DOI: 10.7641/CTA.2014.40012

分布式优化: 算法设计和收敛性分析

洪奕光, 张艳琼†

(中国科学院 数学与系统科学研究院 系统控制重点实验室, 北京 100190)

摘要: 近年来, 随着高科技的蓬勃发展, 特别是云计算和大数据等新兴领域的出现, 分布式优化理论和应用得到了越来越多的重视, 并逐渐渗透到科学研究、工程应用和社会生活的各个方面, 分布式优化是通过多智能体之间的合作协调有效地实现优化的任务, 可用来解决许多集中式算法难以胜任的大规模复杂的优化问题. 如今如何设计出有效的分布式优化算法并对其进行收敛性和复杂性的分析成了优化研究的主要任务之一. 与集中式算法的主要区别在于分布式算法还不得不考虑通讯和协调在优化中起到的重要作用. 本文集中讨论了近年来分布式优化研究中的一些典型热门的研究问题, 从一个侧面介绍了包括无约束优化、带约束优化、以及分布式博弈等方向的部分最新成果. 同时也比较详尽地论述了笔者最近的相关研究成果. 最后, 本文简要地对分布式优化的研究和应用前景进行了展望.

关键词: 分布式优化; 约束; (次)梯度法; Nash均衡

中图分类号: TP13, TP18　　**文献标识码**: A

Distributed optimization: algorithm design and convergence analysis

HONG Yi-guang, ZHANG Yan-qiong†

(Key Lab of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: In recent years, theoretical and practical researchers on distributed optimization are spending more and more efforts to meet the challenge of rapid development of high techniques, particularly the new areas such as the cloud computation and big data sets. Moreover, distributed optimization is increasingly applied to many fields such as scientific researches, engineering applications, and social activities. Distributed optimization considers effective coordination between multiple agents, and can be applied to many large-scale complicated optimization problems that are difficult to be solved by centralized algorithms. How to design an effective distributed optimization algorithm with convergence and limited complexity is one of the most important objectives in optimization research. Being different from the centralized algorithm, a distributed algorithm has to take into account the communication and coordination in the process of optimization. In this paper, we focus on some standard hot topics in recent distributed optimization problems, such as the unconstrained optimization, constrained optimization, and distributed game. On those topics we make a brief survey of recent achievements including some of our research results, and point out prospects for further studies in theoretical research or practical applications.

Key words: distributed optimization; constraints; (sub-)gradient method; Nash equilibrium

1 引言(Introduction)

优化与博弈是运筹学和控制论理论研究中的核心问题之一, 同时还在包括系统科学、人工智能、生物生态、压缩感知、计算机通讯等很多领域有着广泛的应用. 已有成熟的优化与博弈算法大多是集中式的[1–2], 但是随着现代科学的发展, 在生物、物理、社会和工程等众多学科中出现了许多新问题亟待解决, 包括如何有效处理大数据、进行云计算等. 另外, 通信和微电子技术也在迅猛发展, 为此提供了大量的高效、廉价且性能稳定的传感器、处理器以及各种执行器件, 这极大地支持和拓展了各种分布式算法的应用范围, 促进分布式优化的理论研究不断取得新的成果.

现在随着多智能体(multi-agent)系统理论和协调技术的发展, 许多分布式优化算法可以通过借助多智能体网络的方式来实现[3]. 多智能体系统是由一群具备一定的感知、通信、计算和执行能力的智能个体通过通讯等方式关联成的网络系统. 从某种意义上说, 分布式技术与多智能体网络是一对孪生兄弟, 或者是一个事物的两个不同的侧面: 分布式强调技术实现的方法而多智能体强调技术实现的主体. 对复杂大规模系统, 分布式方法比传统集中式方法更为灵活且操作更为方便, 这也使得分布式优化与控制的研究得到了

收稿日期: 2014−01−08; 录用日期: 2014−03−18.
†通信作者. E-mail: yanqiong918@163.com.
基金项目: 国家自然科学基金资助项目(61174071).

迅速发展. 而且已被越来越多的工业和国防应用领域, 包括智能电网、传感器网络、社会网络、信息物理系统(cyber-physical system)等所关注.

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. 在优化理论研究过程中, 优化算法的设计、收敛性的证明、复杂性(包括分析复杂性和算术复杂性)的分析是其中几个关键性的研究问题. 在分布式优化中, 相应的问题正在吸引着来自诸多领域的科技工作者的巨大研究兴趣. 近年来相关的研究人员在分布式优化上已经取得了一系列重要成果(特别是对一些典型分布式算法收敛性等的分析), 并在不同科研领域中的多种期刊和会议上发表.

现在的分布式优化有两大类研究问题: 一类是对性能指标函数的优化, 另一类是对系统动态过程的优化. 由于分布式优化刚刚兴起, 主要的突出理论研究成果还是属于第一类优化中. 在一些重要的现实问题中, 比如资源分配, 传感器网络中的定位等问题, 每个个体往往有一个代价函数, 且整个网络的代价由这些个体的代价函数和来表示. 此网络的目的是通过个体间的局部信息交流而完成整个网络代价函数的优化, 其中每个个体只知道自己的代价函数, 在给定的分布式优化算法下可以得出保证其收敛的条件. 另一类是对系统动态过程的优化, 此类分布式优化问题往往涉及到分布式的(随机)动态规划, 现在虽然已经有些研究, 但大多结果往往是初步的或因所需条件很强难以做深入的理论分析, 因此在本文中提到的分布式优化主要是指第一类的优化.

