

求解无约束一致性优化问题的分布式拟牛顿算法

于慧慧¹, 王永丽¹, 陈勇勇¹, 周秀娟²

(1. 山东科技大学 数学与系统科学学院, 山东 青岛 266590; 2. 山东科技大学 信息科学与工程学院, 山东 青岛 266590)

摘要: 本文主要针对网络中各个节点相互协作, 最大限度地使本地费用函数的总和最小的无约束一致性优化问题, 提出了一类分布式拟牛顿算法。算法仅利用了目标函数的一阶导数信息, 每步通过选取一个满足拟牛顿方程的正定对角矩阵来作为费用函数 Hesse 矩阵逆的校正矩阵, 克服了校正矩阵的非稀疏性对算法分布式实现造成的困难, 减少了计算量和存储空间。在适当条件下, 证明了分布式拟牛顿算法的全局收敛性及局部线性收敛速度, 并通过数值实验验证了算法的优越性。

关键词: 无约束; 一致性优化; 分布式拟牛顿算法; 全局收敛; 线性收敛

中图分类号: O224 文献标志码: A 文章编号: 1672-3767(2016)03-0112-07

Distributed Quasi-Newton Algorithm for Solving Unconstrained Consensus Optimization

YU Huihui¹, WANG Yongli¹, CHEN Yongyong¹, ZHOU Xiujuan²

(1. College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao Shandong 266590, China; 2. College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590, China)

Abstract: A class of distributed quasi-Newton algorithm was proposed for solving the unconstrained consensus optimization, where the network nodes work collaboratively to minimize the sum of their locally known convex costs. The new algorithm used only the first-order derivative information of the cost function. By choosing a positive definite diagonal matrix as the quasi-Newton correction of the inverse of the Hesse matrix of the cost function at each iteration, the proposed algorithm successfully overcame the difficulty caused by the non-sparsity of the correction matrix in the implementation of distribution, and greatly decreased computation and storage pressure. Under suitable conditions, the proposed distributed quasi-Newton algorithm was proved to be globally convergent with local linear convergence rate and its superiority was verified by numerical simulations.

Key words: unconstrained consensus optimization; distributed quasi-Newton algorithm; global convergence; linear convergence

考虑有 n 个节点的连通网络, 设每个节点都有一个本地费用函数 $f_i: \mathbf{R}^p \rightarrow \mathbf{R}$, 而变量 $\mathbf{y} \in \mathbf{R}^p$ 为全局变量。我们的目标是使所有节点的费用函数之和 $f: \mathbf{R}^p \rightarrow \mathbf{R}, f(\mathbf{y}) := \sum_{i=1}^n f_i(\mathbf{y})$ 最小, 即每个节点共同协作解决如下全局优化问题:

$$\mathbf{y}^* := \arg \min_{\mathbf{y}} f(\mathbf{y}) = \arg \min_{\mathbf{y}} \sum_{i=1}^n f_i(\mathbf{y}). \quad (1)$$

收稿日期: 2016-01-05

作者简介: 于慧慧(1991—), 女, 山东菏泽人, 硕士研究生, 主要从事分布式优化算法的研究. E-mail: yxh622@163.com

王永丽(1977—), 女, 山东栖霞人, 博士, 副教授, 主要从事非线性优化理论与算法、分布式优化算法、并行计算的研究, 本文通信作者. E-mail: wangyongli@sdkd.net.cn

其中 \mathbf{y}^* 表示问题(1)的最优解。在现实生活中,这种形式的问题有很多,例如,大数据分析^[1],分布式传感器网络^[2-5]和分布式控制^[6]等。

求解问题(1)的分布式方法分为两类,一类是基于费用函数一阶信息的方法,该类方法中研究较多的是分布式梯度法^[7-10],分布式交替方向乘子法^[11-13]和分布式对偶平均法^[14-15]。虽然这些算法之间存在着很大的差异,但基本思想都可以归结为先进行本地下降步,然后与相邻节点进行变量交换和信息整合。由于该类计算仅涉及到一阶信息,导致了对下降方向曲率估计的不准确性,使得收敛速度较慢。另一类分布式方法是基于费用函数二阶信息的方法,即分布式牛顿类算法,主要的文献有^[16-17]等。目前,对分布式牛顿类算法的研究较少,这主要是因为牛顿法的分布式实现比较困难且计算目标函数的二阶信息计算量较大。但考虑到牛顿类算法比大多数一阶方法具有更快的收敛速度,越来越多的学者开始了对该类算法的研究。

