doi:10.16055/j.issn.1672-058X.2020.0002.002

发散步长准则下的增量聚合梯度算法*

钱晓慧，王湘美**
(贵州大学 数学与统计学院,贵阳 550025)

摘　要:针对目标函数是若干光滑函数之和的优化问题,提出采用发散步长准则的增量聚合梯度算法。与增量梯度算法一样,增量聚合梯度算法的每次迭代也只需要计算其中一个函数的梯度。目前关于增量聚合梯度算法的研究主要是采用常值步长的增量聚合梯度算法,这一算法要求目标函数二阶连续可微且强凸,且常值步长的选取依赖最优点的二阶导数;而发散步长准则不依赖目标函数。在目标函数的梯度有界且李普希兹连续假设条件下,证明了采用发散步长的增量聚合梯度算法的收敛性;最后,通过数值例子验证了算法的收敛性,并与采用相同步长准则的增量梯度算法进行比较;数值结果表明对于某些优化问题,增量聚合梯度算法比采用相同步长的增量梯度算法更有效。

关键词:光滑优化;增量梯度算法;增量聚合梯度算法;发散步长

中图分类号:O224　　**文献标志码**:A　　**文章编号**:1672-058X(2020)02-0006-06

0　引　言

主要研究的是优化问题:

$$\min_{x\in R^n} f(x):=\sum_{i\in I} f_i(x) \tag{1}$$

其中 $I=\{1,2,\cdots,m\}$,对任意的 $i\in I$,$f_i:R^n\to\mathbf{R}$ 为一阶连续可微函数。这个问题出现在许多应用中,包括最小二乘问题[1-3]或更一般的参数估计问题[4]、无线传感器网络中的分布式优化问题[5]、机器学习问题[6-7]、神经网络训练问题[5-8]和函数期望值的最小化等许多实际问题。对于此类问题,可采用梯度法、共轭梯度法和拟牛顿法,但运用此类算法需要计算整个函数的梯度∇f。考虑到目标函数 f 是多个函数之和,一些学者提出了增量方法。例如,Bersekas 等[9]提出了增量梯度算法。设 x_k 为当前迭代点,α_k 为步长,迭代方式为 $x_{k+1}=\varphi_m$,其中 $\varphi_0=x_k$,$\varphi_i=\varphi_{i-1}+\alpha_k\nabla f_i(\varphi_{i-1})$ $(i\in I)$。在文献[9]中,作者研究了采用以下发散步长准则下的增量梯度算法的收敛性,步长 α_k 满足:

$$\alpha_k>0,\sum_{k=0}^{\infty}\alpha_k=\infty,\sum_{k=0}^{\infty}\alpha_k^2<\infty \tag{2}$$

另外,Blatt 等[5]提出了增量聚合梯度算法,算法每次以最近 m 个点的梯度之和为迭代方向。在以下假设条件下,作者证明了当常值步长充分小时迭代点列收敛到最小值点:

(1) 每个 f_i 的梯度在 R^n 上 Lipschitz 连续并且 $\|\nabla f_i(x)\|$ 有界。

(2) 目标函数 $f(x)$ 有唯一全局最小点 x^*,且 $f(x)$ 的 Hessian 矩阵$\nabla^2 f(x)$在 x^* 的某邻域内连续且正定。

(3) 若$\lim\limits_{k\to\infty} f(x_k)=f(x^*)$或$\lim\limits_{k\to\infty}\|\nabla f(x)\|=0$,则有$\lim\limits_{k\to\infty} x_k=x^*$。

收稿日期:2019-06-25;**修回日期**:2019-07-20.

* **基金项目**:国家自然科学基金(11661019);贵州省科技计划项目(20161039).

作者简介:钱晓慧(1994—),女,四川自贡人,硕士研究生,从事运筹学与控制论研究.

** **通讯作者**:王湘美(1972—),女,湖南湘潭人,博士,从事计算数学研究. Email:xmwang2@gzu.edu.cn.

在文献[5]中，常值步长的选取依赖于$\nabla^2 f(\boldsymbol{x}^*)$的最大、最小特征值，所以对目标函数要求高，适用范围受到限制。考虑文献[9]中步长规则选取，注意到发散步长准则的应用不依赖于目标函数，适用范围更广。因此，在文献[5][9]的基础上，本文将研究发散步长准则下的增量聚合梯度算法。

