第 35 卷第 7 期
2018 年 7 月

控 制 理 论 与 应 用
Control Theory & Applications

Vol. 35 No. 7
Jul. 2018

DOI: 10.7641/CTA.2018.80205

# 网络化分布式凸优化算法研究进展

谢 佩[1], 游科友[1†], 洪奕光[2], 谢立华[3]

(1. 清华大学 自动化系, 北京 100084; 2. 中国科学院数学与系统科学研究院 系统科学研究所, 北京 100080;

3.南洋理工大学 电气与电子工程学院, 新加坡 639798)

**摘要**: 分布式优化是指利用网络化多自主体之间的协作来求解的一类优化问题, 其在大规模数值计算、机器学习、资源分配、传感器网络等方面具有重要的研究意义和应用价值. 自主体之间的协作通常基于代数图来描述, 且图的结构对分布式优化算法的设计与性能有显著影响. 本文针对凸优化问题, 基于平衡图和非平衡图的情形, 简要讨论了分布式优化算法的最新研究进展, 并对今后的发展趋势和应用进行展望.

**关键词**: 分布式优化; 网络化算法; 网络结构; 多自主体系统

**引用格式**: 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

**中图分类号**: TP13　　**文献标识码**: A

# A survey of distributed convex optimization algorithms over networks

XIE Pei[1], YOU Ke-you[1†], HONG Yi-guang[2], XIE Li-hua[3]

(1. Department of Automation, Tsinghua University, Beijing 100084, China;
2. Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100080, China;
3. School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore)

**Abstract:** The distributed optimization problem is cooperatively solved by a network of agents, which has significant applications in the large-scale numerical computation, machine learning, scheduling, sensor networks and etc. The interaction among agents is usually described by an algebraic graph, whose structure greatly affects the design and analysis of distributed optimization algorithms. This work focuses on the convex optimization problem and reviews the-state-of-the-art research on distributed optimization algorithms under both balanced graphs and unbalanced graphs, respectively. We also provide some remarks on the future directions and applications of distributed optimization.

**Key words:** distributed optimization; networked algorithms; network structure; multi-agent system

**Citation:** 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

## 1 引言(Introduction)

大量工程实践、管理决策中的实际问题都可以建模为一个数学优化问题. 优化理论在自动控制、调度、系统集成、数值计算、机器学习、压缩感知等方面发挥着重要作用[1]. 随着数据规模的急剧增长, 集中式优化算法因受限于单机的计算瓶颈而难以应对大规模优化问题. 多机协作的分布式算法可大幅降低单机的计算负担, 逐渐被研究者关注. 同时, 大规模传感器网络的发展为分布式优化算法提供了丰富的应用场景, 如智能电网、车联网、无人机编队等.

多自主体系统(multi-agent system)通常指具备感知、通信、计算和执行能力的多个自主体组成的大规模系统, 常被用作分布式优化算法的实现载体. 基于多自主体系统实现的分布式优化算法效率, 不仅取决于每个自主体的计算能力, 还取决于自主体之间的协作. 为简化描述, 后文自主体亦简称为节点.

在优化算法研究中, 收敛性分析、复杂性分析等是研究的核心问题. 分布式优化算法将任务分配给多个节点进行计算, 节点间通过信息传递实现协作. 因此, 如何将优化问题转化为适合多自主体系统分布式求解的形式是首要问题.

根据优化问题的要素, 分布式优化问题包括决策

收稿日期: 2018–03–25; 录用日期: 2018–07–02.
†通信作者. E-mail: youky@tsinghua.edu.cn; Tel.: +86 10-62789635.
本文责任编委编委: 陈增强.
“2017年智能控制研讨会”论文, 黄琳院士推荐.
国家自然科学基金优秀青年基金项目(61722308), 清华大学自主科研计划项目资助.
Supported by the National Natural Science Foundation of China for Excellent Young Scholar (61722308) and the Tsinghua University Initiative Scientific Research Program.

变量的分布式处理、目标函数的分布式处理和约束条件的分布式处理. 其中决策变量的分布式处理适用于高维决策变量的清醒, 如大规模线性规划[2]等问题. 解决此类问题多是基于Guass-Seidel迭代[3], 将高维决策向量拆分为多个低维决策向量, 并通过局部优化和节点之间的协作来求解全局最优. 常见算法包括交替方向乘子法(alternating direction method of multipliers, ADMM)[4]、坐标块下降法[5–6]等.

目标函数的分布式处理的应用范围更为广泛. 例如, 机器学习问题中由训练误差构成的损失函数可以看成多个局部损失函数的加和, 进而利用多个节点进行分布式处理. 若待优化的目标函数存在正则项, 可将其平均分配到每个节点来实现分布式处理. 类似应用还包括协同定位[7]、资源分配[8]等问题.

约束条件的分布式处理主要针对具有大量约束集的优化问题, 可将约束集视作多个局部约束的交集, 设计分布式算法使得每个节点的决策变量最终收敛到同一个值且同时满足各自的局部约束.

本文主要讨论一类分布式凸优化问题, 其目标函数为多个局部目标函数的加和, 而约束是多个局部约束集的交集. 事实上, 很多优化问题都可以转化为此类问题进行求解. 具体形式为

$$\begin{cases} \min\limits_{x\in X}\ f(x)=\sum\limits_{i=1}^{n} f_i(x),\\ \text{s.t. } x\in \bigcap\limits_{i=1}^{n} X_i, \end{cases} \tag{1}$$

其中: $f_i : X(\subseteq \mathbb{R}^m) \to \mathbb{R}$是节点$i$的代价函数且只能被节点$i$获得, $X_i$是局部约束集且只可被节点$i$获得, $f$为总代价函数, 节点之间通过协作来求解优化问题. 本文主要针对$f_i$为凸函数且$X, X_i$均为凸集的情况展开讨论. 对于非凸分布式优化问题目前也有一些研究成果, 如文献[9–10]等. 但限于篇幅, 本文仅在应用与展望一节进行简要介绍.

求解问题(1)的基本思路是每个节点$i$通过与相邻节点交换决策变量$x_i^k$, 并融合其局部代价函数与约束集来更新$x_i^k$, 最终实现所有节点的决策向量渐近收敛到同一最优解, 即$x_i^k \to x^* \in X^*$, 其中$X^*$是优化问题的最优解集合. 因此其亦被称为基于状态趋同[11–19]的分布式优化算法.

分布式优化问题的一个重要特点是算法设计依赖于节点组成的通信网络拓扑结构. 在相关研究中, 各节点间的协作关系通常使用代数图论进行描述. 例如, 当节点之间双向传递信息时, 可使用无向图来描述通信网络; 当节点单向传递信息时, 需要使用有向图来描述通信网络. 当通信链路存在失效、延迟、更替等现象时, 应使用图序列来描述时变的通信网络. 针对不同的通信网络, 往往需要设计不同形式的分布式算法. 因此, 本文将基于通信网络结构简述分布式优化算法的研究进展.

文章余下部分内容组织如下: 在第2节, 将分别针对无约束优化和带约束优化问题介绍适用于平衡通信网络的几种常用分布式优化算法; 在第3节中, 介绍几类消除非平衡通信网络不对称性的改进算法; 在第4节, 针对时变非平衡网络、随机网络、异步通信网络等, 介绍分布式优化算法的研究进展; 在第5节, 介绍分布式优化的其他研究, 包括收敛性分析理论、算法加速、算法性能评估等重要问题; 最后, 分布式优化算法的应用和展望将在第6节给出.

