m \|\delta_j(0)\| + \end{aligned}$$

$$\begin{aligned}
 &g^2(k) \sum_{i=1}^m \pi_i(k+1) \frac{2LCQ\gamma}{z_i(k)(\gamma-\xi)} + \\
 &\beta(k) \left(\tilde{H} + \frac{1}{2}\right) \sum_{j=1}^m \pi_i(k+1) \left(\frac{g(k)}{z_i(k)} \times \right. \\
 &(4C_{\nabla f} + 2L) + \beta(k) \left(\tilde{H} + \frac{1}{2}\right) + \\
 &2 \sum_{j=1}^m a_{ij}(k) \|x_j(k) - x_\mu^*\|) - 2g(k) \sum_{i=1}^m \frac{\pi_i(k+1)}{z_i(k)} \times \\
 &(f_{i,\mu_i}(\bar{x}(k)) - f_{i,\mu_i}(x_\mu^*)). \tag{28}
 \end{aligned}$$

记

$$\begin{aligned}
 D_1 &= \sup_k \max_i \frac{1}{z_i^2(k)} \sum_{i=1}^m \pi_i(k+1) C_{\nabla f}^2, \\
 D_2 &= \sup_k \max_i 2LC \sum_{i=1}^m \pi_i(k+1) \frac{g(k)}{z_i(k)} L \sum_{j=1}^m \|\delta_j(0)\|, \\
 D_3 &= \sup_k \max_i \sum_{i=1}^m \pi_i(k+1) \frac{2LQ_2}{z_i(k)(1-\xi)}, \\
 D_4 &= \sup_k \max_i \left(\tilde{H} + \frac{1}{2}\right) \sum_{j=1}^m \pi_i(k+1) \times \\
 &\left(\frac{g(k)}{z_i(k)} (4C_{\nabla f} + 2L) + \beta(k) \left(\tilde{H} + \frac{1}{2}\right)\right), \\
 D_5 &= \inf_k \min_i 2 \frac{\pi_i(k+1)}{z_i(k)}.
 \end{aligned}$$

令

$$\begin{aligned}
 \iota(k) &= \sum_{i=1}^m \pi_i(k) \|x_i(k) - x_\mu^*\|^2, \\
 \chi(k) &= D_5 g(k) (f_\mu(\bar{x}(k)) - f_\mu(x_\mu^*)), \zeta(k) = 0, \\
 \psi(k) &= g^2(k) (D_1 + D_3) + \xi^k D_2 + \beta(k) D_4.
 \end{aligned}$$

由引理5有 $\sum_{k=1}^\infty g(k) (f_\mu(\bar{x}(k)) - f_\mu(x_\mu^*)) < \infty$ 几乎必然成立和 $\sum_{i=1}^m \pi_i(k) \|x_i(k) - x_\mu^*\|^2$ 几乎必然收敛. 结合

假设 $\sum_{k=0}^\infty g(k) = \infty$ 有

$$\liminf_{k \rightarrow \infty} f_\mu(\bar{x}(k)) = f_\mu(x_\mu^*), \text{ a.s.} \tag{29}$$

由 $\{\bar{x}(k)\}$ 有界和式(29), 故存在子列 $\{\bar{x}(k_l)\} \subseteq \{\bar{x}(k)\}$, 使得

$$\lim_{k \rightarrow \infty} f_\mu(\bar{x}(k_l)) = f_\mu(x_\mu^*), \text{ a.s.},$$

由函数 $f_\mu(x)$ 的连续性, 故存在 $x_\mu^* \in X^*$, 使得

$$\lim_{k \rightarrow +\infty} \bar{x}(k_l) = x_\mu^* \text{ a.s.}$$

根据引理6, 有 $\lim_{k \rightarrow +\infty} \|x_i(k) - \bar{x}(k)\| = 0$ a.s. 可知 $x_i(k)$ 和 $\bar{x}(k)$ 收敛到同一个最优值. 因此可以推断, $\lim_{k \rightarrow +\infty} \bar{x}(k) = x_\mu^*$ a.s. 及 $\lim_{k \rightarrow +\infty} x_i(k) = x_\mu^*$ a.s.

证毕.

定理 2 在量化器不饱和及假设1-4成立的条件

下, 令 $\tilde{x}(k) = \frac{\sum_{\tau=0}^k g(\tau) \bar{x}(\tau)}{\sum_{\tau=0}^k g(\tau)}$ 则有

$$\begin{aligned}
 &E[f(\tilde{x}(k))] - f(x^*) \leq \\
 &\frac{\iota^*(0) + \sum_{\tau=0}^k \psi(\tau)}{D_5 \sum_{\tau=0}^k g(\tau)} + \sum_{i=1}^m \mu_i \sqrt{n} L_0(f_i),
 \end{aligned}$$

其中 $\iota^*(k) = \sum_{i=1}^m \pi_i(k) \|x_i(k) - x^*\|^2$.