文章结构如下，第一节，介绍预备知识；第二节介绍增量聚合梯度算法，并证明采用发散步长准则的增量聚合梯度算法的收敛性；第三节是数值实验及其算法比较。

1　预备知识

在本节，将介绍光滑优化的一些定义及引理. 设$\boldsymbol{R}^n$表示n维欧式空间，$\|\cdot\|$为欧氏范数，$\mathbf{N}$表示自然数集。以下总假设$\boldsymbol{f}:\boldsymbol{R}^n\to\mathbf{R}$为一阶连续可微函数，并记函数$\boldsymbol{f}$在$\boldsymbol{x}\in\boldsymbol{R}^n$点的梯度为$\nabla\boldsymbol{f}(\boldsymbol{x})$

定义 1　如果存在常数$L>0$，使得

$$\|\nabla\boldsymbol{f}(\boldsymbol{x})-\nabla\boldsymbol{f}(\boldsymbol{y})\|\leqslant L\|\boldsymbol{x}-\boldsymbol{y}\| \quad \forall\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{R}^n$$

成立，则称$\nabla\boldsymbol{f}$是模为L的 Lipschitz 连续的。

为证明算法的收敛性，需用到以下引理：

引理 1　若$\nabla\boldsymbol{f}$是模为L的 Lipschitz 连续，则有

$$\boldsymbol{f}(\boldsymbol{y})-\boldsymbol{f}(\boldsymbol{x})\leqslant\frac{L}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2+[\nabla\boldsymbol{f}(\boldsymbol{x})]^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) \quad \forall\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{R}^n$$

引理 2　设序列$\{Y_k\}\{W_k\}\{Z_k\}$满足$\{W_k\}$为非负序列，则

$$Y_{k+1}\leqslant Y_k-W_k+Z_k,\ \forall k\in\mathbf{N}$$

以及$\sum\limits_{k=0}^{\infty}Z_k<+\infty$，则有$\lim\limits_{k\to\infty}Y_k=-\infty$，或者$\{Y_k\}$收敛且有$\sum\limits_{k=0}^{\infty}W_k<\infty$。

由不等式$2ab\leqslant a^2+b^2(a,b\in\mathbf{R})$，容易证明以下结论：

引理 3　设$\sum\limits_{k=0}^{\infty}\alpha_k^2<\infty, j\in\mathbf{N}$，则有$\sum\limits_{k=0}^{\infty}\alpha_k\alpha_{k+j}<\infty$。

2　算法及收敛性

研究优化问题式(1)，即

$$\min_{\boldsymbol{x}\in\boldsymbol{R}^n}\boldsymbol{f}(\boldsymbol{x}):=\sum_{i\in I}\boldsymbol{f}_i(\boldsymbol{x})$$

其中任意$\boldsymbol{i}\in\boldsymbol{I}$，函数$\boldsymbol{f}_i:\boldsymbol{R}^n\to\mathbf{R}$为一阶连续可微函数。为求解式(1)，Batt 等在文献[5]中提出了增量聚合梯度算法。

算法 1　（增量聚合梯度算法）

Step 0　给定任意m个初始点$\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m\in\boldsymbol{R}^n$，以及参数$\varepsilon\geqslant 0$，令$k=m$，计算聚合梯度$\boldsymbol{d}_m=\sum\limits_{i=1}^{m}\nabla\boldsymbol{f}_i(\boldsymbol{x}_i)$。

Step 1　选取步长$\alpha_k>0$。

Step 2　当$k\geqslant m$时，令$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k-\alpha_k\boldsymbol{d}_k$。$\boldsymbol{d}_{k+1}=\boldsymbol{d}_k-\nabla\boldsymbol{f}_{(k)_m+1}(\boldsymbol{x}_{k+1-m})+\nabla\boldsymbol{f}_{(k)_m+1}(\boldsymbol{x}_{k+1})$，$k=k+1$（其中$(k)_m+1$表示$k\bmod m+1$）。

Step 3　若$\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|\leqslant\varepsilon$，算法终止；否则，转 Step1。

正如引言中所提出的文献[5]研究的是常值步长的增量聚合梯度算法，即$\alpha_k=\alpha$，为保证算法的收敛性，α取充分小正数，它依赖于最优点处 Hessian 矩阵的最大、最小特征值。据了解，采用发散步长的增量聚合梯度算法的使用范围更广。因此，本文将研究采用发散步长的增量聚合梯度算法，即步长α_k满足$\alpha_k>0$，$\sum\limits_{k=0}^{\infty}\alpha_k=\infty$，$\sum\limits_{k=0}^{\infty}\alpha_k^2<\infty$。另外，将用到以下假设：