在优化理论发展过程中, 凸优化因为其基本且简单一直受到广泛的关注[1], 很多实际的优化问题都转化或近似成凸优化问题来做. 由于篇幅有限, 本文将以凸优化为基础, 集中讨论分布式优化的3个重要问题: 1) 无约束凸优化; 2) 带约束凸优化; 3) 博弈和Nash均衡. 因为分布式优化领域较广, 在随后的3节中将对这3个方面中的一些研究方向的部分研究进行简要的介绍, 并特别介绍笔者近年来取得的一些研究成果, 希望抛砖引玉能引起大家对分布式优化的研究兴趣.

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

正如凸优化是集中式优化中的基本问题, 无约束分布式凸优化也是分布式优化中最基本的问题和研究的出发点. 为了方便先简要介绍一下凸集和凸函数[1]. 令集合K是欧式空间中的一个集合, 如果对任意的$x, y \in K$和$0 < c < 1$, 均有$cx + (1-c)y \in K$, 则称K为凸集. 如果对任意的$x, y \in K$, $c_i \geqslant 0(i = 1, 2)$, 均有$c_1x + c_2y \in K$, 则称$K$为凸锥. 称由集合$K_1, \cdots, K_n$中所有元素的有限凸组合组成的集合为由集合$K_1, \cdots, K_n$生成的凸包, 记为$\mathrm{co}\{K_1, \cdots, K_n\}$. 有限个点生成的凸包称为凸多面体. 令$f(x)$为一个实函数, 如果对任意的$x, y$和$0 < c < 1$, 均有$f(cx + (1-c)y) \leqslant cf(x) + (1-c)f(y)$, 则称$f$为凸函数. 如果进一步还有$f(cx + (1-c)y) = cf(x) + (1-c)f(y)$当且仅当$x = y$, 则称$f$为严格凸函数. 如果$-f$是(严格)凸函数, 称$f$为(严格)凹函数.

一类得到广泛研究的分布式优化问题是整个网络的代价函数是所有个体代价函数的和函数[4]. 这类问题的目的是需要设计分布式算法最小化该和函数. 其研究的方法主要是分为基于次梯度(subgradient) 的方法和非次梯度方法两大类.

在无约束优化问题中一个典型而简单的分布式优化问题是分布式凸交计算, 为此在这一子节中先讨论它. 对于该问题, 通常可以用梯度方法对它进行研究, 在它的算法中梯度就是对凸集的投影. 随后讨论一般的基于(次)梯度方法的结果, 最后是对非(次)梯度方法的介绍.

2.1 分布式凸交计算(Distributed convex intersection computation)

这几年, 分布式凸交计算越来越受到研究人员的重视, 并在包括图像重构中原像的恢复、凸投射的最佳逼近和目标定位等实际问题中有大量的应用[5–6]. 比如, 在目标定位问题中, 目标源以一定的功率发射某一信号, 网络中的多个传感器会接收到随着传输距离变大而衰减的带量测噪声的信号, 目标定位的目的是通过网络中传感器接收到的信号以找到目标源的位置. 对应于接收到的信号, 每个传感器都会有一个(凸)感知区域. 当这些感知区域具有非空交集时, 目标定位问题可以转化为一个凸交计算问题[6]. 近几十年来, 凸交计算问题(也称凸可行性问题)得到广泛的研究. 当初的研究通常是集中式的, 随后这几年开始了分布式凸交计算的研究.

下面先给出凸交计算问题的精确数学描述. 令$X_1, \cdots, X_n$是Hilbert空间中的闭凸集, 其交集非空, 即$\bigcap\limits_{i=1}^{n} X_i \neq \varnothing$. 凸交计算问题的目的是找到非空交集中的一个点$x^*$:

$$x^* \in \bigcap_{i=1}^{n} X_i = X_0. \tag{1}$$

在文献[7]中得到广泛研究的求解凸交计算问题的方法是对于任意给定的初始点, 依次向这n个凸集做凸投射而生成一个循环投射估计序列, 进而考虑此估计序列的收敛性. 当所有的凸集X_i是闭子空间时, 文献[8]证明了此序列依范数收敛到其凸交X_0中的一个点, 且此收敛点是初始点到凸集X_0上的投射点. 文献[9]从凸集的正则性角度出发, 充分讨论了此循环投射序列的收敛速率问题. 前面提到的文献讨论的都是

集中式算法.

近几年来, 设计分布式算法求解凸交计算问题也得到广泛研究. 分布式凸交计算问题的具体描述是: 考虑一个由n个个体组成的网络, 个体i对应于凸集X_i, 且个体i仅知道集合X_i, 而不知道对应于其他个体的集合$X_j, j \neq i$. 此网络的目的是通过个体之间的相互合作找到非空交集中的一个点.

对于任意时刻$t \geqslant 0$, 网络中个体之间的动态连接关系可以用有向图$\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$来表示. 如果边$(j, i) \in \mathcal{E}_t$, 则称个体$j$是个体$i$在$t$时刻的邻居, 将个体$i$在$t$时刻的所有邻居构成的集合用$\mathcal{N}_i(t)$来表示, 本文假设所有的个体都是它自己的邻居. 文献[4]首先给出如下的离散时间分布式投射同步算法:

$$x_i(k+1) = P_{X_i}(\sum_{j=1}^{n} a_{ij}(k)x_j(k)),\ i = 1, \cdots, n, \tag{2}$$

其中: $x_i(k)$是个体i在时刻k的状态估计, $a_{ij}(k)$表示边(j, i)在时刻k的权重值(如果$(j, i) \in \mathcal{E}_k$, $a_{ij}(k) > 0$; 否则$a_{ij}(k) = 0$), 并且$P_{X_i}(y)$表示y 到集合X_i的投影点. 之后, 在双随机性和一致联合强连通的假设条件下, 证明该算法将同步收敛到X_0中的一个点.