证 类似于定理1的证明过程, 将式(28)中的 x_μ^* 替换为 x^* 不等式依然成立, 将不等式从 k 时刻迭代回0时刻, 可得

$$E[\iota^*(k+1)|F(k)] \leq \iota^*(0) + \sum_{\tau=0}^k \psi(\tau) - \sum_{\tau=0}^k \chi^*(\tau),$$

其中 $\chi^*(k) = D_5 g(k) (f_\mu(\bar{x}(k)) - f_\mu(x^*))$. 对上式两边同时取期望, 由于 $E[E[\iota^*(k+1)|F(k)]] \geq 0$,

$$\sum_{\tau=0}^k E[\chi^*(\tau)] \leq \iota^*(0) + \sum_{\tau=0}^k \psi(\tau),$$

即

$$\begin{aligned}
 &D_5 \sum_{\tau=0}^k g(\tau) (E[f_\mu(\bar{x}(\tau))] - f_\mu(x_\mu^*)) \leq \\
 &\iota^*(0) + \sum_{\tau=0}^k \psi(\tau).
 \end{aligned}$$

上式两端同时除以 $D_5 \sum_{\tau=0}^k g(\tau)$,

$$\begin{aligned}
 &\frac{D_5 \sum_{\tau=0}^k g(\tau) (E[f_\mu(\bar{x}(\tau))] - f_\mu(x_\mu^*))}{D_5 \sum_{\tau=0}^k g(\tau)} \leq \\
 &\frac{\iota^*(0) + \sum_{\tau=0}^k \psi(\tau)}{D_5 \sum_{\tau=0}^k g(\tau)}.
 \end{aligned}$$

利用函数的凸性有

$$E[f_\mu(\tilde{x}(k))] - f_\mu(x^*) \leq \frac{\iota^*(0) + \sum_{\tau=0}^k \psi(\tau)}{D_5 \sum_{\tau=0}^k g(\tau)}. \tag{30}$$

由引理1(a)有 $f(x) \leq f_\mu(x) \leq f(x) + \sum_{i=1}^m \mu_i \sqrt{n} L_0(f_i)$, 结合式(30)定理得证. 证毕.

注 6 由假设3-4可知, 定理2中不等式右边的第1项在 k 趋向 ∞ 时会趋于0, 第2项是由于本文的算法采用了随机无梯度方法而产生的. 由定理2可知算法产生的序列 $E[f(\tilde{x}(k))]$ 能收敛到问题1最优值的邻域, 其半径由 μ_i 和 $L_0(f_i)$ 所决定,

在 $\mu_i \rightarrow 0, \forall i \in V$ 的情况下, 当 $k \rightarrow \infty, E[f(\tilde{x}(k))]$ 收敛到 $f(x^*)$.

定理 3 在 $\mu_i \rightarrow 0, \forall i \in V$ 的情况下, 若步长 $g(k) = \frac{1}{(k+1)^r}$, 算法的收敛速率为 $O(\frac{1}{k^{1-r}})$. 若步长 $g(k) = \frac{1}{k+1}$, 算法的收敛速率为 $O(\frac{1}{\ln k})$.

证 由定理 2 有

$$\lim_{\mu_i \rightarrow 0, i \in V} E[f(\tilde{x}(k))] - f(x^*) \leq \frac{\iota^*(0) + \sum_{\tau=0}^k \psi(\tau)}{D_5 \sum_{\tau=0}^k g(\tau)}. \quad (31)$$

选择步长 $g(k) = \frac{1}{(k+1)^r}, \frac{1}{2} < r \leq 1$, 由

$$\sum_{\tau=0}^k \frac{1}{(\tau+1)^r} \geq \int_0^{k+1} \frac{ds}{(s+1)^r} = \begin{cases} \frac{1}{1-r} ((k+2)^{1-r} - 1), & \frac{1}{2} < r < 1, \\ \ln(k+2), & r = 1. \end{cases}$$

结合式(31)可得

$$\lim_{\mu_i \rightarrow 0, i \in V} E[f(\tilde{x}(k))] - f(x^*) \leq \begin{cases} \frac{(1-r)(\iota^*(0) + \sum_{\tau=0}^k \psi(\tau))}{D_5((k+2)^{1-r} - 1)}, & \frac{1}{2} < r < 1, \\ \frac{\iota^*(0) + \sum_{\tau=0}^k \psi(\tau)}{D_5 \ln(k+2)}, & r = 1, \end{cases}$$

其中 $\iota^*(0) + \sum_{\tau=0}^k \psi(\tau)$ 有界. 证毕.