（H1）$\nabla\boldsymbol{f}_i$是模为$L(L>0)$的 Lipschitz 连续，即

$$\|\nabla\boldsymbol{f}_i(\boldsymbol{x})-\nabla\boldsymbol{f}_i(\boldsymbol{y})\|\leqslant L\|\boldsymbol{x}-\boldsymbol{y}\|,\ \forall\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{R}^n$$

（H2）存在实数$C>0$，对任意的$\boldsymbol{i}\in\boldsymbol{I}$，$k\in\mathbf{N}$，使得

$$\|\nabla\boldsymbol{f}_i(\boldsymbol{x}_k)\|\leqslant C$$

由算法的 Step 2，有

$$\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|=\alpha_k\|\boldsymbol{d}_k\|,\ \forall k\in\mathbf{N} \tag{3}$$

若假设(H1)和(H2)成立，则有

$$\|\boldsymbol{d}_k\|\leqslant mC,\ \forall k\geqslant 2m \tag{4}$$

$$\|\nabla\boldsymbol{f}(\boldsymbol{x})\|\leqslant mC,\boldsymbol{x}\in\boldsymbol{R}^n \tag{5}$$

$\nabla\boldsymbol{f}$是模为mL的 Lipschitz 连续，即

$$\|\nabla\boldsymbol{f}(\boldsymbol{x})-\nabla\boldsymbol{f}(\boldsymbol{y})\|\leqslant mL\|\boldsymbol{x}-\boldsymbol{y}\|,\ \forall\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{R}^n \tag{6}$$

关于算法 1，有以下引理：

引理 4　假设(H1)(H2)成立，则算法 1 产生的点列$\{\boldsymbol{x}_k\}$满足

$$\boldsymbol{f}(\boldsymbol{x}_{k+1})\leqslant$$

$$\boldsymbol{f}(\boldsymbol{x}_k)-\alpha_k\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|^2+$$

$$\beta\sum_{i=1}^{m-1}(m-i)\alpha_k\alpha_{k-i}+\gamma\alpha_k^2$$
$$2m\leqslant k\in\mathbf{N} \tag{7}$$

其中 $\beta=m^2C^2L,\gamma=\dfrac{m^3C^2L}{2}$。

证明 设 $k\geqslant 2m$,由算法 1 的 Step 2 知

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k-\alpha_k\boldsymbol{d}_k=\boldsymbol{x}_k-\alpha_k\sum_{i=1}^{m}\nabla\boldsymbol{f}_{(k-i)_m+1}(\boldsymbol{x}_{k-i+1})$$

又记

$$h_k=\sum_{i=2}^{m}[\nabla\boldsymbol{f}_{(k-i)_m+1}(\boldsymbol{x}_{k-i+1})-\nabla\boldsymbol{f}_{(k-i)_m+1}(\boldsymbol{x}_k)] \tag{8}$$

则有

$$\boldsymbol{d}_k=\nabla\boldsymbol{f}(\boldsymbol{x}_k)+h_k \tag{9}$$

结合式(8)以及假设条件(H1),有

$$\|h_k\|=\|\sum_{i=2}^{m}[\nabla\boldsymbol{f}_{(k-i)_m+1}(\boldsymbol{x}_{k-i+1})-\nabla\boldsymbol{f}_{(k-i)_m+1}(\boldsymbol{x}_k)]\|\leqslant$$
$$\sum_{i=2}^{m}\|\nabla\boldsymbol{f}_{(k-i)_m+1}(\boldsymbol{x}_{k-i+1})-\nabla\boldsymbol{f}_{(k-i)_m+1}(\boldsymbol{x}_k)\|\leqslant$$
$$L\sum_{i=2}^{m}\|\boldsymbol{x}_{k-i+1}-\boldsymbol{x}_k\|$$

由式(3)(4),有

$$\|\boldsymbol{x}_{k-i}-\boldsymbol{x}_k\|\leqslant\|\boldsymbol{x}_{k-i}-\boldsymbol{x}_{k-i+1}\|+$$
$$\|\boldsymbol{x}_{k-i+1}-\boldsymbol{x}_{k-i+2}\|+\cdots+\|\boldsymbol{x}_{k-1}-\boldsymbol{x}_k\|\leqslant mC\sum_{i=1}^{m-1}\alpha_{k-i}$$

