分布式协同优化的研究现状与展望

杨涛*, 柴天佑

东北大学流程工业综合自动化国家重点实验室, 沈阳 110819

* E-mail: yangtao@mail.neu.edu.cn

收稿日期: 2020-02-14; 接受日期: 2020-03-13; 网络版发表日期: 2020-04-24

国家自然科学基金重大项目(批准号: 61991403, 61991404, 61991400)资助

摘要 分布式协同优化通过多智能体之间的合作协同有效地实现优化任务. 分布式协同优化理论和应用是控制科学与工程的重要发展方向之一. 近年来, 随着云计算、大数据、移动互联网、人工智能等新兴技术的发展, 分布式协同优化正面临新的挑战 and 机遇. 本文回顾和总结了近年来分布式协同优化研究中的一些热点研究方向, 包括分布式加速优化算法、分布式非凸优化算法、分布式无梯度优化算法. 并以智能制造、能源互联网、分布式机器学习三个实际应用为导向, 对分布式协同优化未来重点研究方向进行了展望.

关键词 分布式协同优化, 信息物理系统, 智能制造, 能源互联网, 分布式机器学习

1 引言

信息物理系统(cyber-physical systems, CPS)是一个综合计算、网络和物理环境的多维复杂系统, 通过计算、通信与控制技术的有机融合与深度协作, 实现大型工程系统的实时感知、动态控制和信息服务. CPS是美国基金会在2008年提出的^[1], CPS是指将计算资源与物理资源紧密融合与协同, 使得系统的适应性、自治力、效率、功能、可靠性、安全性和可用性远超过今天的系统.

随着物联网、云计算、大数据、人工智能等技术的发展, 信息和物理系统被进一步融合贯通, 形成了高级、复杂的信息物理系统. CPS理念被广泛应用于交

通、能源、国防、制造、医疗、电力、农业等方面, 如智慧城市、智慧工厂、智能电网、智慧交通等^[2,1].

CPS是当今信息科学与技术的重要研究方向之一. 针对CPS的分布式协同优化理论和应用已经成为当代控制科学与工程的重要发展方向之一. 对复杂的信息物理系统, 分布式协同优化与传统集中式优化相比较具有如下特点: (1) 与优化问题相关的信息分布存储在每个智能体中, 因此更隐私; (2) 每个智能体不需要将数据传输到中心节点, 只需要与邻居智能体进行信息交互, 因此更加节约通信成本; (3) 不存在单点故障问题, 极大地提高了系统的鲁棒性; (4) 不依赖于中心节点, 增强了网络的可扩展性.

分布式协同优化的基本结构如图1所示, 每个智能

1) 中国自动化学会. 自动化学科发展路线图. 中国科学技术出版社, 排版中

引用格式: 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: 1414-1425

Yang T, Chai T Y. Research status and prospects of distributed collaborative optimization (in Chinese). Sci Sin Tech, 2020, 50: 1414-1425, doi: 10.1360/SST-2020-0040

图1 分布式协同优化的基本结构图

Figure 1 Basic structure diagram of distributed collaborative optimization.

体(节点)都有一个局部目标函数, 全局目标函数是这些局部目标函数的和, 每个节点通过与邻居节点进行信息交互, 最终协同实现全局优化的目标.

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^N f_i(x), \quad (1)$$

即每个智能体的自身状态收敛到全局最优解.

分布式协同优化的研究可以追溯到20世纪80年代开创性的工作^[3,4]. 近年来, 随着物联网、大数据、移动互联网等新兴技术的发展, 分布式协同优化在智能电网^[5~11]、智能交通^[12~15]、智能制造^[16~23]等工业领域有着广泛应用. 分布式协同优化引起了诸多领域研究人员的巨大兴趣. 研究者在分布式优化算法的设计、收敛性分析、应用等方面取得了一系列显著成果, 详见综述文章^[24~38]和其中的参考文献. 上述综述文章对早期的分布式协同优化算法进行了详细的归纳和总结.

早期的分布式协同优化算法大多采用衰减步长, 因此收敛速度较慢, 极大地浪费了通信和计算资源. 最近的研究热点之一是设计分布式加速算法来提高收敛速度. 另外现有的研究工作大多只考虑凸优化问题, 然而在实际应用中, 比如机器学习、电力系统等, 很多问题都是非凸的. 因此, 设计针对非凸问题的分布式协同优化算法是近年来的另一个研究热点方向. 再有现有的分布式协同优化算法大多都是基于一阶梯度信息, 有的还用到了二阶Hessian信息, 但是在许多实际问题中, 比如深度神经网络的训练, 这些信息难以获取. 因此, 设计分布式无梯度的优化算法也是一个重要研究方向. 本文力图从以上三个方面对分布式协同优化领域的最新研究工作进行归纳和总结, 并以实际应用前景为导向, 对分布式协同优化的未来重要发展方向进行展望.

本文组织结构如下: 第2章介绍图论和凸分析的基本知识; 第3章介绍基于衰减步长的分布式优化算法;

第4章介绍分布式加速优化算法; 第5章梳理关于分布式非凸优化的研究成果; 第6章介绍分布式无梯度优化算法; 最后, 第7章以实际的应用前景为导向, 浅析分布式协同优化的未来重要发展方向.

2 预备知识

本节主要介绍图论和凸分析的基本概念.

2.1 图论

首先介绍图论的基本知识^[39]. 通常采用时变有向图 $\mathcal{G}(k) = (\mathcal{V}, \mathcal{E}(k))$ 来描述网络中节点的连通情况, 其中, $\mathcal{V} = \{1, 2, \dots, N\}$ 表示 N 个节点构成的集合, $\mathcal{E}(k) \subseteq \mathcal{V} \times \mathcal{V}$ 表示边集合, 即 $(j, i) \in \mathcal{E}(k)$ 当且仅当节点 i 可以接受节点 j 的信息. 如果节点 i 可以接受节点 j 的信息, 则称 j 是 i 的入邻居, i 是 j 的出邻居. 节点 i 的入邻居和出邻居构成的集合分别记作 $N_i^{\text{in}}(k) = \{j \in \mathcal{V} | (j, i) \in \mathcal{E}(k)\}$ 和 $N_i^{\text{out}}(k) = \{j \in \mathcal{V} | (i, j) \in \mathcal{E}(k)\}$. 令非负矩阵 $A(k) = [a_{ij}(k)]$ 为有向图 $\mathcal{G}(k)$ 的权重矩阵, 其中, $a_{ij}(k) > 0$ 当且仅当 $(j, i) \in \mathcal{E}(k)$. 如果对所有节点 i 和所有时刻 k , $\sum_{j=1}^N a_{ij}(k) = \sum_{j=1}^N a_{ji}(k)$ 均成立, 则称图 $\mathcal{G}(k)$ 为(加权)平衡图. 如果对所有时刻 k , $A(k) = A^T(k)$ 均成立, 则称图 $\mathcal{G}(k)$ 为无向图. 如果对所有时刻 k , 边集合 $\mathcal{E}(k)$ 均不变, 则称图 $\mathcal{G}(k)$ 为固定图 $\mathcal{G}$. 对于固定有向图 $\mathcal{G}$, 如果存在 $i_1, \dots, i_r$ 满足 $(i_q, i_{q+1}) \in \mathcal{E}$, $q = 1, \dots, r-1$, 则称节点 i_1 到节点 i_r 存在一条有向路径. 若固定有向图 $\mathcal{G}$ 中任意两个节点之间都存在一条有向路径, 则称有向图 $\mathcal{G}$ 是强连通的. 若固定无向图 $\mathcal{G}$ 中任意两个节点之间都存在一条路径, 则称无向图 $\mathcal{G}$ 是连通的. 对于时变有向图 $\mathcal{G}(k)$, 如果存在一个正整数 B , 对于任意 $t \geq 0$, 图 $\mathcal{G}_{(B)} = (\mathcal{V}, \cup_{k_0=tB}^{(t+1)B-1} \mathcal{E}(k_0))$ 是强连通的, 则称时变有向图 $\mathcal{G}(k)$ 一致联合强连通或 B -强连通.