为使得量化器不饱和可以一直采用较大的量化水平, 但会造成通信资源的浪费. 现在笔者给出随迭代时间变化的量化水平更新规则以节约通信成本.

定理 4 在假设 1-4 成立和随机无梯度有界的条件下, 若 $K(k)$ 满足

$$K(1) \geq \frac{g(0)}{\beta(0)} C_{\nabla f} - \frac{1}{2},$$

$$K(k) \geq \frac{\xi^k}{\beta(k)} (2C \sum_{j=1}^m \|\delta_j(0)\|) + \frac{g(k)}{\beta(k)} (C_z C_{\nabla f} + \frac{2CQ\gamma}{\gamma - \xi}) + \frac{3}{\beta(k)} (\tilde{H} + \frac{1}{2}) - \frac{1}{2},$$

其中: $k = 2, 3, \dots, C_z = \sup_k \max_i \frac{1}{z_i(k)}$, 则量化器 Q_K 不饱和.

证 下面通过数学归纳法证明量化器 Q_K 不饱和, 即量化误差 $\Delta(k_{ij}) \leq \frac{1}{2}, k_{ij} \in k_i^l$. 若 $k = 1 \in k_i^l$ 时, 由

$\xi_{ij}(0) = \hat{x}_{ij}(0) = \tilde{x}_{ij}(0)$ 可得

$$\begin{aligned} & \left\| \frac{x_i(1) - \xi_{ij}(0)}{\beta(0)} \right\| = \\ & \frac{1}{\beta(0)} \left\| P_X \left[\sum_{j=1}^m a_{ij}(0)(\hat{x}_j(0)) - g(0) \frac{\tilde{\nabla} f_{i,\mu_i}(k)}{z_i(k)} \right] - \right. \\ & \left. (\hat{x}_i(0)) \right\| \leq \frac{g(0)}{\beta(0)} C_{\nabla f} \leq \\ & K(1) + \frac{1}{2}. \end{aligned}$$

则有 $\Delta(1) \leq \frac{1}{2}$, 即量化器在 $k = 1$ 时刻不饱和.

假设量化器在 k 之前的触发时刻 k_{ij} 都未饱和, 即 $\Delta(k_{ij}) \leq \frac{1}{2}$, 若有 $k+1 \in k_i^l$, 即 $k+1$ 为事件触发时刻. 由于

$$\begin{aligned} \xi_{ij}(k_{ij}) &= \hat{x}_{ij}(k_{ij}) = \tilde{x}_{ij}(k_{ij}) = \\ & x_i(k_{ij}) + \tilde{e}_{ij}(k_{ij}) + \hat{e}_{ij}(k_{ij}) \end{aligned}$$

有

$$\begin{aligned} & \left\| \frac{x_i(k+1) - \xi_{ij}(k_{ij})}{\beta(k)} \right\| = \\ & \frac{1}{\beta(k)} \left\| P_X \left[\sum_{j=1}^m a_{ij}(k)(x_j(k) + \tilde{e}_{ji}(k) + \right. \right. \\ & \left. \left. \hat{e}_{ji}(k)) - g(k) \frac{\tilde{\nabla} f_{i,\mu_i}(\tilde{v}_i(k))}{z_i(k)} \right] - \right. \\ & \left. (x_i(k_{ij}) + \tilde{e}_{ij}(k_{ij}) + \hat{e}_{ij}(k_{ij})) \right\| \leq \\ & \frac{1}{\beta(k)} (\max_{ij} \|x_j(k) - x_i(k_{ij})\| + \|\tilde{e}_{ji}(k) + \hat{e}_{ji}(k)\| + \\ & g(k) C_z C_{\nabla f} + \|\tilde{e}_{ji}(k_{ij}) + \hat{e}_{ji}(k_{ij})\|) \leq \\ & \frac{1}{\beta(k)} (\max_{ij} (\|x_j(k) - x_i(k)\| + \|x_i(k) - x_i(k_{ij})\|) + \\ & \|\tilde{e}_{ji}(k) + \hat{e}_{ji}(k)\| + g(k) C_z C_{\nabla f} + \\ & \|\tilde{e}_{ji}(k_{ij}) + \hat{e}_{ji}(k_{ij})\|) \leq \\ & \frac{1}{\beta(k)} (\max_{ij} (\|x_j(k) - \bar{x}(k)\| + \|x_i(k) - \bar{x}(k)\|) + \\ & 3(\|\tilde{e}(k) + \hat{e}(k)\|) + g(k) C_z C_{\nabla f}). \end{aligned}$$