所以

$$\|h_k\|\leqslant mCL\sum_{i=1}^{m-1}(m-i)\alpha_{k-i} \tag{10}$$

另外,根据式(6)和引理 1 有

$$\boldsymbol{f}(\boldsymbol{x}_{k+1})-\boldsymbol{f}(\boldsymbol{x}_k)=$$
$$\boldsymbol{f}(\boldsymbol{x}_k-\alpha_k\boldsymbol{d}_k)-\boldsymbol{f}(\boldsymbol{x}_k)=$$
$$\int_0^1[\nabla\boldsymbol{f}(\boldsymbol{x}_k-\xi\alpha_k\boldsymbol{d}_k)]^{\mathrm{T}}(-\alpha_k\boldsymbol{d}_k)\mathrm{d}\xi\leqslant$$
$$[\nabla\boldsymbol{f}(\boldsymbol{x}_k)]^{\mathrm{T}}(-\alpha_k\boldsymbol{d}_k)+$$
$$\int_0^1\|\nabla\boldsymbol{f}(\boldsymbol{x}_k-\xi\alpha_k\boldsymbol{d}_k)-\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|\|\alpha_k\boldsymbol{d}_k\|\mathrm{d}\xi\leqslant$$
$$-\alpha_k[\nabla\boldsymbol{f}(\boldsymbol{x}_k)]^{\mathrm{T}}\boldsymbol{d}_k+\|\alpha_k\boldsymbol{d}_k\|\int_0^1 mL\xi\|\alpha_k\boldsymbol{d}_k\|\mathrm{d}\xi\leqslant$$
$$-\alpha_k[\nabla\boldsymbol{f}(\boldsymbol{x}_k)]^{\mathrm{T}}\boldsymbol{d}_k+\frac{mL}{2}\alpha_k^{\ 2}\|\boldsymbol{d}_k\|^2$$

于是有

$$\boldsymbol{f}(\boldsymbol{x}_{k+1})-\boldsymbol{f}(\boldsymbol{x}_k)\leqslant$$
$$-\alpha_k[\nabla\boldsymbol{f}(\boldsymbol{x}_k)]^{\mathrm{T}}(\nabla\boldsymbol{f}(\boldsymbol{x}_k)+h_k)+\frac{m^3C^2L}{2}\alpha_k^{\ 2}\leqslant$$
$$-\alpha_k\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|^2-\alpha_k[\nabla\boldsymbol{f}(\boldsymbol{x}_k)]^{\mathrm{T}}h_k+\frac{m^3C^2L}{2}\alpha_k^{\ 2}\leqslant$$
$$-\alpha_k\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|^2+mC\alpha_k\|h_k\|+\frac{m^3C^2L}{2}\alpha_k^{\ 2}\leqslant$$
$$-\alpha_k\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|^2+m^2C^2L\sum_{i=1}^{m-1}(m-i)\alpha_k\alpha_{k-i}+\frac{m^3C^2L}{2}\alpha_k^2$$

证毕。

下面证明算法 1 的收敛性.

定理 1 假设(H1)和(H2)成立,$\{\boldsymbol{x}_k\}$ 由算法 1 采用发散步长准则式(2)产生,则有 $\lim\limits_{k\to\infty}\boldsymbol{f}(\boldsymbol{x}_k)=-\infty$,或者 $\{\boldsymbol{f}(\boldsymbol{x}_k)\}$ 收敛且 $\lim\limits_{k\to\infty}\nabla\boldsymbol{f}(\boldsymbol{x}_k)=0$,

证明 由引理 4 中的式(7)知,存在 $\beta,\gamma>0$,使得

$$\boldsymbol{f}(\boldsymbol{x}_{k+1})\leqslant$$
$$\boldsymbol{f}(\boldsymbol{x}_k)-\alpha_k\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|^2+$$
$$\beta\sum_{i=1}^{m-1}(m-i)\alpha_k\alpha_{k-i}+\gamma\alpha_k^2,\ \forall k\geqslant 2m$$

根据引理 3 以及 $\sum\limits_{k=0}^{\infty}\alpha_k^2<\infty$(式(2)),可知

$$\beta\sum_{k=2m}^{\infty}\sum_{i=1}^{m-1}(m-i)\alpha_k\alpha_{k-i}<\infty$$

因此

$$\sum_{k=2m}^{\infty}(\beta\sum_{i=1}^{m-1}(m-i)\alpha_k\alpha_{k-i}+\gamma\alpha_k^2)<\infty \tag{11}$$

由引理 2($Y_k:=\boldsymbol{f}(\boldsymbol{x}_k)$,$W_k:=\alpha_k\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|^2$,$\{Z_k\}:=\beta\sum\limits_{i=1}^{m-1}(m-i)\alpha_k\alpha_{k-i}+\gamma\alpha_k^2$,有 $\lim\limits_{k\to\infty}\boldsymbol{f}(\boldsymbol{x}_k)=-\infty$,或者 $\{\boldsymbol{f}(\boldsymbol{x}_k)\}$ 收敛且

$$\sum_{k=0}^{\infty}\alpha_k\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|^2<\infty \tag{12}$$

