使用网络节点信息传递策略的分布式优化新算法

马文龙[1], 付敏跃[2†], 张焕水[1]

(1. 山东大学 控制科学与工程学院, 山东 济南 250100; 2. 纽卡斯尔大学 电子工程与计算学院, 澳大利亚 新南威尔士州 2308)

摘要: 本文基于统计学习中众所周知的信度传播理论来研究非线性凸优化问题的分布式算法. 通过对优化问题中的网络图中节点上和节点之间的计算以及信息传递过程的深入研究, 结合信度传播理论得出适合分布式优化算法的信息传递策略. 在集中式经典牛顿法和原始对偶方法框架下, 所提分布式算法通过网络中的信息传递策略来完成设计. 所提的分布式牛顿–拉夫森算法在无圈连通图情形下是集中式牛顿法的分布式实现. 所提分布式原始对偶算法在无圈图情形下有集中式原始对偶算法的收敛效果, 且对于有圈连通图也有较好的适应性和鲁棒性. 仿真实验说明了我们所提信息传递策略和算法的收敛效果和适合的应用场景.

关键词: 多自主体系统; 凸优化; 牛顿–拉夫森方法; 原始对偶方法; 信度传播

引用格式: 马文龙, 付敏跃, 张焕水. 使用网络节点信息传递策略的分布式优化新算法. 控制理论与应用, 2021, 38(12): 2001 – 2009

DOI: 10.7641/CTA.2021.00596

New distributed optimization algorithm using network node information transfer strategy

MA Wen-long[1], FU Min-yue[2†], ZHANG Huan-shui[1]

(1. School of Control Science and Engineering, Shandong University, Jinan Shandong 250100, China;
2. School of Electrical Engineering and Computing, University of Newcastle, NSW 2308, Australia)

Abstract: This paper is based on the well-known belief propagation theory in statistical learning to study distributed algorithms for nonlinear convex optimization problems. Through in-depth research on the calculation and information transfer process on and between nodes in the network graph of the optimization problem, combined with the belief propagation theory, an information transfer strategy suitable for distributed optimization algorithms is obtained. Under the framework of the centralized classic Newton method and the primal-dual method, the proposed distributed algorithm completes the design using the information transmission in the network. The proposed distributed Newton–Raphson algorithm is a distributed implementation of the centralized Newton method in the case of acyclic connected graphs. The proposed distributed primitive dual algorithm has a similar convergence property of the centralized primal-dual algorithm in the case of acyclic graphs, and it also has better adaptability and robustness for cyclic connected graphs. The simulation experiment illustrates the convergence property and suitable application scenarios of our proposed information transmission strategy and algorithms.

Key words: multi-agent systems; convex optimization; Newton–Raphson method; primal-dual method; belief propagation

Citation: MA Wenlong, FU Minyue, ZHANG Huanshui. New distributed optimization algorithm using network node information transfer strategy. *Control Theory & Applications*, 2021, 38(12): 2001 – 2009

1 引言

本文研究由稀疏无向连通图$\mathcal{G}=\{\mathcal{V},\mathcal{E}\}$来刻画的分布式优化问题:

$$\min_{x\in\mathbb{R}^p} f(x) \doteq \sum_{i=1}^{n} f_i(x), \tag{1}$$

其中: $\mathcal{V}=\{1,2,\cdots,n\}$, $\mathcal{E}\subset\mathcal{V}\times\mathcal{V}$. 网络中各节点具有私有成本函数$f_i(x):\mathbb{R}^p\to\mathbb{R}$. $f(x)$是总成本函数, $x\in\mathbb{R}^p$是要优化的全局决策变量. 这个问题已经在多自主控制[1]、无线传感器网络中的传感器融合[2]、分布式学习[3–4]中有多种应用.

收稿日期: 2020−09−06; 录用日期: 2021−05−18.

†通信作者. E-mail: minyue.fu@newcastle.edu.au.

本文责任编委: 俞立.

国家自然科学基金项目(61633014, 61573221, U1701264)资助.

Supported by the National Natural Science Foundation of China (61633014, 61573221, U1701264).

在分布式的意义下, 优化问题(1)可以等价改写为

$$\begin{aligned} &\min_{\boldsymbol{x}\in\mathbb{R}^{np}} F(\boldsymbol{x}) \doteq \sum_{i=1}^{n} f_i(x_i),\\ &\text{s.t.}\ \ x_i = x_j,\ \forall i,j\in\mathcal{V}, \end{aligned} \tag{2}$$

这里$\boldsymbol{x}=\mathrm{col}\{x_1,x_2,\cdots,x_n\}\in\mathbb{R}^{np}$. 本文所关注和追求的算法是仅通过每个节点的计算以及和与邻居节点间的通信来实现的分布式优化算法.

在现有大量的分布式优化算法的文献中, 分布式优化算法主要是在经典优化算法如梯度类方法, 或对偶方法类的基础上结合趋同矩阵来重构的分布式迭代算法. 在梯度方法类中具有代表性的文献主要有, 文献[5]中的分布式(次)梯度法(distributed subgradient method, DSM), 文献[6]中考虑提前量的分布式涅斯捷罗夫加速梯度法(Nesterov gradient method), 文献[7]利用多步迭代状态信息的分布式重球(heavy-ball)梯度方法, 文献[8]中利用优化函数的二阶梯度信息的近似牛顿算法. 文献[9]中通过动态系统来迭代构建趋近于全局牛顿–拉夫森梯度方向的分布式算法. 另一类在原始对偶算法的文献中具代表性的主要有, 文献[10]中借助趋同矩阵文献提出的分布式拉格朗日算法, 以及大量的如文献[11–13]中研究的分布式的交替方向乘子法(alternating direction method of multipliers, ADMM)算法. 现存的这两大类分布式优化算法中, 其内嵌的趋同矩阵是算法实现分布式的关键, 并使整体的分布式优化算法设计具有简单明确, 收敛性易于证明的特点. 但这同时也由于这类趋同矩阵(随机矩阵或双随机局阵)的使用降低了分布式算法的收敛速度, 使得这样的分布式算法很难达到相应的中心式算法的收敛速度.

在本文中, 不同于上述的两类代表性文献中的直接使用趋同矩阵的分布式优化算法设计, 本文深入到问题所涉及的网络拓扑图中节点间的信息传递方式来考虑算法设计. 这里, 本文基于在通讯和编码理论中广泛研究的信度传播(belief propagation)[14–15]理论, 直接设计节点间的迭代信息传递策略. 这一策略确保可以通过有限步和邻居节点间信息交换来达到精确趋同(无圈图)或渐进趋同(有圈图), 同样这一策略适当修改可以用于线性方程组分布式求解. 本文在作者之前两篇文章[16–17]的基础上归纳改进出的两个分布式优化算法进一步凸显了这一策略在分布式优化算法设计的有效性. 信度传播理论和算法主要是利用网络中节点与节点之间相互传递信息而更新当前整个网络中的状态, 该算法是一种迭代的方法且所有信息的传播可以并行实现, 经过多次迭代后, 所有节点的信度不再发生变化, 就称此时每一个节点的状态即为最优状态, 对于无环连通图, 信度传播算法可以收敛到其最优解. 这一算法主要用在解决概率图模型概率推断问题和通信学习中. 在本文中, 作者借鉴这一理论来设计本文优化算法中替换大量文献中常用的趋同矩阵的关键的工具.