结合引理 6 有

$$\begin{aligned} & \left\| \frac{x_i(k+1) - \xi_{ij}(k_{ij})}{\beta(k)} \right\| \leq \\ & \frac{\xi^k}{\beta(k)} (2C \sum_{j=1}^m \|\delta_j(0)\|) + \frac{3}{\beta(k)} (\tilde{H} + \frac{1}{2}) + \\ & \frac{g(k)}{\beta(k)} (C_z C_{\nabla f} + \frac{2CQ\gamma}{\gamma - \xi}) \leq \\ & K(k) + \frac{1}{2}, \end{aligned}$$

故有

$$\|\Delta_{ij}(k+1)\| \leq \frac{1}{2},$$

即量化器在 $k + 1$ 时刻不饱和, 故定理2成立.

5 数值仿真

考虑由10个智能体组成的多智能体系统:

$$\begin{aligned} \min_x f(x) &= \sum_{i=1}^{10} f_i(x), \\ \text{s.t. } x &\in X. \end{aligned}$$

假设 $x \in \mathbb{R}, X = [-2, 2]$. 智能体 i 的局部目标函数如下:

$$f_i(x) = \sum_{j=1}^{p_i} \frac{1}{2} \|a_{ij}x - b_{ij}\| + \lambda \|x\|_\infty,$$

其中: $p_i = 10, \forall i \in V, \lambda = 10, a_{ij}$ 和 b_{ij} 都是按照正态分布随机生成的. 通信网络按照图1给出了时变的一般不平衡有向图连接, 其中任何单个图都不连通的, 但它们的联合图是连通的. 假设在整个迭代过程中, 图 $G(k)$ 在图1中的两个图上连续变化.

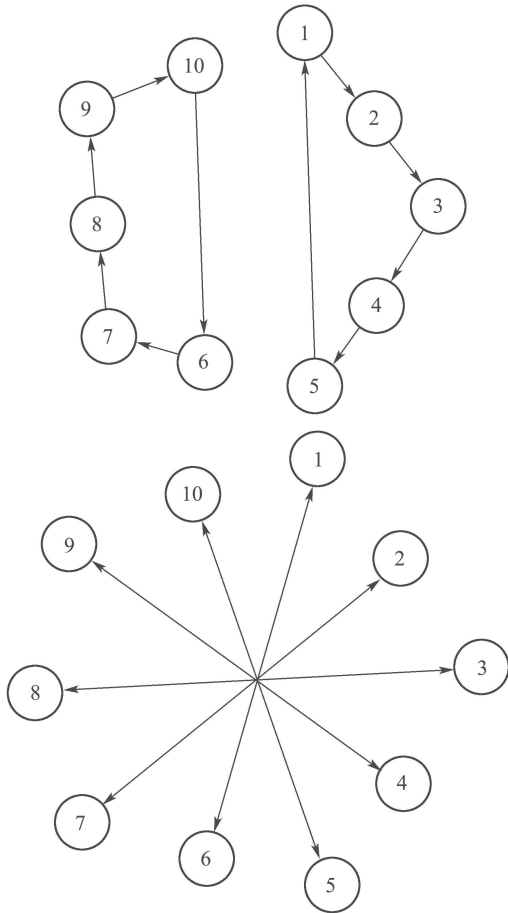


图1 时变联合强连通有向图

Fig. 1 Time-varying joint strongly connected directed graph

由引理1的结论

$$f_i(x) \leq f_{i,\mu_i}(x) \leq f_i(x) + \sqrt{n}\mu_i L_0(f_i),$$

给定 $f_{i,\mu_i}(x) - f_i(x) \leq \epsilon$, 有 $\mu_i \leq \frac{\epsilon}{\sqrt{n}L_0(f_i)}$. 在仿真

例子中, $L_0(f_i) = \frac{a_i}{2} + \lambda$, 其中: $\lambda = 10, n = 10, a_i$ 服

从正态分布. 令 $\epsilon = 0.1, L_0 \approx \lambda$, 有 $\mu_i \leq 0.025$. 令所有的 $f_{i,\mu_i}(x)$ 的光滑参数 $\mu_i = 0.05, \forall i \in V$, 量化级数 $K = 100$, 触发条件中常数 $\tilde{H} = 5$, 选取收缩函数为 $\beta = \frac{1}{k^{1.1}}$, 步长为 $g(k) = \frac{1}{k}$.