为完成定理 1 的证明,下面需证明当 $\{\boldsymbol{f}(\boldsymbol{x}_k)\}$ 收敛时,有 $\lim\limits_{k\to\infty}\nabla\boldsymbol{f}(\boldsymbol{x}_k)=0$。设 $\{\boldsymbol{f}(\boldsymbol{x}_k)\}$ 收敛,如果存在常数 $\eta>0$ 和一个整数 K,对所有的 k 有 $k>K$ 成立,$\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|\geqslant\eta$,则

$$\sum_{k=K}^{\infty}\alpha_k\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|^2\geqslant\eta^2\sum_{k=K}^{\infty}\alpha_k=\infty$$

与式(12)矛盾,因此

$$\liminf_{k\to\infty}\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|=0 \tag{13}$$

下面用反证法证明 $\limsup\limits_{k\to\infty}\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|=0$。

假设 $\limsup\limits_{k\to\infty}\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|>0$,则必存在常数 $\sigma>0$,有无穷多个 k 满足 $\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|>\sigma$,又由式(13)知有无穷多个 k 满足 $\|\nabla\boldsymbol{f}(\boldsymbol{x}_k)\|<\dfrac{\sigma}{2}$,因此,存在无限的

整数集 $N_1 \subset \mathbf{N}$,使得对任意的 $k \in N_1$,存在一个整数 $l(k)>k$,满足

$$\|\nabla \boldsymbol{f}(\boldsymbol{x}_k)\| < \frac{\sigma}{2}, \quad \|\nabla \boldsymbol{f}(\boldsymbol{x}_{l(k)})\| > \sigma \tag{14}$$

$$\frac{\sigma}{2} \leqslant \|\nabla \boldsymbol{f}(\boldsymbol{x}_j)\| \leqslant \sigma, k<j<l(k) \tag{15}$$

于是对任意的 $k \in N_1$,有

$$\|\nabla \boldsymbol{f}(\boldsymbol{x}_{l(k)})\| - \|\nabla \boldsymbol{f}(\boldsymbol{x}_k)\| \geqslant \frac{\sigma}{2} \tag{16}$$

另外,

$$\begin{aligned}
&\|\nabla \boldsymbol{f}(\boldsymbol{x}_{k+1})\| - \|\nabla \boldsymbol{f}(\boldsymbol{x}_k)\| \leqslant \\
&\quad \|\nabla \boldsymbol{f}(\boldsymbol{x}_{k+1}) - \nabla \boldsymbol{f}(\boldsymbol{x}_k)\| \leqslant \\
&\quad mL\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| \leqslant \\
&\quad mL\alpha_k \|\boldsymbol{d}_k\| \leqslant \\
&\quad m^2 CL\alpha_k
\end{aligned}$$

因为 $\lim\limits_{k\to\infty} \alpha_k = 0$,所以存在 $K_1 \in N_1$,使得对 $k \geqslant K_1$,有 $m^2 CL\alpha_k < \frac{\sigma}{4}$。

再由式(14)或式(15),有

$$\|\nabla \boldsymbol{f}(\boldsymbol{x}_{k+1})\| \geqslant \frac{\sigma}{2}, k \in N_1$$

因此

$$\begin{aligned}
&\|\nabla \boldsymbol{f}(\boldsymbol{x}_k)\| \geqslant \|\nabla \boldsymbol{f}(\boldsymbol{x}_{k+1})\| - m^2 CL\alpha_k \geqslant \\
&\quad \frac{\sigma}{2} - \frac{\sigma}{4} = \frac{\sigma}{4}, K_1 \leqslant k \in N_1
\end{aligned} \tag{17}$$

结合式(17)以及式(7),有

$$\begin{aligned}
&\left(\frac{\sigma}{4}\right)^2 \sum_{j=k}^{l(k)-1} \alpha_j \leqslant \\
&\quad \boldsymbol{f}(\boldsymbol{x}_k) - \boldsymbol{f}(\boldsymbol{x}_{l(k)}) + \\
&\quad \sum_{j=k}^{l(k)-1} \left(\beta \sum_{i=1}^{m-1} (j-i)\alpha_j \alpha_{j-i} + \gamma \alpha_j^2\right) \\
&\quad K_1 \leqslant k \in N_1
\end{aligned}$$