本文研究分布式求解问题(1)的拟牛顿方法,只需计算目标函数的一阶信息,通过选取一个满足拟牛顿方程的正定对角矩阵来作为费用函数 Hesse 矩阵逆的校正矩阵,来克服校正矩阵的非稀疏性对算法分布式实现造成的困难,从而实现拟牛顿法的分布式计算,减少计算量和存储空间。同时,在适当条件下证明分布式拟牛顿算法的收敛性,并通过数值试验验证算法的有效性。应当指出的是,这里的分布式是指每个节点仅知道其本地费用函数,且只能够与相邻节点进行信息交换。

1 预备知识

首先给出关于问题(1)的一些预备知识。

假设给定网络是一个无向网络 $G = (V, E)$, 其中, V 是节点集, E 是边集, 即所有节点对 $\{i, j\}$ 可以通过一个交流链路进行信息交换。对网络中每个节点的费用函数 f_i , 作如下假设。

假设 1 假设函数 $f_i: \mathbf{R}^p \rightarrow \mathbf{R}$ 是系数为 $\mu > 0$ 的强凸函数, 即

$$f_i(\mathbf{z}) \geq f_i(\mathbf{y}) + \nabla f_i(\mathbf{y})^T(\mathbf{z} - \mathbf{y}) + \frac{\mu}{2} \|\mathbf{z} - \mathbf{y}\|^2, \quad \forall \mathbf{y}, \mathbf{z} \in \mathbf{R}^p. \tag{2}$$

并且具有常数为 L 的 Lipschitz 连续梯度, 即

$$\|\nabla f_i(\mathbf{y}) - \nabla f_i(\mathbf{z})\| \leq L \|\mathbf{y} - \mathbf{z}\|. \tag{3}$$

假设 2 网络 $G = (V, E)$ 是无向简单连通图, 即不存在自循环和多重链路。

令 O_i 表示与节点 i 相邻的节点集合, $\bar{O}_i = O_i \cup \{i\}$ 。将一个对称的双随机矩阵 \mathbf{W} 与该网络相关联, 其中 \mathbf{W} 的元素都是非负的, 并且行(列)相加等于 1。更详细的假设如下:

假设 3 矩阵 $\mathbf{W} = \mathbf{W}^T \in \mathbf{R}^{n \times n}$ 是元素为 w_{ij} 的随机矩阵, 其中 w_{ij} 的定义如下:

$$w_{ij} = \begin{cases} > 0 & \{i, j\} \in E \\ = 0 & \{i, j\} \notin E, i \neq j \end{cases}, \quad \text{并且 } w_{ii} = 1 - \sum_{j \in O_i} w_{ij}.$$

且存在常数 w_{\min}, w_{\max} 使得

$$0 < w_{\min} \leq w_{ii} \leq w_{\max} < 1.$$

定义 1^[16] 对于矩阵 $\mathbf{M} = (\mathbf{M}_{ij}) \in \mathbf{R}^{np \times np}$, 其中 $\mathbf{M}_{ij} \in \mathbf{R}^{p \times p}$, 定义 \mathbf{M} 的范数如下:

$$\|\mathbf{M}\| := \max_{j=1, \dots, n} \sum_{i=1}^n \|\mathbf{M}_{ij}\|_2$$

类似的, 对向量 $\mathbf{x} \in \mathbf{R}^{np}, \mathbf{x}_i \in \mathbf{R}^p$, 定义:

$$\|\mathbf{x}\| := \sum_{i=1}^n \|\mathbf{x}_i\|_2$$

其中 $\|\cdot\|_2$ 为欧式范数。

2 分布式拟牛顿法

牛顿法的基本思想是利用目标函数的二次 Taylor 展开来逼近目标函数, 但在实际问题中, 尤其是大规

模的高维问题中,如果所基于的网络是稀疏的,则目标函数的二阶导也可能是稀疏的,从而使得所求解的问题具有特殊结构。但遗憾的是,即使目标函数的 Hesse 矩阵是稀疏的,但其逆却是稠密的。为了减少计算量,避免 Hesse 矩阵逆的计算,我们利用目标函数及其一阶导数信息来近似 Hesse 矩阵的逆矩阵,通过构造特殊的拟牛顿矩阵,实现拟牛顿矩阵的分布式计算。