将本文算法与分布式gradient-push, 原有的不带事件触发和量化机制的随机无梯度投影方法在相同参数设置下进行对比. 图2比较了3种算法误差 $f(x) - f(x^*)$ 的收敛情况, 结果表明3种算法都能收敛到最优解.

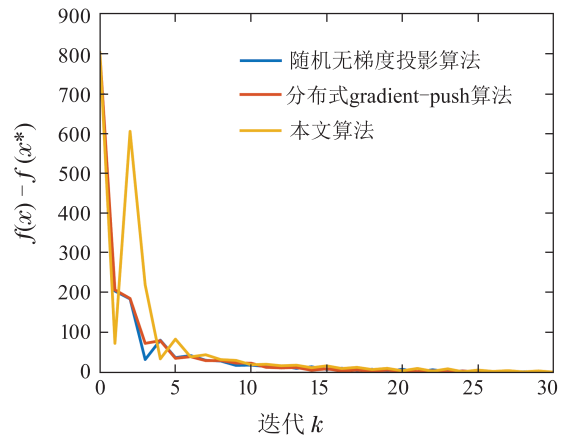


图2 误差 $f(x) - f(x^*)$ 的收敛情况

Fig. 2 Convergence of error $f(x) - f(x^*)$

相比于gradient-push算法, 虽然gradient-push算法有更快的收敛速率, 如图3所示, 但带投影算子的gradient-push的收敛性尚未得到证明, 且算法要求权重矩阵是列随机矩阵, 这在构造权重矩阵时要求每个智能体知道节点的出度.

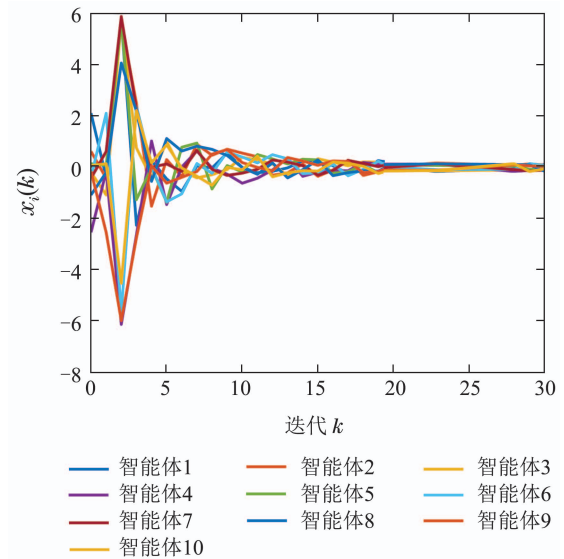


图3 Gradient-push算法状态值 $x_i(k)$ 的收敛情况

Fig. 3 Convergence of state value $x_i(k)$ of Gradient-push algorithm

而本文的算法是针对约束优化问题, 且仅要求权重矩阵是行随机矩阵, 在基于广播的通信机制中具有更好的实用性. 相比于原有的不带事件触发和量化机制的随机无梯度投影方法, 两种算法的收敛速率变化不大, 如图4-5所示. 但本文的算法不要求智能体准确地接收到邻居的状态信息, 且应用了量化通信和事件触发机制, 降低了对通信网络的带宽要求且减少了通信次数, 如图6所示, 其中*为事件触发时刻. 图7显示了不同参数 μ_i 对状态值收敛情况的影响, 可以看出, 选取不同 μ_i 值, 虽然不会影响到智能体的收敛速率, 但 μ_i 过大时, 会影响到状态值收敛的精确度.

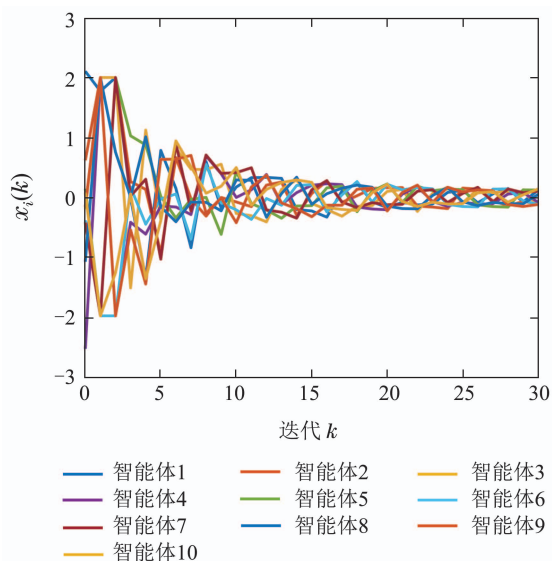


图 4 Gradient-free projection算法状态值 $x_i(k)$ 的收敛情况
Fig. 4 Convergence of state value $x_i(k)$ of Gradient-free projection algorithm