2.2 凸分析

接下来介绍一下凸分析的基本概念^[40]. 对于函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 如果对任意 $x, y \in \mathbb{R}^n$ 和 $0 \leq \theta \leq 1$, $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$, 则称函数 f 为凸函数. 对于连续可微函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 如果存在常数 $\mu > 0$

使得 $(\nabla f(y) - \nabla f(x))^T(y - x) \geq \mu \|y - x\|^2$ 对任意 $x, y \in \mathbb{R}^n$ 成立, 则称函数 f 为强凸函数. 对于连续可微函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 如果存在常数 $\mu > 0$ 使得 $(\nabla f(y) - \nabla f(x))^T(y - x) \geq \mu \|y - x\|^2$ 对任意 $y \in \mathbb{R}^n$ 成立, 则称函数 f 关于点 x 是有限强凸的. 对于连续可微函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 如果存在常数 $L > 0$ 使得 $\|\nabla f(y) - \nabla f(x)\| \leq L \|y - x\|$ 对任意 $x, y \in \mathbb{R}^n$ 成立, 则称函数 f 为 L -光滑或简称光滑. 对于连续可微函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 令 $X^* = \operatorname{argmin}_{x \in \mathbb{R}^n} f(x)$, $f^* = \min_{x \in \mathbb{R}^n} f(x)$, 如果存在常数 $\mu > 0$ 使得 $\frac{1}{2} \left\| \nabla f(x) \right\|^2 \geq \mu(f(x) - f^*)$ 对任意 $x \in \mathbb{R}^n$ 成立, 则称函数 f 满足Polyak-Łojasiewicz (PL)条件. 函数满足PL条件, 并不意味着强凸或者凸, 也可能非凸.

3 基于衰减步长的分布式优化算法

本节简单回顾了早期的基于衰减步长的分布式优化算法. 基于一致性算法和(次)梯度下降方法, Nedić和Ozdaglar^[41]以及Johansson等人^[42]分别提出如下的分布式梯度下降(distributed gradient descent, DGD)算法:

$$x_i(k+1) = \sum_{j=1}^N w_{ij}(k)x_j(k) - \alpha(k)s_i(k), \quad (2)$$

其中, $w_{ij}(k)$ 是通信网络在时刻 k 边 (j, i) 的权重, 权重矩阵 $W(k) = [w_{ij}(k)]$ 为双随机矩阵(行和、列和均为1), $s_i(k)$ 是局部目标函数 $f_i(x)$ 的次梯度, 步长 $\alpha(k) > 0$ 非增, 且满足

$$\sum_{k=0}^{\infty} \alpha(k) = \infty, \sum_{k=0}^{\infty} \alpha^2(k) < \infty. \quad (3)$$

在时变无向通信网络一致联合连通的条件下, 作者证明了算法的收敛性, 即所有节点的决策变量都收敛到式(1)的一个最优解.

DGD算法(式(2))只适用于无向图. 但是在实际情况中, 每个节点的传送速率不同, 通信网络的连接也不是双向的. 因此随后研究者考虑了通信网络是有向图的情况. Nedić和Olshevsky^[43]提出了push-sum算法. 该算法去除了权重矩阵双随机的要求, 只要求权重矩阵为列随机矩阵(列和为1). 如果时变有向图 B -强连通, 且非增步长满足式(3)的条件, 算法收敛到一个最优解.

上述两个分布式协同优化算法随后被广泛推广, 包括带约束的优化问题, 通信网络存在延时、丢包的情况等, 详见综述文章[24~38]. 但是这些算法大多采用衰减步长, 因此收敛速度较慢. 如果采用固定步长, 收敛速度有所提高, 但是只收敛到最优解附近, 而不能收敛到最优解^[44,45].

4 分布式加速优化算法

最近的研究热点是设计快速收敛到最优解的分布式加速优化算法. 这些算法大多采用固定步长, 另外要求局部目标函数光滑且强凸. 下面分别介绍适用于无向图和有向图的分布式加速优化算法.

4.1 无向图

EXTRA算法. Shi等人^[46]提出分布式精确收敛的一阶算法EXTRA (exact first-order algorithm). 节点 i 的更新规则如下:

$$\begin{aligned} x_i(k+2) = & x_i(k+1) + \sum_{j=1}^N w_{ij}x_j(k+1) \\ & - \sum_{j=1}^N \tilde{w}_{ij}x_j(k) - \alpha(\nabla f_i(x_i(k+1)) \\ & - \nabla f_i(x_i(k))), \end{aligned}$$

其中, $\alpha > 0$ 是定步长, 权重矩阵 $W = [w_{ij}]$ 和 $\tilde{W} = [\tilde{w}_{ij}]$ 为双随机矩阵. 与DGD算法(式(2))不同, EXTRA算法用过去两次迭代的状态和梯度信息来更新节点的状态. EXTRA算法可以看作DGD算法和一个累积校正项的和, 其中累积校正项用以校正由于固定步长引起的误差. 当固定无向图连通, 两个权重矩阵满足一定条件, 定步长小于一定的常数, 局部目标函数光滑, 全局目标函数关于最优点是有限强凸的, 作者证明了EXTRA算法的线性收敛性.

DIGing算法. 接下来介绍基于分布式不精确梯度方法和梯度跟踪技术(distributed inexact gradient and gradient tracking, DIGing)的一类加速优化算法.