2.1 问题的等价变形

引入辅助函数 $\Phi: \mathbf{R}^{np} \rightarrow \mathbf{R}$, 令 $\mathbf{x} = (x_1, \dots, x_n) \in \mathbf{R}^{np}$, 并记 $\mathbf{Z} \in \mathbf{R}^{np \times np}$ 为 \mathbf{W} 和单位矩阵 $\mathbf{I} \in \mathbf{R}^{p \times p}$ 的 Kronecker 积: $\mathbf{Z} = \mathbf{W} \otimes \mathbf{I}$, 则由文献[18]可知问题(1)等价于如下问题:

$$\min_{\mathbf{x} \in \mathbf{R}^{np}} \Phi(\mathbf{x}) = \alpha \sum_{i=1}^n f_i(\mathbf{x}_i) + \frac{1}{2} \mathbf{x}^T (\mathbf{I}_{np \times np} - \mathbf{Z}) \mathbf{x}. \quad (4)$$

其中 $\alpha > 0$ 为罚参数, $\mathbf{I}_{np \times np}$ 表示 $np \times np$ 的单位矩阵。

记 $F(\mathbf{x}) = \sum_{i=1}^n f_i(\mathbf{x}_i)$, 则有 $\nabla F(\mathbf{x}) = (\nabla f_1(\mathbf{x}_1), \dots, \nabla f_n(\mathbf{x}_n))$ 。于是,问题(4)目标函数的梯度为:

$$\nabla \Phi(\mathbf{x}) = \alpha \nabla F(\mathbf{x}) + (\mathbf{I}_{np \times np} - \mathbf{Z}) \mathbf{x}. \quad (5)$$

2.2 算法框架

求解问题(4)的拟牛顿法的基本迭代格式为: $\mathbf{x}^{k+1} = \mathbf{x}^k + \epsilon \mathbf{d}^k$ 。

其中 ϵ 为搜索步长,搜索方向为 $\mathbf{d}^k = -\mathbf{H}^k \nabla \Phi(\mathbf{x}^k)$ 。

\mathbf{H}^k 是目标函数 Hesse 矩阵逆矩阵的近似,满足正定性和如下拟牛顿条件:

$$\mathbf{H}^k \mathbf{y}^k = \mathbf{s}^k. \quad (8)$$

其中 $\mathbf{y}^k := \nabla \Phi(\mathbf{x}^k) - \nabla \Phi(\mathbf{x}^{k-1}) \in \mathbf{R}^{np}$, $\mathbf{s}^k := \mathbf{x}^k - \mathbf{x}^{k-1} \in \mathbf{R}^{np}$ 。由(8)式可以看出,若 \mathbf{H}^k 无任何特殊结构,则问题(4)的分布式求解就会非常困难。为了能够方便的对问题(4)进行分布式求解,我们借鉴文献[19]中对角稀疏拟牛顿算法的思想。

令 k 表示迭代次数, $k = 0, 1, \dots$ 。 $\mathbf{x} = (x_1, \dots, x_n) \in \mathbf{R}^{np}$ 表示第 k 次迭代得到的对最优值 \mathbf{x}^* 的近似。设 $\mathbf{H}^k \in \mathbf{R}^{np \times np}$ 为块对角正定矩阵,满足拟牛顿条件(8)。记其每个对角块为 $\mathbf{h}_i^k \in \mathbf{R}^{p \times p}$, 且为正定矩阵。为了简化计算,本文中取每个对角块 $\mathbf{h}_i^k \in \mathbf{R}^{p \times p}$ 为对角阵,且要求每个 \mathbf{h}_i^k 都近似满足拟牛顿条件,即:

$$\min \| \mathbf{h}_i^k \mathbf{y}_i^k - \mathbf{s}_i^k \|^2, \quad i = 1, 2, \dots, n \quad (9)$$