另外, 虽然大多数优化问题是从离散时间角度研究的, 但是用连续时间动力学研究优化的收敛性也是值得注意的一个方向. 文献[10]首次研究了连续时间的非线性多智能体系统在最一般的$[t, \infty)$联合连通拓扑假设下的优化趋同, 提出如下的连续时间分布式投射同步算法:

$$\begin{aligned}\dot{x}_i(t) = &\sum_{j=1}^{n} a_{ij}(t)(x_j(t) - x_i(t)) + \\ &P_{X_i}(x_i(t)) - x_i(t),\end{aligned} \tag{3}$$

其中$i = 1, \cdots, n$. 并且该文首次给出了在无向图时, $[t, \infty)$联合连通是该算法收敛的充分必要条件.

以上关于分布式凸交计算问题都是基于能够获得精确投射点, 然而在实际应用中, 投射点的计算都不可避免存在一定的误差, 实际得到的往往只是精确投射点的一个近似. 文献[11]首先引入近似投影(approximate projection)的概念, 将点v投射到集合K带近似角度θ的近似投影集定义为下面集合:

$$P_K^a(v, \theta) = \begin{cases} C_K(v, \theta) \cap H_K^+(v), & v \notin K, \\ \{v\}, & v \in K, \end{cases} \tag{4}$$

其中:

$$C_K(v, \theta) = v + \{z | \langle z, P_K(v) - v\rangle \geqslant |z||v|_K \cos\theta\},$$
$$H_K^+(v) = \{z | \langle P_K(v) - z, P_K(v) - v\rangle \geqslant 0\},$$

并且给出了如下的近似投影同步算法:

$$x_i(k+1) = \sum_{j=1}^{n} a_{ij}(k)(P_j^a(k)),\ i = 1, \cdots, n, \tag{5}$$

这里$P_j^a(k) \in P_{X_i}^a(x_i(k),\ \theta_{ik})$是状态$x_i(k)$向凸集$X_i$带近似角度$\theta_{ik}$的近似投影. 与文献[4]提出的算法相比, 上述的近似投影算法在不要求邻接矩阵的双随机性情形下给出了系统达到全局最优同步关于近似投影程度的鲁棒性条件.

上面所提到的都是基于投影算子的一些确定性分布式优化算法, 而一些随机的算法也得到广泛研究. 比如, 考虑到在现实的机器网络中, 每个机器人携带的能量有限且代价高, 在文献[12]中从节省能量的角度提出一种随机休眠算法来计算凸交问题. 其算法的主要思想是个体在对非空交集中的点进行估计时, 所有的个体独立地以一定的概率p向自己的集合做凸投射, 以概率$1-p$保持上一步的状态不变, 具体算法如下:

$$\begin{cases} x_i(k+1) = \sum_{j=1}^{n} a_{ij}(k)y_j(k), \\ d_j(k) = P_{X_j}(x_j(k)) - x_j(k), \\ y_j(k) = \begin{cases} x_j(k) + \alpha_k d_j(k), & \text{以概率}p\,; \\ x_j(k), & \text{以概率}1-p\,. \end{cases} \end{cases} \tag{6}$$

然后给出了所有个体几乎处处渐近收敛到非空集合中某一个点的充分条件.

上面谈到的凸交问题主要强调计算, 其中的多智能体是虚拟的进程, 因而其动力学取成最简单的形式. 实际上, 实现分布式凸交过程的智能体还可以具有复杂的物理动态模型. 比如文献[13]讨论了一群机器人如何合作解决凸交优化问题, 每个机器人作为智能体只知道自己对目标的感知区域(通常被描述成一个凸集), 而它们的目的是找到它们感知集合中的目标. 为此, 该文研究了在固定和切换拓扑下这些机器人如何寻找并且避开碰撞, 并在一定的不确定下, 个体达到寻找凸交的目的.

2.2 次梯度方法(Sub-gradient method)

上面讨论的凸交可以看做是一个简单特殊的优化问题, 下面讨论一般的分布式凸优化问题, 它们有着更为广泛或潜在的实际应用背景. 对于一般的分布式无约束问题:

$$\min f(x) = \sum_{i=1}^{n} f_i(x), \tag{7}$$

其中: f_i是个体i的代价函数且只能被i观测到, f为整个网络的代价函数, 而该网络的目的就是通过个体间的相互协作来最小化f. 文献[14]提出了基于次梯度的分布式优化算法:

$$\begin{aligned} x_i(k+1) = &\sum_{j=1}^{n} a_{ij}(k)x_j(k) - \alpha d_i(k), \\ &i = 1, \cdots, n. \end{aligned} \tag{8}$$

这里: α是常步长, $d_i(k) \in \partial f_i(x_i(k))$是$f_i$在$x_i(k)$处

的次梯度. 然后在权重平衡图和次梯度有界的假设下, 该文给出了系统达到和代价函数最优值逼近的一个估计. 对于此类基于次梯度的常步长算法往往不能保证算法的收敛性, 或者说即使收敛性能够得到但不能保证一定能收敛到最优解集上. 为了保证算法收敛的最优性, 步长渐近收敛到0的条件是必要的, 即 $\lim\limits_{k\to\infty}\alpha_k=0$; 但是该条件又会带来收敛速度慢的弊端, 为了克服这个缺点, 文献[15]提出来下面一种动态系统优化算法来解决问题(7):

$$\begin{cases} v_i(t)=\sum\limits_{j=1}^{n}a_{ij}(t)(x_j(t)-x_i(t)),\\ \dot{x}_i(t)=v_i(t)+\sum\limits_{j=1}^{n}a_{ij}(t)(y_j(t)-y_i(t))-d_i(t),\\ \dot{y}_i(t)=-v_i(t), \end{cases} \tag{9}$$

其中$d_i(t)\in\partial f_i(x_i(t))$. 并且证明了上述算法在固定拓扑和连通性假设下渐近同步收敛到问题(7) 的最优解.