在具体的分布式优化算法设计中, 首先, 基于一、二阶梯度信息和通过信度传播理论修改设计的信息传播策略策略给出牛顿迭代的分布式算法, 这个策略使每个节点有限步精确获取全局的牛顿梯度下降方向, 这样的双层迭代算法在无圈连通图情况下可以做到中心式牛顿迭代算法的收敛; 其次, 作者采取原始对偶的迭代框架设计算法来应对有圈图情形, 通过适当修改信息传递策略使之能用来分布式迭代求解最优性条件转化来的线性方程组. 在无圈图情况下, 本文的信息传递策略能精确求解原始对偶框架下的子优化问题; 对于有圈图, 本文的整体算法也有好的收敛和适应性. 仿真实验进一步说明分布式原始对偶算法对有圈图情形的有效性, 对比现有使用二次梯度信息算法[8], 更体现本文分布式原始对偶算法的优势.

文章主要内容包括问题描述, 算法设计、仿真实验、结论与展望和参考文献. 本文采用$\mathrm{diag}\{\cdot\}$表示对角矩阵; $\mathrm{col}\{\cdot\}$表示列向量; $|\cdot|$表示对“$\cdot$”中元素取绝对值; $\otimes$为克罗内克积(kronecker product); $\mathbf{1}_n$为元属均为1的向量; $\boldsymbol{I}_{p\times p}$为单位矩阵; $\boldsymbol{A}\succeq\mathbf{0}(\succ\mathbf{0})$表示矩阵$\boldsymbol{A}$为半正定(正定)矩阵; $\doteq$为定义符号.

2 问题描述

2.1 网络模型和相关假设

本文考虑一个由二元组$\mathcal{G}=\{\mathcal{V},\mathcal{E}\}$刻画的(稀疏)网络图, 这是具有$n$个节点(或自主体)和$m$个边的无向连通简单图, $\mathcal{V}$和$\mathcal{E}$分别是节点和边的集合. $\mathcal{N}_i=\{j\in\mathcal{V}|(i,j)\in\mathcal{E}\}$是节点$i$的邻居. 假设在边$e_{ij}$的表示中$i<j$, 边点关联矩阵$\boldsymbol{C}\in\mathbb{R}^{m\times n}$中每一列代表一个节点, 每一行对应一个边. 在此矩阵的列中, 所有边均根据字典顺序排序. 则$\boldsymbol{C}$的元素可以定义为

$$\boldsymbol{C}_{((i,j),s)} \doteq \begin{cases} 1, & e_{ij}\in\mathcal{E},\ s=i,\\ -1, & e_{ij}\in\mathcal{E},\ s=j,\\ 0, & \text{其他}. \end{cases} \tag{3}$$

图$\mathcal{G}$的拉普拉斯矩阵$\boldsymbol{L}\doteq\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}=\boldsymbol{D}-\boldsymbol{A}$, 是对称半正定矩阵, 且满足$\boldsymbol{L}\mathbf{1}=\mathbf{0}$. 这里$\boldsymbol{A}\in\mathbb{R}^{n\times n}$是节点邻接矩阵, $\boldsymbol{D}$是以图$\mathcal{G}$中节点的度(d_i为节点i的度)为元素的对角矩阵.

假设 1 图$\mathcal{G}=\{\mathcal{V},\mathcal{E}\}$是连通、无向、简单图.

2.2 问题描述

在图$\mathcal{G}=\{\mathcal{V},\mathcal{E}\}$中, 本文的目标是通过网络中各节点的协作来求解函数$f(x)$或$F(\boldsymbol{x})$的最小值, 即分布式解决问题(1)或等价问题(2). $\mathbb{R}^p$中的x_i是保存在节点i上的全局决策变量x的副本, 由于图是双向连通的, 因此问题(1, 2)中的优化问题还可等价于

$$\{x_i^*\}=\arg\min_{\{x_i\}_{i=1}^n}\sum_{i=1}^n f_i(x_i), \tag{4a}$$

$$\text{s.t. } x_i - x_j = \mathbf{0}_p,\ \forall i,j \in \mathcal{N}_i, \tag{4b}$$

这里$x_1^* = x_2^* = \cdots = x^*$, x^*为问题(1)的最优值解. 利用式(3)中的关联矩阵$\boldsymbol{C}$, 记$\boldsymbol{x}^* \doteq \mathbf{1}_n \otimes x^*$, $C \doteq \boldsymbol{C} \otimes \boldsymbol{I}_{p\times p}$, 这里$C \in \mathbb{R}^{mp\times np}$, $\boldsymbol{x}^* \in \mathbb{R}^{np}$. 本文得到优化问题(4a)的向量形式

$$\begin{aligned} &\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in\mathbb{R}^{np}} F(\boldsymbol{x}),\\ &\text{s.t. } C\boldsymbol{x} = \mathbf{0}_{mp}. \end{aligned} \tag{5}$$

在本文中, 考虑到算法设计以及保证最优解的存在唯一性, 有如下对目标函数的假设:

假设 2 对于$\mathcal{V}$中任意节点i, $f_i(\cdot)$是二阶连续可微的闭恰当函数, 还具有严格凸和光滑的性质.

注 1 在假设(2)下, 存在$0 < \gamma \leqslant \varGamma \leqslant \infty$, 使得$\gamma \boldsymbol{I}_{p\times p} \preceq \nabla^2 f_i(x) \preceq \varGamma \boldsymbol{I}_{p\times p}$. 这样$f(x)$, $F(\boldsymbol{x})$ 也是严格凸的二阶连续可微的光滑函数, 二阶梯度也是正且有界的.

3 算法设计

在本章中, 利用网络节点私有目标函数的一、二阶梯度信息, 本文依据信度传播理论在之前工作[16]的基础上修改改进的节点间的信息传递策略来设计分布式优化算法. 在所提的两个分布式算法中, 分布式牛顿法是由这一信息传递策略结合牛顿–拉夫森方法构成; 进一步修改这一信息传递策略, 应用在原始对偶框架下得出分布式原始对偶算法.

3.1 分布式牛顿算法

在本章节中, 在图$\mathcal{G}$为无圈图的情况下考虑问题(1). 依据经典的牛顿–拉夫森迭代法,

$$\boldsymbol{x}(k+1) = \psi(\boldsymbol{x}(k)), \tag{6}$$

这里$\psi(\boldsymbol{x}(k)) \doteq (h(\boldsymbol{x}(k)))^{-1} g(\boldsymbol{x}(k))$. 定义全局变量

$$\begin{aligned} &h(\boldsymbol{x}(k)) \doteq \sum_{i=1}^n \nabla^2 f_i(x_i(k)),\\ &g(\boldsymbol{x}(k)) \doteq \sum_{i=1}^n [h_i(x_i(k))x_i(k) - \nabla f_i(x_i(k))]. \end{aligned}$$

为了获取全局变量$h(\boldsymbol{x}(k))$, $g(\boldsymbol{x}(k))$进而计算$\psi(\boldsymbol{x}(k))$. 类似之前文献[16]采用的信息传递策略, 从如下任意节点$i \in \mathcal{V}$私有目标函数的一、二阶信息

$$\begin{aligned} &g_i(x_i(k)) \doteq h_i(x_i(k))x_i(k) - \nabla f_i(x_i(k)),\\ &h_i(x_i(k)) \doteq \nabla^2 f_i(x_i(k)) \end{aligned}$$

出发来设计信息传递策略. 令$s_{i\to j}(l)$和$t_{i\to j}(l)$表示在迭代l时由节点i传递给节点j的信息. 考虑如下的信息传递策略和计算过程:

1) 初始化($l=0$),

$$\begin{cases} s_{i\to j}(0) = g_i(x_i(k)),\\ t_{i\to j}(0) = h_i(x_i(k)). \end{cases} \tag{7}$$

2) 在迭代$l=1,2,\cdots,d$中, 对于任意$i\in\mathcal{V}$, 计算

$$\tilde{s}_i(l) = g_i(x_i(k)) + \sum_{j\in\mathcal{N}_i} s_{j\to i}(l-1), \tag{8a}$$

$$\tilde{t}_i(l) = h_i(x_i(k)) + \sum_{j\in\mathcal{N}_i} t_{j\to i}(l-1). \tag{8b}$$

3) 对于$j \in \mathcal{N}_i$, 计算

$$\begin{cases} s_{i\to j}(l) = \tilde{s}_i(l) - s_{j\to i}(l-1),\\ t_{i\to j}(l) = \tilde{t}_i(l) - t_{j\to i}(l-1). \end{cases} \tag{9}$$

这里给出在无圈图$\mathcal{G}$行情下上述信息传递策略和计算过程的详细分析并综合这一策略所能达到的结果. 考虑无圈图$\mathcal{G}$的一个分解$\mathcal{G} = \mathcal{G}_i \cup (i,j) \cup \mathcal{G}_j$, 移除边$(i,j)$后即可得到两个不相交的子图$\mathcal{G}_i$和$\mathcal{G}_j$, $\mathcal{V}_i$和$\mathcal{V}_j$是相应的点集. 令$\mathcal{V}_i(l)$是属于$\mathcal{V}$中的距离节点i至少l步远的点集($\forall l > 0$且$i \notin \mathcal{V}_i(l)$).

综合分析上面的信息传递策略和计算过程, 对于每一个节点i本质上是在迭代过程中逐步不重复地收集到$\mathcal{V}_i(l)$中节点信息. 对于这一策略所能达到的结果, 本文用数学归纳法来分析如下: 首先在$l=1$时, i传给j的信息除了自身信息$g_i(x_i(k))$还应包含$s_{m\to i}(0)$, $m\in\mathcal{N}_i \setminus j$. 又由于$m\in\mathcal{N}_i \setminus j$等价为$m\in\mathcal{V}_i(1)\setminus\mathcal{V}_j$, 从而有

$$\begin{aligned} s_{i\to j}(1) = g_i(x_i(k)) + \sum_{m\in\mathcal{N}_i\setminus j} s_{m\to i}(0) = \\ g_i(x_i(k)) + \sum_{m\in\mathcal{V}_i(1)\setminus\mathcal{V}_j} g_m(x_m(k)), \end{aligned} \tag{10}$$

其次考虑$l=2$时, $s_{i\to j}(2)$除了自身信息$g_i(x_i(k))$还应该包含$s_{m\to i}(1)$, $m\in\mathcal{N}_i\setminus j$, 由$l=1$的情况, $s_{m\to i}(1)$除了自身信息$g_m(x_m(k))$还应该包含$s_{v\to m}(0)$, $v \in \mathcal{N}_m \setminus i$. 这样$v$的集合又可以看做是集合$\mathcal{V}_i(2)\setminus\mathcal{V}_j$, 从而有

$$\begin{aligned} s_{i\to j}(2) = g_i(x_i(k)) + \sum_{m\in\mathcal{N}_i\setminus j} s_{m\to i}(1) = \\ g_i(x_i(k)) + \sum_{m\in\mathcal{V}_i(2)\setminus\mathcal{V}_j} g_m(x_m(k)). \end{aligned} \tag{11}$$

重复以上内容, 通过数学归纳, 对于$\forall l > 0$, 得出

$$s_{i\to j}(l) = g_i(x_i(k)) + \sum_{m\in\mathcal{V}_i(l)\setminus\mathcal{V}_j} g_m(x_m(k)). \tag{12}$$

对于$\forall j \in \mathcal{N}_i$, 则在$\mathcal{G}_j$中与节点$j$距离$l-1$步远的节点与节点$i$的距离为$l$, 这样就有

$$\mathcal{V}_i(l)\setminus\mathcal{V}_j \sqcup \mathcal{V}_j(l-1)\setminus\mathcal{V}_i \sqcup \{j\} = \mathcal{V}_i(l).$$

这样联系式(9)(12), 可得

$$\begin{aligned} \tilde{s}_i(l) = s_{i\to j}(l) + s_{j\to i}(l-1) = \\ g_i(x_i(k)) + \sum_{m\in\mathcal{V}_i(l)} g_m(x_m(k)). \end{aligned}$$

同理可得到$\tilde{t}_i(l)$的类似等式, 综上可得

$$\tilde{s}_i(l) = g_i(x_i(k)) + \sum_{j\in\mathcal{V}_i(l)} g_j(x_j(k)), \tag{13a}$$

$$\tilde{t}_i(l) = h_i(x_i(k)) + \sum_{j\in\mathcal{V}_i(l)} h_j(x_j(k)). \tag{13b}$$

这样在固定迭代k和任意确定的i, 在内层迭代l逐步增

大时, $\tilde{s}_i(l), \tilde{t}_i(l)$收集到的信息在逐步增多. 当$l \geqslant d$($d$为图的直径)时, $\tilde{s}_i(l) = g(\boldsymbol{x}(k)), \tilde{t}_i(l) = h(\boldsymbol{x}(k))$. 又有

$$g(\boldsymbol{x}(k)) = \sum_{i=1}^{n}[h_i(x_i(k))x_i(k) - \nabla f_i(x_i(k))], \tag{14}$$

$$h(\boldsymbol{x}(k)) = \sum_{i=1}^{n}\nabla^2 f_i(x_i(k)). \tag{15}$$

进而这样就保证了$\psi(\boldsymbol{x}(k)) \doteq (h(\boldsymbol{x}(k)))^{-1}g(\boldsymbol{x}(k))$在足够多的内层迭代($l > d$, d为图的直径)时可以为任意节点i获得, 在结合外部迭代(6)易于得出如下结论:

定理 1 在假设1–2以及无圈图的条件下, 如下的牛顿迭代:

$$x_i(k+1) = \psi(\boldsymbol{x}(k)), \tag{16}$$

其中$\psi(\boldsymbol{x}(k))$由每个i通过上述的信息传递策略和计算过程(7)–(9)获得. 则对于任意节点$i \in \mathcal{N}_i$在任意初值下都收敛到唯一的最优值x^*. 特别地, 若选取相同的初值$x_i(0) = \bar{x}(0), \forall i \in \mathcal{V}$, 每个节点的迭代轨迹重合, 也即$x_i(k) = x_j(k) = \bar{x}(k), \forall i, j \in \mathcal{V}$.

注 2 每一个节点i的状态变量$x_i(k)$由相同的全局变量$\psi(\boldsymbol{x}(k))$来驱动, 即每个节点的迭代都是中心式牛顿迭代的再现. 为简洁记, 不再通过仿真实验验证定理1的收敛.

这一分布式牛顿算法是传递策略和计算过程(7)–(9)在经典牛顿法上的直接应用, 在无圈连通图情况下经过足够的内迭代实现了中心式牛顿法的分布式实现. 这一算法对于更一般的连通图(主要针对有圈图), 类似文献[16]也可以推广使用并具有收敛性, 但在本文中, 作者选择在更具鲁棒性的原始对偶方法框架下适当改进这一信息传递策略和计算过程来设计分布式优化算法.

