Pure Mathematics 理论数学, 2023, 13(12), 3525-3536
Published Online December 2023 in Hans. https://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2023.1312367

Hadamard流形上的多目标邻近梯度算法

刘仁金，王湘美*

贵州大学数学与统计学院，贵州 贵阳

收稿日期：2023年11月12日；录用日期：2023年12月13日；发布日期：2023年12月22日

摘 要

邻近梯度算法是求解非光滑优化问题的经典算法。本文将多目标优化问题的邻近梯度算法推广到Hadamard流形上。在一定条件下，证明了算法产生序列的聚点是Pareto稳定点。在目标函数满足Polyak-Lojasiewicz不等式时，得到算法的收敛速度是线性的。所得结果在Hadamard流形上是新的。

关键词

Hadamard流形，邻近梯度算法，Polyak-Lojasiewicz不等式

Proximal Gradient Algorithm for Multiobjective Optimization on Hadamard Manifold

Renjin Liu, Xiangmei Wang*

School of Mathematics and Statistics, Guizhou University, Guizhou University, Guiyang Guizhou

Received: Nov. 12^{th}, 2023; accepted: Dec. 13^{th}, 2023; published: Dec. 22^{nd}, 2023

* 通讯作者。

文章引用: 刘仁金, 王湘美. Hadamard流形上的多目标邻近梯度算法[J]. 理论数学, 2023, 13(12): 3525-3536.
DOI: 10.12677/pm.2023.1312367

Abstract

Proximal gradient algorithm is a classical algorithm for solving nonsmooth optimization problems. In this paper, the multiobjective proximal gradient algorithm is extended to Hadamard manifold. Under certain conditions, it is proved that the cluster point of the sequence generated by the algorithm is Pareto stationary. In the case, when the objective function satisfies the Polyak-Lojasiewicz inequality, the convergence rate of the algorithm is linear.

Keywords

Hadamard Manifold, Proximal Gradient Algorithm, Polyak-Lojasiewicz Inequality

1. 前言

黎曼流形的优化技术可以将欧氏空间中的高维问题转化为较低维问题, 欧氏空间中的非凸问题, 约束优化问题或者病态问题可以通过转化为适当黎曼流形上的凸问题, 无约束问题或者非病态问题. 因此黎曼流形优化理论和算法的研究引起了许多学者的关注, 很多欧氏空间上的经典的优化技术和算法都推广到了黎曼流形上. 例如, 牛顿法, 梯度法, 信赖域算法, 邻近梯度算法, 非精确邻近点算法等, 见 [1–7]及其参考文献.

非光滑优化问题源于许多实际问题, 例如稀疏主成分分析 [8, 9], 稀疏盲去卷积 [10] 和无监督特征选择 [11] 等. 邻近梯度算法因迭代成本低, 是求解非光滑等优化问题的高效算法, 因此受到很多学者的研究. 例如, FISTA在邻近梯度算法中用Nesteror 动量技术生成辅助序列$\{y_k\}$, 得到算法的收敛速度, 提出了加速迭代收缩阈值算法 [12]. Huang [13] 等人将邻近梯度采样方法推广到黎曼流形上, 该方法对求解小规模问题非常有效, 但缺乏收敛性分析. Hosseini和Uschmajew在 [14]中提出了一种具有收敛性分析的黎曼梯度采样方法. 在 [15]中，Hosseini等人结合次梯度和拟牛顿思想, 提出了一种新的黎曼线搜索方法. Liu 等人在 [16]中, 当$g = 0$ 和$F = f$ 是Lipschitz 连续可微时, 研究了黎曼流形上测地凸优化的一阶加速方法，得到最优收敛速度$O(1/k^2)$. Chen 等人在 [17] 提出了一种黎曼近端梯度方法, 该方法适用于流形M是欧几里得空间的子流形的情况, 作者证明了由其黎曼

邻近梯度算法的搜索方向的范数为零, 得到了算法的全局收敛性. Huang 和Wei在 [18] 证明了 [17] 中由黎曼近端梯度方法生成的序列的任何极限点都是临界点. 此外, 他们在 [19] 中将欧氏空间的加速迭代收缩阈值算法推广到黎曼流形上, 得到了可加速的黎曼邻近梯度方法, 并在目标函数满足黎曼Kurdyka-Lojasiewicz 不等式时, 得到算法的收敛速度. Fukuda 等人在 [20, 21] 中提出欧氏空间中求解多目标优化问题的邻近梯度算法的研究. 但是, 邻近梯度算法在流形上求解多目标优化问题还没有相关的研究.