因为 $\{\boldsymbol{f}(\boldsymbol{x}_k)\}$ 收敛,知

$$\lim_{k\to\infty, K_1 \leqslant k \in N_1} \boldsymbol{f}(\boldsymbol{x}_k) - \boldsymbol{f}(\boldsymbol{x}_{l(k)}) = 0$$

步长 α_k 满足式(11),有

$$\lim_{k\to\infty, K_1 \leqslant k \in N_1} \sum_{j=k}^{l(k)-1} \left(\beta \sum_{i=1}^{m-1} (j-i)\alpha_j \alpha_{j-i} + \gamma \alpha_j^2\right) = 0$$

因此

$$\lim_{k\to\infty, K_1 \leqslant k \in N_1} \sum_{j=k}^{l(k)-1} \alpha_j = 0 \tag{18}$$

另一方面,由式(16)有

$$\begin{aligned}
&\frac{\sigma}{2} \leqslant \|\nabla \boldsymbol{f}(\boldsymbol{x}_{l(k)})\| - \|\nabla \boldsymbol{f}(\boldsymbol{x}_k)\| \leqslant \\
&\quad \|\nabla \boldsymbol{f}(\boldsymbol{x}_{l(k)}) - \nabla \boldsymbol{f}(\boldsymbol{x}_k)\| \leqslant \\
&\quad mL\|\boldsymbol{x}_{l(k)} - \boldsymbol{x}_k\| \leqslant \\
&\quad mL\sum_{j=k}^{l(k)-1} \alpha_j \|\boldsymbol{d}_j\| \leqslant \\
&\quad m^2 CL \sum_{j=k}^{l(k)-1} \alpha_j
\end{aligned}$$

所以

$$\liminf_{k\to\infty, K_1 \leqslant k \in N_1} \sum_{j=k}^{l(k)-1} \alpha_j \geqslant \frac{\sigma}{2m^2 CL}$$

这与式(18)矛盾,因此 $\limsup\limits_{k\to\infty} \|\nabla \boldsymbol{f}(\boldsymbol{x}_k)\| = 0$,又 $\liminf\limits_{k\to\infty} \|\nabla \boldsymbol{f}(\boldsymbol{x}_k)\| = 0$,所以 $\lim\limits_{k\to\infty} \|\nabla \boldsymbol{f}(\boldsymbol{x}_k)\| = 0$。证毕。

3 数值结果

在这一节,将用算法 1 求解几个具体优化问题,并通过数值计算结果比较该算法和增量梯度算法[9,Algorithm 3]的收敛效率。其中增量聚合梯度算法的任意 m 个初始点都取其等于 x_0(即 $x_1 = x_2 = \cdots = x_m = x_0$)。特别地,应用算法 1 求解例 1 和例 2 中的稳健估计问题和源定位问题,两个算例均来自参考文献[5],在所有算例中算法计算精度均取为 $\varepsilon = 10^{-5}$,编程软件为 Matlab[10],两个算例中算法都采用相同步长并且皆满足式(2)。

例 1 源定位问题[5]。设传感器 i 等距分布在 $100 * 100$ 场的空间位置 $\boldsymbol{r}_i \in \boldsymbol{R}^2$ 上,每个传感器收集从源点 $\boldsymbol{x}^* \in \boldsymbol{R}^2$ 所发射声信号的噪声测量 y_i. 源定位问题就是根据收集的信号 $\{y_i\}$,找到源的位置 $\boldsymbol{x}^*$。基于远场假设和各向同性声波传播模型[11-13],源位置估计问题可归结为非线性最小二乘问题:

$$\min_{\boldsymbol{x} \in \boldsymbol{R}^2} \boldsymbol{f}(\boldsymbol{x}) = \sum_{i=1}^{m} \boldsymbol{f}_i(\boldsymbol{x})$$

其中 $\boldsymbol{f}_i: \boldsymbol{R}^2 \to \mathbf{R}(i=1,2,\cdots,m)$ 定义为

$$\boldsymbol{f}_i(\boldsymbol{x}) = (y_i - g(\|\boldsymbol{r}_i - \boldsymbol{x}\|^2))^2, \forall \boldsymbol{x} \in \boldsymbol{R}^2$$

函数 $g: \mathbf{R} \to \mathbf{R}$ 定义为

$$g(z) = \begin{cases} \dfrac{A}{z} & z \geqslant A/\varepsilon \\ 2\varepsilon - \dfrac{z\varepsilon^2}{A} & z < A/\varepsilon \end{cases} \quad \forall z \in \mathbf{R} \tag{19}$$