Qu和Li^[47]以及Nedić等人^[48]分别提出了如下算法:

$$\begin{aligned} x_i(k+1) = & \sum_{j=1}^N w_{ij}x_j(k) - \alpha y_i(k), \\ y_i(k+1) = & \sum_{j=1}^N w_{ij}y_j(k) + \nabla f_i(x_i(k+1)) - \nabla f_i(x_i(k)), \end{aligned} \quad (4)$$

其中, $\alpha > 0$ 是定步长, 权重矩阵 $W = [w_{ij}]$ 为双随机矩阵. $x_i(k)$ 的更新规则用到了变量 $y_i(k)$ 而不是全局平均梯度, 因此是分布式不精确梯度方法. 另外, $y_i(k)$ 的更新规则利用了动态一致算法^[49], 使其最终跟踪全局平均梯度. 当局部目标函数光滑且强凸, 定步长小于一定的常数, 固定无向图连通, Qu和Li^[47]基于线性系统不等式思想分析了算法的线性收敛性. 针对一致联合连通的时变无向图, Nedić等人^[48]给出了基于小增益思想的算法线性收敛性的证明. 另外, Nedić等人^[48]还证明了通过恰当选取EXTRA算法的两个权重矩阵, DIGing算法与EXTRA算法等价. 最近, Jakovetić^[50]给出了针对EXTRA算法和DIGing算法的统一的原-对偶分析, 揭示了两种算法的主要区别在于原误差对对偶误差的影响. 当局部目标函数为二次型函数, 固定无向图连通, Yang等人^[51]给出了DIGing算法线性收敛的步长的充分必要条件.

文献^[47,48]提出的算法都要求每个节点采用相同的步长, 最近研究者提出了一系列不同节点可以采用不同步长的分布式加速优化算法. 例如, Xu等人^[52]提出了Aug-DGM (augmented distributed gradient method) 算法, Nedić等人^[53]提出了ATC-DIGing (adapt-then-combine distributed inexact gradient tracking) 算法, Xu等人^[54]提出了AsynDGM (asynchronous distributed gradient method) 算法. 上述算法节点 i 的迭代过程如下:

$$\begin{aligned} x_i(k+1) &= \sum_{j=1}^N w_{ij}(x_j(k) - \alpha_j y_j(k)), \\ y_i(k+1) &= \sum_{j=1}^N w_{ij}(y_j(k) + \nabla f_j(x_j(k+1)) - \nabla f_j(x_j(k))), \end{aligned} \quad (5)$$

其中, $\alpha_j > 0$ 是节点 j 的定步长.

式(4)算法采用CTA (combine-then-adapt) 结构^[55], 即节点的状态与邻居节点的状态先结合, 然后再进行调整. 与式(4)算法相比较, 式(5)算法采用ATC (adapt-then-combine) 结构^[55], 即节点的状态先进行调整, 然后再与邻居节点调整后的状态相结合. 基于ATC结构, 不同节点采用不同的定步长, 因此步长的选择更加灵活. 当局部目标函数光滑且强凸, 定步长小于一定的常数, 针对连通无向图, 文献^[52,53]给出了算法的线性收敛性分析; 针对无向随机图, 基于小增益思想, 文献^[54]证明了算法的线性收敛性.

其他分布式加速优化算法. 通过引入积分项来校

正定步长对分布梯度下降算法带来的误差, Wang等人^[56]、Lei等人^[57]、Yang等人^[58]以及Yao等人^[59]分别提出了连续和离散的分布式PI (proportional-integral) 算法. 若二阶Hessian信息可以获取, Wei等人^[60,61]、Varagnolo等人^[62]、Mokhtari等人^[63]和Bajović等人^[64]分别提出了分布式牛顿算法; Lu和Yang^[65]提出了ZGS (zero-gradient-sum) 算法.

4.2 有向图

上一节介绍的分布式优化算法只适用于无向图, 这一节介绍适用于有向图的分布式加速优化算法.

基于push-sum的算法. EXTRA算法和DIGing算法只适用于无向图, 随后研究者将其推广到有向图. 例如, 将EXTRA算法与push-sum算法相结合, Zeng等人^[66]提出了ExtraPush算法. 将EXTRA算法和特征向量估计方法相结合, Xi和Khan^[67]提出了DEXTRA (directed-EXTRA) 算法. 如果固定有向图强连通, 权重矩阵列随机, 定步长满足一定条件, 上述算法均线性收敛到全局最优解, 但是对定步长的下界都有要求. 因此, 最近Xi等人^[68]提出了加速的分布式有向图优化 (accelerated distributed directed optimization, ADD-OPT) 算法. 该算法的定步长下界为0, 可以更加灵活地选择定步长. 上述算法都只适用有向图为固定图的情况, 针对时变有向图, Nedić等人^[53]将DIGing算法与push-sum算法相结合, 提出了Push-DIGing算法并给出算法线性收敛性的分析.

基于push-pull的算法. 将DIGing算法和push-pull算法结合, Du等人^[69]、Xin等人^[70]和Pu等人^[71]分别提出了针对有向图的分布式加速优化方法. 与基于push-sum的算法不同, 基于push-pull的算法分别用行随机和列随机权重矩阵来更新决策变量和对全局平均梯度的估计变量. 当局部目标函数光滑且强凸, 固定有向图强连通, 定步长小于一定的常数, 上述算法均线性收敛到全局最优解.

连续加速优化算法. 上述的加速方法都是离散算法, 也有研究者提出了适用于有向图的连续分布式加速优化算法. 针对平衡有向图, Gharesifard和Cortes^[72]提出了分布式PI算法, 并给出了基于Lyapunov稳定性分析的证明. Kia等人^[73]、Xie和Lin^[74]分别提出了针对非平衡有向图的分布式PI算法. Chen和Ren^[75]将ZGS算法推广到平衡有向图.

5 分布式非凸优化算法

现有的研究工作大多只考虑凸优化问题, 然而在实际应用中, 比如机器学习、电力系统等, 很多问题都是非凸的. 因此, 如何设计针对非凸问题的分布式优化算法是近年来一个研究热点方向. 非凸优化问题一般不存在全局最优解, 所设计的算法最好的可能就是收敛到平衡点集. 接下来分别介绍适用于无向图和有向图的分布式非凸优化算法.