随着分布式优化的深入发展, 近些年来, 研究人员取得了很多基于(次)梯度的变形算法和研究成果, 比如:

a) 基于量化信息的分布式优化.

在基于量化信息的分布式优化这一问题里, 每个个体由于通讯能力的限制, 不能完全量化得到其他个体的信息, 而只能得到通讯量化后的信息. 大部分基于次梯度的方法[16–17], 虽然给出了固定或者切换拓扑下分布式优化的算法的收敛性分析, 但是在实际应用中由于量化误差的存在, 不能达到精确解, 并且其精确程度与量化的精度有关. 在该问题中, 有一些基本的问题还没有完全解决, 特别是: i) 如何实现(在切换拓扑下)不受量化误差影响的精确优化? ii) 如何在能够实现优化的前提下减少信息传输率? 因此为了达到这些量化优化, 笔者最近采用了编码解码器和量化器尺度变化的技术, 通过编码解码去除每个个体与其邻居个体之间交换信息的误差来实现算法的收敛性, 而且利用尺度变换得到精确的优化解. 进一步笔者还研究了多少比特才能保证实现优化目标.

b) 随机优化.

在一些现实问题中, 比如从社会观点动力学的角度而言, 个体在做决策时往往会相互独立地以一定的概率坚持自己的观点, 以一定概率会受到邻居个体观点的影响. 文献[18]受此启发提出一种随机优化算法来计算优化问题(7):

$$x_i(k+1)=\begin{cases}\sum\limits_{j=1}^{n}a_{ij}(k)x_j(k), & \text{以概率}p;\\ P_{X_i}(x_i(k)), & \text{以概率}1-p,\end{cases} \tag{10}$$

其中X_i是局部目标函数f_i的最优解集, 其信息只能被个体i所得到. 然后该文给出了算法几乎处处渐近同步收敛到问题(7)最优解的充分条件, 并且给出数值试验验证了在一定条件下, 差不多5%算例中随机算法的收敛效果比确定性的还要好.

除此之外, 文献[19]利用了次梯度和一致性算法来解决固定无向拓扑下多智能体的解耦优化问题; 增量次梯度法(incremental subgradient method)[20], 以及包括信息交换过程中存在量测噪音和计算次梯度带来的误差[21], 切换拓扑是独立同分布的随机过程[22]等优化问题也得到广泛研究与发展.

2.3 非传统梯度方法(Nontraditional gradient method)

对于一般基于次梯度的分布式优化算法, 其算法收敛速度只能达到$O(1/\sqrt{k})$. 在最近研究中, 文献[23]提出一种分布式快速梯度算法(fast distributed gradient method)来解决问题(7), 使得当目标函数满足一定的光滑性条件并且邻接矩阵的谱半径全局已知时, 该算法的收敛速度能够达到$O(1/k^2)$. 其算法具体形式如下:

$$\begin{cases} x_i(k)=\sum\limits_{j=1}^{n}a_{ij}y_j(k-1)-\alpha_{k-1}\nabla f_i(y_i(k-1)),\\ y_i(k)=x_i(k)+\beta_{k-1}(x_i(k)-x_i(k-1)), \end{cases} \tag{11}$$

其中系统的连接图是固定的无向连通图.

还有很多分布式优化问题可以用非基于梯度的算法来解决. 一个现在比较常用的方法是交替方向乘子法(alternating direction method of multipliers, ADMM). 标准ADMM算法的主要思想就是将原来的优化问题分解成两个子问题, 序列地求解它们并更新相应的对偶变量, 但其主要缺点在于分解后的两个子问题不能完全分布式求解, 因为对偶变量的更新需要全局状态信息, 这样的话就必须得有一个中心来负责. 为了克服这个缺点, 针对问题(7), 文献[24]提出一种分布式ADMM法. 系统连接图$\mathcal{G}=(\mathcal{V},\mathcal{E})$ 中的每一条边$e_{ij}\in\mathcal{E}$对应一个对偶变量λ_{ij}, 而对于个体i除了要负责更新自己的状态x_i 外, 还要负责一些对偶变量更新, 然后根据事先给定的次序, 个体依次进行更新, 算法具体执行如下:

$$\begin{cases} x_i(k)=\arg\min\limits_{x}\dfrac{\beta}{2}\sum\limits_{j\in P(i)}\|x_j(k+1)-\\ \qquad x-\dfrac{1}{\beta}\lambda_{ji}(k)\|^2+\ f_i(x)+\\ \qquad \dfrac{\beta}{2}\sum\limits_{j\in S(i)}\|x-x_j(k)-\dfrac{1}{\beta}\lambda_{ij}(k)\|^2,\\ \lambda_{ri}(k+1)=\\ \lambda_{ri}(k)-\beta(x_r(k+1)-x_i(k+1)), \end{cases} \tag{12}$$

其中: 常数$\beta > 0$, $r \in P(i)$, 且$P(i)=\{j|e_{ji} \in \mathcal{E}, j < i\}$, $S(i) = \{j|e_{ij} \in \mathcal{E}, i < j\}$. ADMM法较之一般的次梯度算法, 优势在于其收敛速度有很大的提高, 能够达到$O(1/k)$, 但是其不足在于需要提前规定好个体进行状态更新的次序. 除此之外, 还有其他一些非梯度算法也在文献中出现, 比如逐对补偿算法[25]和增广Lagrange算法[26]等.