受以上研究的启发(特别是 [17, 20, 21]), 本文将邻近梯度算法求解多目标优化问题推广到Hadamard 流形上, 在Hadamard流形的切空间上寻找合适的切向量作为搜索方向, 证明了算法产生序列的聚点是Pareto 稳定点. 此外, 借助价值函数, 在满足Polyak-Lojasiewicz(PL) 不等式时, 得到算法的收敛速度是线性的. 研究的结果在Hadamard流形上是新的.

本文的结构如下. 第二节介绍黎曼流形上的一些基本概念和性质. 主要内容在第三节, 包括邻近梯度点算法收敛性和收敛速度分析. 第四节是总结.

2. 预备知识

本节介绍本文中用到的符号和黎曼流形上的基本概念和性质, 可参考文献 [22–24].

设(M, g)是一个m维Hadamard流形, g是M上的黎曼度量. 我们用∇表示M上的Levi-Civita联络. 设$x \in M$. 用T_xM表示点x处的切空间, $TM := \bigcup_{x\in M} T_xM$表示$M$上的切丛. 设$\gamma : [a, b] \to M$ 是连接两点$x, y \in M$的分段光滑曲线($\gamma(a) = x, \gamma(b) = y$). 定义曲线$\gamma$ 的长度为$L(\gamma) := \int_a^b \|\gamma'(t)\| dt$. 两点$x, y \in M$ 的距离定义为$\mathrm{d}(x, y) := \inf_\gamma L(\gamma)$.如果$\nabla_{\gamma'} V = 0$, 向量场$V$称为沿着曲线$\gamma$平行. 特别地, 对于光滑曲线$\gamma$, 如果$\gamma'$自平行, 即$\nabla_{\gamma'}\gamma' = 0$, 则称$\gamma$ 是一条测地线. $\mathrm{d}(x, y) := \inf_\gamma L(\gamma)$. 以点$x$为中心, $r > 0$为半径的闭球记为$B(x, r)$, 即

$$B(x, r) := \{y \in M : \mathrm{d}(y, x) \le r\}.$$

在Hadamard流形上, 对于任意两点$x, y \in M$, 有唯一的一条测地线γ连接, $v \in T_xM$, 指数映射$\exp_x : T_xM \to M$ 定义为

$$\exp_x v = \gamma(1; x, v) = y,$$

且$\gamma(0) = x, \gamma'(0) = v$. 指数逆映射$\exp_x^{-1} : M \to T_xM$ 定义为

$$\exp_x^{-1} y = v,$$

且$\|\exp_x^{-1} y\| = \left\|\exp_y^{-1} x\right\| = \mathrm{d}(x, y)$. 平行移动$\mathcal{P}_{(x,y)} : T_xM \to T_yM$定义为

$$\mathcal{P}_{(x,y)} v = \omega, \quad \omega \in T_yM, v \in T_xM,$$

且保持内积不变, 即

$$\langle v, \omega' \rangle_x = \left\langle \mathcal{P}_{(x,y)} v, \mathcal{P}_{(x,y)} \omega' \right\rangle_y, \quad v, \omega' \in T_xM.$$

自然数集合记为$\mathbb{N}$. 对任意$n\in\mathbb{N}$, $C^n(M)$ 表示定义在M 上的n 阶连续可微函数的集合. 设$h\in C^1(M)$, h的梯度场记为$\mathrm{grad}h$.

设$C\subset M$是Hadamard流形M的非空子集, 对任意$x,y\in C$, 测地线γ有$\gamma(0)=x,\gamma(1)=y,\gamma(t)\in C,t\in[0,1]$, 若满足$\gamma((1-t)x+ty)\in C$, 则称$C$为凸集.