5.1 无向图

首先介绍针对无向图的分布式非凸优化算法. Bianchi和Jakubowicz^[76]提出了基于分布式投影随机梯度下降和随机gossip的分布式非凸优化算法. 在衰减步长的条件下, 证明了算法几乎必然(almost surely)收敛到平衡点集, 但是没有给出收敛速度. Wai等人^[77]提出了分布式Frank-Wolfe算法. 在衰减步长的条件下, 证明了算法以 $O(1/\sqrt{k})$ 的速度收敛到平衡点集. Hong等人^[78]提出了Prox-PDA (proximal primal-dual algorithm) 算法, 并证明了算法以 $O(1/k)$ 的速度收敛到平衡点集. 随后Hajenizhad和Hong^[79]提出了PProx-PDA (perturbed proximal primal-dual algorithm) 算法用以解决带约束的非凸优化问题. 如果参数选取适当, 算法渐近收敛到平衡点集. Zeng和Yin^[80]将DGD算法和Proximal-DGD算

法推广到非凸优化问题, 在固定步长和衰减步长满足一定条件下, 证明了算法的收敛性.

5.2 有向图

接下来介绍适用于有向图的分布式非凸优化算法. Lorenzo等人^[81]基于连续凸逼近方法和动态一致性算法的凸化分解技术, 提出了NEXT算法. Tatarenko和Tour^[82]将push-sum算法推广到非凸优化问题上. 在时变有向图一致联合强连通, 步长衰减的条件下, 上述两种算法均渐近收敛到平衡点集. Scutari和Sun^[83]提出了基于连续凸逼近方法和扰动push-sum算法的SONATA算法. 在时变有向图一致联合强连通的条件下, 如果算法采用小于一定常数的定步长, 则次线性收敛到平衡点集附近; 如果算法采用衰减步长, 则渐近收敛到平衡点集. 随后Tian等人^[84]将SONATA算法和异步push-sum算法相结合提出了ASY-SONATA算法. 在固定有向图强连通, 时变延时有届的条件下, 作者证明了算法次线性收敛到平衡点集.

表1给出了本节综述的各类分布式非凸优化算法的收敛条件和收敛速度. 这些分布式非凸优化算法收敛到平衡点集或者其平衡点集附近, 但是平衡点可能是局部最优解, 也可能是鞍点. 在集中式的框架下, 文献[85,86]指出很多一阶梯度算法, 通过适当的随机初始化, 可以几乎必然避免陷入鞍点. 如何理论证明本

表 1 不同分布式非凸优化算法的收敛条件和收敛速度

Table 1 Convergence conditions and convergence rates of different distributed non-convex optimization algorithms

文献	算法	约束情况	通信网络	步长	收敛速度
[76]	分布式投影随机梯度下降+gossip	带约束	时变无向随机图, 平均图连通	衰减	渐近, 平衡点集
[77]	分布式Frank-Wolfe	带约束	固定无向图连通	衰减	$O(1/\sqrt{k})$, 平衡点集
[78]	Prox-PDA	无约束	固定无向图连通	衰减	$O(1/k)$, 平衡点集
				固定	$O(1/k)$, 平衡点集附近
[79]	PProx-PDA	带约束	固定无向图连通	衰减	$O(1/k)$, 平衡点集
				固定	$O(1/k)$, 平衡点集附近
[80]	DGD, Proximal-DGD	无约束	固定无向图连通	衰减	渐近, 平衡点集
				固定	渐近, 平衡点集附近
[81]	NEXT	带约束	时变有向图 B -强连通	衰减	渐近, 平衡点集
[82]	Push-sum	无约束	时变有向图 B -强连通	衰减	渐近, 平衡点集
[83]	SONATA	带约束	时变有向图 B -强连通	衰减	渐近, 平衡点集
				固定	次线性, 平衡点集附近
[84]	ASY-SONATA	无约束	固定有向图强连通, 时变延时有届	衰减	$O(1/k)$, 平衡点集

节介绍的分布式非凸优化算法也可以几乎必然避免陷入鞍点值得关注.

6 分布式无梯度优化算法

现有的分布式优化算法大多用到了一阶梯度信息, 有的还用到了二阶Hessian信息. 但是在许多实际问题中, 比如深度神经网络的训练, 这些信息难以获取. 在集中式的框架下, Nesterov和Spokoiny^[87]提出了随机无梯度(零阶)算法来估计梯度信息, 有效地克服了上述困难. 基于该算法的思想, 最近研究者提出了一系列分布式无梯度优化算法. 接下来分别介绍针对凸优化问题和非凸优化问题的分布式零阶算法.

6.1 分布式零阶凸优化算法

首先介绍针对凸优化问题的分布式零阶算法. Sahu等人^[88]提出了基于Kiefer-Wolfowitz随即逼近方法的分布式零阶算法. 如果局部目标函数光滑且强凸, 无向随机图的平均图连通, 算法以 $O(1/\sqrt{k})$ 的速度均方(mean square)收敛到全局最优解附近. 上述研究工作考虑的是局部目标函数光滑的情况, 也有研究者研究了局部目标函数非光滑的情况. Yuan和Ho^[89]提出了分布式随机无梯度算法, 时变有向图一致联合强连通, 权重矩阵双随机条件下, 分析了算法的收敛性. 为了去除双随机权重矩阵的限制, Pang和Hu^[90]提出了基于push-pull算法的分布式随机无梯度算法, 并针对固定强连通有向图, 给出了算法的收敛速度. Wang等人^[91]采用随机差分技术提出了分布式零阶算法. 在时变有向图一致联合强连通, 权重矩阵双随机, 局部目标函

数强凸但是非光滑的条件下, 证明了算法几乎必然收敛到全局最优解附近.

6.2 分布式零阶非凸优化算法

接下来介绍针对非凸优化问题的分布式零阶算法. 针对固定无向连通图和星形图, Hajinezhad等人^[92]提出了基于原-对偶零阶方法的分布式非凸优化算法, 证明了算法次线性收敛到平衡点集.

在集中式的框架下, Karimi等人^[93]给出一系列算法在目标函数满足PL (Polyak-Łojasiewicz)条件下线性收敛性的分析. 目标函数满足PL条件, 并不意味着强凸, 也可能非凸. 在该条件下, 所有平衡点都是全局最小值, 但这并不意味着优化问题有唯一最优解. 目前针对全局目标函数满足PL条件的非凸优化问题的分布式算法相对较少. Tang等人^[94]将DIGing算法与两种不同梯度估计方法相结合, 提出了两种不同的分布式零阶优化算法. 如果固定无向图连通, 局部目标函数光滑, 全局目标函数满足PL条件, 所提出的分布式零阶算法具有与分布式一阶算法和集中式算法相当的收敛性.

表2给出了本节综述的各类分布式零阶算法的收敛条件和收敛速度.