注 1 从上面的分析可以看出, 次梯度方法相比于其他优化方法的优势在于其算法形式简单且对于代价函数的要求低(一般只需满足凸性即可), 对于一般的凸优化问题都适用, 其不足在于对步长渐近收敛到0的要求导致了收敛速度慢的弊端. 当代价函数满足光滑性条件时, 其次梯度就是梯度, 故次梯度方法较之梯度方法其区别仅在于对代价函数的光滑性要求. 而对于一般的非传统梯度方法, 其出发点大都是从提高算法收敛速度的角度, 但往往可能会伴随着算法形式的复杂化、维数的增加、代价函数光滑性的高要求或者要提前已知个体间通信拓扑图的一些信息, 这就使得非传统梯度方法具有一定的局限性, 它们所适用的优化问题较之次梯度方法更为特定一些. 比如, ADMM算法的维数会随着网络边的增加而增加, 由此导致在解决大规模非稀疏的网络优化问题时, 对计算机的存储能力要求加大.

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

现实中的很多优化问题是带有约束的, 其约束形式包括受限集、等式及不等式约束. 同样, 在解决分布式优化问题中, 也会碰到不少约束问题. 实际上, 很多非凸的优化问题可以近似为带约束的凸优化问题.

在分布式优化中, 有两类约束用在优化算法中. 一类是优化问题中不得不考虑的客观给定约束, 还有一类是优化问题中自身隐含的约束. 对于客观存在的约束, 研究者不得不去面对和处理, 但在不好处理时就采取方法取代或逼近这些约束条件. 而后一类约束往往是所研究的优化问题中自然满足的, 有时可以主动加入这些辅助约束以提高算法的效率. 本文在随后二节分别介绍.

3.1 给定约束(Given constraints)

针对一般的带受限集的约束优化问题:

$$\min f(x) = \sum_{i=1}^{n} f_i(x),\ \text{s.t.}\ x \in \bigcap_{i=1}^{n} X_i, \tag{13}$$

其中代价函数f_i和约束集X_i只能被个体i观测到. 对于求解这类问题, 文献[4]提出下面的分布式投影次梯度算法:

$$x_i(k+1)=P_{X_i}(\sum_{j=1}^{n} a_{ij}(k)x_j(k)-\alpha d_i(k)), \tag{14}$$

其中: $i=1,\cdots,n$, $d_i(k) \in \partial f_i(\sum_{j=1}^{n} a_{ij}(k)x_j(k))$. 随后它们给出了在如下两种特殊情形下, 即: 1) 系统连接图是完全连通图; 2) 所有个体的受限集都相同(即$X_i = X_j$, $\forall i, j$), 所有个体渐近同步收敛到问题(13)的一个最优解的充分条件. 注意到情形2)相当于约束集信息是全局共享, 但没有将约束集做到完全的分布式. 之后文献[27]将上面结果推广到更一般的情形, 即各个受限集不同且连接图是一致联合强连通. 基于对偶理论的思想, 文献[28]考虑了带等式和不等式约束的受限优化问题, 并且提出了一种分布式原始对偶次梯度算法(distributed prime-dual sub-gradient algorithm).

在文献[28]的基础上, 笔者考虑了带不等式约束的受限优化问题:

$$\min_x \sum_{i=1}^{n} f_i(x),\ \text{s.t.}\ g(x) \leqslant 0,\ x \in X, \tag{15}$$

其中: 代价函数f_i只能被个体i观测到, 而约束函数g和受限集X是全局已知的. 在一些实际应用中, 由于某些计算或量测误差的存在, 精确的梯度往往是达不到的, 类似文献[11]提到的近似投影的概念, 在此基础上, 笔者最近还提出一种分布式近似梯度算法(approximate gradient algorithm), 并且给出了保证该算法收敛的梯度精确性条件.

无论是无约束问题(7)还是约束问题(13)(15)都是在目标函数和约束集满足凸性假设下给出的分析结果. 笔者在文章[29]中考虑了求解有限个非空非凸集合(带"洞"凸集$R_1, \cdots, R_n$)交集的问题, 即要找到非空交集$\bigcap_{i=1}^{n} R_i$中一个点. 这里集合R_i是从一个大的闭凸集X_i中挖去一个小的闭凸集Y_i, 即$R_i := \{x|x \in X_i \cap Y_i^c\}$, 其中$Y_i^c$是集合$Y_i$的余集. 对于这类的非凸优化问题, 用以下受限约束问题来近似求解:

$$\min_x \sum_{i=1}^{n} c_i|x|_{Y_i}^2,\ \text{s.t.} x \in X_0 = \bigcap_{i=1}^{n} X_i, \tag{16}$$

其中$c_1, \cdots, c_n$是一组给定的非负凸组合. 文献[30]提出一种近似投影算法, 对于每个个体i, 先是与自己邻居状态做一个加权平均, 接着向小集合Y_i做一个近似投影, 再向大集合X_i做精确投影. 然后给出了在投影精确度满足一定条件下算法收敛到问题(16)的最优解, 并且该最优解与参数$c_1, \cdots, c_n$的选取有关, 当参数选取恰当时, 该解位于交集$\bigcap_{i=1}^{n} R_i$中. 在分析受限约束问题(16)时, 与文献[4]相比, 本文给出在更一般情形下算法收敛性分析, 即约束集不同且连接图是一致联合强连通.

3.2 辅助约束(Auxiliary constraints)

这种辅助约束方法是将一些"显然"的知识加到算法中去以提高算法的效率, 研究该类问题的文献资料相对比较新.