其中, $\mathbf{y}_i, \mathbf{s}_i$ 分别是与变量 \mathbf{x}_i 相对应的向量 \mathbf{y} 和向量 \mathbf{s} 的子向量, $\mathbf{s}_i^k = \mathbf{x}_i^k - \mathbf{x}_i^{k-1} \in \mathbf{R}^p$,

$$\begin{aligned} \mathbf{y}_i^k &= \nabla \Phi(\mathbf{x}_i^k) - \nabla \Phi(\mathbf{x}_i^{k-1}) = \\ &\alpha (\nabla f_i(\mathbf{x}_i^k) - \nabla f_i(\mathbf{x}_i^{k-1})) + \sum_{j \in O_i} \omega_{ij} ((\mathbf{x}_i^k - \mathbf{x}_j^{k-1}) - (\mathbf{x}_i^{k-1} - \mathbf{x}_j^{k-1})) \in \mathbf{R}^p. \end{aligned}$$

为了保证校正矩阵 \mathbf{H}^k 的正定性及方向 $\mathbf{d}^k = -\mathbf{H}^k \nabla \Phi(\mathbf{x}^k)$ 的下降性,限制 $\mathbf{h}_i^k (i = 1, 2, \dots, n)$ 中对角元素的取值,要求 $h_{ij}^k \in [m, M], j = 1, 2, \dots, p$, 其中 $0 < m \leq M$ 为常数。此时(9)式可以写为:

$$\min \| \mathbf{h}_i^k \mathbf{y}_i^k - \mathbf{s}_i^k \|^2 = \min_{m \leq h_{ij}^k \leq M} \sum_{j=1}^p (h_{ij}^k y_{ij}^k - s_j^k)^2$$

其中 h_{ij} 表示对角矩阵 \mathbf{h}_i 的第 j 个对角元, y_{ij} 表示子向量 \mathbf{y}_i 中第 j 个分量, s_{ij} 表示子向量 \mathbf{s}_i 的第 j 个分量。

由上式不难看出,每步迭代,当 $y_{ij}^k = 0$ 时,可取 $h_{ij}^k = \frac{m+M}{2}$ 。

而当 $y_{ij}^k \neq 0$ 时, \mathbf{h}_i^k 的对角元可按式计算:

$$h_{ij}^k = \begin{cases} \frac{s_j^k}{y_{ij}^k}, & m \leq \frac{s_j^k}{y_{ij}^k} \leq M; \\ M, & M < \frac{s_j^k}{y_{ij}^k}; \\ m, & \frac{s_j^k}{y_{ij}^k} < m. \end{cases} \quad (11)$$

于是,由(7)式可以写出在每个节点 i 处的下降方向:

$$\mathbf{d}_i^k = -\mathbf{h}_i^k \left[\alpha \nabla f_i(\mathbf{x}_i^k) + \sum_{j \in O_i} \omega_{ij} (\mathbf{x}_i^k - \mathbf{x}_j^k) \right] \quad (12)$$

从而每个节点处变量 \mathbf{x}_i 的迭代公式为 $\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \epsilon \mathbf{d}_i^k$ (13)

由上述讨论可得分布式拟牛顿算法(Distributed-Diagonal Quasi-Newton, D-DQN)的框架如下:

分布式拟牛顿算法

对每个节点 i , 给定 $\alpha, \epsilon > 0$ 。

步 1. 初始化: 对每个节点 i , 令 $k = 0$, 给定初始点 $\mathbf{x}_i^0, \mathbf{x}_i^1 \in \mathbf{R}^p$ 。

步 2. 对每个节点 i , 将其迭代点 \mathbf{x}_i^k 传递给其相邻节点 $j \in O_i$, 并从相邻节点接收 \mathbf{x}_j^k 。

步 3. 对每个节点 i , 计算搜索方向:

$$\mathbf{d}_i^k = -\mathbf{h}_i^k \left[\alpha \nabla f_i(\mathbf{x}_i^k) + \sum_{j \in O_i} \omega_{ij} (\mathbf{x}_i^k - \mathbf{x}_j^k) \right]$$

步 4. 对每个节点 i , 更新其近似解:

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \epsilon \mathbf{d}_i^k$$

步 5. 由(10)式、(11)式计算 \mathbf{h}_i^{k+1} 的元素, 令 $k = k + 1$, 返回步 2。

注: 注意到拟牛顿方法每步迭代都涉及到当前节点及其相邻节点前两步的变量信息, 因此在给定初始条件时, 每个节点需要给出两个初始值。在算法的第 2 步, 每个节点需要与其相邻节点进行信息交换。第 3 步中计算搜索方向时涉及到了参数 α , 这一参数的取值在文献[18]中已有研究, 本文在数值实验时取 α 为固定值。算法第 4 步, 本文采用一个固定步长 ϵ , 显然变量的迭代只涉及到自身的函数信息和相邻节点的信息, 实现了问题(1)的分布式计算。