7 总结与展望

本文总结了近年来分布式协同优化的研究成果, 重点介绍了分布式加速优化算法、分布式非凸优化算法和分布式无梯度优化算法的最新结果. 目前分布式协同优化的研究仍处于发展阶段, 现有的分布式协同

表 2 不同分布式零阶算法的收敛条件和收敛速度

Table 2 Convergence conditions and convergence rates of different distributed zeroth-order algorithms

文献	约束情况	目标函数	通信网络	收敛速度
[88]	无约束	局部目标函数强凸且光滑	时变无向随机图, 平均图连通	$O(1/\sqrt{k})$ 均方, 全局最优解附近
[89]	有约束	局部目标函数凸但是非光滑	时变有向图B-强连通	渐近, 全局最优解附近
[90]	有约束	局部目标函数凸但是非光滑	固定有向图强连通	渐近, 全局最优解附近
[91]	无约束	局部目标函数强凸但是非光滑	时变有向图B-强连通	几乎必然, 全局最优解附近
[92]	无约束	局部目标函数非凸且光滑	固定无向图连通	次线性, 平衡点集
		局部目标函数非凸但是非光滑	星形	次线性, 平衡点集
[94]	无约束	局部目标函数光滑, 全局目标函数PL非凸	固定无向图连通	线性, 全局最优解集

优化的研究工作主要以理论的算法收敛性分析为主, 如何将分布式协同优化的理论研究与实际应用相结合是值得深入探索的方向. 本节以智能制造、能源互联网、分布式机器学习三个实际应用为导向, 对分布式协同优化未来的重点研究方向进行展望.

7.1 智能制造

智能制造的发展目标是实现高效化和绿色化. 为了实现上述发展目标, 应采用基于五大系统的发展路线²⁾. 五大系统包括: 生产制造过程智能自主控制系统、制造全流程智能协同优化控制系统、制造全流程智能优化决策系统、生产制造过程安全运行监控与自优化系统和工业过程虚拟制造系统. 其中分布式协同优化在制造全流程智能优化决策系统中起着重要作用.

工业生产全流程是由一个或多个工业装备组成的生产工序, 多个生产工序构成了全流程生产线. 其功能是将进入的原料加工为半成品材料或者产品. 为了实现生产全流程的产品质量、产量、消耗、成本等综合生产指标的优化, 必须协同各个生产工序(即工业过程智能体)来共同实现. 另外生产运行过程中存在生产指标范围、原料、设备等多种动态因素. 这些动态因素可归结为目标的变化以及约束条件的变化. 因此全流程优化问题本质是一个带约束的多冲突目标动态优化问题. 但是现有的分布式协同优化研究工作大多只考虑单一目标静态优化问题. 所以需要开展带约束的多冲突目标动态分布式协同优化的研究工作, 并在此基础上, 开展协同优化与控制一体化的研究工作.

7.2 能源互联网

能源互联网本质上是一个物质、能量与信息深度耦合的系统, 是物理空间、能量空间、信息空间乃至社会空间耦合的多域、多层次关联, 表现出混杂多尺度动态与复杂网络特性, 具有更广阔的开放性和更大的系统复杂性. 分布式能源协同优化是能源互联网信息物理系统架构下的重要特征^[95].

在能源互联网中, 以风电、光伏为代表的新能源电力, 存在发电的随机波动不确定性. 这些不确定性

会导致发电的成本函数不能提前确定, 但是现有的分布式协同优化大多针对已知确定的目标函数. 另外现有分布式能源协同优化的研究工作大多没有考虑通信网络存在的延时、丢包和变拓扑结构的情况^[96]. 为了解决上述问题, 需要开展信息传输受限情况下的分布式能源鲁棒协同优化的研究工作.

能源互联网作为典型的信息物理系统面临着安全威胁, 受到各种类型的网络攻击和物理攻击^[97]. 但是除了Sundaram和Gharesifard^[98]以及Zhao等人^[99]的研究工作, 现有的分布式协同优化的研究工作大多没有考虑到系统受到攻击的情况. 因此, 需要分析现有的分布式协同优化算法在受到攻击情况下的性能, 并在此基础上, 开展弹性分布式协同优化的研究工作.

7.3 分布式机器学习

近年来, 机器学习特别是深度学习, 在推荐系统、图像分类、自然语言处理等领域取得了巨大的成功, 已经成为人工智能领域的重要发展方向. 随着训练参数与样本规模的急剧增长, 一台计算机很难处理大规模的机器学习模型. 另外海量训练数据往往分布存储在多台计算机上, 每个计算机只能获取本地数据集. 分布式机器学习利用多个计算节点加速模型的训练和优化为解决上述问题提供了先决条件^[100].

机器学习尤其是深度学习大多采用随机梯度下降算法(SGD)来训练和优化模型^[101,102]. 但是随机梯度算法往往收敛比较慢, 近年来, 研究者提出一系统机制来提高随机梯度算法的收敛速度^[103~106]. 如何将这些加速方法扩展成分布式随机梯度算法是未来一个重要的研究方向.

在训练大规模神经网络的过程中, 梯度传输的成本是一个很大的瓶颈. 最近研究者提出一系列方法来提高通信效率, 比如量化梯度^[107,108], 多步本地SGD^[109], 带记忆的稀疏SGD^[110], sign-SGD^[111]等. 另外最近研究者提出decentralized learning的方法^[112,113]以此来提高通信效率. 但是现有的算法大多没有考虑到通信网络存在的延时、丢包和变拓扑结构的情况. 如何将上述分布式学习算法推广到信息传输受限的情况也值得关注.

2) 同脚注1)