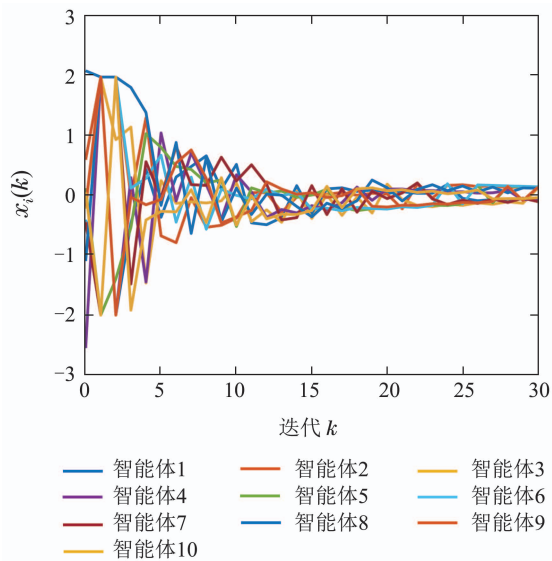


图 5 本文算法状态值 $x_i(k)$ 的收敛情况
Fig. 5 Convergence of state value $x_i(k)$ of the algorithm in this paper

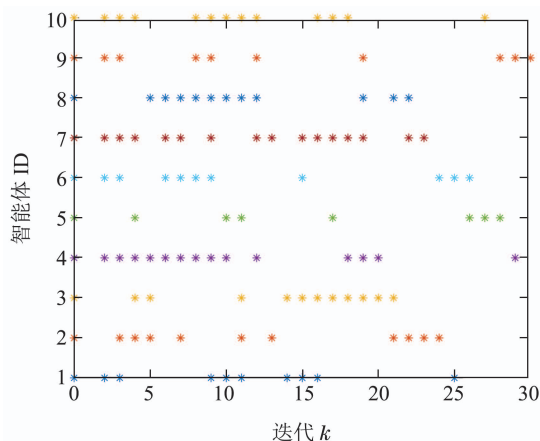


图 6 事件触发时刻数列 k_i^l
Fig. 6 Event trigger time sequence k_i^l

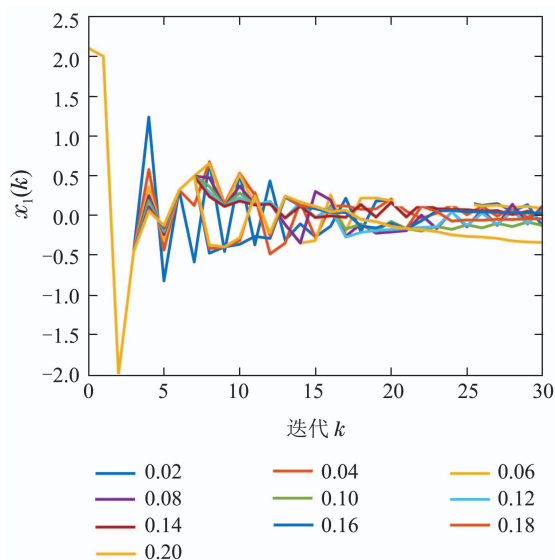


图 7 不同参数 μ 对智能体1的状态收敛情况的影响
Fig. 7 The influence of different parameters μ on the state convergence of agent 1

6 结论

本文结合编译码动态量化规则和事件触发机制提出了一种基于事件触发的分布式量化随机无梯度投影算法, 用于解决分布式约束优化问题. 所提出算法适用于一般行随机的时变不平衡有向图, 这在基于广播的通信机制中具有更好的实用性. 本文证明了当目标函数为凸且Lipschitz连续, 在选取合适的量化水平后, 所有的智能体的状态值收敛到其高斯光滑近似函数的最优解, 序列 $E[f(\tilde{x}(k))]$ 能收敛到问题1最优值的邻域, 并给出了选择不同步长时算法的收敛速率. 数值仿真验证了所提算法的收敛性, 表明了合理的设置事件触发机制和量化通信的规则后, 降低了对通信网络的带宽的要求, 更有效地利用通信资源, 同时不会降低原有分布式随机无梯度算法的收敛速率.