定义1 设$f:M\to R$是M上的连续可微函数, 对任意的$x,y\in M$, 测地线$\gamma(t)\in M,\gamma(0)=x,\gamma(1)=y$, $t\in[0,1]$ 如果

$$f(\gamma(t))\le tf(x)+(1-t)f(y),$$

则称f是M上的测地凸函数.如果

$$f(\gamma(t))< tf(x)+(1-t)f(y),$$

则称f是M上的测地强凸函数.

定义2 设$f:M\to R$是M上的连续可微函数, $\mu>0$, 对任意的$x,y\in M$, 如果

$$f(y)\ge f(x)+\left\langle \mathrm{grad}f(x),\exp_x^{-1}y\right\rangle+\frac{\mu}{2}\left\|\exp_x^{-1}y\right\|^2. \tag{1}$$

则称f是$\mu>0-$强凸函数. 当$\mu=0$时, f是凸函数, $\mu>0$时, f是强凸函数.

定义3 设$f:M\to R$是M上的连续可微函数, $L>0$, 对任意的$x,y\in M$, 如果f的梯度$\mathrm{grad}f$满足

$$\left\|\mathcal{P}_{(x,y)}\mathrm{grad}f(x)-\mathrm{grad}f(y)\right\|\le L\mathrm{d}(x,y),$$

则称f是梯度L-Lipschitz连续的, L为Lipschitz常数.

定义4 [16]设$f:M\to R$是M上的连续可微函数, 满足梯度L-Lipschitz连续, 如果

$$f(y)\le f(x)+\left\langle \mathrm{grad}f(x),\exp_x^{-1}y\right\rangle+\frac{L}{2}\left\|\exp_x^{-1}y\right\|. \tag{2}$$

则称f是测地$L-$光滑函数.

定义5 设$f:M\to R$是M上的连续可微函数, $x\in M,v\in T_xM$, 则f 在x 处的方向导数为

$$f'(x;v)=\lim_{t\to 0^+}\frac{f(\exp_x tv)-f(x)}{t}.$$

考虑一个多目标优化问题

$$\min_{x\in M}F(x)=(F_1(x),...,F_m(x))^T,$$

其中$F_i:M\to R$, $i\in I=\{1,...,m\}$. 定义R^m中的的偏序关系$\le(<)$, 对$u=(u_i),v=(v_i)\in R^m$, 如果$u\le v(u<v)$当且仅当$u_i\le v_i(u_i<v_i)$.

定义6 设$x^*\in M$, 如果不存在$x\in M$, 使$F(x)\le F(x^*)$, 且$F(x)\ne F(x^*)$, 则称x^*是F的Pareto最优点.

定义7 设$x^* \in M$, 如果不存在$x \in M$, 使$F(x) < F(x^*)$, 则称x^*是F的弱Pareto最优点,.

注1. F的Pareto最优点是F的弱Pareto最优点, 反之不一定成立.

定义8 [22] 设$x' \in M$, 如果

$$\max_{i \in I} F_i{}'(x'; v) \geq 0, \ \ \forall v \in T_{x'}M. \tag{3}$$

则称x'为F的Pareto稳定点.

3. 黎曼流形上的多目标邻近梯度算法

3.1. 算法及收敛性分析

本节给出Hadamard流形上的多目标邻近梯度算法及其收敛性的证明. 考虑如下的多目标优化问题：

$$\min_{x \in M} F(x) = f(x) + g(x), \tag{4}$$

其中$F : M \to R^m$是向量值函数($F = (F_i)_{i \in I}$), $F_i : M \to R$, $g : M \to R$是测地凸函数($g = (g_i)_{i \in I}$), $f : M \to R$ 是测地L_i-光滑函数($f = (f_i)_{i \in I}$), 令

$$L = \max_{i \in I} L_i. \tag{5}$$

设$x \in M$, 对任意的$v \in T_xM$, 定义函数$\psi_x : T_xM \to R$如下:

$$\psi_x(v) = \max_{i \in I} \left\{ \langle \operatorname{grad} f_i(x), v \rangle + g_i(\exp_x v) - g_i(x) \right\}. \tag{6}$$

易知ψ_x 是一个凸函数, 且$\psi_x(0) = 0$.