参考文献

- 1 Cyber-physical systems. Program Announcements & Information. The National Science Foundation, 4201 Wilson Boulevard, Arlington, Virginia 22230, USA. 2008-09-30. Retrieved 2009-07-21
- 2 Chai T Y. Development directions of automation science and technology (in Chinese). *Acta Automat Sin*, 2018, 44: 1923–1930 [柴天佑. 自动化科学与技术发展方向. *自动化学报*, 2018, 44: 1923–1930]
- 3 Tsitsiklis J N. Problems in decentralized decision making and computation. Dissertation for the Doctoral Degree. Cambridge: Massachusetts Institute of Technology, 1984
- 4 Bertsekas D P, Tsitsiklis J N. *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs: Prentice-Hall, 1989
- 5 Kraning M, Chu E, Lavaei J, et al. Dynamic network energy management via proximal message passing. *Found Trends Optim*, 2014, 1: 70–122
- 6 Bidram A, Lewis F L, Davoudi A. Distributed control systems for small-scale power networks: Using multiagent cooperative control theory. *IEEE Contr Sys Mag*, 2016, 34: 56–77
- 7 Molzahn D K, Dorfler F, Sandberg H, et al. A survey of distributed optimization and control algorithms for electric power systems. *IEEE Trans Smart Grid*, 2017, 8: 2941–2962
- 8 Yu X, Xue Y. Smart grids: A cyber-physical systems perspective. *Proc IEEE*, 2016, 104: 1058–1070
- 9 Guo F, Wen C, Song Y. *Distributed Control and Optimization Technologies in Smart Grid Systems*. Florida: CRC Press, 2017
- 10 Yao Y Y, Tian F Z, Mei F, et al. Dynamical economic dispatch using distributed barrier function-based optimization algorithm. *Sci China Tech Sci*, 2019, 62: 2104–2112
- 11 Mei S W, Zhu J Q. Mathematical and control scientific issues of smart grid and its prospects (in Chinese). *Acta Automat Sin*, 2013, 39: 119–131 [梅生伟, 朱建全. 智能电网中的若干数学与控制科学问题及其展望. *自动化学报*, 2013, 39: 119–131]
- 12 Reilly J, Bayen A M. Distributed optimization for shared state systems: Applications to decentralized freeway control via subnetwork splitting. *IEEE Trans Intell Transp Syst*, 2015, 16: 3465–3472
- 13 Lagarrigue F L, Annaswamy A, Engell S, et al. Systems and control for the future of humanity, research agenda: Current and future roles, impact and grand challenges. *Ann Rev Control*, 2017, 43: 1–64
- 14 Ning B. A number of scientific and technical problems in intelligent transportation (in Chinese). *Sci Sin Inf*, 2018, 48: 1264–1269 [宁滨. 智能交通中的若干科学和技术问题. *中国科学: 信息科学*, 2018, 48: 1264–1269]
- 15 Yuan H N, Guo G. Vehicle cooperative optimization scheduling in transportation cyber physical systems (in Chinese). *Acta Automat Sin*, 2019, 45: 143–152 [原豪男, 郭戈. 交通信息物理系统中的车辆协同运行优化调度. *自动化学报*, 2019, 45: 143–152]
- 16 Chai T Y. Industrial process control systems: research status and development direction (in Chinese). *Sci Sin Inf*, 2016, 46: 1003–1015 [柴天佑. 工业过程控制系统研究现状与发展方向. *中国科学: 信息科学*, 2016, 46: 1003–1015]
- 17 Chai T Y. Artificial intelligence research challenges in intelligent manufacturing processes (in Chinese). *Bull Natl Nat Sci Found China*, 2018, 32: 251–256 [柴天佑. 制造流程智能化对人工智能的挑战. *中国科学基金*, 2018, 32: 251–256]
- 18 Gui W H, Yue W C, Xie Y F, et al. A review of intelligent optimal manufacturing for aluminum reduction production (in Chinese). *Acta Automat Sin*, 2018, 44: 1957–1970 [桂卫华, 岳伟超, 谢永芳, 等. 铝电解生产智能优化制造研究综述. *自动化学报*, 2018, 44: 1957–1970]
- 19 Qian F, Du W L, Zhong W M, et al. Problems and challenges of smart optimization manufacturing in petrochemical industries (in Chinese). *Acta Automat Sin*, 2017, 43: 893–901 [钱锋, 杜文莉, 钟伟民, 等. 石油和化工行业智能优化制造若干问题及挑战. *自动化学报*, 2017, 43: 893–901]
- 20 Guan X P, Chen C L, Yang B, et al. Towards the integration of sensing, transmission and control for industrial network systems: Challenges and recent developments (in Chinese). *Acta Automat Sin*, 2019, 45: 25–36 [关新平, 陈彩莲, 杨博, 等. 工业网络系统的感知-传输-控制一体化: 挑战 and 进展. *自动化学报*, 2019, 45: 25–36]
- 21 Kusiak A. Smart manufacturing must embrace big data. *Nature*, 2017, 544: 23–25
- 22 Yuan Y, Tang X, Zhou W, et al. Data driven discovery of cyber physical systems. *Nat Commun*, 2019, 10: 4894
- 23 Yuan Y, Ma G, Cheng C, et al. A general end-to-end diagnosis framework for manufacturing systems. *Natl Sci Rev*, 2020, 7: 418–429
- 24 Hong Y G, Zhang Y Q. Distributed optimization: Algorithm design and convergence analysis (in Chinese). *Contr Theory Appl*, 2014, 31: 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: 850–857]

- 25 Yi P, Hong Y G. Distributed cooperative optimization and its applications (in Chinese). *Sci Sin Math*, 2016, 46: 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: 1547–1564]
- 26 Long Y S, Liu S, Xie L H. Distributed constrained stochastic optimal consensus (in Chinese). *Sci Sin Math*, 2016, 46: 1487–1498 [龙昱岫, 刘帅, 谢立华. 带集合约束的分布式随机凸优化. *中国科学: 数学*, 2016, 46: 1487–1498]
- 27 Xie P, You K Y, Hong Y G, et al. A survey of distributed convex optimization algorithms over networks (in Chinese). *Contr Theory Appl*, 2018, 35: 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: 918–927]
- 28 Wang L, Lu K H, Guan Y Q. Distributed optimization via multi-agent systems (in Chinese). *Contr Theory Appl*, 2019, 36: 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: 1820–1883]
- 29 Zhang J Q, You K Y. Research progress of distributed optimization algorithms (in Chinese). *CAA-PAMI Newsletter*, 2019, 1: 3–9 [张家琦, 游科友. 分布式优化算法的研究进展. *中国自动化学会模式识别与机器智能专业委员会通讯*, 2019, 1: 3–9]
- 30 Yang B, Johansson M. Distributed optimization and games: A tutorial overview. In: Bemporad A, Heemels M, Johansson M, eds. *Networked Control Systems*. London: Springer, 2010. 109–148
- 31 Boyd S. Distributed optimization and statistical learning via the alternating direction method of multipliers. *FNT Machine Learning*, 2011, 3: 1–122
- 32 Sayed A. Adaptation, learning, and optimization over networks. *FNT Machine Learning*, 2014, 7: 311–801
- 33 Nedić A. Convergence rate of distributed averaging dynamics and optimization in networks. *Found Trends Syst Contr*, 2015, 2: 1–100
- 34 Nedić A, Liu J. Distributed optimization for control. *Annu Rev Contr Robot Autonmat Syst*, 2018, 1: 77–103
- 35 Nedić A, Olshevsky A, Rabbat M G. Network topology and communication-computation tradeoffs in decentralized optimization. *Proc IEEE*, 2018, 106: 953–976
- 36 Giselsson P, Rantzer A. *Large-Scale and Distributed Optimization*. Switzerland: Springer International Publishing AG, 2018
- 37 Qin J, Ma Q, Shi Y, et al. Recent advances in consensus of multi-agent systems: A brief survey. *IEEE Trans Ind Electron*, 2017, 64: 4972–4983
- 38 Yang T, Yi X, Wu J, et al. A survey of distributed optimization. *Annu Rev Control*, 2019, 47: 278–305
- 39 Godsi, C, Royle, G F. *Algebraic Graph Theory*. New York: Springer-Verlag, 2011. 207
- 40 Nesterov Y. *Introductory Lectures on Convex Optimization: A Basic Course*. New York: Springer US, 2004. 87
- 41 Nedić A, Ozdaglar A. Distributed subgradient methods for multi-agent optimization. *IEEE Trans Automat Contr*, 2009, 54: 48–61
- 42 Johansson B, Keviczky T, Johansson M, et al. Subgradient methods and consensus algorithms for solving convex optimization problems. In: *Proceedings of the 47th IEEE Conference on Decision and Control (CDC)*. IEEE, 2008. 4185–4190
- 43 Nedić A, Olshevsky A. Distributed optimization over time-varying directed graphs. *IEEE Trans Automat Contr*, 2015, 60: 601–615
- 44 Matei I, Baras J S. Performance evaluation of the consensus-based distributed subgradient method under random communication topologies. *IEEE J Sel Top Signal Process*, 2011, 5: 754–771
- 45 Yuan K, Ling Q, Yin W. On the convergence of decentralized gradient descent. *SIAM J Optim*, 2015, 26: 1835–1854
- 46 Shi W, Ling Q, Wu G, et al. EXTRA: An exact first-order algorithm for decentralized consensus optimization. *SIAM J Optim*, 2015, 25: 944–966
- 47 Qu G, Li N. Harnessing smoothness to accelerate distributed optimization. *IEEE Trans Control Netw Syst*, 2018, 5: 1245–1260
- 48 Nedić A, Olshevsky A, Shi W. Achieving geometric convergence for distributed optimization over time-varying graphs. *SIAM J Optim*, 2017, 27: 2597–2633
- 49 Zhu M, Martínez S. Discrete-time dynamic average consensus. *Automatica*, 2010, 46: 322–329
- 50 Jakovetic D. A unification and generalization of exact distributed first-order methods. *IEEE Trans Signal Inf Process over Networks*, 2019, 5: 31–46
- 51 Yang T, George J, Qin J, et al. Distributed least squares solver for network linear equations. *Automatica*, 2020, 113: 108798
- 52 Xu J, Zhu S, Soh Y C, et al. Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes. In: *Proceedings of the 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015. 2055–2060
- 53 Nedić A, Olshevsky A, Shi W, et al. Geometrically convergent distributed optimization with uncoordinated step-sizes. In: *Proceedings of the IEEE American Control Conference*. IEEE, 2017. 3950–3955
- 54 Xu J, Zhu S, Soh Y C, et al. Convergence of asynchronous distributed gradient methods over stochastic networks. *IEEE Trans Automat Contr*, 2018, 63: 434–448