## 2 平衡图下的分布式凸优化算法(Distributed convex optimization algorithms over balanced graphs)

采用图$\mathcal{G}=\{\mathcal{V},\mathcal{E},A\}$描述节点间的连接情况, 其中: $\mathcal{V}:=\{1,\cdots,n\}$表示$n$个节点构成的集合; $\mathcal{E}$表示相邻节点构成的边集,即$(i,j)\in\mathcal{E}$当且仅当节点$i$从节点$j$直接获得信息, 并记$\mathcal{N}_i=\{j|(i,j)\in\mathcal{E}\}$为节点$i$的邻居; 非负矩阵$A=[a_{ij}]\in\mathbb{R}^{n\times n}$为图$\mathcal{G}$的权重矩阵, 满足$a_{ij}>0$当且仅当$(i,j)\in\mathcal{E}$. 如果对于任意$i,j\in\mathcal{V}$, 有$\sum\limits_{k=1}^{n} a_{ik}=\sum\limits_{k=1}^{n} a_{kj}$, 则称图$\mathcal{G}$为平衡图, 反之为非平衡图. 此外平衡图中常假定$A$为双随机矩阵, 即$\sum\limits_{k=1}^{n} a_{ik}=\sum\limits_{k=1}^{n} a_{kj}=1$. 若$A^{\mathrm{T}}=A$, 图$\mathcal{G}$为无向图. 若对任意$i,j\in\mathcal{V}$, 存在一条从$j$到$i$的路径使得路径上的每条边属于$\mathcal{E}$, 则称图$\mathcal{G}$为强连通. 本节将在强连通的假定下, 针对无向图和平衡图不加区分地介绍相关的分布式优化算法研究成果.

### 2.1 分布式无约束优化(Distributed unconstrained optimization)

分布式无约束优化问题的目标函数为局部凸函数之和, 且无局部约束, 即$X_1=\cdots=X_n=\mathbb{R}^m$. 特别地, 在本节中对目标函数仅假定凸性, 在第5.2节中将进一步讨论光滑性对算法收敛速度的影响. 因篇幅有限, 本节仅讨论文献中4种典型的求解思路, 即分布式次梯度方法、分布式对偶平均算法、辅助约束乘子法和非光滑惩罚项法.

Nedic等[20]将分布式趋同算法和次梯度方法结合, 提出如下分布式次梯度方法:

$$x_i^{k+1}=\sum_{j\in\mathcal{N}_i} a_{ij}x_j^k-\alpha^k d_i^k, \tag{2}$$

其中: $d_i^k$是$f_i(x)$在$x_i^k$处的次梯度, 且$\|d_i^k\|$一般假设一致有界; $\alpha^k>0$是满足一定条件的步长. 该工作证明了当$A=[a_{ij}]$为双随机矩阵, 图$\mathcal{G}$强连通, 且步长满足如下条件时:

$$\sum_{k=1}^{\infty}\alpha^k=\infty,\ \alpha^k\to 0, \tag{3}$$

所有节点的决策变量将收敛到问题(1)的同一个最优解$x^* \in X^*$, 即$x_i^k \to x^*$. 算法的收敛速度为$O(1/\sum_{i=1}^{k}\alpha^i)$, 当步长为$\alpha^k = 1/k^{0.5+\epsilon}(\epsilon \to 0^+)$, 达到理论最快收敛速度为$O(1/\sqrt{k})$, 这与集中式次梯度方法的最快收敛速度吻合. 当取定常步长时, 算法将收敛到最优解的一个邻域. 值得一提的是, 若优化问题存在全局约束, 可将算法(2)每次迭代的计算结果投影全局约束集上进行处理[21], 此算法称为分布式次梯度投影算法. 当目标函数为强凸时, Liu等[22]通过选取合适的步长, 可将分布式次梯度投影算法的收敛速度提升为$O(1/k)$.

Dunchi等[23]则提出了如下分布式对偶平均(dual averaging)算法:

$$z_i^{k+1} = \sum_{j\in\mathcal{N}_i} a_{ij}z_j^k + g_i^k, \tag{4a}$$

$$x_i^{k+1} = \mathrm{prox}_X^{\psi}(z_i^{k+1}, \alpha^k), \tag{4b}$$

其中: $g_i^k$是$f_i$在对偶变量$z_i^k$处的次梯度, $\alpha^k$为步长, 近端函数(proximal function)$\psi$为强凸, 即

$$\psi(y) \geqslant \psi(x) + \langle\nabla\psi(x), y-x\rangle + \frac{1}{2}\|x-y\|^2.$$

对应的近端算子(proximal operator)定义为

$$\mathrm{prox}_X^{\psi}(z,\alpha) = \underset{x\in X}{\mathrm{argmin}}\{\langle z, x\rangle + \frac{1}{\alpha}\psi(x)\}.$$

若选择合适步长, 算法执行过程中各节点的历史平均状态$\hat{x}_i^t = \frac{1}{t}\sum_{k=1}^{t}x_i^k$将以速度$\mu(\mathcal{G},\psi)\log t/\sqrt{t}$收敛到原问题的最优解, 其中$\mu(\mathcal{G},\psi)$只取决于通信拓扑结构和近端函数$\psi$.

通过引入一个辅助变量, 可以将无约束情形下的问题(1)写成如下带等式约束的优化问题:

$$\begin{cases}\min \sum_{i=1}^{n} f_i(x_i),\\ \text{s.t. } x_i = z,\ \forall i \in \mathcal{V}.\end{cases} \tag{5}$$

针对该问题, 可引入带惩罚项的Lagrange函数, 并使用对偶解耦或ADMM算法[24]对其进行求解. 若通信图强连通, 问题(1)还等价于

$$\begin{cases}\min \sum_{i=1}^{n} f_i(x_i),\\ \text{s.t. } x_i = x_j,\ \forall (i,j) \in \mathcal{E}.\end{cases} \tag{6}$$

每个等式约束对应于通信图中的一条边, 也同时对应一个拉格朗日乘子$\lambda_{ij}$. 在此问题基础上, Wei等[25]提出了交替方向乘子法的分布式实现, 并证明其收敛速度能够达到$O(1/k)$. 但该算法仅适用于无向图. 有关ADMM的研究很多, 由于篇幅有限本文不详细介绍, 可参考Boyd等的综述[24].

类似地, Zhang等[26]也采用了问题(6)的形式, 但将约束表示为非光滑惩罚项. 该工作证明了当$\rho$大于一个显式下界时, 问题(6)与下面的惩罚问题等价:

$$\min \sum_{i=1}^{n} f_i(x_i) + \frac{\rho}{2}\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_i}|x_i - x_j|. \tag{7}$$

通过用次梯度方法求解, 其得到如下分布式算法:

$$x_i^{k+1} = x_i^k - \rho\alpha^k\sum_{j\in\mathcal{N}_i}\mathrm{SGN}(x_i^k - x_j^k) - \alpha^k d_i^k, \tag{8}$$

其中SGN$(\cdot)$为1–0符号函数. 此算法具有和分布式次梯度方法基本相同的收敛速度, 其优势是只需利用相邻节点状态的偏序关系而不需准确的相对状态信息.

### 2.2 分布式带约束优化(Distributed constrained optimization)

分布式带约束优化问题中, 每个节点包含局部约束$x \in X_i$. 其中最简单的情况是目标函数为常数, 即求解可行性问题$x \in \bigcap_{i=1}^{n} X_i$. 由于局部约束集合$X_i$被假设为凸集, 该问题又称为分布式凸交问题, 在协同定位问题[27]中有着广泛的应用.

Nedic等[28]最早给出了解决此问题的离散时间分布式投射同步算法, 具体形式如下:

$$x_i^k = P_{X_i}(\sum_{j=1}^{n} a_{ij}x_j^k), \tag{9}$$

其中$P_{X_i}(\cdot)$表示投影到凸集$X_i$的算子. 在无向图和强连通图的假设下, 该算法将使得各节点渐进收敛到所有局部凸约束的交集(若非空). 然而除了规则的凸集(如球、超平面等), 大部分凸集的投影计算复杂度较高. 为减少投影计算带来的运算复杂性, Lou等[29]提出了以近似投影方法, 或是在迭代时以一定概率使用投影算子, 减少投影算子使用的频率[30].

当约束$X_i$具有特定结构时, 如当$X_i$为线性等式约束时, 其本质上为分布式求解线性方程的问题. 但常用的求解方法如Kaczmarz算法[31–32]的执行方式是串行的, 因此You等[33]将Kaczmarz算法和趋同算法结合来分布式地求解线性方程. 当$X_i$为凸不等式约束时, Polyak等[34]则提出随机近似投影算法来求解.