引理1. 设$x \in M$, 对任意的$v \in T_xM$, 下列等式成立:

$$\psi'_x(0; v) = \max_{i \in I} F'_i(x; v). \tag{7}$$

证明 因为$\psi_x(0) = 0$, 由方向导数的定义, 则有

$$\begin{aligned}
\psi'_x(0; v) &= \lim_{t \to 0^+} \frac{\psi_x(tv) - \psi(0)}{t} \\
&= \lim_{t \to 0^+} \max_{i \in I} \frac{\{\langle \operatorname{grad} f_i(x), tv \rangle + g_i(\exp_x tv) - g_i(x)\}}{t} \\
&= \max_{i \in I} \lim_{t \to 0^+} \frac{\{\langle \operatorname{grad} f_i(x), tv \rangle + g_i(\exp_x tv) - g_i(x)\}}{t} \\
&= \max_{i \in I} F'_i(x; v).
\end{aligned}$$

设$x \in M, l > 0$, ,对任意的$v \in T_xM$,定义函数$\phi_{l,x} : T_xM \to R$如下:

DOI: 10.12677/pm.2023.1312367

$$\phi_{l,x}(v) = \psi_x(v) + \frac{l}{2}\|v\|^2. \tag{8}$$

ψ_x由(6)定义, 可知$\phi_{l,x}$是强凸函数, 且$\phi_{l,x}(0) = 0$.

由以上所定义函数, 我们将定义每次邻近梯度算法的搜索方向$v_k = v_l(x_k)$ 如下:

$$v_l(x) = \underset{v \in T_x M}{\operatorname{argmin}} \phi_{l,x}(v). \tag{9}$$

设$\beta_l(x)$表示(9)的最优值, 即

$$\beta_l(x) = \min_{v \in T_x M} \phi_{l,x}(v) = \phi_{l,x}(v_l(x)). \tag{10}$$

注2. 1. 因为$\phi_{l,x}$是强凸函数, 则(8)有唯一解. 因此, $v_l(x_k)$ 是良定的.

2. 因为$\phi_{l,x}(0) = 0$, 有$\phi_{l,x}(v_l(x_k)) \leq 0$.

3. 由([25] Maximun Theorem)知映射$\beta_l(x)$是连续映射. 此外,由([26] Corollary 8.1), 且$v_l(x)$是唯一的, 所以是$v_l(x)$ 也是连续的.

当f_i是测地L_i-光滑函数, $L = \max_{i \in I} L_i$, g_i是测地凸函数时，给出如下的多目标邻近梯度算法, 下文称之为算法1.

算法1.

Step 1. 初始化; $l > L$, $x_0 \in M, k = 0$.

Step 2. 计算子问题(9): 令$x = x_k, v_k = v_l(x_k)$.

Step 3. 如果$v_k = 0$, 则停止.

Step 4. 令$x_{k+1} = \exp_{x_k} v_k, k = k + 1$, *Step 2.*

下面引理将表述$v_l(\cdot)$, $\beta_l(\cdot)$与问题(4) 中F 的Pareto 稳定点之间的关系.

引理2. 设$v_l(x)$, $\beta_l(x)$分别为(9)和(10) 所定义, x 是F的Pareto 稳定点的充分必要条件是$v_l(x) = 0$, $\beta_l(x) = 0$.

证明 假设x是F的Pareto稳定点, $v_l(x) \neq 0$, 或者$\beta_l(x) < 0$. 由注2 中前两点, $v_l(x) \neq 0$当且仅当$\beta_l(x) < 0$. 因此,

$$\beta_l(x) = \psi_x(v_l(x)) + \frac{l}{2}\|v_l(x)\|^2 < 0. \tag{11}$$

因为ψ_x是凸函数, 且$\psi_x(0) = 0$, 则有

$$\begin{aligned}
\psi_x(tv_l(x)) &= \psi_x(tv_l(x) + (1-t)\psi_x(0)) \\
&= t\psi_x(v_l(x)) + (1-t)\psi_x(0) \\
&= t\psi_x(v_l(x)) \\
&< -\frac{tl}{2}\|v_l(x)\|^2 \qquad \forall t \in (0,1),
\end{aligned}$$

最后的不等式由(11)得. 因此, 对任意的$t \in (0, 1)$, 有

$$\frac{\psi_x(tv_l(x))}{t} \le -\frac{tl}{2}\|v_l(x)\|^2.$$

因为$v_l(x) \neq 0, l > 0$, 结合(7),有

$$\psi'_x(0; v_l(x)) = \max_{i\in I} F'_i(x; v_l(x)) \le -\frac{l}{2}\left\|v_l(x)^2\right\| < 0,$$

与x是F的Pareto稳定点矛盾.