- 55 Sayed A. Diffusion adaption over networks. Academic Press Library in Signal Processing, 2014, 3: 323–453
- 56 Wang J, Elia N. Control approach to distributed optimization. In: Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton). IEEE, 2010. 557–561
- 57 Lei J, Chen H F, Fang H T. Primal–dual algorithm for distributed constrained optimization. *Syst Control Lett*, 2016, 96: 110–117
- 58 Yang T, Wan Y, Wang H, et al. Global optimal consensus for discrete-time multi-agent systems with bounded controls. *Automatica*, 2018, 102: 104–111
- 59 Yao L, Yuan Y, Sundaram S, et al. Distributed finite-time optimization. In: Proceedings of the 14th IEEE International Conference on Control and Automation (ICCA). IEEE, 2018. 147–154
- 60 Wei E, Ozdaglar A, Jadbabaie A. A distributed newton method for network utility maximization—I: Algorithm. *IEEE Trans Automat Contr*, 2013, 58: 2162–2175
- 61 Wei E, Ozdaglar A, Jadbabaie A. A distributed newton method for network utility maximization—Part II: Convergence. *IEEE Trans Automat Contr*, 2013, 58: 2176–2188
- 62 Varagnolo D, Zanella F, Cenedese A, et al. Newton-Raphson consensus for distributed convex optimization. *IEEE Trans Automat Contr*, 2016, 61: 994–1009
- 63 Mokhtari A, Ling Q, Ribeiro A. Network newton distributed optimization methods. *IEEE Trans Signal Process*, 2017, 65: 146–161
- 64 Bajović D, Jakovetić D, Krejić N, et al. Newton-like method with diagonal correction for distributed optimization. *SIAM J Optim*, 2017, 27: 1171–1203
- 65 Lu J, Tang C Y. Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case. *IEEE Trans Automat Contr*, 2012, 57: 2348–2354
- 66 Zeng J, Yin W. Extrapush for convex smooth decentralized optimization over directed networks. *J Comput Math*, 2017, 35: 381–394
- 67 Xi C, Khan U A. DEXTRA: A fast algorithm for optimization over directed graphs. *IEEE Trans Automat Contr*, 2017, 62: 4980–4993
- 68 Xi C, Xin R, Khan U A. ADD-OPT: Accelerated distributed directed optimization. *IEEE Trans Automat Contr*, 2018, 63: 1329–1339
- 69 Du W, Yao L, Wu D, et al. Accelerated distributed energy management for microgrids. In: Proceedings of the IEEE Power and Energy Society General Meeting. 2018
- 70 Xin R, Khan U A. A linear algorithm for optimization over directed graphs with geometric convergence. *IEEE Control Syst Lett*, 2018, 2: 325–330
- 71 Pu S, Shi W, Xu J, et al. A push-pull gradient method for distributed optimization in networks. In: Proceedings of the 57th IEEE Conference on Decision and Control (CDC). IEEE, 2018. 3385–3390
- 72 Gharesifard B, Cortes J. Distributed continuous-time convex optimization on weight-balanced digraphs. *IEEE Trans Automat Contr*, 2014, 59: 781–786
- 73 Kia S S, Cortés J, Martínez S. Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication. *Automatica*, 2015, 55: 254–264
- 74 Xie Y, Lin Z. Global optimal consensus for multi-agent systems with bounded controls. *Syst Control Lett*, 2017, 102: 104–111
- 75 Chen W, Ren W. Event-triggered zero-gradient-sum distributed consensus optimization over directed networks. *Automatica*, 2016, 65: 90–97
- 76 Bianchi P, Jakubowicz J. Convergence of a multi-agent projected stochastic gradient algorithm for non-convex optimization. *IEEE Trans Automat Contr*, 2012, 58: 391–405
- 77 Wai H T, Lafond J, Scaglione A, et al. Decentralized frank-wolfe algorithm for convex and nonconvex problems. *IEEE Trans Automat Contr*, 2017, 62: 5522–5537
- 78 Hong M, Hajinezhad D, Zhao M. Prox-PDA: The proximal primal-dual algorithm for fast distributed nonconvex optimization and learning over networks. In: Proceedings of the 34th International Conference on Machine Learning (ICML). 2017
- 79 Hajinezhad D, Hong M. Perturbed proximal primal-dual algorithm for nonconvex nonsmooth optimization. *Math Program*, 2019, 176: 207–245
- 80 Zeng J, Yin W. On nonconvex decentralized gradient descent. *IEEE Trans Signal Process*, 2018, 66: 2834–2848
- 81 Lorenzo P D, Scutari G. NEXT: In-network nonconvex optimization. *IEEE Trans Signal Inf Process Netw*, 2016, 2: 120–136
- 82 Tatarenko T, Touri B. Non-convex distributed optimization. *IEEE Trans Automat Contr*, 2017, 62: 3744–3757
- 83 Scutari G, Sun Y. Distributed nonconvex constrained optimization over time-varying digraphs. *Math Program*, 2019, 176: 497–544
- 84 Tian Y, Sun Y, Scutari G. ASY-SONATA: Achieving linear convergence in distributed asynchronous multiagent optimization. In: Proceedings of