参考文献:

- [1] YI Peng, HONG Yiguang. Distributed cooperative optimization and its applications. *Science China-mathematics*, 2016, 46(10): 1547 – 1564.
(衣鹏, 洪奕光. 分布式合作优化及其应用. *中国科学: 数学*, 2016, 46(10): 1547 – 1564.)
- [2] HONG Yiguang, ZHANG Yanqiong. Distributed optimization: Algorithm design and convergence analysis. *Control Theory & Applications*, 2014, 31(7): 850 – 857.
(洪奕光, 张艳琼. 分布式优化: 算法设计和收敛性分析. *控制理论与应用*, 2014, 31(7): 850 – 857.)
- [3] XIE Pei, YOU Keyou, HONG Yiguang, et al. A survey of distributed convex optimization algorithms over networks. *Control Theory & Applications*, 2018, 35(7): 918 – 927.
(谢佩, 游科友, 洪奕光, 等. 网络化分布式凸优化算法研究进展. *控制理论与应用*, 2018, 35(7): 918 – 927.)
- [4] WANG Long, LU Kaihong, GUAN Yongqiang. Distributed optimization via multi-agent systems. *Control Theory & Applications*, 2019, 36(11): 1820 – 1883.
(王龙, 卢开红, 关永强. 分布式优化的多智能体方法. *控制理论与应用*, 2019, 36(11): 1820 – 1883.)
- [5] YANG Tao, CHAI Tianyou. Research status and prospects of distributed collaborative optimization. *Science China-mathematics*, 2020, 50(11): 1414 – 1425.
(杨涛, 柴天佑. 分布式协同优化的研究现状与展望. *中国科学: 技术科学*, 2020, 50(11): 1414 – 1425.)
- [6] YANG T, YI X L, WU J F, et al. A survey of distributed optimization. *Annual Reviews in Control*, 2019, 47(1): 278 – 305.
- [7] NEDIĆ A, LIU J. Distributed optimization for control. *Annual Review of Control Robotics & Autonomous Systems*, 2018, 1(1): 77 – 103.
- [8] LYU X C, NI W, TIAN H, et al. Distributed online optimization of fog computing for selfish devices with out-of-date information. *IEEE Transactions on Wireless Communications*, 2018, 17(11): 7704 – 7717.
- [9] SCUTARI G, FACCHINEI F, SONG P, et al. Decomposition by partial linearization: Parallel optimization of multi-agent systems. *IEEE Transactions on Signal Processing*, 2014, 62(3): 641 – 656.
- [10] KIA S S. Distributed optimal in-network resource allocation algorithm design via a control theoretic approach. *Systems & Control Letters*, 2017, 107(1): 49 – 57.
- [11] NEDIĆ A, ASUMAN O. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 2009, 54(1): 48 – 61.
- [12] NEDIĆ A, LIU J. On convergence rate of weighted-averaging dynamics for consensus problems. *IEEE Transactions on Automatic Control*, 2017, 62(2): 766 – 781.
- [13] SRIVASTAVA K, NEDIĆ A, STIPANOVIĆ D M. Distributed constrained optimization over noisy networks. *Proceedings of the 49th IEEE Conference on Decision and Control (CDC)*. Atlanta: IEEE, 2010: 1945 – 1950.
- [14] LIN P, REN W, SONG Y D. Distributed multi-agent optimization subject to nonidentical constraints and communication delays. *Automatica*, 2016, 65(1): 120 – 131.
- [15] NEDIĆ A, OLSHEVSKY A. Distributed optimization over time-varying directed graphs. *IEEE Transactions on Automatic Control*, 2015, 60(3): 601 – 615.
- [16] XIE P, YOU K Y, TEMPO R, et al. Distributed convex optimization with inequality constraints over time-varying unbalanced digraphs. *IEEE Transactions on Automatic Control*, 2018, 63(12): 4331 – 4337.
- [17] RABBAT M, NOWAK R D. Quantized incremental algorithms for distributed optimization. *IEEE Journal on Selected Areas in Communications*, 2005, 23(4): 798 – 808.
- [18] NEDIĆ A, OLSHEVSKY A, OZDAGLAR A, et al. Distributed subgradient methods and quantization effects. *Proceedings of the 47th IEEE Conference on Decision and Control*. Cancun: IEEE, 2008: 4177 – 4184.
- [19] TAHERI H, MOKHTARI A, HASSANI H, et al. Quantized push-sum for gossip and decentralized optimization over directed graphs. *Proceedings of the 37th International Conference on Machine Learning (ICML)*. Online: PMLR, 2020: 9324 – 9333.
- [20] LI T, XIE L H. Distributed consensus over digital networks with limited bandwidth and time-varying topologies. *Automatica*, 2011, 47(9): 2006 – 2015.
- [21] YI P, HONG Y G. Quantized subgradient algorithm and data-rate analysis for distributed optimization. *IEEE Transactions on Control of Network Systems*, 2014, 1(4): 380 – 392.
- [22] ZHANG J Q, YOU K Y. Distributed optimization with binary relative information over deterministically time-varying graphs. *Proceedings of the 57th IEEE Conference on Decision and Control (CDC)*. Miami, Beach: IEEE, 2018: 2747 – 2752.
- [23] YU G, SHEN Y J. Event-triggered distributed optimisation for multi-agent systems with transmission delay. *IET Control Theory and Applications*, 2019, 13(14): 2188 – 2196.
- [24] HAYASHI N K, SUGIURA T, KAJIYAMA Y, et al. Distributed event-triggered algorithm for unconstrained convex optimisation over weight-balanced directed networks. *IET Control Theory and Applications*, 2020, 14(2): 253 – 261.
- [25] LI H Q, LIU S, SOH Y C, et al. Event-triggered communication and data rate constraint for distributed optimization of multiagent systems. *IEEE Transactions on Systems, Man, Cybernetics: Systems*, 2018, 48(11): 1908 – 1919.
- [26] NESTEROV Y. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 2017, 17(2): 527 – 566.
- [27] CHEN X M, GAO C. Strong consistency of random gradient-free algorithms for distributed optimization. *Optimal Control Applications and Methods*, 2017, 38(2): 247 – 265.
- [28] WANG L F, WANG Y H, HONG Y G. Distributed online optimization with gradient-free design. *Proceedings of the 31th Chinese Control Conference (CCC)*. Guangzhou: IEEE, 2019: 5677 – 5682.
- [29] XIE J, LIANG C X, XIAO Y C. Reactive power optimization for distribution network based on distributed random gradient-free algorithm. *Energies*, 2018, 11(3): 534.
- [30] PANG Y P, HU G Q. Randomized gradient-free distributed optimization methods for a multiagent system with unknown cost function. *IEEE Transactions on Automatic Control*, 2020, 65(1): 333 – 340.
- [31] WANG D, ZHOU J, WANG Z H, et al. Random gradient-free optimization for multiagent systems with communication noises under a time-varying weight balanced digraph. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, 2020, 50(1): 281 – 289.
- [32] YUAN D M, XU S Y, LU J W. Gradient-free method for distributed multi-agent optimization via push-sum algorithms. *International Journal of Robust and Nonlinear Control*, 2015, 25(10): 1569 – 1580.
- [33] DING J J, YUAN D M, JIANG G P, et al. Distributed quantized gradient-free algorithm for multi-agent convex optimization. *Proceedings of the 29th Chinese Control and Decision Conference*. Guangzhou: IEEE, 2017: 6431 – 6435.
- [34] LI H Q, LU Q G, HUANG T W. Distributed projection subgradient algorithm over time-varying general unbalanced directed graphs. *IEEE Transactions on Automatic Control*, 2019, 64(3): 1309 – 1316.
- [35] LU H H. “Relative continuity” for non-Lipschitz nonsmooth convex optimization using stochastic (or deterministic) mirror descent. *Inform Journal on Optimization*, 2019, 1(4): 288 – 303.