假设$v_l(x) = 0, \beta_l(x) = 0$, x不是F的Pareto稳定点. 由$\beta_l(x)$的定义, 有

$$\phi_{l,x}(v) = \psi_x(v) + \frac{l}{2}\left\|v_l(x)^2\right\| \ge \beta_l(x) = 0 \quad v \in T_xM.$$

取$t \in (0, 1)$, 有

$$\frac{\psi_x(tv) + \frac{l}{2}\|tv)\|^2}{t} \ge 0 \quad v \in T_xM.$$

取$t \to 0$, 对上式取极限, 结合(7), 有

$$\max_{i\in I} F'_i(x; v) \ge 0,$$

与x不是F的Pareto稳定点矛盾. 综上得证.

下面, 我们将给出f_i是测地L_i-光滑函数, $L = \max_{i\in I} L_i$, 迭代步长为1时的邻近梯度算法, 并证明算法产生的序列聚点存在时, 为Pareto稳定点.

为证明算法的收敛性，首先证明下面的引理.

引理3. 设$\{v_k\}$是由度算法1产生的序列, 序列$\{F_i(x_k)\}_{i\in I}$有下界, 则有

$$\lim_{k\to\infty} \|v_k\| = 0,$$

证明 因为f_i是测地L_i-光滑函数, $L = \max_{i\in I} L_i$, 所以

$$f_i(x_{k+1}) \le f_i(x_k) + \langle \mathrm{grad} f_i(x_k), v_k\rangle + \frac{L}{2}\|v_k\|^2,$$

其中$v_k \in T_{x_k}M, x_{k+1} = \exp_{x_k} v_k$. 进而

$$\begin{aligned}
f_i(x_{k+1}) + g_i(x_{k+1}) &= f_i(x_k) + g_i(x_k) + f_i(x_{k+1}) - f_i(x_k) + g_i(x_{k+1}) - g_i(x_k)\\
&\le f_i(x_k) + g_i(x_k) + \langle \mathrm{grad} f_i(x_k), v_k\rangle + g_i(x_{k+1}) - g_i(x_k) + \frac{L}{2}\|v_k\|^2\\
&\le f_i(x_k) + g_i(x_k) + \psi_{x_k}(v_k) + \frac{L}{2}\|v_k\|^2.
\end{aligned}$$

由(8)知$\phi_{l,x}(v) = \psi_x(v) + \frac{l}{2}\|v\|^2$是强凸函数, 且$\phi_{l,x}(0) = 0$, 有$\phi_{l,x}(v_l(x_k)) \le 0$, 所以

$$\psi_{x_k}(v_k) \le -\frac{l}{2}\|v_k\|^2,$$

进而有

$$f_i(x_{k+1}) + g_i(x_{k+1}) \le f_i(x_k) + g_i(x_k) + \frac{L-l}{2}\left\|v_k\right\|^2.$$

又因为$\{F_i(x_k)\}_{i\in I}$有下界, 则存在$(\hat{F}_i)_{i\in I} \in R^m$, 使得

$$\hat{F}_i \le F_i(x_k) = f_i(x_k) + g_i(x_k),$$

将上式从$k = 0$ 到$k = \hat{k}$ 做累加, 得

$$f_i(x_{\hat{k}+1}) + g_i(x_{\hat{k}+1}) \le f_i(x_0) + g_i(x_0) + \frac{L-l}{2}\sum_{k=0}^{\hat{k}}\left\|v_k\right\|^2.$$

又因为$l > L$, 则有

$$\sum_{k=0}^{\hat{k}}\left\|v_k\right\|^2 \le \frac{2}{l-L}(f_i(x_0) + g_i(x_0) - \hat{F}_i).$$

令$\hat{k} \to \infty$, 可得

$$\sum_{k=0}^{\hat{k}}\left\|v_k\right\|^2 < \infty.$$

可得

$$\lim_{k\to\infty}\|v_k\| = 0.$$

定理1. 如果x^*是算法产生的序列$\{x_k\}$的聚点, 则x^*是F 的Pareto 稳定点. 此外, 如果$\{F_i(x_k)\}_{i\in I}$有下界, 则算法产生的序列$\{x_k\}$的聚点存在, 且都是F 的Pareto稳定点.