3.2 分布式原始对偶算法

在更具鲁棒性的原始对偶框架下, 兼顾考虑有圈图的情形, 从等价问题(5)出发, 结合文献[17]来完成算法的设计. 令问题(5)的增广拉格朗日函数$\mathcal{L}_\rho: \mathbb{R}^{np} \times \mathbb{R}^{mp} \to \mathbb{R}$(参数$\rho > 0$)为

$$\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}) \doteq F(\boldsymbol{x}) + \rho \boldsymbol{y}^{\mathrm{T}}C\boldsymbol{x} + \frac{\rho}{2}\boldsymbol{x}^{\mathrm{T}}L\boldsymbol{x}, \tag{17}$$

这里$\boldsymbol{y} \in \mathbb{R}^{mp}$为对偶变量, $L \doteq C^{\mathrm{T}}C = \boldsymbol{L} \otimes \boldsymbol{I}_{p\times p}$. 增广项$\frac{\rho}{2}\boldsymbol{x}^{\mathrm{T}}L\boldsymbol{x}$为违反趋同一致性限制的惩罚项. 在假设2下, 增广拉格朗日函数有如下推论.

推论 1[18] 由假设2, $\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y})$具有惟一的鞍点, 即存在惟一的解对$(\boldsymbol{x}^*, \boldsymbol{y}^*)$对于所有的$\boldsymbol{x} \in \mathbb{R}^{np}, \boldsymbol{y} \in \mathbb{R}^{mp}$满足

$$\mathcal{L}_\rho(\boldsymbol{x}^*, \boldsymbol{y}) \leqslant \mathcal{L}_\rho(\boldsymbol{x}^*, \boldsymbol{y}^*) \leqslant \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}^*).$$

基于增广拉格朗日函数(17), 原始对偶框架如下:

$$\boldsymbol{x}(k+1) = \arg\min_{\boldsymbol{x}} \mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{y}(k)) + \frac{\epsilon}{2}\|\boldsymbol{x} - \boldsymbol{x}(k)\|^2, \tag{18a}$$

$$\boldsymbol{y}(k+1) = \boldsymbol{y}(k) + \rho C\boldsymbol{x}(k+1), \tag{18b}$$

把$\boldsymbol{y}^{\mathrm{T}}(k)C$看作一个整体, 记$\boldsymbol{q}(k) \doteq C^{\mathrm{T}}\boldsymbol{y}(k)$为新对偶迭代变量. 则原始对偶迭代可以改写为

$$\begin{cases} \boldsymbol{x}(k+1) = \arg\min_{\boldsymbol{x}} \mathcal{L}_{\rho,\epsilon}(\boldsymbol{x}, \boldsymbol{q}(k)), \\ \boldsymbol{q}(k+1) = \boldsymbol{q}(k) + \rho L\boldsymbol{x}(k+1). \end{cases} \tag{19}$$

这里

$$\mathcal{L}_{\rho,\epsilon}(\boldsymbol{x}, \boldsymbol{q}(k)) = F(\boldsymbol{x}) + \rho \boldsymbol{q}^{\mathrm{T}}(k)\boldsymbol{x} + \frac{\rho}{2}\boldsymbol{x}^{\mathrm{T}}L\boldsymbol{x} + \frac{\epsilon}{2}\|\boldsymbol{x} - \boldsymbol{x}(k)\|^2,$$

最后一项只用于提高迭代稳定性.

首先$\mathcal{L}_{\rho,\epsilon}(\boldsymbol{x}, \boldsymbol{q}(k))$的二阶近似形式可写为

$$\begin{aligned}&\mathcal{L}_{\rho,\epsilon}(\boldsymbol{x}, \boldsymbol{q}(k)) \approx \\ &F(\boldsymbol{x}(k)) + \boldsymbol{q}^{\mathrm{T}}(k)\boldsymbol{x}(k) + \frac{\rho}{2}\boldsymbol{x}^{\mathrm{T}}(k)L\boldsymbol{x}(k) + \\ &(\nabla F(\boldsymbol{x}(k))^{\mathrm{T}} + \boldsymbol{q}^{\mathrm{T}}(k) + \rho \boldsymbol{x}^{\mathrm{T}}(k)L)\Delta\boldsymbol{x} + \\ &\frac{1}{2}\Delta\boldsymbol{x}^{\mathrm{T}}(\nabla^2 F(\boldsymbol{x}(k)) + \rho L + \epsilon \boldsymbol{I}_{np\times np})\Delta\boldsymbol{x},\end{aligned} \tag{20}$$

其中$\Delta\boldsymbol{x} \doteq \boldsymbol{x} - \boldsymbol{x}(k)$,

$$\nabla F(\boldsymbol{x}(k)) \doteq (\nabla f_1(x_1(k)) \;\; \nabla f_2(x_2(k)) \;\; \cdots \;\; \nabla f_n(x_n(k)))^{\mathrm{T}},$$

$$\nabla^2 F(\boldsymbol{x}(k)) \doteq \mathrm{diag}\{\nabla^2 f_1(x_1(k)), \nabla^2 f_2(x_2(k)), \cdots, \nabla^2 f_n(x_n(k))\}.$$

令$\mathcal{L}_{\rho,\epsilon}(\Delta\boldsymbol{x}, \boldsymbol{q}(k))$关于$\Delta\boldsymbol{x}$的梯度为零, 记这时的$\Delta\boldsymbol{x}$为$\Delta\boldsymbol{x}(k)$, 可得如下方程:

$$(\nabla^2 F(\boldsymbol{x}(k)) + \rho L + \epsilon \boldsymbol{I}_{np\times np})\Delta\boldsymbol{x}(k) = -(\nabla F(\boldsymbol{x}(k)) + \boldsymbol{q}(k) + \rho L\boldsymbol{x}(k)).$$

令

$$\boldsymbol{b}(k) \doteq -(\nabla F(\boldsymbol{x}(k)) + \boldsymbol{q}(k) + \rho L\boldsymbol{x}(k)),$$

$$\boldsymbol{H}(k) \doteq \nabla^2 F((k)) + \rho L + \epsilon \boldsymbol{I}_{np\times np},$$

则上述等式变成方程组求解问题

$$\boldsymbol{H}(k)\Delta\boldsymbol{x}(k) = \boldsymbol{b}(k). \tag{21}$$

对于$\boldsymbol{H}(k) = \{h_{ij}(k)\}$, $h_{ij} \in \mathbb{R}^{p\times p}$为分块矩阵, 有如下的推论.

推论 2 矩阵$\boldsymbol{H}(k)$是严格主对角占优的.

证 在$\boldsymbol{H}(k)$的构成中, $\boldsymbol{L} = \{l_{ij}\}$是对称半正定的拉普拉斯矩阵, $\forall i \in \mathcal{V}$, $l_{ii} = \sum_{\forall j\neq i}|l_{ij}|$, 再由假设2可知, $\nabla^2 F(\boldsymbol{x})$是正定对角矩阵. 又由于$\epsilon > 0$, 则$\forall i \in \mathcal{V}$,

$$h_{ii} = \nabla^2 f_i(x_i(k)) + l_{ii}\boldsymbol{I}_{p\times p} + \epsilon \boldsymbol{I}_{p\times p} \succ \sum_{j\neq i}|l_{ij}|\boldsymbol{I}_{p\times p} = \sum_{\forall j\neq i}|h_{ij}|,$$

从而结论成立.