3 收敛性分析

引理 1 若假设 1~3 成立, 则有 $\|\mathbf{d}^k\| \leq M \|\nabla \Phi(\mathbf{x}^k)\|$ 。

证明 因为 \mathbf{H}_k 为块对角矩阵, 由本文中范数的定义, 可知:

$$\|\mathbf{H}^k\| = \max_{j=1, \dots, n} \|\mathbf{H}_{jj}^k\|_2 = \max_{j=1, \dots, n} (\lambda(\mathbf{H}_{jj}^k)^T \mathbf{H}_{jj}^k)^{1/2} \leq M。$$

因此, 由 \mathbf{d}^k 的定义可得

$$\|\mathbf{d}^k\| \leq \|\mathbf{H}^k\| \cdot \|\nabla \Phi(\mathbf{x}^k)\| \leq M \|\nabla \Phi(\mathbf{x}^k)\| \quad (14)$$

下面我们将证明, 存在步长 ϵ 使得算法 D-DQN 产生的序列收敛到问题(4)的解。由假设 1 可知, $\nabla \Phi(\mathbf{x})$ Lipchitz 连续, 且其 Lipchitz 常数为 $\tilde{L} = \alpha L + 2(1 - \omega_{\min})$, 即

$$\|\nabla \Phi(\mathbf{x}) - \nabla \Phi(\mathbf{y})\| \leq \tilde{L} \|\mathbf{x} - \mathbf{y}\|, \mathbf{x}, \mathbf{y} \in \mathbf{R}^{np} \quad (15)$$

且 $\frac{\tilde{\mu}}{2} \|\mathbf{x} - \mathbf{x}^*\| \leq \Phi(\mathbf{x}) - \Phi(\mathbf{x}^*) \leq \frac{1}{\tilde{\mu}} \|\nabla \Phi(\mathbf{x})\|^2$ (16)

其中 $\tilde{\mu} = \alpha \mu$, $\mathbf{x} \in \mathbf{R}^{np}$ 。

定理 1 若假设 1~3 成立, 并取步长 $\epsilon = \frac{m}{M^2 \tilde{L}}$, 则算法 D-DQN 产生的序列 $\{\mathbf{x}^k\}$ 收敛到问题(4)的极小值点 \mathbf{x}^* , 即 $\lim_{k \rightarrow \infty} \mathbf{x}^k = \mathbf{x}^*$, 且至少具有线性收敛速度, 即:

$$\Phi(\mathbf{x}^{k+1}) - \Phi(\mathbf{x}^*) \leq \left(1 - \frac{m^2 \tilde{\mu}}{2LM^2}\right) (\Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*)), k = 0, 1, \dots$$

证明 由算法的迭代过程与题设可得

$$\begin{aligned} \Phi(\mathbf{x}^{k+1}) - \Phi(\mathbf{x}^*) &= \Phi(\mathbf{x}^k + \epsilon \mathbf{d}^k) - \Phi(\mathbf{x}^*) = \\ &= \Phi(\mathbf{x}^k) + \int_0^1 \nabla^T \Phi(\mathbf{x}^k + t\epsilon \mathbf{d}^k) \epsilon \mathbf{d}^k dt - \Phi(\mathbf{x}^*) + \epsilon \nabla^T \Phi(\mathbf{x}^k) \mathbf{d}^k - \epsilon \nabla^T \Phi(\mathbf{x}^k) \mathbf{d}^k \leq \end{aligned}$$

$$\begin{aligned} \Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*) + \epsilon \int_0^1 \|\nabla^T \Phi(\mathbf{x}^k + t\epsilon \mathbf{d}^k) - \nabla^T \Phi(\mathbf{x}^k)\| \cdot \|\mathbf{d}^k\| dt + \epsilon \nabla^T \Phi(\mathbf{x}^k) \mathbf{d}^k &\leq \\ \Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*) + \epsilon \int_0^1 \tilde{L} t \epsilon \|\mathbf{d}^k\|^2 dt + \epsilon \nabla^T \Phi(\mathbf{x}^k) \mathbf{d}^k &= \\ \Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*) + \frac{\tilde{L}}{2} \epsilon^2 \|\mathbf{d}^k\|^2 + \epsilon \nabla^T \Phi(\mathbf{x}^k) \mathbf{d}^k & \end{aligned}$$