对于一般的带约束优化问题(1), Nedic等[28]提出了分布式次梯度投影算法, 即在分布式次梯度法基础上增加了局部约束集的投影操作, 具体形式如下:

$$x_i^{k+1} = P_{X_i}(\sum_{j=1}^{n} a_{ij}x_j^k - \alpha^k d_i^k), \tag{10}$$

其中$\alpha^k$为满足条件(3)的步长, 并证明了在一些特定情况下算法的收敛性. Liu等[35]则在此基础上提出一种可使用常数步长的算法, 与式(10)中的衰减步长相比具有更快的收敛速度. 尽管次梯度投影法可以解决带约束分布式优化问题, 但面对复杂约束集$X_i$, 投影计算开销较大, 从而降低了算法的实用性. 因此针对

特定的约束形式时, 需要采取特定的处理方法.

例如, 针对包含全局不等式约束的优化问题, 研究者提出了基于趋同的分布式原始对偶(primal dual)方法[36–37]. 针对局部不等式约束, Lee等[38]则提出了在不等式约束上近似投影的方法, 降低计算精确投影的成本; You等[39]则提出了一种适合于无向图的分布式原始对偶方法, 以处理局部不等式约束.

此外, 还可以使用惩罚项方法, 将带约束问题转化为带惩罚项的无约束问题, 详见Aybat等[40]的工作.

## 3 非平衡图下的分布式凸优化算法(Distributed convex optimization algorithms over unbalanced graphs)

在非平衡通信网络$\mathcal{G}=\{\mathcal{V},\mathcal{E},A\}$中, 节点间通信是单向的且可能存在$i,j$使得$\sum\limits_{k=1}^{n}a_{ik}\neq\sum\limits_{k=1}^{n}a_{kj}$. 一般假定$A$仅为行随机矩阵, 即$\sum\limits_{k=1}^{n}a_{ik}=1$. 给定节点$i$, 向其发送信息的节点被称为“入邻居”, 接收节点$i$传出消息的节点被称为“出邻居”. 节点$i$的入邻居和出邻居构成的集合分别被记作

$$\mathcal{N}_i^{\rm in}=\{j|(i,j)\in\mathcal{E}\},\ \mathcal{N}_i^{\rm out}=\{j|(j,i)\in\mathcal{E}\}.$$

在非平衡图中直接使用仅适用于平衡图的分布式算法往往不能得到原问题最优解. 例如, Xi等[41] 解释了分布式次梯度法(2)在非平衡图下本质上是在优化局部函数的加权和, 即$\sum\limits_{i=1}^{n}\pi_if_i(x)$, 其中$\pi'A=\pi'$. 因此算法设计的关键是处理这种非平衡性. 本节将讨论4类处理非平衡性的算法, 即基于权重矩阵左特征向量的算法、基于Push-Sum协议的算法、基于盈余变量的算法、基于上境图形式的算法.

### 3.1 基于权重矩阵左特征向量的算法 (Perron vector based algorithms)

分布式次梯度算法(2)在非平衡图中本质在优化加权目标函数$\hat{F}(x)=\sum\limits_{i=1}^{n}\pi_if_i(x)$, 其中$\sum\limits_{i=1}^{n}\pi_i=1$, 且$[\pi_1,\cdots,\pi_n]A=[\pi_1,\cdots,\pi_n]$是权重矩阵$A$关于特征值1的左特征向量. 注意到

$$\sum_{i=1}^{n}f_i(x)=\sum_{i=1}^{n}\pi_i\frac{f_i(x)}{\pi_i},$$

那么在非平衡图中, 只需要将$f_i(x)/\pi_i$视作新的局部函数, 再使用分布式次梯度法即可抵消权重$\pi_i$带来的影响. 具体而言, 每个节点仅需执行如下迭代算法即可保证收敛到原问题最优解:

$$x_i^{k+1}=\sum_{i=1}^{n}a_{ij}x_j^k-\alpha^k\frac{d_i^k}{\pi_i}. \tag{11}$$

算法(11)中的$\pi_i$由全局通信图的权重矩阵决定, 单个节点并不知道$\pi_i$的确切值. 于是, Morral等[42]提出了分布式估计$\pi_i$的方法, 再与式(11)相结合, 即可解决非平衡图下的分布式优化问题.

此算法直观,形式简洁,但局限性也明显, 即只能应用在固定的非平衡图上. 因为一旦$A$随时间变化, $\pi_i$不是常数而无法估计, 算法便不可行.

### 3.2 基于Push-Sum协议的算法(Push-Sum based algorithms)

Push-Sum协议最早被提出用于解决有向图中的平均趋同问题[43]. 其核心思想是每个节点向外发送其加权后状态, 以消除网络不平衡性. 基于Push-Sum协议, Tsianos等[44]最早将分布式对偶平均算法拓展到非平衡图中, 设计了收敛较快的分布式优化算法.

Nedic等[45]则将分布式次梯度法和Push-Sum协议相结合, 提出了如下subgradient-push算法:

$$w_i^{k+1}=\sum_{j\in\mathcal{N}_i^{\rm in}}\frac{x_j^k}{d_j^{\rm out}}, \tag{12a}$$

$$y_i^{k+1}=\sum_{j\in\mathcal{N}_i^{\rm in}}\frac{y_j^k}{d_j^{\rm out}}, \tag{12b}$$

$$z_i^{k+1}=\frac{w_i^{k+1}}{y_i^{k+1}}, \tag{12c}$$

$$x_i^{k+1}=w_i^{k+1}-\alpha^{k+1}\partial f_i(z_i^{k+1}), \tag{12d}$$

其中: $d_j^{\rm out}$为节点$j$的出度, 即其“出邻居”集合$\mathcal{N}_j^{\rm out}$中的元素个数, $\partial f_i(z_i^{k+1})$为$f_i$在$z_i^{k+1}$处的次梯度. Ndeic等证明了该算法在强连通非平衡图网络中的收敛性. 由于算法(12)的形式比较复杂, Nedic等[46]针对目标函数为强凸的情况亦给出了上述算法的随机化形式, 减小了算法的计算量.

基于Push-Sum协议的分布式优化算法要求节点可获取自身的出度,但这在某些通信模式下不一定能实现. 同时, 该算法不易推广至带约束的优化问题.

### 3.3 基于盈余变量的算法 (Surplus based algorithms)

Xi等[41]提出了一种基于盈余变量(surplus)的算法. 该算法通过引入了盈余变量$y_i^k$, 并令其与$x_i^k$同时迭代更新, 使得$y_i^k$收敛到0的同时, $x_i^k$也能收敛到局部目标和函数的最小值点. 具体形式如下:

$$x_i^{k+1}=\sum_{j=1}^{n}a_{ij}x_j^k+\epsilon y_i^k-\alpha^k\nabla f_i(x_i^k), \tag{13a}$$

$$y_i^{k+1}=x_i^k-\sum_{j=1}^{n}a_{ij}x_j^k+\sum_{j=1}^{n}b_{ij}y_j^k-\epsilon y_i^k, \tag{13b}$$

其中: $B=[b_{ij}]\in\mathbb{R}^{n\times n}$为列随机矩阵, $\epsilon>0$为某个合理选择的参数. 盈余变量$y_i^k$在向外发送时控制发出变量的总“份额”为1, 即保证$B$为列随机矩阵, 而节点在接收$x_j^k$时控制总“份额”为1, 即保证$A$为行随

机矩阵. 为此, 引入如下记号:

$$\begin{cases} z_i^k = \begin{cases} x_i^k, & i \in \{1, \cdots, n\}, \\ y_{i-n}^k, & i \in \{n+1, \cdots, 2n\}, \end{cases} \\ g_i^k = \begin{cases} \nabla f_i(x_i^k), & i \in \{1, \cdots, n\}, \\ 0_m, & i \in \{n+1, \cdots, 2n\}, \end{cases} \\ M = \begin{bmatrix} A & \epsilon I \\ I - A & B - \epsilon I \end{bmatrix}, \end{cases} \tag{14}$$

则算法(13)可以写成如下形式:

$$z_i^k = \sum_{j=1}^{2n} [M]_{ij} z_j^k - \alpha^k g_i^k. \tag{15}$$

式(15)具有分布式梯度下降算法的形式, 但$M$不是双随机矩阵甚至不是非负矩阵. 然而$M$具有与双随机矩阵相似的一些性质, 由此可以证明算法(13)在非平衡图下$x_i^k$将收敛到问题最优解, 且盈余变量$y_i^k$收敛到0. 此算法也要求节点可获得自身的出度信息.