这样迭代(19)可以写为如下迭代:

$$\boldsymbol{x}(k+1)=\boldsymbol{x}(k)+\Delta\boldsymbol{x}(k), \tag{22a}$$

$$\boldsymbol{H}(k)\Delta\boldsymbol{x}(k)=\boldsymbol{b}(k), \tag{22b}$$

$$\boldsymbol{q}(k+1)=\boldsymbol{q}(k)+\rho L\boldsymbol{x}(k+1). \tag{22c}$$

迭代过程(22)构成了问题(5)完整的原始和对偶迭代框架. 方程组(22b)的分布式求解作为这一框架的内层迭代为整个问题的关键. 本文用增广矩阵$[\boldsymbol{H}(k):\boldsymbol{b}(k)]$表示所要处理的方程组. 依据推论2, $\boldsymbol{H}(k)$是稀疏的主对角严格占优的对称正定矩阵.

$$\boldsymbol{b}(k)=\mathrm{col}\{b_1(k),b_2(k),\cdots,b_n(k)\},$$
$$b_i(k)=-\nabla f_i(x_i(k))-q_i(k)-\rho(d_ix_i-\sum_{j\in\mathcal{N}_i}x_j).$$

对于方程组具体求解, 要有别于前述牛顿算法中的信息传递策略和计算过程, 结合文献[17]研究基础, 改进的信息传递策略和计算过程如下: 在固定的原始对偶外层迭代k情况下, 定义$h_{i\to j}(t)$和$b_{i\to j}(t)$为在t步迭代中两个用来传递从节点i到节点$j\in\mathcal{N}_i$信息的变量. 选取初始值

$$h_{i\to j}(0)=h_{ii}(k)\in\mathbb{R}^{p\times p},\ b_{i\to j}(0)=b_i(k)\in\mathbb{R}^p.$$

对于$\forall i\in\mathcal{V},j\in\mathcal{N}_i$, 对$[\boldsymbol{H}(k):\boldsymbol{b}(k)]$进行下述操作:

$$\tilde{h}_i(k,t)=h_{ii}(k)-\sum_{v\in\mathcal{N}_i}h^{-1}_{v\to i}(t-1)h_{vi}(k)h_{iv}(k), \tag{23a}$$

$$\tilde{b}_i(k,t)=b_i(k)-\sum_{v\in\mathcal{N}_i}h^{-1}_{v\to i}(t-1)h_{iv}(k)\times b_{v\to i}(t-1), \tag{23b}$$

$$\Delta x_i(k,t)=\tilde{h}_i^{-1}(k,t)\tilde{b}_i(k,t), \tag{23c}$$

$$h_{i\to j}(t)=\tilde{h}_i(k,t)+h^{-1}_{j\to i}(t-1)h_{ji}(k)h_{ij}(k), \tag{23d}$$

$$b_{i\to j}(t)=\tilde{b}_i(k,t)+h^{-1}_{j\to i}(t-1)h_{ij}(k)\times b_{j\to i}(t-1). \tag{23e}$$

这是一个完整的信息传递和计算的迭代过程, 相对于前述的迭代过程(7)–(9)的只是在之前项中添加了求逆的运算. 添加求逆运算的作用在整体方程组解的不变性, 添加的求逆运算的合法性由后续的推论3保证. 中间变量$\tilde{h}_i(k,t)$和$\tilde{b}_i(k,t)$分别是对$\boldsymbol{H}(k)$的第i行主元和$\boldsymbol{b}(k)$中对应第i分量的在内层迭代为t时的变化结果的记录. 并记它们的向量形式为

$$\tilde{\boldsymbol{H}}(k,t)\doteq\mathrm{diag}\{\tilde{h}_1(k,t),\tilde{h}_2(k,t),\cdots,\tilde{h}_n(k,t)\},$$
$$\tilde{\boldsymbol{b}}(k,t)\doteq\mathrm{col}\{\tilde{b}_1(k,t),\tilde{b}_2(k,t),\cdots,\tilde{b}_n(k,t)\}.$$

在网络图$\mathcal{G}$为无圈连通图时, 当内层迭代t足够多(大于图的直径)时, 由式(23c)计算得到的解为精确解, 这时方程组$[\boldsymbol{H}(k):\boldsymbol{b}(k)]$和$[\tilde{\boldsymbol{H}}(k,t):\tilde{\boldsymbol{b}}(k,t)]$之间只相差若干的行线性初等变换, 是相互等价的. 在网络图$\mathcal{G}$为有圈连通图时, 足够多的内迭代也是相当于在对方程组$[\boldsymbol{H}(k):\boldsymbol{b}(k)]$进行行初等变换, 保持方程组解的不变性, 但不能完全消去所以的方程组中矩阵$\boldsymbol{H}(k)$的非对角元素, 只是使矩阵$\boldsymbol{H}(k)$的主对角元素站优的权重在随着内迭代逐步加大. 这种情况下记录的$[\tilde{\boldsymbol{H}}(k,t):\tilde{\boldsymbol{b}}(k,t)]$作为方程组的解在随着迭代逐步逼近方程组$[\boldsymbol{H}(k):\boldsymbol{b}(k)]$的解.

对这一内层迭代过程式(7)–(9)的分析可以直接总结出定理2.

推论 3[17] 在$\boldsymbol{H}(k)$严格主对角占优且迭代初值为$h_{i\to j}(0)=h_{ii}(k), b_{i\to j}(0)=b_i(k), \tilde{h}_i(0)=h_{ii}(k)$的条件下, 对于$\forall i\in\mathcal{V}, j\in\mathcal{N}_i, t=0,1,\cdots$, 有

$$h_{i\to j}(t)\succ|h_{ij}(k)|,\ \tilde{h}_i(k,t)\succ\mathbf{0}_{p\times p}.$$

证 由初始条件并用数学归纳法易于证明推论的成立, 详情可参考文献[17].

定理 2 在假设1–2下, 设所涉及的$\mathcal{G}$的直径为d, 迭代过程(23)具有如下的结论:

1) 若图$\mathcal{G}$为无圈图, 经过$t\geqslant d$次的迭代过程, 方程组$[\boldsymbol{H}(k):\boldsymbol{b}(k)]$等价为$[\tilde{\boldsymbol{H}}(k,t):\tilde{\boldsymbol{b}}(k,t)]$, 且有

$$\Delta\boldsymbol{x}(k,t)=\tilde{\boldsymbol{H}}(k,t)^{-1}\tilde{\boldsymbol{b}}(k,t)=\Delta\boldsymbol{x}^*(k)=\boldsymbol{H}(k)^{-1}\boldsymbol{b}(k). \tag{24}$$

当$0<t<d$时, $\Delta\boldsymbol{x}(k,t)=\tilde{\boldsymbol{H}}^{-1}(k,t)\tilde{\boldsymbol{b}}(k,t)$,

$$\lim_{t\to d}\Delta\boldsymbol{x}(k,t)=\Delta\boldsymbol{x}^*(k)=\boldsymbol{H}(k)^{-1}\boldsymbol{b}(k). \tag{25}$$

2) 若所涉及的图$\mathcal{G}$为有圈图时, 经过t次的迭代过程, $\Delta\boldsymbol{x}(k,t)=\tilde{\boldsymbol{H}}^{-1}(k,t)\tilde{\boldsymbol{b}}(k,t)$且

$$\lim_{t\to\infty}\Delta\boldsymbol{x}(k,t)=\Delta\boldsymbol{x}^*(k). \tag{26}$$

定理2的具体证明可以类比参考文献[17]中定理的证明. 本文后续给出了仿真实验来说明迭代过程(23)在无圈和有圈连通图情况下的效果和区别. 同时定理2确保了对于有圈图, 可以截断有限步的内部迭代过程来完成原始对偶框架下算法的设计.