由 \mathbf{d}^k 的定义, $\nabla^T \Phi(\mathbf{x}^k) \mathbf{d}^k = -\nabla^T \Phi(\mathbf{x}^k) \mathbf{H}_k \nabla \Phi(\mathbf{x}^k) \leq -m \|\nabla \Phi(\mathbf{x}^k)\|^2$

于是,由引理 1 可得

$$\begin{aligned} \Phi(\mathbf{x}^{k+1}) - \Phi(\mathbf{x}^*) &\leq \Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*) + \frac{\epsilon^2}{2} \tilde{L} M^2 \|\nabla \Phi(\mathbf{x}^k)\|^2 - \epsilon m \|\nabla \Phi(\mathbf{x}^k)\|^2 = \\ \Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*) + \left(\frac{\epsilon^2}{2} \tilde{L} M^2 - \epsilon m\right) \|\nabla \Phi(\mathbf{x}^k)\|^2 & \end{aligned} \quad (17)$$

令 $q(\epsilon) = \frac{\epsilon^2 \tilde{L} M^2}{2} - \epsilon m$, 易知 $q(0) = 0, q'(\epsilon) = \tilde{L} \epsilon M^2 - m$, 且 $q''(\epsilon) > 0$, 因此函数 $q(\epsilon)$ 的最小值点为:

$\epsilon^* = m/\tilde{L}M^2$, 其最优值为

$$q(\epsilon^*) = -\frac{m^2}{2M^2\tilde{L}}. \quad (18)$$

由(17)和(18)可得: $\Phi(\mathbf{x}^{k+1}) - \Phi(\mathbf{x}^*) \leq \Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*) - \frac{m^2}{2M^2\tilde{L}} \|\nabla \Phi(\mathbf{x}^k)\|^2$

且 $\Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*) \leq \frac{1}{\tilde{\mu}} \|\nabla \Phi(\mathbf{x}^k)\|^2$, 于是可得

$$-\frac{m^2}{2M^2\tilde{L}} \|\nabla \Phi(\mathbf{x}^k)\|^2 \leq -(\Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*)) \frac{m^2 \tilde{\mu}}{2M^2\tilde{L}}$$

故 $\Phi(\mathbf{x}^{k+1}) - \Phi(\mathbf{x}^*) \leq \left(1 - \frac{m^2 \tilde{\mu}}{2M^2\tilde{L}}\right) (\Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*))$. 由 $\tilde{\mu} = \alpha\mu \leq \alpha L < \tilde{L}$, 可知 $\tilde{\mu}/\tilde{L} < 1$, 且

$m \leq M$, 于是有 $\xi := 1 - \frac{m^2 \tilde{\mu}}{2M^2\tilde{L}} \in (0, 1)$, 从而

$$\Phi(\mathbf{x}^{k+1}) - \Phi(\mathbf{x}^*) \leq \xi (\Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*)).$$

且有 $\lim_{k \rightarrow \infty} \Phi(\mathbf{x}^k) = \Phi(\mathbf{x}^*)$.

因为 $\Phi \in C(\mathbf{R}^n)$, 结合上述结果可知序列 $\{\mathbf{x}^k\}$ 收敛到 \mathbf{x}^* , 即: $\lim_{k \rightarrow \infty} \mathbf{x}^k = \mathbf{x}^*$, 证毕。

定理 1 的证明表明对任意步长 $\epsilon \in \left(0, \frac{m}{M^2\tilde{L}}\right]$, 算法 D-DQN 全局收敛。且当步长取较大值时, 算法收敛

较快。

4 数值实验

本节通过数值实验来验证 D-DQN 算法求解问题(1)的有效性, 并将本文提出的算法与分布式梯度法(DGN)在迭代次数和运行时间两方面进行比较。实验运行环境为 Pentium(R)E5500 2.77GHz 双核处理器, 内存 2.00 GB。所有算法程序均用 Matlab2015a 编写。