在式(19)中,A 和 ε 是表示与信号强度相关的常数,容易验证函数 $\boldsymbol{f}_i(i=1,2,\cdots,m)$ 一阶连续可

微。在数值模拟实验中,源点 $\boldsymbol{x}^*=(60,60)$,$A=1\ 000$,$\varepsilon=1$,$m=16$。取 $\boldsymbol{r}_i$ 为 100 * 100 网格上均匀分布的 16 个点,噪声测量 y_i 是根据高斯分布产生的,其均值为 $g(\|\boldsymbol{r}_i-\boldsymbol{x}^*\|^2)$,方差为 1。例 1 的实验结果见表 1,两种算法中的步长为 $\alpha_k=\frac{1}{(k+1)^{\frac{101}{200}}}$。

表 1 源定位

Table 1 Source localization

初始点($\boldsymbol{x}_0$)	算法	$\boldsymbol{x}_k$	$\|\nabla \boldsymbol{f}(\boldsymbol{x}_k)\|$	迭代次数(k)
$\boldsymbol{x}_0=(50,50)$	[9,Algorithm3]	(69.797 7,50.697 2)	$2.631\ 0e^{-05}$	50 000
	算法 1	(69.797 9,50.696 4)	$9.999\ 8e^{-06}$	48 893
$\boldsymbol{x}_0=(60,60)$	[9,Algorithm3]	(69.797 9,50.697 4)	$2.664\ 3e^{-05}$	50 000
	算法 1	(69.797 9,50.696 5)	$9.999\ 4e^{-06}$	40 223
$\boldsymbol{x}_0=(80,80)$	[9,Algorithm3]	(69.798 2,50.697 8)	$3.452\ 5e^{-05}$	50 000
	算法 1	(69.798 4,50.697 8)	$1.135\ 1e^{-05}$	50 000

例 2 稳健估计问题[5]。传感器网络用部署大量低成本传感器来密集监视某个区域的某种形态。由于低成本传感器的可靠性有限,系统必须设计成对单个传感器的可能故障具有鲁棒性。这意味着在评估任务中一些传感器会产生不合理的测量结果,即异常值。在文献[12]中,作者建议使用稳健的统计数据来减轻数据中异常值的影响[13-14]。稳健估计使用的是赋予异常值较少权重的目标函数,可以实现此目的的一个常用函数是“Fair”函数 $g:\mathbf{R}\to\mathbf{R}$,定义为

$$g(x)=c^2\left[\frac{|x|}{c}-\ln\left(1+\frac{|x|}{c}\right)\right],\ \forall x\in\mathbf{R} \quad (20)$$

与参考文献[12]类似,模拟了一个用于测量污染水平的传感器网络,并假设一定比例的传感器损坏且提供不合理的测量结果。每个传感器都收集污染水平的单个噪声测量值,通过最小化目标函数式(21)来确定平均污染水平的估计值。这里研究的稳健估计问题转化为优化问题:

$$\min_{x\in\mathbf{R}} f(x)=\sum_{i=1}^{m} f_i(x) \quad (21)$$

其中 $f_i:\mathbf{R}\to\mathbf{R}$ 定义为

$$f_i(x)=\frac{1}{m}g(x-y_i),\ \forall x\in\mathbf{R}$$

其中 y_i 是由传感器 i 采集的测量值,容易验证函数 $f_i(i=1,2,\cdots,m)$ 一阶连续可微。现假设本次模拟有 20 个传感器,即 $m=20$。为了反映传感器故障的可能性,一半的样本按均值为 10,方差为 1 的高斯分布生成,另一半是按均值为 10,方差为 10 的高斯分布生成。其中式(19)中的系数 $c=10$。例 2 的实验结果见表 2,两种算法的步长皆为 $\alpha_k=\frac{1}{k+10}$。

表 2 稳健估计

Table 2 Robust estimation

初始点(x_0)	算法	x_k	$\|\nabla \boldsymbol{f}(\boldsymbol{x}_k)\|$	迭代次数(k)
$x_0=0$	[9,Algorithm3]	9.739 2	0.085 6	50 000
	算法 1	9.846 4	$5.579\ 2e^{-06}$	264
$x_0=5$	[9,Algorithm3]	9.806 4	0.031 9	50 000
	算法 1	9.846 4	$9.151\ 3e^{-06}$	227
$x_0=15$	[9,Algorithm3]	8.882 8	0.029 3	50 000
	算法 1	9.846 4	$7.447\ 2e^{-05}$	207