其中一类是加入图(同步)的条件: 回顾上面提到

的分布式ADMM法[24], 针对连接图中的每条边$e_{ij}\in\mathcal{E}$都相应的引入一个约束条件$x_i=x_j$, 使得原无约束问题(7)等价于下面带约束问题:

$$\min\sum_{i=1}^{n}f_i(x_i),\ \text{s.t.}\ x_i=x_j,\ \forall e_{ij}\in\mathcal{E}. \tag{17}$$

由于此类约束条件具有一定的解耦性才使得对偶变量的更新能够分布式进行. 不同于文献[24], 文献[31]从Laplace矩阵角度, 引入约束条件$Lx=0$. 因为当系统连接图是固定强连通时, 条件$Lx=0$与$x_i=x_j,\ i,j\in\{1,\cdots,n\}$是等价的, 所以问题(7)等价于下面带约束问题:

$$\min_{x}\sum_{i=1}^{n}f_i(x_i),\ \text{s.t.}\ Lx=0, \tag{18}$$

其中$x=(x_1^{\mathrm{T}},\cdots,x_n^{\mathrm{T}})^{\mathrm{T}}$. 然后用类似文献[15]提出的用动态系统求解优化问题的思想, 在讨论有向图时, 文献[30]给出了下面连续时间优化算法:

$$\begin{cases}\dot{x}=-\alpha Lx-Lz-\nabla\tilde{f}(x),\\ \dot{z}=Lx,\end{cases} \tag{19}$$

其中: $z=(z_1^{\mathrm{T}},\cdots,z_n^{\mathrm{T}})^{\mathrm{T}}$, $\alpha>0$是控制参数. 它们给出了系统连接图是双随机强连通, 且α满足一定条件时, 算法(19)将收敛到问题(18)的一个最优解. 对于这类加入有关图信息约束的算法在分布式研究中还存在一定的局限性, 主要原因是它不能直接处理切换图的情况.

另一类利用了最优点的零梯度和性质. 对于无约束问题(7)而言, 其最优解x^*满足零梯度和性质, 即$\sum_{i=1}^{N}f_i(x^*)=0$. 文献[25]将零梯度和条件作为约束引入到问题中, 提出一种gossip算法, 并且在算法中一直保持该约束条件成立. 其算法的大体思想是在第k步迭代时任选一对个体(i,j)进行通信进而更新各自的状态, 而对于其他没被选中的个体其状态保持不变. 个体(i,j)在更新状态时必须满足梯度和为零的性质不变, 即

$$\begin{aligned}&f_i'(x_i(k))+f_j'(x_j(k))=\\&f_i'(x_i(k+1))+f_j'(x_i(k+1)),\end{aligned}$$

且$x_s(k+1)=x_s(k),\ \forall s\notin\{i,j\}$. 定义图$G_\infty=(V,E_\infty)$, 如果一对个体$(i,j)$被无限次重复选取进行通信, 则令边$e_{ij}\in E_\infty$. 文献[25]给出当图$G_\infty$是连通图时, 上述的gossip算法同步收敛到问题(7)的一个最优解.

虽说目前关于分布式带约束优化的研究已经取得一定的突破与成果, 但是总体上还处于发展阶段, 其理论研究往往局限于约束集和目标函数的凸性假设条件下. 对于一般的非凸优化问题, 现有的工作大都是进行凸放松或者凸近似, 将其转化为凸优化问题来求解以避开非凸带来的问题求解难度, 由此使得如何选择一种有效的凸近似方法变得尤其重要. 其次, 对于上面两小节介绍的一些分布式的优化算法来求解约束问题时, 大多数算法只能给出收敛性保证, 具体收敛速率的估计还不能得到. 总的来说分布式带约束优化的研究还存在很多的问题亟需解决.

4 分布式博弈(Distributed game)

在许多实际优化问题中, 还不得不考虑优化目标中多方对抗或合作的因素, 因此越来越多的控制工程师投入到分布式博弈优化算法的研究中. 分布式博弈问题有着重要的军用和民用价值, 也是多移动传感器网络或机器人网络等领域中的重要研究课题. 其中涉及到多组具有通信、计算、移动能力的个体如何分布式作用而实现系统的Nash均衡或其他群体任务.

有关分布式博弈的研究也很广[29,32], 包括网络演化博弈、分布式任务分配等. 势博弈已经研究了多年, 被广泛应用于交通拥塞博弈、区域覆盖、资源分配等实际问题中. 从分布式优化的角度出发, 笔者更多关注的是如何将势博弈的优势与其有效地结合. 文献[33]利用对偶理论, 提出一种基于状态的势博弈设计策略来解决分布式优化问题, 并且将其运用于电动车充电管理问题中.

下面主要就分布式零和博弈对相关工作进行介绍. 研究零和博弈的背景有两类: 第1类是现实问题本身就是零和博弈, 而第2类是将约束问题转换为该博弈问题.

在这里主要讨论一个最简单的情况, 它涉及到分布式计算鞍点[34–35]. 为此先给一个凸凹目标函数的定义, $\phi(\cdot,\cdot):X\times Y(\subset\mathbb{R}^{m_1}\times\mathbb{R}^{m_2})\to\mathbb{R}$是一个(严格)凸凹函数如果它是第1个变量的(严格)凸函数和第2个变量的(严格)凹函数. 给定一组点$(\hat{x},\hat{y})$, 记$\partial_x\phi(\hat{x},\hat{y})$为凸函数$\phi(\cdot,\hat{y})$在$\hat{x}$处的次梯度, 记$\partial_y\phi(\hat{x},\hat{y})$为凹函数$\phi(\hat{x},\cdot)$在$\hat{y}$处的次梯度. $(x^*,y^*)\in X\times Y$被称为是函数ϕ在$X\times Y$上的鞍点, 如果下式成立:

$$\phi(x^*,y)\leqslant\phi(x^*,y^*)\leqslant\phi(x,y^*),\ \forall x\in X,\ y\in Y. \tag{20}$$