基于盈余变量的方法最早由Cai等[47–48]提出, 并用来解决非平衡图中的平均趋同问题. 在固定非平衡图中, Xi等[49]还在算法(13)的基础上加入投影算子, 解决了带全局约束的分布式优化问题.

### 3.4 基于上境图形式的算法(Algorithms using epigraphs)

分布式优化问题包含局部约束时, 前3类针对非平衡图的算法无法直接使用. 尽管分布式ADMM算法可解决带局部约束的优化问题, 但要求通信图是无向的. You等[39]将分布式次梯度法和Polyak随机投影算法[50]结合, 提出了一种结构简单的分布式算法, 用来求解在非平衡网络中包含由凸函数表示的不等式局部约束的优化问题, 但目标函数是线性的. 受此启发, Xie等[51]将目标函数为凸函数加和的分布式优化问题转化成其上境图形式进行求解. 通过引入辅助变量来刻画局部目标函数的上界, 将目标函数转化为由辅助变量构成的单一线性函数, 而原先的局部目标函数则转化局部约束. 最后通过改进You等[39]的算法来解决非平衡图下带局部约束的优化问题. 值得一提的是, 该算法通过引入额外的固定方向Polyak投影, 提高了算法的收敛稳定性.

## 4 一般通信网络下的分布式凸优化算法(Distributed convex optimization algorithms over general networks)

前文介绍的分布式优化算法大多数仅针对固定的通信网络, 即网络结构不随时间变化. 然而在实际情形中, 多自主体系统节点间通信时可能存在丢包、延时和链路更替等情况, 有的时候通信形式还可能是随机的甚至异步的. 本节将针对时变非平衡网络、随机通信网络和异步通信网络介绍相应的分布式优化算法的研究成果.

### 4.1 时变非平衡通信网络(Time-varying unbalanced communication networks)

多自主体间时变非平衡通信网络采用有向图序列$\mathcal{G}(k) = \{\mathcal{V}, \mathcal{E}^k, A^k\}$表示, 其中: $\mathcal{E}^k$表示在$k$时刻节点间的连接情况, 非负矩阵$A^k = [a_{ij}^k] \in \mathbb{R}^{n\times n}$表示$k$时刻图的权重, 且$A^k$在任意$k$时刻为行随机矩阵.

Nedic等[45]在针对时变非平衡图网络设计分布式优化算法的过程中, 对图的连通性提出了如下假设:

**假设 1** 存在正整数$B$, 使得对于任意$l \geqslant 0$, 联合图$\mathcal{G}_B^l = \{\mathcal{V}, \mathcal{E}_B^l\}$是强连通的, 其中联合图的边集为$\mathcal{E}_B^l = \bigcup_{k=lB}^{(l+1)B-1} \mathcal{E}^k$.

在假设1下, Nedic等[52]使用Push-Sum协议的思想改进了分布式带跟踪技术的非精确梯度算法(distributed inexact gradient with tracking technique, DIGing), 设计了在时变非平衡图中快速收敛的分布式优化算法. 在此算法的基础上, Nedic等[53]又提出了自适应组合DIGing分布式优化算法, 使得每个节点迭代时可以采用不同的步长.

Xie等[54]提出分布式次梯度Polyak投影算法, 其在假设1下可以利用时变非平衡图网络解决带局部约束的优化问题. 算法的核心是将原问题转化为上境图形式, 并利用转化后的单一线性目标函数梯度恒定这一特点, 使得非平衡性不再影响算法的收敛结果.

在时变非平衡通信网络中, 切换图是一种特殊的通信形式. 切换图随时间变化, 但可以看作一个有限状态机. 在切换图中, Lou等[55]提出了求解零和博弈最优点的算法, 其本质为求解带约束优化问题的鞍点.

### 4.2 随机通信网络(Stochastic communication networks)

尽管非平衡时变通信网络允许节点间连接结构随时间变化, 但通常要求在给定时间段内是联合连通的(即假设1). 然而在某些随机网络[56]中无法保证此类连通性假设. 随机通信网络常采用随机过程描述节点连接状态, 如Burt等[57]中提到的反熵(anti-entropy)网络. 为研究随机网络下的分布式优化, Tahbaz等[58]提出了随机图模型和相关连通性假设, 文献[59–60]也进行了类似的定义, 即如下假设:

**假设 2** 随机图序列$\mathcal{G}^k$的权重矩阵$A^k$是独立同分布的, 且满足存在$\gamma > 0$, 使得$a_{ij}^k \geqslant \gamma$若$a_{ij}^k > 0$. 同时记$\bar{A} = \mathrm{E}(A^k)$, $\bar{\mathcal{E}} = \{(i,j)|\bar{A}_{ij} > 0\}$, 那么随机图序列的期望$\bar{\mathcal{G}} = \{\mathcal{V}, \bar{\mathcal{E}}, \bar{A}\}$是强连通的.

若假设2成立且$\bar{\mathcal{G}}$为平衡图, Lobel等[61]证明了分布式次梯度法在随机网络中的渐近收敛性, 同时还给出了算法步长为常数时收敛误差的上界.

另一种随机网络为反熵(anti-entropy)无向网络: 在给定时刻$k$, 仅单个节点$I_k$被激活, 与它的邻近节点$J_k$交换信息. Lee等[62]针对此类网络, 提出了异步平均随机投影算法, 用来解决带约束的分布式优化问题. 该算法允许每个节点存在无穷个局部约束, 如带不确定参数的鲁棒约束.

### 4.3 异步通信网络(Asynchronous communication networks)

在同步通信网络中, 节点使用统一的时钟来保证算法迭代的同步. 同步的分布式优化算法需等待最慢节点计算结束再进行下次迭代, 故算法执行效率受限于最慢节点的处理速度. 而异步通信网络的分布式算法中节点不需要等待其他节点的计算结果即可进行迭代, 可提升算法执行效率. Hannah等[63]给出了异步算法相对于同步算法性能提升的定量分析.

异步网络本质上是存在未知信息延迟的时变通信网络, 异步的分布式优化算法典型形式如下:

$$x_i^{k+1} = h_i^k(x_i^k, [x_{ij}^{k-\tau_{ij}}]_{j\in\mathcal{N}_i^{\text{in}}}), \tag{16}$$

其中$\tau_{ij}$为信息从节点$j$传递到节点$i$的时间延迟. 在设计异步算法时, 通常会对延迟$\tau_{ij}$做一定的假设.

当通信延迟$\tau_{ij}$存在上界时, Tsitsiklis等[64]证明了异步网络下特定线性系统的收敛性. 针对此假设下的异步通信网络, 研究者们提出了异步交叉方向乘子法[65]、异步近端坐标下降法[66]、异步近端原始对偶方法[67]、异步牛顿法[68]等.

当通信延迟$\tau_{ij}$不存在上界而是使用随机过程进行描述时, Sun等[69]提出一种异步坐标下降法, 其在通信延迟符合泊松过程时概率意义下收敛. Xu等[70]将适应组合机制(adapt-then-combine)应用在异步梯度算法中, 证明了定常步长下异步算法的收敛性.

其他工作还包括Modi等[71]提出的异步等式约束分布式优化算法(asynchronous distributed constraint optimization with quality guarantees, ADOPT)算法及其改进算法[72], 其核心思想是在通信网络中采用深度优先搜索遍历各节点状态的可行性和最优性. 异步分布式优化问题中还有些基于事件触发的算法, 如Zhong等[73]的工作.

## 5 分布式优化其他研究(Other topics on distributed optimization)

前3节围绕不同网络通信形式介绍了典型的分布式优化算法, 本节将探讨分布式优化其他研究热点问题, 包括分布式优化的收敛性分析、算法加速和算法总体性能分析.