证明 设x^*是$\{x_k\}$的聚点, 则存在$\{x_k\}$的子列$\{x_{k_j}\}$收敛到x^*, 由引理2 可知有$v_{k_j} = v_l(x_{k_j}) \to v_l(x^*) = 0$, 即$x^*$是$F$的Pareto 稳定点. 由引理4 可知, 当$\{F_i(x_k)\}_{i\in I}$有下界, 则有$\lim_{k\to 0}\|v_k\| = 0$. 又由引理2知$v_k$ 对应的x^*是F的Pareto稳定点.

3.2. 算法收敛速度分析

本节将介绍在算法1中, 用F的价值函数, 当目标函数满足$Polyak - Lojasiewicz$不等式下得到邻近梯度算法的收敛速度.

定义9 [27, Simple merit function] 定义问题(4)的简单价值函数$u_0 : M \to R$ 为

$$u_0(x) = \sup_{v\in T_xM}\min_{i\in I}\{F_i(x) - F_i(\exp_x v)\}, \forall x \in M. \tag{12}$$

定义10 [27, Regularized merit function] 设$l > 0$, 定义问题(4)的正则化价值函数$u_l : M \to R$ 为

$$u_l(x) = \max_{v\in T_xM}\min_{i\in I} - \phi_{l,x}(v), \forall x \in M. \tag{13}$$

由(8)知

$$u_l(x) = \max_{v \in T_x M} \min_{i \in I} \left\{ \langle \operatorname{grad} f_i(x), -v \rangle + g_i(x) - g_i(\exp_x v) - \frac{l}{2} \|v\|^2 \right\}, \forall x \in M. \tag{14}$$

注3. u_0中的$(F_i)_{i\in I}$和u_l中的$(f_i)_{i\in I}$, $(g_i)_{i\in I}$是问题(4)中对应的函数.

定理2. [27, Theorem3.1, Theorem3.3,Theorem3.2] 设$x \in M$, $v \in T_x M$, u_0, u_l分别为简单, 正则化价值函数, 则下列说法成立:

(i). 对任意的$x \in M$, $u_0(x) \geq 0$. x是F的弱Pareto最优点, 当且仅当$u_0(x) = 0$.

(ii). 对任意的$x \in M$, $u_l(x) \geq 0$. x是F的Pareto稳定点, 当且仅当$u_l(x) = 0$.

(iii). 对任意的$x \in M$, $l \geq L > 0$, 则$u_l(x) \leq u_L(x) \leq \frac{l}{L} u_l(x)$

注4. 定理2是欧氏空间中价值函数 [27, Theorem3.1, Theorem3.3,Theorem3.2]的直接推广.

定义11 设对任意的$x \in M$, $v \in T_x M$, $c > 0$, 如果不等式

$$L u_L(x) \geq c u_0(x). \tag{15}$$

成立, 则称问题(4)满足$Polyak-Lojasiewicz$不等式. 当问题(4)为单目标函数, 且满足Polyak-Lojasiewicz不等式时, 即当$m = 1, g_1 = 0$, $c > 0$成立

$$\frac{1}{2} \|\operatorname{grad} f_1(x)\|^2 \geq c(f_1(x) - f_1^*), \forall x \in M, \tag{16}$$

其中f_1^*是f_1的最优值.

定理3. 如果问题(4)满足$Polyak-Lojasiewicz$不等式, 则算法1产生的序列$\{v_k\}$使得$\{u_0(x_k)\}$线性收敛, 即

$$u_0(x_{k+1}) \leq (1 - \frac{c}{l}) u_0(x_k).$$

证明 因为f_i是测地L_i-光滑函数, $L = \max_{i \in I} L_i$, 有

$$F_i(x_{k+1}) - F_i(x_k) \leq \left\langle f_i(x_k), \exp_{x_k}^{-1} x_{k+1} \right\rangle + g_(x_{k+1}) - g_i(x_k) + \frac{l}{2} \left\| \exp_{x_k}^{-1} x_{k+1} \right\|^2,$$