算法 1 分布式原始对偶算法.

初始化: 选取$\boldsymbol{x}(0)=\mathbf{1}_n\otimes x_1(0), \boldsymbol{q}(0)=\mathbf{0}_{np}, \rho>0, \epsilon>0$.

主要迭代:

1) 对于$k=1,2,\cdots$.

2) 对于任意$i\in\mathcal{E}$: 依据式(23)通过邻居节点信息传递计算$b_i(k)$, 令$h_{i\to j}(0)=h_{ii}, b_{i\to j}(0)=b_i(k), j\in\mathcal{N}_i$.

3) 对于$t=1,2,\cdots,T$.

4) 对于任意i, 根据式(23a)–(23c)计算$\tilde{h}_i(k,t)$, $\tilde{b}_i(k,t)$和$\Delta x_i(k,t)$.

5) 对于任意$j\in\mathcal{N}_i$, 依据式(23d)和式(23e)计算$h_{i\to j}(t)$和$b_{i\to j}(t)$并把它们传递给节点j.

6) 对于i, 更新$x_i(k+1)=x_i(k)+\Delta x_i(k,t)$.

7) 对于i, 更新

$$q_i(k+1)=q_i(k)+\rho(d_ix_i-\sum_{j\in\mathcal{N}_i}x_j).$$

定理 3 考虑上面的原始对偶算法, 在满足假设1–2下, 初值$\boldsymbol{x}(0)=\mathbf{1}_n\otimes x_1(0)$, 存在$0<\xi<1$, 整体原始对偶的迭代满足如下条件:

$$\|\boldsymbol{x}(k+1)-\boldsymbol{x}^*\|\leqslant\xi\|\boldsymbol{x}(k)-\boldsymbol{x}^*\|, \tag{27a}$$

$$\boldsymbol{q}(k+1)=\boldsymbol{q}(k)+\rho L\boldsymbol{x}(k+1), \tag{27b}$$

则算法显然是收敛的, 且收敛速度依赖内层迭代的次数和图的结构.

证 对于算法的迭代, 式(27b)是自然满足的, 式(27a)取决于$\Delta\boldsymbol{x}(k,t)$ 是否是梯度下降方向. 由算法构造,

$$\Delta\boldsymbol{x}(k,t)=\tilde{\boldsymbol{H}}^{-1}(k,t)\tilde{\boldsymbol{b}}(k,t), \tag{28}$$

$$\nabla\mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{q}(k))|_{\boldsymbol{x}=\boldsymbol{x}(k)}=-\boldsymbol{b}(k). \tag{29}$$

上面两向量相乘为常数, 即

$$(\nabla\mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{q}(k))|_{\boldsymbol{x}=\boldsymbol{x}(k)})^{\mathrm{T}}\Delta\boldsymbol{x}(k,t)=$$
$$-\boldsymbol{b}(k)^T\tilde{\boldsymbol{H}}^{-1}(k,t)\tilde{\boldsymbol{b}}(k,t). \tag{30}$$

由推论3可知$\tilde{\boldsymbol{H}}^{-1}(k,t)$为对角元属均为正值的正定对角矩阵. 而$\boldsymbol{b}(k)$变成$\tilde{\boldsymbol{b}}(k,t)$只需一系列行列式为1的行初等变换矩阵. 这样有

$$\boldsymbol{b}^{\mathrm{T}}(k)\tilde{\boldsymbol{H}}^{-1}(k,t)\tilde{\boldsymbol{b}}(k,t)=\tilde{\boldsymbol{b}}^{\mathrm{T}}(k)\tilde{\boldsymbol{H}}^{-1}(k,t)\tilde{\boldsymbol{b}}(k,t)>0, \tag{31}$$

从而

$$(\nabla\mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{q}(k))|_{\boldsymbol{x}=\boldsymbol{x}(k)})^{\mathrm{T}}\Delta\boldsymbol{x}(k,t)<0, \tag{32}$$

这样$\Delta\boldsymbol{x}(k,t)$是梯度下降方向.

由推论1中的鞍点存在且唯一, 原始对偶算法的整体收敛性是显然的. 若图是无圈图且有足够多的内迭代(大于等于图的直径), 算法会有更好的收敛速度.

注 3 定理3定性地说明了分布式原始对偶算法的收敛性, 定理中得$\xi>0$的上确界依赖于内迭代的所能达到的精度, 有待深入研究.

4 仿真实验

因为所提的分布式牛顿算法意义明确且与经典的中心式牛顿法具有一样的收敛效果, 这里就略去对其的仿真实验. 本文中的仿真实验主要针对所提的分布式原始对偶算法. 为方便表述和计算, 这里考虑$p=1$的标量情况.

4.1 对于方程组分布式求解的仿真实验

本文所提的分布式原始对偶算法中方程组的分布式求解是其中的关键步骤, 定理2给出了这一关键内层迭代过程式(23)在无圈和有圈连通图情况下的收敛效果. 本文先对这一定理结果通过简单算例来做仿真验证.

首先, 先给出两个拉普拉斯矩阵$\boldsymbol{L}_1$和$\boldsymbol{L}_2$

$$\boldsymbol{L}_1=\begin{pmatrix}2&-1&-1&0&0\\-1&3&0&-1&-1\\-1&0&1&0&0\\0&-1&0&1&0\\0&-1&0&0&1\end{pmatrix},$$

$$\boldsymbol{L}_2=\begin{pmatrix}2&-1&-1&0&0\\-1&3&0&-1&-1\\-1&0&1&0&0\\0&-1&0&2&-1\\0&-1&0&-1&2\end{pmatrix}.$$

这里: $\boldsymbol{L}_1$表征一个包含5个节点的连通树图, $\boldsymbol{L}_2$表征的有圈连通图与$\boldsymbol{L}_1$的连通树图的唯一区别是在其最后两个叶节点添加了连线.

其次利用上面的连通树图和有圈连通图来构造方程组, 具体如下: 方程组$\boldsymbol{H}_1\boldsymbol{x}=\boldsymbol{b}_1$和$\boldsymbol{H}_2\boldsymbol{x}=\boldsymbol{b}_2$中$\boldsymbol{H}_1$和$\boldsymbol{H}_2$分别由对$\boldsymbol{L}_1$和$\boldsymbol{L}_2$进行对角线元素加倍, 非对角线元素趋绝对值的处理得到. 令$\boldsymbol{b}_1=(4\ \ 6\ \ 2\ \ 2\ \ 2)^{\mathrm{T}}$, $\boldsymbol{b}_2=(4\ \ 6\ \ 2\ \ 4\ \ 4)^{\mathrm{T}}$, 这样构造出的两个方程组的解均为

$$\boldsymbol{x}=(\frac{2}{3}\ \ \frac{2}{3}\ \ \frac{2}{3}\ \ \frac{2}{3}\ \ \frac{2}{3})^{\mathrm{T}}.$$

最后使用定理2给出的迭代过程式(23)来求解上述的两个方程组, 下图是仿真结果.

图1–2的仿真实验显示了本文修改过的信息传递和计算的迭代过程式(23)单独拿出来作为求解方程组方法的迭代方法对于无圈连通图和有圈连通图都是有效和收敛的, 且两者的对比可得有圈图要比无圈图花费更多的迭代才能达到相同的精度.