注:由上述两个算例的数据结果可知,在相同步长准则下,对某些优化问题,采用算法 1 的效率要高于[9,Algorithm 3]。

4 结 论

本文基于求解大规模优化问题时采用常值步长的增量聚合梯度法,提出并研究了采用发散步长准则的增量聚合梯度算法。与常值步长相比较,发散步长不依赖目标函数,应用范围更广。文章最后的数值结果表明对于某些优化问题,增量聚合梯度算法比采用相同步长准则的增量梯度算法更有效,这为求解大规模解优化问题提供了一种解决方法。今后,还将继续研究聚合梯度法的其他应用和分析算法的收敛速度。

参考文献(References):

[1] BERTSEKAS D P. Incremental Least Squares Methods and the Extended Kalman Filter[J]. SIAM Journal on Optimization,1996,6(3):807—822

[2] BERTSEKAS D P. A New Class of Incremental Gradient Methods for Least Squares Problems [J]. SIAM Journal on Optimization, 1997, 7(4):913—926

[3] MORIYAMA H, YAMASHITA N, FUKUSHIMA M. The Incremental Gauss-newton Algorithm with Adaptive Stepsize Rule[J]. Computational Optimization & Applications, 2003,

26(2): 107— 141

[4] GURBUZBALABAN M, OZDAGLAR A, PARRILO P. A Globally Convergent Incremental Newton Method[J]. Mathematical Programming,2015,151(1):283—313

[5] BLATT D, HERO A, GAUCHMAN H. A Convergent Incremental Gradient Method with a Constant Step Size[J]. SIAM Journal on Optimization,2007,18(1):29—51

[6] BOTTOU L, CUN Y L. On-line Learning for Very Large Data Sets[J]. Applied Stochastic Models in Business & Industry,2005,21(2):137—151

[7] ROUX N L, SCHMIDT M, BACH F R. A Stochastic Gradient Method with an Exponential Convergence Rate for Finite Training Sets [J]. Advances in Neural Information Processing Systems,2012(4):2663—2671

[8] GRIPPO L. A Class of Unconstrained Minimization Methods for Neural Network Training[J]. Optimization Methods and Software,1994, 4(2):135—150

[9] BERTSEKAS D P, TSITSIKLIS J N. Gradient Convergence in Gradient Methods with Errors[J]. SIAM Journal on Optimization ,2000, 10(3):627—642

[10] TREFETHEN. Spectral Methods in MATLAB[M]. New York: SIAM, 2000

[11] CHEN J C, YAO K, HUDSON R E. Source Localization and Beam Forming [J]. IEEE Signal Processing Magazine, 2002(19):30—39

[12] RABBAT M G, NOWAK R D. Decentralized Source Localization and Tracking [C]//Montreal: IEEE International Conference on Acoustics, 2004

[13] HUBER P. Robust Statistics [M]. New York: John Wiley & Sons, 1981

[14] POLYAK B T. Introduction to Optimization[M]. New York: Optimization Software Income,1987

[15] REY W J J. Introduction to Robust and Quasi-Robust Statistical Methods[M]. Berlin: Springer Verlag, 1983

Incremental Aggregated Gradient Algorithm with the Divergence Step Size

QIAN Xiao-hui, WANG Xiang-mei

(School of Mathematics and Statistics, Guizhou University, Guiyang 550025, China)

Abstract: For optimizing that objective function is the sum of several smooth functions, the incremental aggregated gradient algorithm employing the divergence step size is proposed. As the same as the incremental gradient algorithm, each iteration of the incremental aggregated gradient algorithm only needs to calculate a single gradient of a function. At present, the studies on the incremental aggregated gradient algorithm mainly employ the constant step size under the assumption that the objective function is twice continuously differentiable and strongly convex, and the choice of the constant step size depends on the second order derivative of the optimal point. However, the divergence step size is independent of the objective function. The convergence of the incremental aggregated gradient algorithm with the divergence step size is verified under the assumption that the gradients of the objective function are bounded and Lipschitz continuous. Finally, some numerical examples are provided to illustrate the theoretical result and to compare with the incremental gradient algorithm with the same step size. The numerical results show that, for some optimizations, the incremental aggregated gradient algorithm is more effective than the incremental gradient algorithm with the same step size.

Key words: smooth optimization; incremental gradient algorithm; incremental aggregated gradient algorithm; divergence step size

责任编辑:李翠薇

引用本文/Cite this paper:

钱晓慧, 王湘美. 发散步长准则下的增量聚合梯度算法[J]. 重庆工商大学学报(自然科学版),2020,37(2):6—11

QIAN X H, WANG X M. Incremental Aggregated Gradient Algorithm with the Divergence Step Size[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2020,37(2):6—11