令$x_{k+1} = \exp_{x_k} v_k$, $v_k \in T_{x_k} M$, 则有

$$\begin{aligned} F_i(x_{k+1}) - F_i(x_k) &\leq \left\langle \operatorname{grad} f_i(x_k), \exp_{x_k}^{-1} x_{k+1} \right\rangle + g_(x_{k+1}) - g_i(x_k) + \frac{l}{2} \left\| \exp_{x_k}^{-1} x_{k+1} \right\|^2 \\ &= \langle f_i(x_k), v_k \rangle + g(\exp_{x_k} v_k) - g_i(x_k) + \frac{l}{2} \|v_k\|^2 \\ &\leq \max_{v_k \in T_{x_k} M} \left\{ \langle \operatorname{grad} f_i(x_k), v_k \rangle + g(\exp_{x_k} v_k) - g_i(x_k) + \frac{l}{2} \|v_k\|^2 \right\} \\ &= -u_l(x_k), \end{aligned}$$

由定理2中的(iii), 得$L u_l(x) \leq L u_L(x) \leq l u_l(x)$, 又因为$L u_L(x) \geq c u_0(x)$, 所以

$$F_i(x_{k+1}) - F_i(x_k) \leq -u_l(v_k) \leq -\frac{c}{l} u_0(x_k).$$

则对任意的$x \in M$, 有

$$F_i(x_{k+1}) - F_i(x) \leq F_i(x_k) - F_i(x) - \frac{c}{l}u_0(x_k),$$

进而

$$\sup_{v_k \in T_{x_k}M} \min_{i \in I} F_i(x_{k+1}) - F_i(x) \leq \sup_{v_k \in T_{x_k}M} \min_{i \in I} F_i(x_{k+1}) - F_i(x) - \frac{c}{l}u_0(x_k).$$

由u_0的定义, 所以有

$$u_0(x_{k+1}) \leq (1 - \frac{c}{l})u_0(x_k).$$

4. 总结

本文将多目标邻近梯度算法推广到了Hadamard流形上, 证明了算法产生序列$\{x_k\}$ 的聚点是F的Pareto稳定点. 此外, 在算法1中, 当目标函数满足$Polyak - Lojasiewicz$不等式时, 利用价值函数证明了算法的收敛速度是线性的. 这个结果在Hadamard流形上是新的.

基金项目

国家自然科学基金(12161017)和贵州省科技计划项目(ZK[2022]110)。

参考文献

[1] Adler, R.L., Dedieu, J.P., Margulies, J.Y., *et al.* (2002) Newton's Method on Riemannian Manifolds and a Geometric Model for the Human Spine. *IMA Journal of Numerical Analysis*, **22**, 359-390. https://doi.org/10.1093/imanum/22.3.359

[2] Li, C. and Wang, J. (2006) Newton's Method on Riemannian Manifolds: Smale's Point Estimate Theory under the γ-Condition. *IMA Journal of Numerical Analysis*, **26**, 228-251. https://doi.org/10.1093/imanum/dri039

[3] Gabay, D. (1982) Minimizing a Differentiable Function over a Differential Manifold. *Journal of Optimization Theory and Applications*, **37**, 177-219. https://doi.org/10.1007/BF00934767

[4] Wang, J.H., Li, C., Lopez, G. and Yao, J.C. (2015) Convergence Analysis of Inexact Proximal Point Algorithms on Hadamard Manifolds. *Journal of Global Optimization*, **61**, 553-573. https://doi.org/10.1007/s10898-014-0182-2

[5] Wang, X.M. (2018) Subgradient Algorithms on Riemannian Manifolds of Lower Bounded Curvatures. *Optimization*, **67**, 179-194. https://doi.org/10.1080/02331934.2017.1387548

[6] Absil, P.A., Baker, C.G. and Gallivan, K.A. (2007) Trust-Region Methods on Riemannian Manifolds. *Foundations of Computational Mathematics*, **7**, 303-330. https://doi.org/10.1007/s10208-005-0179-9

[7] Huang, W. and Wei, K. (2022) Riemannian Proximal Gradient Methods. *Mathematical Programming*, **194**, 371-413. https://doi.org/10.1007/s10107-021-01632-3