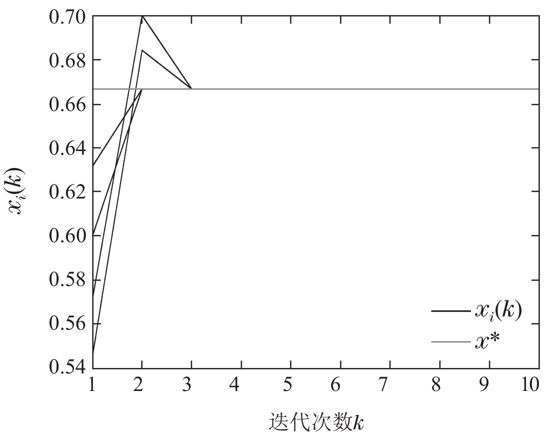

图 1 无圈图时方程组求解迭代过程

Fig. 1 The iterative process of solving equations in acyclic graphs

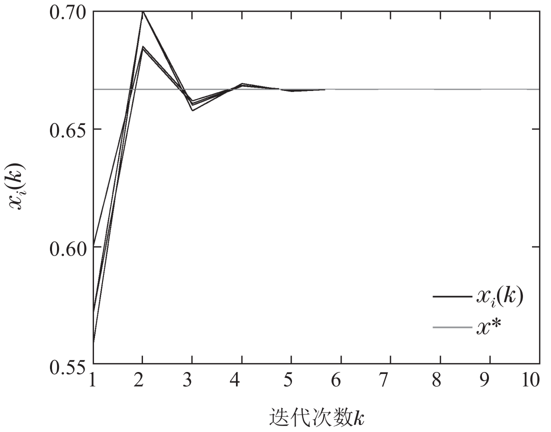

图 2 有圈图时方程组求解迭代过程

Fig. 2 The iterative process of solving equations in cyclic graphs

4.2 对于分布式原始对偶算法的仿真实验

考虑一个二次优化问题的例子验证所提原始对偶(primal dual, PD)算法并对比文献[8]中的利用泰勒展开的近似网络牛顿算法(network Newton, NN).

二次优化函数为

$$F(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-\boldsymbol{h}^{\mathrm{T}}\boldsymbol{x},\ \boldsymbol{A}\succ 0.$$

这里对角正定矩阵$\boldsymbol{A}\in\mathbb{R}^n\times\mathbb{R}^n$, 向量

$$\boldsymbol{h}=(h_1\ \ h_2\ \ \cdots\ \ h_n)^{\mathrm{T}}\in\mathbb{R}^n.$$

$f_i(x_i)=\dfrac{1}{2}a_{ii}x_i^2-h_ix_i$. 网络为含有$n=5$个点的连通有圈图, 其拉普拉斯矩阵为前述的$\boldsymbol{L}_2$, $\boldsymbol{A}$为由向量$(2\ \ 8\ \ 14\ \ 20\ \ 26)^{\mathrm{T}}$生成的对角矩阵. 向量

$$\boldsymbol{h}=(-14\ \ -14\ \ -14\ \ -14\ \ -14)^{\mathrm{T}}.$$

若考虑一致性限制条件, $F(\boldsymbol{x})$的最小值为

$$\boldsymbol{x}^*=(1\ \ 1\ \ 1\ \ 1\ \ 1)^{\mathrm{T}}.$$

在仿真实验中, 初值$\forall i, x_i(0)=10$, 内层迭代选定为$T=2$. 在所提原始对偶算法(简记为PD–2)中, $\boldsymbol{q}(0)$选为零向量, $\rho=10, \epsilon=2$; 在文献[8]中的网络牛顿算法(简记为NN–2)中, $\alpha=10^{-2}, \epsilon=\dfrac{1}{2}$, 基于$\boldsymbol{L}_2$的双随机矩阵$\boldsymbol{W}=\{w_{ij}\}$如下设定: $w_{ij}=\dfrac{1}{6}$, 当

$$j\in\mathcal{N}_i,\ w_{ii}=1-\sum_{j\in\mathcal{N}_i}w_{ij}.$$

为公平对比, 本文对文献[8]中的主要迭代过程额外添加趋同矩阵$\boldsymbol{W}$使之变为

$$x_{i,t+1}=\sum_{j\in\mathcal{N}_i}W_{ij}x_{j,t}+\epsilon d_{i,t}^{(\mathrm{T})},\qquad(33)$$

其中$d_{i,t}^{(\mathrm{T})}$为由泰勒展开截取的牛顿方向. 均方误差的平方根(square root of mean square error, SMSE)定义为

$$\mathrm{SMSE}(k)\doteq\sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{\|x_i(k)-x^*\|^2}{\|x_i(0)-x^*\|^2}}.$$

表 1 固定外层迭代, 内层迭代为2的情况下的计算成本、信息交换次数以及储存成本

Table 1 Fixed outer layer iteration, the calculation cost, the number of information exchanges and the storage cost when the inner layer iteration is 2

算法	PD–2	NN–2
计算成本	$\mathrm{O}(n^2)$	$\mathrm{O}(n^2)$
任意确定节点i的信息交换次数	$5\mathcal{N}_i$	$4\mathcal{N}_i$
任意确定节点i的储存成本	$5\mathcal{N}_i$	$3\mathcal{N}_i$

收敛效果和对比如图3–4所示.

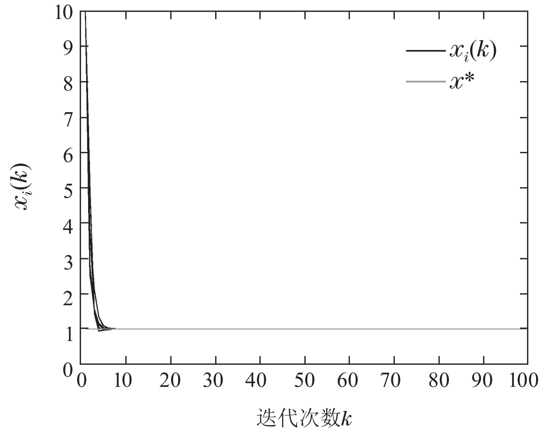

图 3 PD–2中x_i的收敛效果

Fig. 3 The convergence effect of state x_i in PD–2

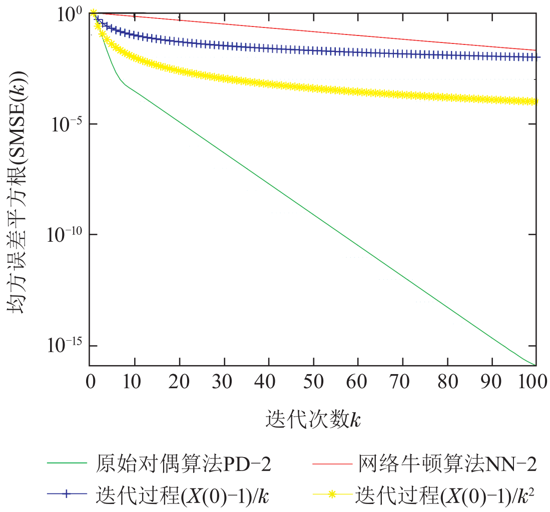

图 4 均方误差平方根的收敛效果

Fig. 4 The convergence effect of the square root of the mean square error (SMSE)

首先, 这说明了作者所提的分布式原始对偶算法(PD–2)在有圈图截断内层迭代($T=2$)的情况下依然

有好的收敛精度, 且收敛速度要快于$\frac{1}{k^2}$. 其次, 所提的分布式原始对偶算法(PD–2)与文献[8]网络牛顿算法(NN–2)的优化成本如表1所示, 在计算量级, 信息交换次数和存储成本相差都不大, 但文献中(NN–2)的线性收敛速度更接近$\frac{1}{k}$, 与本文所提算法(PD–2)有明显的差距. 总之, 本文的算法对于一般的有圈连通图在适当多花费少量的信息交换和储存成本的情况下, 在收敛速度上具有更大的优势.