一个策略博弈通常用$M=(I,W,U)$来表示, 其中I是所有参与方的集合: $W=W_1\times\cdots\times W_n$, n是所有参与者的数目, W_i是参与者i的策略集合, $U=(u_1,\cdots,u_n)$, $u_i:W\to\mathbb{R}$是参与者i的收益函数[2]. 如果$\sum_{i=1}^{n}u_i=0$, 则称M是零和博弈. 如果对所有的$i\in V$和所有的$w_i\in W_i$, 均有

$$u_i(w_i^*,w_{-i}^*)\geqslant u_i(w_i,w_{-i}^*), \tag{21}$$

则称$w^*=(w_1^*,\cdots,w_n^*)$是一个Nash均衡点, 其中w_{-i}^*表示除了参与者i外其他所有参与者的策略. 一个二人零和博弈$(n=2,u_1+u_2=0)$的所有Nash均衡点集

合和收益函数u_2的所有鞍点集合是相同的.

至今, 有关多智能体系统的分布式零和博弈问题的论文很少. 当然在这方面的研究也有基于次梯度的方法. 比如文献[36]提出了一个基于两个子网络的对抗和连续时间动力学的寻求Nash 均衡点的算法, 他们的结果主要是对固定权重平衡图得出的. 在此基础上, 文献[37]提出一个在切换拓扑下寻求Nash均衡点的离散算法, 并且对权重非平衡图也进行了研究.

尽管分布式博弈的严格理论结果还不算多, 但是分布式博弈的思想已经开始在很多工程问题(如多机器人搜捕或智能电网自律调度[38])得到应用.

5 结束语(Conclusions)

本文对当今分布式优化中的一些热点或重要问题, 介绍了研究背景和主要内容, 揭示了研究思想. 特别是对无约束和有约束优化的算法和收敛性做了简要的探讨, 并比较集中地介绍了笔者近年来在此方面所取得的工作进展. 一些未来可能的研究方向将包括: 1) 在拓扑结构方面, 如何对非平衡图的分布式优化给出有效的算法; 2) 在带有约束的优化方面, 如何利用多智能体的优势来化解约束方面的复杂性; 3) 考虑到现实的不确定性等因素, 如何考虑随机因素对优化的影响或如果给出有效的分布式随机优化算法来解决一些复杂的收敛性问题; 4) 考虑到通信约束, 如何分析量化、时延等通信问题对优化的精度和速度的影响; 5) 在计算复杂性方面, 如何有效地给出分布式算法的复杂性分析.

还值得一提的相关重要问题是如何充分地利用控制理论的优势将控制方法与分布式优化进行密切结合研究新问题. 在这方面, 一些有意义的初步结果也已经得到, 比如文献[39]就将分布式优化与PI 控制设计相结合提出一种连续时间的比例积分分布式优化算法, 并基于Lyapunov稳定性理论得到算法收敛性分析. 除此之外, 文献[40]也从类似的角度出发, 在充分采用Lyapunov方法的基础上提出了基于内模(internal model)的分布式优化算法来处理带时变干扰的动态优化问题. 当然在控制与优化的有效结合方面仍然还有许多问题需要探索和认识.

总之, 分布式优化方面的理论研究刚刚起步, 广泛的应用研究也在开展, 不仅仅限于本文所讨论的几个方面. 许多重要的理论与实际问题亟待解决, 新的研究领域还有待开拓.

参考文献(References):

[1] BOYD S P, VANDENBERGHE L. *Convex Optimization* [M]. New York: Cambridge University Press, 2004.

[2] OSBORNE M J. *A Course in Game Theory* [M]. Cambridge, MA: MIT Press, 1994.

[3] 洪奕光, 翟超. 多智能体系统动态协调与分布式控制设计 [J]. 控制理论与应用, 2011, 28(10): 1506 – 1512.
(HONG Yiguang, ZHAI Chao. Dynamic coordination and distributed control design of multi-agent systems [J]. *Control Theory & Applications*, 2011, 28(10): 1506 – 1512.)

[4] 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.

[5] COMBETTES P L. The convex feasibility problem in image recovery [J]. *Advances in Imaging and Electron Physics*, 1996, 95(25): 155 – 270.

[6] MENG W, XIAO W, XIE L. An efficient EM algorithm for energy-based multisource localization in wireless sensor networks [J]. *IEEE Transactions on Instrumentation and Measurement*, 2011, 60(3): 1017 – 1027.

[7] ARONSZAJN N. Theory of reproducing kernels [J]. *Transactions of the American mathematical society*, 1950, 68(3): 337 – 404.

[8] HALPERIN I. The product of projection operators [J]. *Acta Scientiarum Mathematicarum (Szeged)*, 1962, 23(1): 96 – 99.

[9] DEUTSCH F, HUNDAL H. The rate of convergence for the cyclic projections algorithm III: regularity of convex sets [J]. *Journal of Approximation Theory*, 2008, 155(2): 155 – 184.

[10] SHI G, JOHANSSON K H, HONG Y. Reaching an optimal consensus: dynamical systems that compute intersections of convex sets [J]. *IEEE Transactions on Automatic Control*, 2013, 58(3): 610 – 622.

[11] LOU Y, SHI G, JOHANSSON K H, et al. Reaching optimal consensus for multi-agent systems based on approximate projection [C] //*Proceedings of the 10th World Congress on Intelligent Control and Automation (WCICA)*. Beijing: IEEE, 2012: 2794 – 2800.

[12] 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.