### 5.1 收敛性分析(Convergence analysis)

分布式优化算法的收敛需要节点状态同时满足趋同性、可行性和最优性. 其中趋同性分析基于代数图论[74]、控制理论[75]和矩阵论[76]等理论, 而可行性和最优性的分析则利用了凸分析[77]、梯度分析[78]和对偶理论[79], 上鞅定理[80]是分析随机算法收敛性的常用工具. 影响分布式优化算法的收敛速度的因素包括通信网络结构、迭代步长等. 值得注意的是, 分布式优化算法的收敛性分析与优化问题的结构和通信图的结构密切相关.

停止准则是优化算法的核心问题之一. 在集中式算法中, 常通过设置最大迭代次数、考察目标函数下降速度或判断目标函数梯度是否近似零来决定停止算法迭代. 然而在分布式优化算法中, 由于单个节点无法直接获取所有其他节点的信息, 将难以通过局部信息为节点设定停止准则. 目前有一些文献[81–85]针对分布式趋同算法设计了基于局部信息的停止准则.

### 5.2 算法加速(Accelerated algorithms)

对于非光滑的凸优化问题, 集中式次梯度方法最快收敛速度只能达到$O(1/\sqrt{k})$, 这也是分布式次梯度方法的最优收敛速度. 然而, 当优化问题满足某些连续和光滑条件时, 往往可以设计收敛速度更快的算法. 例如, 可假设$f_i(x)$可微且满足Lipshitz条件:

$$f_i(x) \leqslant f_i(y) + (y-x)'\nabla f_i(x) + \frac{L}{2}\|y-x\|^2.$$

在有些情形下还假设$f_i(x)$强凸, 即存在$\mu > 0$使得

$$f_i(x) \geqslant f_i(y) + (y-x)'\nabla f_i(x) + \frac{\mu}{2}\|y-x\|^2.$$

比值$\eta = L/\mu$被称为$f_i(x)$的条件数, 若函数不是强凸, 则$\mu = 0$(或$\eta = +\infty$). 若目标函数Lipshitz可微且强凸, Nesterov等[78]证明了一阶算法的最快收敛速度为$O((1-\eta^{-0.5})^k)$; 若目标函数非强凸($\eta = +\infty$), 最快收敛速度为$O(1/k^2)$. 显然该收敛速度比一般性的次梯度算法要快. 为加快收敛速度, Nesterov 提出了外插 (extrapolation) 技术, 而该思想也被广泛应在在很多分布式优化算法中. 例如 Jakovetic 等[86]使用该技术, 提出了如下分布式 Nesterov 梯度算法 (distributed Nesterov gradient, D–NG):

$$x_i^k = \sum_{j=1}^n a_{ij}^k y_j^{k-1} - \alpha^{k-1}\nabla f_i(y_i^{k-1}), \tag{17a}$$

$$y_i^k = x_i^k + \frac{k-1}{k+2}\left(x_i^k - x_i^{k-1}\right). \tag{17b}$$

该算法的收敛速度达到$O(\log k/k)$, 这虽然比分布式次梯度算法的 $O(1/\sqrt{k})$ 要快, 但却比集中式算法的 $O(1/k^2)$要慢. 为改进这一速度, Jakovetic等[86]通过在单次迭代中多次执行趋同算法, 提出了带趋同迭代的分布式 Nesterov 梯度 (distributed Nesterov gradient with consensus iterations)算法, 并使用定步长将算法收敛速度提升为$O(\log k/k^2)$. 但该算法第$k$次迭代需进行$k$次趋同迭代和通信, 带来了额外的通信成本.

当局部目标函数可以拆分为光滑局部函数和一个

非光滑公共函数时, Chen等[87]通过引入proximal算子处理非光滑项, 并使用D–NC算法处理光滑项, 将非光滑问题的收敛速度提高到了$O(1/k)$, 但同样每次迭代需要进行多次趋同通信.

D–NG和D–NC算法在一定程度上改进了分布式次梯度方法的收敛速度, 但难以达到集中式Nesterov算法的收敛速度. 因此有研究者通过引入目标函数梯度的外插来进一步提升收敛速度. 如Shi等[88]提出了分布式精确一阶算法(exact first-order algorithm, EXTRA), 其矩阵形式如下:

$$x^{k+2}=(I+A)x^{k+1}-Hx^{k}-\alpha\left(\nabla f(x^{k+1})-\nabla f(x^{k})\right), \tag{18}$$

其中矩阵$H$为矩阵$I$和矩阵$A$的线性组合. 此算法的优点是步长$\alpha$可以为常数, 收敛速度达到$O(1/k)$, 且对强凸的目标函数收敛速度能达到$O(\tau_{\rm E}^{k})$, 其中$0<\tau_{\rm E}<1$. EXTRA算法具有很快的收敛速度, 但仅适用于平衡图. 此后研究者将其推广到了非平衡图中. 例如, 将EXTRA算法和特征向量估计结合, Xi等[89–90]提出了有向图EXTRA(directed EXTRA)算法和加速的有向图分布式优化(accelerated distributed directed optimization)算法; Zeng等[91]则将EXTRA算法与Push-Sum算法相结合, 提出了EXTRA–PUSH算法.

和EXTRA思想类似, Xu等[91]提出另一类基于梯度差分的分布式优化算法, 即DIGing算法[52], 在强凸问题中具有几何收敛速率. 算法的矩阵形式如下:

$$x^{k+1}=A(x^{k}-Dy^{k}), \tag{19a}$$

$$y^{k+1}=A(y^{k}+\nabla f(x^{k+1})-\nabla f(x^{k})), \tag{19b}$$

其中$D$为非负对角矩阵. 相对于EXTRA, 该算法的优势在于每个节点可以选择不同的常数步长.

最近, Qu等[92]提出一种收敛速度更快的分布式优化算法. 该算法结合D–NG算法和EXTRA算法的优点, 同时利用变量外插和梯度外插, 并将其线性组合作为更新结果. 分别针对强凸函数和非强凸函数, 设计了两种快速Nesterov梯度算法, 收敛速度分别为$O(\tau_{\rm A}^{k})$和$O(1/k^{1.4})$, 其中$0<\tau_{\rm A}<\tau_{\rm E}<1$. 但与D–NC不同, 该算法不需要在每次迭代中为实现状态的趋同而进行多次通信.

若在分布式优化算法执行过程中允许交换二阶信息, 可使用牛顿法. 但牛顿法需要使用全局信息来计算目标函数Hessian矩阵的逆, 单个节点无法直接完成. 因此, Ribeiro课题组[68]构造了基于趋同约束的拉格朗日函数, 并提出了分布式BFGS拟牛顿法来求解拉格朗日函数的鞍点. 在无向图中, 该算法的对偶变量的收敛速度可达$O(1/k)$.

### 5.3 分布式优化算法性能评估(Performance evaluation of distributed optimization algorithms)

分布式优化算法的性能指标可归结为各节点均以某精度收敛到最优解所花费的时间. 其影响因素包括节点数目, 通信结构, 单次迭代计算时间, 总迭代次数等. 以增加网络节点为例, 节点数目的增加可降低单个节点的计算负荷. 然而这也增加了通信成本, 亦可能导致总迭代次数的增加. 正如Tsianos等[93]所描述, 分布式优化算法的性能是计算成本和通信成本的平衡. Nedic等[94]对常见分布式优化算法的性能和通信方式之间的关系进行了详细讨论. 仅就通信性能来说, 其耗时与网络的拓扑结构、通信带宽、传输数据量等多个因素相关, 这加大了评估算法性能的难度.

## 6 应用与展望(Application and expectation)

本文针对分布式凸优化问题介绍了不同通信图结构下的研究进展, 并对此领域的研究热点进行了简要讨论. 事实上, 针对非凸问题的分布式优化算法也有相关研究. 例如, 使用分布式的方法训练深度神经网络[95–96], 使用对偶方法分布式求解非凸问题[97]等. 而当约束集为非凸集时, Lin等[9–10]提出了一种无向图网络下的最优趋同算法. 但限于篇幅, 非凸问题的研究成果本文不再详细讨论. 分布式优化算法在实际问题中的应用大致归纳如下:

• 分布式估计问题. 分布式优化可应用于统计参数的估计[98–99], 以及物理系统中的估计问题, 如传感器网络的协同定位[100–101]等.