实验生成节点个数为 n 的连通网络, 考虑如下形式一致性问题

$$\min_{\mathbf{x}} f(\mathbf{x}) := \sum_{i=1}^n f_i(\mathbf{x}) = \sum_{i=1}^n \frac{1}{2} (\mathbf{x} - \mathbf{a}_i)^T \mathbf{B}_i (\mathbf{x} - \mathbf{a}_i)$$

其中决策变量 $\mathbf{x} \in \mathbf{R}^p$, $\mathbf{B}_i \in \mathbf{R}^{p \times p}$ 为正定对称矩阵。矩阵 \mathbf{B}_i 的生成规则如下: 首先生成矩阵 $\hat{\mathbf{B}}_i, i = 1, 2, \dots, n$, 其值服从相对独立的标准正态分布; 然后提取矩阵 $\frac{1}{2} (\hat{\mathbf{B}} + \hat{\mathbf{B}}^T)$ 的特征值矩阵 $\hat{\mathbf{Q}} \in \mathbf{R}^{p \times p}$; 最后令

$B_i = \hat{Q} \text{diag}(\hat{c}_i) \hat{Q}^T$, 其中 $\hat{c}_i \in \mathbf{R}^p$ 的分量随机独立产生于 $[1, 101]$ 。对 $i = 1, 2, \dots, n$, 向量 $a_i \in \mathbf{R}^p$ 的分量取值范围为 $[1, 11]$ 。实验中权重矩阵的生成采用 Metropolis 权重规则, 罚参数取固定值: $\alpha = 0.0001, M = 10, m = 8$, 步长 $\epsilon = \frac{m}{M^2 L} \approx 0.21$ 。

采用终止条件: $e = \frac{1}{n} \sum_{i=1}^n \frac{\|x_i^k - x^*\|^2}{\|x^*\|^2}, x^* \neq 0$ 。当误差 $e < 10^{-5}$ 时, 算法终止。

当全局变量维数 $p = 3$, 图 1、图 2 分别取不同网络节点数时, 分布式梯度下降法(DGN)和分布式拟牛顿算法(D-DQN)迭代次数与误差的关系曲线。可以看出, 分布式拟牛顿算法的收敛速度要快于分布式梯度下降方法, 且在刚开始迭代时表现出非常快的下降速度, 从而体现出分布式拟牛顿法在求解大规模稀疏一致性优化问题中的优越性。

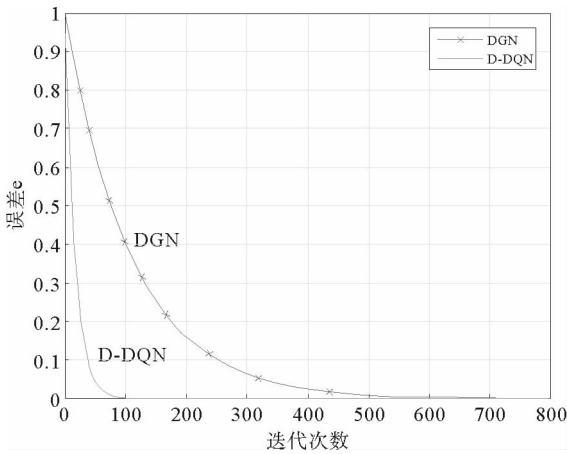


图 1 当 $n=100$ 时, D-DQN 算法与 DGN 算法的比较
Fig. 1 The comparison between the D-DQN and DGN for $n=100$

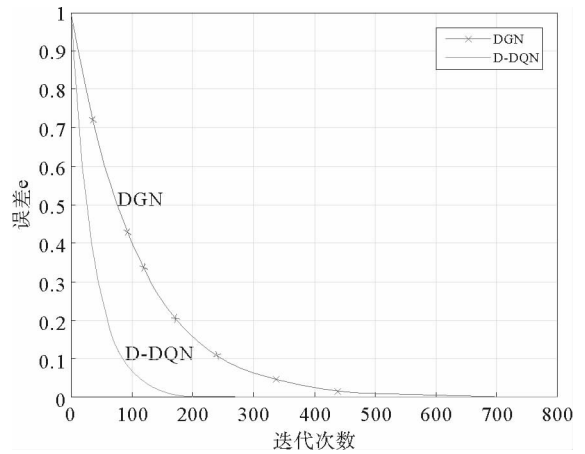


图 2 当 $n=200$ 时, D-DQN 算法与 DGN 算法的比较
Fig. 2 The comparison between the D-DQN and DGN for $n=200$