- [36] PANG Y P, HU G Q. Distributed nash equilibrium seeking with limited cost function knowledge via a consensus-based gradient-free method. *IEEE Transactions on Automatic Control*, 2020, 66(1): 1832 – 1839.
- [37] YE M J, HU G Q. Distributed seeking of time-varying nash equilibrium for non-cooperative games. *IEEE Transactions on Automatic Control*, 2015, 60(11): 3000 – 3005.
- [38] NEDIĆ A, OLSHEVSKY A, SHI W. Achieving geometric convergence for distributed optimization over time-varying graphs. *SIAM Journal on Optimization*, 2017, 27(4): 2597 – 2633.
- [39] CHELLAPPA R, THEODORIDIS S. *Academic Press Library in Signal Processing, Volume 7: Array, Radar and Communications Engineering*. London: Academic Press, 2017.
- [40] NEDIĆ A, OZDAGLAR A, PARRILO P A. Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 2010, 55(4): 922 – 938.
- [41] ALETTI G, CRIMALDI I, GHIGLIETTI A. Synchronization of reinforced stochastic processes with a network-based interaction. *Annals of Applied Probability*, 2017, 27(6): 3787 – 3844.

作者简介:

谢奕彬 硕士研究生, 目前研究方向为多智能体系统的分布式优化, E-mail: maxieyibin@mail.scut.edu.cn;

高文华 博士, 副教授, 目前研究方向为随机时滞系统的鲁棒控制和多智能体系统的分布式优化, E-mail: whgao@scut.edu.cn.