• 分布式统计学习问题. 分布式优化可应用于机器学习和统计学习领域, 如分布式ADMM算法在优化训练样本损失中的应用[24]、分布式优化算法在Logistic回归及Lasso[102]、模型预测控制[103]等问题中的应用.

• 网络控制系统协作问题. 分布式优化在多自主体的协作控制中发挥重要作用, 如无人机编队问题[104]、机器人目标搜索[105]、多机器人系统控制问题[106–108]等.

• 大规模资源调度问题. 大规模资源调度问题通常是非凸的, 常采用拉格朗日松弛[109]或者智能优化算法[110]等技术进行分布式处理.

通信和协作是分布式优化算法的核心, 选择合适的通信构架将直接影响分布式优化算法的性能. 需要在未来着重研究的问题包括如何同步各节点执行算法的时钟, 如何设计高效的通信协议和数据压缩算法, 如何针对具备特定结构的问题改进算法, 如何增强算法适应复杂多变的网络结构的能力, 以及如何使用分布式算法更精确地求解非凸问题等.

## 参考文献(References):


[1] HONG Yiguang, ZHANG Yanqiong. Distributed optimization: algorithm design and convergence analysis [J]. *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 [J]. 控制理论与应用, 2014, 31(7): 850 – 857.)

[2] MURTAGH B A, SAUNDERS M A. Large-scale linearly constrained optimization [J]. *Mathematical Programming*, 1978, 14(1): 41 – 72.

[3] GRIPPO L, SCIANDRONE M. On the convergence of the block nonlinear Gauss-Seidel method under convex constraints [J]. *Operations Research Letters*, 2000, 26(3): 127 – 136.

[4] FUKUSHIMA M. Application of the alternating direction method of multipliers to separable convex programming problems [J]. *Computational Optimization and Applications*, 1992, 1(1): 93 – 111.

[5] BERTSEKAS D P. *Nonlinear Programming* [M]. Belmont: Athena Scientific, 1999.

[6] TSENG P. Convergence of a block coordinate descent method for nondifferentiable minimization [J]. *Journal of Optimization Theory and Applications*, 2001, 109(3): 475 – 494.

[7] LANGENDOEN K, REIJERS N. Distributed localization in wireless sensor networks: a quantitative comparison [J]. *Computer Networks*, 2003, 43(4): 499 – 518.

[8] AGASSOUNON W, MARTINOLI A. Efficiency and robustness of threshold-based distributed allocation algorithms in multi-agent systems [C] //*The 1st International Joint Conference on Autonomous Agents and Multiagent Systems: Part 3*. Bologna, Italy: ACM, 2002: 1090 – 1097.

[9] LIN P, REN W, GAO H. Distributed velocity-constrained consensus of discrete-time multi-agent systems with nonconvex constraints, switching topologies, and delays [J]. *IEEE Transactions on Automatic Control*, 2017, 62(11): 5788 – 5794.

[10] LIN P, REN W, YANG C, et al. Distributed consensus of second-order multiagent systems with nonconvex velocity and control input constraints [J]. *IEEE Transactions on Automatic Control*, 2018, 63(4): 1171 – 1176.

[11] DEGROOT M H. Reaching a consensus [J]. *Journal of the American Statistical Association*, 1974, 69(345): 118 – 121.

[12] REN W, BEARD R W, ATKINS E M. A survey of consensus problems in multi-agent coordination [C] //*American Control Conference*. Portland, OR, USA: IEEE, 2005: 1859 – 1864.

[13] OLFATI-SABER R, MURRAY R M. Consensus problems in networks of agents with switching topology and time-delays [J]. *IEEE Transactions on Automatic Control*, 2004, 49(9): 1520 – 1533.

[14] HORN R A, JOHNSON C R. *Matrix Analysis* [M]. Cambridge: Cambridge University Press, 2012.

[15] OLFATI-SABER R, FAX J A, MURRAY R M. Consensus and cooperation in networked multi-agent systems [J]. *Proceedings of the IEEE*, 2007, 95(1): 215 – 233.

[16] LIN P, JIA Y. Average consensus in networks of multi-agents with both switching topology and coupling time-delay [J]. *Physica A: Statistical Mechanics and its Applications*, 2008, 387(1): 303 – 313.

[17] REN W, BEARD R W. Consensus algorithms for double-integrator dynamics [M] //*Distributed Consensus in Multi-vehicle Cooperative Control: Theory and Applications*. London: Springer, 2008: 77 – 104.

[18] YOU K, LI Z, XIE L. Consensus condition for linear multi-agent systems over randomly switching topologies [J]. *Automatica*, 2013, 49(10): 3125 – 3132.

[19] CAO Y, YU W, REN W, et al. An overview of recent progress in the study of distributed multi-agent coordination [J]. *IEEE Transactions on Industrial Informatics*, 2013, 9(1): 427 – 438.

[20] NEDIC A, OZDAGLAR A. Distributed subgradient methods for multi-agent optimization [J]. *IEEE Transactions on Automatic Control*, 2009, 54(1): 48 – 61.

[21] JOHANSSON B, KEVICZKY T, JOHANSSON M, et al. Subgradient methods and consensus algorithms for solving convex optimization problems [C] //*The 47th IEEE Conference on Decision and Control*. Cancun, Mexico: IEEE, 2008: 4185 – 4190.

[22] LIU S, QIU Z, XIE L. Convergence rate analysis of distributed optimization with projected subgradient algorithm [J]. *Automatica*, 2017, 83(9): 162 – 169.

[23] DUCHI J C, AGARWAL A, WAINWRIGHT M J. Dual averaging for distributed optimization: convergence analysis and network scaling [J]. *IEEE Transactions on Automatic Control*, 2012, 57(3): 592 – 606.

[24] BOYD S, PARIKH N, CHU E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. *Foundations and Trends in Machine Learning*, 2011, 3(1): 1 – 122.

[25] WEI E, OZDAGLAR A. Distributed alternating direction method of multipliers [C] //*The 51st IEEE Conference on Decision and Control*. Maui, Hawaii, USA: IEEE, 2012: 5445 – 5450.

[26] ZHANG J Q, YOU K Y, BASAR T. Distributed discrete-time optimization by exchanging one bit of information [J]. *arXiv:* arXiv: 1709.08360, 2017.

[27] LANGENDOEN K, REIJERS N. Distributed localization in wireless sensor networks: a quantitative comparison [J]. *Computer Networks*, 2003, 43(4): 499 – 518.

[28] NEDIC A, OZDAGLAR A, PARRILO P A. Constrained consensus and optimization in multi-agent networks [J]. *IEEE Transactions on Automatic Control*, 2010, 55(4): 922 – 938.

[29] LOU Y, SHI G, JOHANSSON K H, et al. Approximate projected consensus for convex intersection computation: convergence analysis and critical error angle [J]. *IEEE Transactions on Automatic Control*, 2014, 59(7): 1722 – 1736.

[30] LOU Y, SHI G, JOHANSSON K H, et al. Convergence of random sleep algorithms for optimal consensus [J]. *Systems & Control Letters*, 2013, 62(12): 1196 – 1202.

[31] STROHMER T, VERSHYNIN R. A randomized Kaczmarz algorithm with exponential convergence [J]. *Journal of Fourier Analysis and Applications*, 2009, 15(2): 262 – 278.

[32] ELDAR Y C, NEEDELL D. Acceleration of randomized Kaczmarz method via the Johnson-Lindenstrauss lemma [J]. *Numerical Algorithms*, 2011, 58(2): 163 – 177.

[33] YOU K, SONG S, TEMPO R. A networked parallel algorithm for solving linear algebraic equations [C] //*The 55th IEEE Conference on Decision and Control*. Las Vegas, NV, USA: IEEE, 2016: 1727 – 1732.

[34] POLYAK B T. A new method of stochastic approximation type [J]. *Avtomatika i telemekhanika*, 1990, 51(7): 98 – 107.

[35] LIU Q, YANG S, HONG Y. Constrained consensus algorithms with fixed step size for distributed convex optimization over multiagent networks [J]. *IEEE Transactions on Automatic Control*, 2017, 62(8): 4259 – 4265.

[36] YUAN D, XU S, ZHAO H. Distributed primal-dual subgradient method for multiagent optimization via consensus algorithms [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2011, 41(6): 1715 – 1724.