[8] Jolliffe, I.T., Trendafilov, N.T. and Uddin, M. (2003) A Modified Principal Component Technique Based on the LASSO. *Journal of Computational and Graphical Statistics*, **12**, 531-547. https://doi.org/10.1198/1061860032148

[9] Genicot, M., Huang, W. and Trendafilov, N.T. (2015) Weakly Correlated Sparse Components with Nearly Orthonormal Loadings. *Geometric Science of Information: Second International Conference*, Palaiseau, 28-30 October 2015, 484-490. https://doi.org/10.1007/978-3-319-25040-3_52

[10] Zhang, Y., Lau, Y., Kuo, H., *et al.* (2017) On the Global Geometry of Sphere-Constrained Sparse Blind Deconvolution. *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 21-26 July 2017, 4894-4902.

[11] Tang, J. and Liu, H. (2012) Unsupervised Feature Selection for Linked Social Media Data. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, Beijing, 12-16 August 2012, 904-912.

[12] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, **2**, 183-202. https://doi.org/10.1137/080716542

[13] Huang, W. (2013) Optimization Algorithms on Riemannian Manifolds with Applications. The Florida State University, Tallahassee, FL.

[14] Hosseini, S. and Uschmajew, A. (2017) A Riemannian Gradient Sampling Algorithm for Nonsmooth Optimization on Manifolds. *SIAM Journal on Optimization*, **27**, 173-189. https://doi.org/10.1137/16M1069298

[15] Hosseini, S., Huang, W. and Yousefpour, R. (2018) Line Search Algorithms for Locally Lipschitz Functions on Riemannian Manifolds. *SIAM Journal on Optimization*, **28**, 596-619. https://doi.org/10.1137/16M1108145

[16] Liu, Y., Shang, F., Cheng, J., *et al.* (2017) Accelerated First-Order Methods for Geodesically Convex Optimization on Riemannian Manifolds. *Advances in Neural Information Processing Systems*, **30**.

[17] Chen, S., Ma, S., So, A.M.-C., *et al.* (2020) Proximal Gradient Method for Nonsmooth Optimization over the Stiefel Manifold. *SIAM Journal on Optimization*, **30**, 210-239. https://doi.org/10.1137/18M122457X

[18] Huang, W. and Wei, K. (2019) An Extension of Fast Iterative Shrinkage-thresholding to Riemannian Optimization for Sparse Principal Component Analysis. https://doi.org/10.48550/arXiv.1909.05485

[19] Huang, W. and Wei, K. (2022) Riemannian Proximal Gradient Methods. *Mathematical Programming*, **194**, 371-413. https://doi.org/10.1007/s10107-021-01632-3

[20] Tanabe, H., Fukuda, E.H. and Yamashita, N. (2019) Proximal Gradient Methods for Multiobjective Optimization and Their Applications. *Computational Optimization and Applications*, **72**, 339-361. https://doi.org/10.1007/s10589-018-0043-x

[21] Tanabe, H., Fukuda, E.H. and Yamashita, N. (2023) Convergence Rates Analysis of a Multiobjective Proximal Gradient Method. *Optimization Letters*, **17**, 333-350. https://doi.org/10.1007/s11590-022-01877-7

[22] Do Carmo, M.P. and Flaherty Francis, J. (1992) Riemannian Geometry. Birkhäuser.

[23] 陈维桓, 李兴校. 黎曼几何引论[M]. 北京: 北京大学出版社, 2002: 12.

[24] Boothby, W.M. (1986) An Introduction to Differentiable Manifolds and Riemannian Geometry. Academic Press, Cambridge, MA.

[25] Padlewska, B. and Darmochwal, A. (1990) Topological Spaces and Continuous Functions. *Formalized Mathematics*, **1**, 223-230.

[26] Hogan, W.W. (1973) Point-to-Set Maps in Mathematical Programming. *SIAM Review*, **15**, 591-603. https://doi.org/10.1137/1015073

[27] Tanabe, H., Fukuda, E.H. and Yamashita, N. (2023) New Merit Functions for Multiobjective Optimization and Their Properties. *Optimization*, 1-38. https://doi.org/10.1080/02331934.2023.2232794