4.3 分布式原始对偶算法在小世界网络场景的仿真实验

小世界网络(small world network, SWN)是一种常见的随机图, 现实中大量的现象如网络中的导航菜单、电网、代谢处理网络、脑神经网络、选民网络, 社交影响网络等都被证明是小世界网络. 这种网络模型两个显著特点是: 1) 平均最短路径长度较小; 2) 节点聚集效果比较显著. 这样的图边比较多但图的直径比较小. 这一小节作者把所提出的分布式原始对偶算法应用在小世界网络中, 来仿真其收敛效果.

具体的网络图如图5所示, 是一个具有$N = 200$个节点, 400条边的连通图(有圈). 每个节点的目标函数为

$$f_i(x) = \frac{1}{2}q_i x^2 + p_i x,\ \nabla f_i(x) = q_i > 0,$$

这其中q_i和p_i分别在$[0.01, 800]$和$[-1, 1]$中随机选取. $\rho = 1, \epsilon = 1$, 内层迭代选取为2. 具体的仿真效果如图6所示.

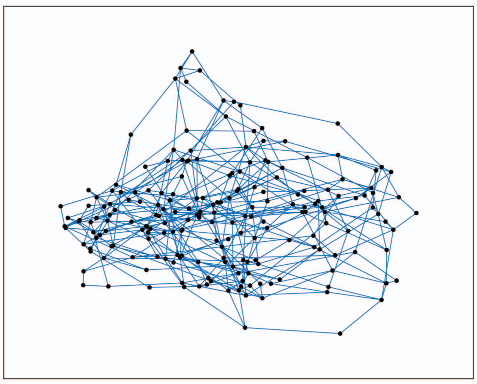

图 5 小世界网络

Fig. 5 Small world network

图6进一步说明了本文的分布式原始对偶算法的收敛性, 且与第4.2小节中对比网络规模, 本文算法的收敛速度和精度并没随图的规模扩大而有大的变化. 这说明了本文所提的这个算法对小世界网络有好的收敛效果并对其规模有较好的适应性.

5 结论与展望

本文为分布式优化问题提供了一种有别于常见的直接内嵌使用常规趋同矩阵的算法设计方法. 作者的设计方法在无圈连通图时可以做到精确趋同的特性使所提出的分布式牛顿算法做到集中式算法的收敛效果, 对于一般性的有圈连通图时, 拉格朗日原始对偶框架下的分布式算法在截断内层迭代的情况下有较好的收敛性. 在后续研究作者希望致力于分析出内层迭代的次数和收敛速度之间的定量关系来进一步说明这样的算法设计方法的灵活性和鲁棒性.

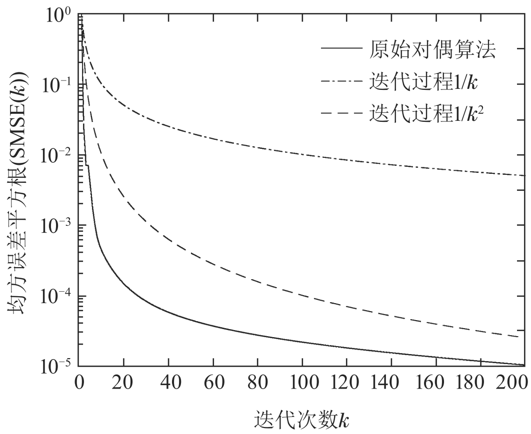

图 6 均方误差的平方根(SMSE)的收敛效果

Fig. 6 The convergence effect of the square root of the mean square error (SMSE)

参考文献:

[1] JOHANSSON B. *On Distributed Optimization in Networked Systems*. Stockholm: Royal Institute of Technology, 2008.

[2] ZHU S Y, CHEN C L, LI W S, et al. Distributed optimal consensus filter for target tracking in heterogeneous sensor networks. *IEEE Transactions on Cybernetics*, 2013, 43(6): 1963 – 1976.

[3] PREDD J B, KULKARNI S R, POOR H, et al. A collaborative training algorithm for distributed learning. *IEEE Transactions on Information Theory*, 2009, 55(4): 1856 – 1871.

[4] HAN T, CHI M, GUAN Z H, et al. Distributed three dimensional formation containment control of multiple unmanned aerial vehicle systems. *Asian Journal of Control*, 2017, 19(3): 1103 – 1113.

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

[6] QU G N, LI N. Accelerated distributed nesterov gradient descent for convex and smooth functions. *2017 IEEE 56th Annual Conference on Decision and Controlv (CDC)*. Melbourne, VIC, Australia: IEEE, 2017: 2260 – 2267.

[7] XIN R, KHAN U A. Distributed heavy-ball: A generalization and acceleration of first-order methods with gradient tracking. *arXiv: 1808.02942*, 2018.

[8] MOKHTARI A, LING Q, RIBEIRO A. Network newton distributed optimization methods. *IEEE Transactions on Signal Processing*, 2017, 65(1): 146 – 161.

[9] VARAGNOLO D, ZANELLA F, CENEDESE A, et al. Newton-raphson consensus for distributed convex optimization. *IEEE Transactions on Automatic Control*, 2016, 61(4): 994 – 1009.

[10] JAKOVETIC D, MOURA J M, XAVIER J. Linear convergence rate of a class of distributed augmented lagrangian algorithms. *IEEE Transactions on Automatic Control*, 2015, 60(4): 922 – 936.

[11] WEI S, LING Q, YUAN K, et al. On the linear convergence of the ADMM in decentralized consensus optimization. *IEEE Transactions on Signal Processing*, 2014, 62(7): 1750 – 1761.

[12] CHANG T H, HONG M Y, WANG X F. Multi-agent distributed optimization via inexact consensus ADMM. *IEEE Transactions on Signal Processing*, 2014, 63(2): 482 – 497.

[13] CHANG T H. A proximal dual consensus ADMM method for multi-agent constrained optimization. *IEEE Transactions on Signal Processing*, 2016, 64(14): 3719 – 3734.

[14] WEISS Y, FREEMAN W T. Correctness of belief propagation in gaussian graphical models of arbitrary topology. *Neural Information Processing Systems*, 1999, 13(10): 673 – 679.

[15] MALIOUTOV D M, JOHNSON J K, WILLSKY A S. Walk-sums and belief propagation in gaussian graphical models. *Journal of Machine Learning Research*, 2006, 7: 2031 – 2064.

[16] MA W L, FU M Y, CUI P, et al. Finite-time average consensus based approach for distributed convex optimization. *Asian Journal of Control*, 2020, 22: 323 – 333.

[17] MA W L, ZHANG H S, FU M Y. Distributed convex optimization based on ADMM and belief propagation methods. *Asian Journal of Control*, 2021, 23: 1040 – 1051.

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

作者简介:

马文龙 博士研究生, 目前研究方向为多自主体网络趋同控制和分布式优化, E-mail: sqsyxiaoma@163.com.cn;

付敏跃 教授, IEEE院士, 目前研究方向为控制系统、信号处理和通信, E-mail: mingyue.fu@newcastle.edu.au;

张焕水 教授, 博士生导师, 目前研究方向为最优估计和控制、时滞系统、随机系统、信号处理和无线传感器联网系统, E-mail: hszhang@sdu.edu.cn.