[37] CHANG T H, NEDIC A, SCAGLIONE A. Distributed constrained optimization by consensus-based primal-dual perturbation method [J]. *IEEE Transactions on Automatic Control*, 2014, 59(6): 1524 – 1538.

[38] LEE S, ZAVLANOS M M. Approximate projection methods for decentralized optimization with functional constraints [J]. *IEEE Transactions on Automatic Control*, 2017, DOI: 10.1109/TAC.2017.2778696.

[39] YOU K, TEMPO R, XIE P. Distributed algorithms for robust convex optimization via the scenario approach [J]. *IEEE Transactions on Automatic Control*, 2018, DOI: 10.1109/TAC.2018.2828093.

[40] AYBAT N, WANG Z, IYENGAR G. An asynchronous distributed proximal gradient method for composite convex optimization [C] //*The 32nd International Conference on Machine Learning*. Lille, France: JMLR, 2015: 2454 – 2462.

[41] XI C, KHAN U A. Directed-distributed gradient descent [C] //*The 53rd Annual Allerton Conference on Communication, Control, and Computing*. Monticello, IL, USA: IEEE, 2015: 1022 – 1026.

[42] MORRAL G. Distributed estimation of the left Perron eigenvector of non-column stochastic protocols for distributed stochastic approximation [C] //*American Control Conference*. Boston, MA, USA: IEEE, 2016: 3352 – 3357.

[43] KEMPE D, DOBRA A, GEHRKE J. Gossip-based computation of aggregate information [C] //*The 44th Annual IEEE Symposium on Foundations of Computer Science*. Cambridge, MA, USA: IEEE, 2003: 482 – 491.

[44] TSIANOS K I, LAWLOR S, RABBAT M G. Push-sum distributed dual averaging for convex optimization [C] //*The 51st IEEE Conference on Decision and Control*. Maui, HI, USA: IEEE, 2012: 5453 – 5458.

[45] NEDIC A, OLSHEVSKY A. Distributed optimization over time-varying directed graphs [J]. *IEEE Transactions on Automatic Control*, 2015, 60(3): 601 – 615.

[46] NEDIC A, OLSHEVSKY A. Stochastic gradient-push for strongly convex functions on time-varying directed graphs [J]. *IEEE Transactions on Automatic Control*, 2016, 61(12): 3936 – 3947.

[47] CAI K, ISHII H. Average consensus on general strongly connected digraphs [J]. *Automatica*, 2012, 48(11): 2750 – 2761.

[48] XU Y, CAI K, HAN T, et al. A fully distributed approach to resource allocation problem under directed and switching topologies [C] //*The 10th Asian Control Conference*. Kota Kinabalu, Malaysia: IEEE, 2015: 1 – 6.

[49] XI C, KHAN U A. Distributed subgradient projection algorithm over directed graphs [J]. *IEEE Transactions on Automatic Control*, 2017, 62(8): 3986 – 3992.

[50] NEDIC A. Random algorithms for convex minimization problems [J]. *Mathematical Programming*, 2011, 129(2): 225 – 253.

[51] XIE P, YOU K, SONG S, et al. Distributed random-fixed projected algorithm for constrained optimization over digraphs [J]. *IFAC-PapersOnline*, 2017, 50(1): 14436 – 14441.

[52] NEDIC A, OLSHEVSKY A, SHI W. Achieving geometric convergence for distributed optimization over time-varying graphs [J]. *SIAM Journal on Optimization*, 2017, 27(4): 2597 – 2633.

[53] NEDIC A, OLSHEVSKY A, SHI W, et al. A geometrically convergent distributed optimization with uncoordinated step-sizes [C] //*American Control Conference*. Seattle, WA, USA: IEEE, 2017: 3950 – 3955.

[54] XIE P, YOU K, TEMPO R, et al. Distributed convex optimization with inequality constraints over time-varying unbalanced digraphs [J]. *IEEE Transactions on Automatic Control*, 2018. DOI: 10.1109/TAC.2018.2816104.

[55] LOU Y, HONG Y, XIE L, et al. Nash equilibrium computation in subnetwork zero-sum games with switching communications [J]. *IEEE Transactions on Automatic Control*, 2016, 61(10): 2920 – 2935.

[56] HAAS Z J, HALPERN J Y, LI L. Gossip-based ad hoc routing [J]. *IEEE/ACM Transactions on Networking (ToN)*, 2006, 14(3): 479 – 491.

[57] BURT R S. Bandwith and echo: trust, information, and gossip in social networks [M] //CASELLA A, RAUSCH J E. *Integrating the Study of Networks and Markets*. New York: Sage, 2001.

[58] TAHBAZ-SALEHI A, JADBABAIE A. A necessary and sufficient condition for consensus over random networks [J]. *IEEE Transactions on Automatic Control*, 2008, 53(3): 791 – 795.

[59] HATANO Y, MESBAHI M. Agreement over random networks [J]. *IEEE Transactions on Automatic Control*, 2005, 50(11): 1867 – 1872.

[60] WU C W. Synchronization and convergence of linear dynamics in random directed networks [J]. *IEEE Transactions on Automatic Control*, 2006, 51(7): 1207 – 1210.

[61] LOBEL I, OZDAGLAR A. Distributed subgradient methods for convex optimization over random networks [J]. *IEEE Transactions on Automatic Control*, 2011, 56(6): 1291 – 1306.

[62] LEE S, NEDIC A. Asynchronous gossip-based random projection algorithms over networks [J]. *IEEE Transactions on Automatic Control*, 2016, 61(4): 953 – 968.

[63] HANNAH R, YIN W. More iterations per second, same quality–why asynchronous algorithms may drastically outperform traditional ones [J]. *arXiv:* arXiv: 1708.05136, 2017.

[64] TSITSIKLIS J, BERTSEKAS D, ATHANS M. Distributed asynchronous deterministic and stochastic gradient optimization algorithms [J]. *IEEE Transactions on Automatic Control*, 1986, 31(9): 803 – 812.

[65] WEI E, OZDAGLAR A. On the $O(1 = k)$ convergence of asynchronous distributed alternating direction method of multipliers [C] //*Global Conference on Signal and Information Processing*. Austin, TX, USA: IEEE, 2013: 551 – 554.

[66] PENG Z, XU Y, YAN M, et al. ARock: an algorithmic framework for asynchronous parallel coordinate updates [J]. *SIAM Journal on Scientific Computing*, 2016, 38(5): A2851 – A2879.

[67] WU T, YUAN K, LING Q, et al. Decentralized consensus optimization with asynchrony and delays [J]. *IEEE Transactions on Signal and Information Processing over Networks*, 2018, 4(2): 293 – 307

[68] EISEN M, MOKHTARI A, RIBEIRO A. A decentralized quasi-Newton method for dual formulations of consensus optimization [C] //*The 55th IEEE Conference on Decision and Control*, 2016: 1951 – 1958.

[69] SUN T, HANNAH R, YIN W. Asynchronous coordinate descent under more realistic assumptions [C] //*Advances in Neural Information Processing Systems*. Long Beach, CA, USA: NIPS, 2017: 6183 – 6191.

[70] XU J, ZHU S, SOH Y C, et al. Convergence of asynchronous distributed gradient methods over stochastic networks [J]. *IEEE Transactions on Automatic Control*, 2018, 63(2): 434 – 448.

[71] MODI P J, SHEN W M, TAMBE M, et al. ADOPT: Asynchronous distributed constraint optimization with quality guarantees [J]. *Artificial Intelligence*, 2005, 161(1/2): 149 – 180.

[72] SILAGHI M C, YOKOO M. Nogood based asynchronous distributed optimization (ADOPT ng) [C] //*The 5th International Joint Conference on Autonomous Agents and Multiagent Systems*. Hakodate, Japan: ACM, 2006: 1389 – 1396.

[73] ZHONG M, CASSANDRAS C G. Asynchronous distributed optimization with event-driven communication [J]. *IEEE Transactions on Automatic Control*, 2010, 55(12): 2735 – 2750.

[74] GODSIL C, ROYLE G F. *Algebraic Graph Theory* [M]. Berlin: Springer Science & Business Media, 2013.