- the 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton). IEEE, 2018. 543–551
- 85 Lee J D, Panageas I, Piliouras G, et al. First-order methods almost always avoid strict saddle points. *Math Program*, 2019, 176: 311–337
- 86 Jin C, Ge R, Netrapalli P, et al. How to escape saddle points efficiently. In: *Proceedings of the 34th International Conference on Machine Learning (ICML)*. 2017
- 87 Nesterov Y, Spokoiny V. Random gradient-free minimization of convex functions. *Found Comput Math*, 2017, 17: 527–566
- 88 Sahu A K, Jakovetić D, Bajovic D, et al. Distributed zeroth order optimization over random networks: A Kiefer-Wolfowitz stochastic approximation approach. In: *Proceedings of the 57th IEEE Conference on Decision and Control (CDC)*. IEEE, 2018. 4951–4958
- 89 D M Yuan, Ho D W C. Randomized gradient-free method for multiagent optimization over time-varying networks. *IEEE Trans Neural Netw Learning Syst*, 2015, 26: 1342–1347
- 90 Pang Y, Hu G. Randomized gradient-free distributed optimization methods for a multiagent system with unknown cost function. *IEEE Trans Automat Contr*, 2020, 65: 333–340
- 91 Wang Y, Zhao W, Hong Y, et al. Distributed subgradient-free stochastic optimization algorithm for nonsmooth convex functions over time-varying networks. *SIAM J Control Optim*, 2019, 57: 2821–2842
- 92 Hajinezhad D, Hong M, Garcia A. ZONE: Zeroth-order nonconvex multiagent optimization over networks. *IEEE Trans Automat Contr*, 2019, 64: 3995–4010
- 93 Karimi H, Nutini J, Schimit M. Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. 2016. 795–811
- 94 Tang Y, Li N. Distributed zero-order algorithms for nonconvex multi-agent optimization. In: *Proceedings of the 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2019. 781–786
- 95 Su Q Y. *Energy Internet and We-Energy*. Singapore: Springer, 2018
- 96 Yang T, Wu D, Fang H, et al. Distributed energy resource coordination over time-varying directed communication networks. *IEEE Trans Control Netw Syst*, 2019, 6: 1124–1134
- 97 Liu T, Tian J, Wang J Z, et al. Integrated security threats and defense of cyber-physical systems (in Chinese). *Acta Automat Sin*, 2019, 45: 5–24 [刘炆, 田决, 王稼舟, 等. 信息物理融合系统综合安全威胁与防御研究. *自动化学报*, 2019, 45: 5–24]
- 98 Sundaram S, Gharesifard B. Distributed optimization under adversarial nodes. *IEEE Trans Automat Contr*, 2019, 64: 1063–1076
- 99 Zhao C, He J, Wang Q. Resilient distributed optimization algorithm against adversary attacks. *IEEE Trans Autom Control*, 2019, doi: 10.1109/TAC.2019.2954363
- 100 Bach F. Distributed machine learning over networks. In: *Semi-plenary Talk of the 58th IEEE Conference on Decision and Control*. 2019
- 101 Bottou L, Curtis F E, Nocedal J. Optimization methods for large-scale machine learning. *SIAM Review*, 2018, 60: 223–311
- 102 Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge: MIT Press, 2016
- 103 Johnson R, Zhang T. Accelerating stochastic gradient descent using predictive variance reduction. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS)*. 2013. 315–323
- 104 Kingma D P, Ba J. Adam: A method for stochastic optimization. In: *Proceedings of the International Conference on Learning Representations (ICLR)*. 2015
- 105 Ruder S. An overview of gradient descent optimization algorithms. arXiv: 1609.04747v2
- 106 Zhu Z. Katyusha: The first direct acceleration of stochastic gradient methods. *J Mach Learn Res*, 2018, 18: 1–51
- 107 Tang H, Gan S, Zhang C, et al. Communication compression for decentralized training. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS)*. 2018. 7663–7673
- 108 Alistarh D, Grubic D, Li J, et al. QSGD: Communication-efficient SGD via gradient quantization and encoding. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)*. 2018. 1707–1718
- 109 Stich S U. Local SGD converges fast and communicates little. In: *International Conference on Learning Representations (ICLR)*. 2019
- 110 Stich S U, Cordonnier J B, Jaggi M. Sparsified SGD with memory. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS)*. 2018. 4452–4463
- 111 Bernstein J, Wang Y, Azizzadenesheli K, et al. SIGNSGD: Compressed optimisation for non-convex problems. In: *Proceedings of the 35th International Conference on Machine Learning (ICML)*. 2018
- 112 Lian X, Zhang C, Zhang H, et al. Can decentralized algorithms outperform centralized algorithms? A case study for decentralized parallel

- stochastic gradient descent. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS). 2017. 5336–5346
- 113 Assran M, Nicolas L, Ballas N, et al. Stochastic gradient push for distributed deep learning. In: Proceedings of the 36th International Conference on Machine Learning (ICML). 2019

Research status and prospects of distributed collaborative optimization

YANG Tao & CHAI TianYou

State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China

Distributed collaborative optimization is the effective implementation of optimization tasks through cooperation and collaboration of multiple agents. The theory and application of distributed collaborative optimization is one of the important development directions in Control Science and Engineering. In recent years, with the development of emerging technologies such as cloud computing, big data, mobile internet, and artificial intelligence, distributed collaborative optimization has been facing new challenges and opportunities. This paper reviews and summarizes some hot research directions of distributed collaborative optimization in recent years, including distributed accelerated optimization algorithms, distributed non-convex optimization algorithms, and distributed gradient-free optimization algorithms. Moreover, guided by three practical applications of intelligent manufacturing, Energy Internet and distributed machine learning, key future research directions of distributed collaborative optimization are prospected.

distributed collaborative optimization, cyber-physical systems, intelligent manufacturing, energy internet, distributed machine learning

doi: [10.1360/SST-2020-0040](https://doi.org/10.1360/SST-2020-0040)