5 结论

本文针对无约束一致性优化问题(1), 提出了一类新的分布式拟牛顿算法, 算法通过选取一个近似满足拟牛顿方程的正定对角矩阵来近似费用函数 Hesse 矩阵的逆, 实现了拟牛顿算法的分布式实现。在取固定步长的情况下, 证明了算法的全局收敛性及局部线性收敛速度, 并通过数值实验验证了算法的有效性。实验结果表明, 分布式拟牛顿算法在求解大规模稀疏一致性优化问题方面具有较好的应用前景。

参考文献:

- [1] DANESHMAND A, FACCHINEI F, KUNGURTSEV V, et al. Hybrid random/deterministic parallel algorithms for non-convex big data optimization[J]. Journal of Inverstigative Medicine, 2014, 60(4): 671-675.
- [2] SCHIZAS I D, RIBEIRO A, GIANNAKIS G B. Consensus in ad hoc WSNs with noisy links—Part I: Distributed estimation of deterministic signals[J]. IEEE Transactions on Signal Processing, 2008, 56(1): 350-364.
- [3] KAR S, MOURA J M F, RAMANAN K. Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication[J]. IEEE Transactions on Information Theory, 2012, 58(6): 3575-3605.
- [4] SAYED A H, LOPES C G. Adaptive Processing over distributed networks[J]. Leice Transactions on Fundamentals of Electronics Communications and Computer Sciences, 2007, E90-A(8): 1504-1510.
- [5] CATTIVELLI F S, SAYED A H. Diffusion LMS strategies for distributed estimation[J]. IEEE Transactions on Signal Processing, 2010, 58(3): 1035-1048.

- [6]MOTA J F C, XAVIER J M F, AGUIAR P M Q, et al. Distributed optimization with local domains: Applications in MPC and network flows[J]. IEEE Transactions on Automatic Control, 2015, 60(7): 2004-2009.
- [7]NEDIĆ A, OZDAGLAR A. Distributed subgradient methods for multi-agent optimization[J]. IEEE Transactions on Automatic Control, 2009, 54(1): 48-61.
- [8]JAKOVETIC D, XAVIER J, MOURA J M F. Fast distributed gradient methods[J]. IEEE Transactions on Automatic Control, 2014, 59(5): 1131-1146.
- [9]YUAN K, LING Q, YIN W. On the convergence of decentralized gradient descent[DB/OL]. (2015-07-01)[2016-01-05]. <http://arXiv preprint arXiv:1310.7063>.
- [10]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.
- [11]LING Q, RIBEIRO A. Decentralized linearized alternating direction method of multipliers[C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014: 5447-5451.
- [12]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.
- [13]SHI W, LING Q, YUAN K, et al. On the linear convergence of the ADMM in decentralized consensus optimization[J]. IEEE Transactions on Signal Processing, 2014, 62(7): 1750-1761.
- [14]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.
- [15]TSIANOS K I, LAWLOR S, RABBAT M G. Push-sum distributed dual averaging for convex optimization[C]//IEEE 51st Annual Conference on Decision and Control (CDC), 2012: 5453-5458.
- [16]BAJOVIC D, JAKOVETIC D, KREJIC N, et al. Newton-like method with diagonal correction for distributed optimization [DB/OL](2015-09-05)[2016-01-05]. <http://arXiv preprint arXiv:1509.01703>.
- [17]MOKHTARI A, LING Q, RIBEIRO A. An approximate Newton method for distributed optimization[C]//IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.
- [18]JAKOVETIC D, MOURA J M F, XAVIER J. Distributed Nesterov-like gradient algorithms[C]//IEEE 51st Annual Conference on Decision and Control (CDC), 2012: 5459-5464.
- [19]时贞军, 孙国. 无约束优化问题的对角稀疏拟牛顿法[J]. 系统科学与数学, 2006, 26(1): 101-112.
SHI Zhenjun, SUN Guo. A diagonal-sparse quasi-Newton method for unconstrained optimization problem[J]. Journal of Systems Science and Mathematical Sciences, 2006, 26(1): 101-112.

(责任编辑: 傅 游)