[75] CHEN C T. *Linear System Theory and Design* [M]. Oxford: Oxford University Press, 1995.

[76] HORN R A, JOHNSON C R. *Matrix Analysis* [M]. Cambridge: Cambridge University Press, 1990.

[77] BOYD S, VANDENBERGHE L. *Convex Optimization* [M]. Cambridge: Cambridge University Press, 2004.

[78] NESTEROV Y. *Introductory Lectures on Convex Optimization: a Basic Course* [M]. Boston: Kluwer Academic Publishers, 2004.

[79] BERTSEKAS D P. *Nonlinear Programming* [M]. Belmont: Athena Scientific, 1999.

[80] ROBBINS H, SIEGMUND D. A convergence theorem for nonnegative almost supermartingales and some applications [M] //ROBBINS H, SIEGMUND D. *Optimizing Methods in Statistics*. New York: Springer, 1985: 111 – 135.

[81] YADAV V, SALAPAKA M V. Distributed protocol for determining when averaging consensus is reached [C] //*The 45th Annual Allerton Conference on Communication, Control, and Computing*. Monticello, IL, USA: IEEE, 2007: 715 – 720.

[82] MATTERN F. Algorithms for distributed termination detection [J]. *Distributed Computing*, 1987, 2(3): 161 – 175.

[83] MANITARA N E, HADJICOSTIS C N. Distributed stopping for average consensus in directed graphs via a randomized event-triggered strategy [C] //*The 6th International Symposium on Communications, Control and Signal Processing*. Athens, Greece: IEEE, 2014: 483 – 486.

[84] BAZ D. A method of terminating asynchronous iterative algorithms on message passing systems [J]. *Parallel Algorithms and Applications*, 1996, 9(1/2): 153 – 158.

[85] XIE P, YOU K, WU C. How to stop consensus algorithms, locally? [C] //*The 56th IEEE Conference on Decision and Control*. Melbourne, Australia: IEEE, 2017: 4544 – 4549.

[86] JAKOVETIC D, XAVIER J, MOURA J M F. Fast distributed gradient methods [J]. *IEEE Transactions on Automatic Control*, 2014, 59(5): 1131 – 1146.

[87] CHEN A I, OZDAGLAR A. A fast distributed proximal-gradient method [C] //*The 50th Annual Allerton Conference on Communication, Control, and Computing*. Monticello, IL, USA: IEEE, 2012: 601 – 608.

[88] SHI W, LING Q, WU G, et al. Extra: An exact first-order algorithm for decentralized consensus optimization [J]. *SIAM Journal on Optimization*, 2015, 25(2): 944 – 966.

[89] XI C, KHAN U A. DEXTRA: a fast algorithm for optimization over directed graphs [J]. *IEEE Transactions on Automatic Control*, 2017, 62(10): 4980 – 4993.

[90] XI C, XIN R, KHAN U A. ADD–OPT: accelerated distributed directed optimization [J]. *IEEE Transactions on Automatic Control*, 2018, 63(5): 1329 – 1339.

[91] ZENG J, YIN W. Extrapush for convex smooth decentralized optimization over directed networks [J]. *Journal of Computational Mathematics*, 2017, 56(6): 381 – 394.

[92] QU G, LI N. Accelerated distributed Nesterov gradient descent [J]. *arXiv:* arXiv: 1705.07176, 2017.

[93] TSIANOS K, LAWLOR S, RABBAT M G. Communication/computation tradeoffs in consensus-based distributed optimization [C] //*Advances in Neural Information Processing Systems*. Reno, CA, USA: NIPS, 2012: 1943 – 1951.

[94] NEDIC A, OLSHEVSKY A, RABBAT M G. Network topology and communication-computation tradeoffs in decentralized optimization [J]. *Proceedings of the IEEE*, 2018, 106(5): 953 – 976.

[95] DEAN J, CORRADO G, MONGA R, et al. Large scale distributed deep networks [C] //*Advances in Neural Information Processing Systems*. Reno, CA, USA: NIPS, 2012: 1223 – 1231.

[96] LIAN X, HUANG Y, Li Y, et al. Asynchronous parallel stochastic gradient for nonconvex optimization [C] //*Advances in Neural Information Processing Systems*. Montreal, Canada: NIPS, 2015: 2737 – 2745.

[97] HOUSKA B, FRASCH J, DIEHL M. An augmented Lagrangian based algorithm for distributed nonconvex optimization [J]. *SIAM Journal on Optimization*, 2016, 26(2): 1101 – 1127.

[98] BANKS H T, KUNISCH K. *Estimation Techniques for Distributed Parameter Systems* [M]. Boston: Birkhauser, 1989.

[99] CATTIVELLI F S, SAYED A H. Diffusion LMS strategies for distributed estimation [J]. *IEEE Transactions on Signal Processing*, 2010, 58(3): 1035 – 1048.

[100] LANGENDOEN K, REIJERS N. Distributed localization in wireless sensor networks: a quantitative comparison [J]. *Computer Networks*, 2003, 43(4): 499 – 518.

[101] MOORE D, LEONARD J, RUS D, et al. Robust distributed network localization with noisy range measurements [C] //*The 2nd International Conference on Embedded Networked Sensor Systems*. Baltimore, MD, USA: ACM, 2004: 50 – 61.

[102] MATEOS G, BAZERQUE J A, GIANNAKIS G B. Distributed sparse linear regression [J]. *IEEE Transactions on Signal Processing*, 2010, 58(10): 5262 – 5276.

[103] VENKAT A N, HISKENS I A, RAWLINGS J B, et al. Distributed MPC strategies with application to power system automatic generation control [J]. *IEEE Transactions on Control Systems Technology*, 2008, 16(6): 1192 – 1206.

[104] OLFATI-SABER R, MURRAY R M. Graph rigidity and distributed formation stabilization of multi-vehicle systems [C] //*The 41st IEEE Conference on Decision and Control*. Las Vegas, NV, USA: IEEE, 2002, 3: 2965 – 2971.

[105] LI S, KONG R, GUO Y. Cooperative distributed source seeking by multiple robots: algorithms and experiments [J]. *IEEE/ASME Transactions on Mechatronics*, 2014, 19(6): 1810 – 1820.

[106] REN W, SORENSEN N. Distributed coordination architecture for multi-robot formation control [J]. *Robotics and Autonomous Systems*, 2008, 56(4): 324 – 333.

[107] QIU Z, HONG Y, XIE L. Optimal consensus of Euler-Lagrangian systems with kinematic constraints [J]. *IFAC–PapersOnLine*, 2016, 49(22): 327 – 332.

[108] ZHANG Y, DENG Z, HONG Y. Distributed optimal coordination for multiple heterogeneous Euler-Lagrangian systems [J]. *Automatica*, 2017, 79(5): 207 – 213.

[109] KUTANOGLU E, DAVID WU S. On combinatorial auction and Lagrangean relaxation for distributed resource scheduling [J]. *IIE Transactions*, 1999, 31(9): 813 – 826.

[110] EASTON F F, MANSOUR N. A distributed genetic algorithm for deterministic and stochastic labor scheduling problems [J]. *European Journal of Operational Research*, 1999, 118(3): 505 – 523.

## 作者简介:

**谢　佩**　(1992–), 男, 清华大学自动化系博士研究生, 研究方向为分布式优化、学习及控制等, E-mail: xie-p13@mails.tsinghua.edu.cn;

**游科友**　(1985–), 男, 清华大学自动化系副教授, 博士生导师, 青年千人和优青, 研究方向为网络化系统的学习、优化与控制及其应用, E-mail: youky@tsinghua.edu.cn;

**洪奕光**　(1966–), 男, 中科院数学与系统科学研究所研究员, IEEE Fellow和杰青, 研究方向为多自主体系统、分布式优化、非线性系统、社会网络、机器人等, E-mail: yghong@issa.ac.cn;

**谢立华**　(1964–), 男, 新加坡南洋理工大学电气与电子工程学院教授, IEEE Fellow和IFAC Fellow, 研究方向为鲁棒控制、传感器网络、网络控制系统、估计理论和信号处理, E-mail: elhxie@ntu.edu.sg.