[13] MENG Z, YANG T, SHI G, et al. Cooperative set aggregation for multiple Lagrangian systems [J/OL]. arXiv: 1402.2634, 2014.

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

[15] WANG J, ELIA N. Control approach to distributed optimization [C] //*Proceedings of the 48th Annual Conference on Communication, Control, and Computing*. Allerton: IEEE, 2010: 557 – 561.

[16] NEDIC A, OLSHEVSKY A, OZDAGLAR A, et al. Distributed subgradient methods and quantization effects [C] //*Proceedings of the 47th IEEE Conference on Decision and Control*. Mexico: IEEE, 2008: 4177 – 4184.

[17] RABBAT M G, NOWAK R D. Quantized incremental algorithms for distributed optimization [J]. *IEEE Journal on Selected Areas in Communications*, 2005, 23(4): 798 – 808.

[18] SHI G, JOHANSSON K H. Randomized optimal consensus of multi-agent systems [J]. *Automatica*, 2012, 48(12): 3018 – 3030.

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

[20] NEDIC A, BERTSEKAS D P. Incremental subgradient methods for nondifferentiable optimization [J]. *SIAM Journal on Optimization*, 2001, 12(1): 109 – 138.

[21] SRIVASTAVA K, NEDIC A. Distributed asynchronous constrained stochastic optimization [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2011, 5(4): 772 – 790.

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

[23] JAKOVATIC D, XAVIER J, MOURA J M. Fast distributed gradient methods [J/OL]. arXiv: 1112.2972, 2011.

[24] WEI E, OZDAGLAR A. Distributed alternating direction method of multipliers [C] //*Proceedings of IEEE Conference on Decision and Control*. Maui: IEEE, 2012: 5445 – 5450.

[25] LU J, TANG C Y, REGIER P R, et al. Gossip algorithms for convex consensus optimization over networks [J]. *IEEE Transactions on Auotmatic Control*, 2011, 56(12): 2917 – 2923.

[26] JAKOVATIC D, XAVIER J, MOURA J M. Cooperative convex optimization in networked systems: augmented Lagrangian algorithms with directed gossip communication [J]. *IEEE Transactions on Signal Processing*, 2011, 59(8): 3889 – 3902.

[27] LIN P, REN W. Distributed subgradient projection algorithm for multi-agent optimization with nonidentical constraints and switching topologies [C] //*Proceedings of IEEE Conference on Decision and Control*. Maui: IEEE, 2012: 6813 – 6818.

[28] ZHU M, MARTINEZ S. On distributed convex optimization under inequality and equality constraints via primal-dual subgradient methods [J]. *IEEE Transactions on Automatic Control*, 2012, 57(1): 151 – 164.

[29] FRIHAUF P, KRSTIC M, BASAR T. Nash equilibrium seeking in noncooperative games [J]. *IEEE Transactions on Automatic Control*, 2012, 57(5): 1192 – 1207.

[30] ZHANG Y, LOU Y, HONG Y. An approximate projection algorithm for distributed set intersection computation [C] //*Proceedings of the 32nd Chinese Control Conference*. Xi'an: IEEE, 2013: 7216 – 7221.

[31] GHARESIFARD B, CORTES J. Distributed continuous-time convex optimization on weight-balanced digraphs [J]. *IEEE Transactions on Automatic Control*, 2014, 59(3): 781 – 786.

[32] STANKOVIC M S, JOHANSSON K H, STIPANOVIC D M. Distributed seeking of Nash equilibria with applications to mobile sensor networks [J]. *IEEE Transactions on Automatic Control*, 2012, 57(4): 904 – 919.

[33] YI P, ZHANG Y, HONG Y. Design games to solve distributed optimization problem with application in electric vehicle charge management [C] //*Proceedings of the 32nd Chinese Control Conference*. Xi'an: IEEE, 2013: 6873 – 6878.

[34] DURR H B, EBENBAUER C. A smooth vector field for saddle point problems [C] //*Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC–ECC)*. Orlando: IEEE, 2011: 4654 – 4660.

[35] MAISTRISKII D. Gradient methods for finding saddle points [J]. Matekon 13, 1977: 3 – 22.

[36] GHARESIFARD B, CORTES J. Distributed convergence to Nash equilibria in two-network zero-sum games [J]. *Automatica*, 2013, 49(6): 1683 – 1692.

[37] LOU Y, HONG Y. Distributed convergence to Nash equilibrium of antagonistic optimization networks [C] //*Proceedings of the 32nd Chinese Control Conference*. Xi'an: IEEE, 2013: 6976 – 6980.

[38] 卢强, 盛成玉, 陈颖. 巨型风电并网系统的协同自律控制 [J]. 控制理论与应用, 2011, 28(10): 1491 – 1495.
(LU Qiang, SHENG Chengyu, CHEN Ying. Coordinated autonomous control strategy for power systems with large-scale wind power plants [J]. *Control Theory & Applications*, 2011, 28(10): 1491 – 1495.)

[39] DROGE G, KAWASHIMA H, EGERSTEDT M. Continuous-time proportional-integral distributed optimization for networked systems [J/OL]. arXiv: 1309.6613, 2014.

[40] WANG X, YI P, HONG Y. Dynamic optimization for multi-agent systems with external disturbances [J]. *Control Theory & Technology*, 2014, 12(2): 132 – 138.

作者简介:

洪奕光 (1966–), 男, 第4届(1997年)《关肇直奖》获奖论文作者, 研究员, 研究领域为多智能体系统、分布式优化、非线性系统动态分析和控制设计、社会网络、机器人等, E-mail: yghong@iss.ac.cn;

张艳琼 (1987–), 女, 博士研究生, 研究领域为多智能体系统和分布式优化, E-mail: yanqiong918@163.com.