Advances in Applied Mathematics 应用数学进展, 2022, 11(3), 1107-1115
Published Online March 2022 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2022.113119

Stiefel流形上非光滑优化的一种带外推的可变度量邻近梯度算法

张金超

河北工业大学，理学院，天津

收稿日期：2022年2月18日；录用日期：2022年3月14日；发布日期：2022年3月21日

摘　要

本文针对Stiefel流形上一类目标函数为光滑损失函数与非光滑函数之和的非凸优化问题，提出了一种基于收缩的可变度量惯性邻近梯度算法。所提出的算法在已有的加速黎曼邻近梯度算法基础上，引入了对角Barzilai-Borwein类步长策略，该策略可以更好的捕获问题的局部几何信息，进一步加速算法的收敛。理论上，证明了算法全局收敛到稳定点。最后，本文给出了稀疏主成分分析问题的数值结果，验证了该方法的有效性。

关键词

非凸非光滑优化，变尺度，惯性邻近梯度算法，Stiefel流形

A Variable Metric Proximal Gradient Method with Extrapolation for Nonsmooth Optimization over the Stiefel Manifold

Jinchao Zhang

School of Science, Hebei University of Technology, Tianjin

Received: Feb. 18th, 2022; accepted: Mar. 14th, 2022; published: Mar. 21st, 2022

Abstract

In this paper, we propose a retraction based variable metric inertial proximal gradient method for solving a class nonconvex optimization problem over the Stiefel manifold whose objective function is the summation of a smooth cost function and a nonsmooth function. Based on existing iner-

文章引用: 张金超. Stiefel 流形上非光滑优化的一种带外推的可变度量邻近梯度算法[J]. 应用数学进展, 2022, 11(3): 1107-1115. DOI: 10.12677/aam.2022.113119

tial Riemannian proximal gradient method, the proposed method introduces a metric changing called diagonal Barzilai-Borwein step-size strategy at each iteration, which can better capture the local geometric of this class problem and accelerate the convergence of the algorithm. Theoretically, we show that the proposed method globally converges to a stationary point. Numerical results on solving sparse PCA problem is reported to demonstrate the efficiency of the proposed method.

Keywords

Nonconvex Nonsmooth Optimization, Variable Metric, Inertial Proximal Gradient Method, Stiefel Manifold

1. 引言

本文主要考虑以下具有 Stiefel 流形约束的一类非凸非光滑优化问题：

$$\begin{aligned} &\min \ F(x) := f(x) + g(x) \\ &\text{s.t.} \quad x \in \mathcal{M} \end{aligned} \tag{1}$$

其中函数 $f : \mathcal{M} \to \mathbb{R}$ 是可微的但可能是非凸的，函数 $g : \mathcal{M} \to \mathbb{R}$ 是连续的但也可能是非光滑的。可行集 $\mathcal{M} := \mathcal{S}_{n,p} = \left\{ X : X \in \mathbb{R}^{n\times p}, X^{\mathrm{T}} X = I_p \right\}$ 称为 Stiefel 流形，也称为正交约束，其中 I_p 记为 p 阶单位阵($p \le n$)。

Stiefel 流形上的非光滑优化问题由于其在各个领域的广泛应用引起了许多研究者的注意。例如，在统计与数据科学中的稀疏主成分分析问题[1]；物理中的压缩模问题[2]；图像领域中的盲去卷积问题[3]以及机器学习中的无监督特征选取[4]等问题，其本质都是 Stiefel 流形上的非光滑优化问题。更多的相关应用，我们推荐读者参考文献[5]。

通常情况下，由于其目标函数的不可微性和流形约束的非凸性，问题(1)是较难求解的。对于部分问题已经证明是 NP-难的[6]。求解问题(1)的数值算法相较于目标函数是光滑形式的算法是有限的，目前已有的算法包括次梯度类算法[7] [8]、算子分裂算法[9] [10]以及邻近梯度算法[11] [12]。然而，上述所提到的算法存在着寻找下降方向成本高、缺乏收敛性分析、对参数敏感等问题。众所周知，对于目标函数为光滑损失函数与非光滑连续函数之和的复合结构，当其对应的邻近算子容易求解时，邻近梯度算法是求解该类问题的最有效方法之一。然而，由于问题(1)中 Stiefel 流形的约束，此时邻近算子是没有显式解的。针对这一问题，最近 Chen 等[13]提出了一种基于收缩的邻近梯度算法，其中下降方向是由限制在 Stiefel 流形切空间中的邻近子问题决定。尽管该子问题没有显示解，作者通过将其转换为一个非线性方程系统，进而采用正则的半光滑牛顿法[14]高效求解。随后，Huang 等[15]将快速迭代收缩阈值算法(FISTA)从欧氏空间推广到黎曼空间中。进一步，Huang 等[16]在 2021 年针对一般的黎曼流形研究了邻近梯度算法及其 Nesterov 加速版本，作者进一步借助 Kurdyka-Łojasiewicz 不等式性质分析了所提出算法的局部收敛率。虽然上述文献中均假设目标函数中光滑部分是梯度 Lipschitz 连续的，然而在实际问题中，其 Lipschitz 常数的精确计算并不是容易的。

受文献[17] [18]的启发，本文基于上述研究成果，提出了一种可变度量的惯性邻近梯度算法求解问题(1)，所提出的算法采用对角化的 Barzilai-Borwein (BB)步长估计 Lipschitz 常数，同时结合 Nseterov 动量

项进一步加速算法收敛，并给出收敛性分析。最后，在稀疏主成分分析问题中的数值结果验证了我们的方法是有效的。

2. 预备知识

本节引入一些流形优化的基本定义和概念。

定义 1 (收缩映射) 一个光滑映射 $\mathrm{Retr}_X : T_X\mathcal{M} \to \mathcal{M}$ 满足

1) $\mathrm{Retr}_X(0) = X, \forall X \in \mathcal{M}$，其中 0 定义为 $T_X\mathcal{M}$ 的零元素；

2) 对于任意的 $X \in \mathcal{M}$，有如下极限成立：

$$\lim_{0 \leftarrow \xi \in T_X\mathcal{M}} \frac{\left\| \mathrm{Retr}_X(\xi) - (X+\xi) \right\|_F}{\|\xi\|_F} = 0,$$

则 Retr_X 被称为收缩映射。

定义 2 (广义 Clarke 次微分)对于一个定义在 $\mathcal{M}$ 上的局部 Lipschitz 函数 F，则 F 在 $X \in \mathcal{M}$ 处方向 V 上的黎曼方向导数定义为：

$$F^{\circ}(X,V) = \lim_{Y \to X, t \downarrow 0} \sup \frac{F \circ \phi^{-1}\left(\phi(Y) + tD\phi(X)[V]\right) - F\phi^{-1}\left(\phi(Y)\right)}{t},$$

其中 (ϕ, U) 是在 X 处的坐标卡。则函数 F 在 $X \in \mathcal{M}$ 处的广义梯度或者 Clarke 次微分表示为 $\hat{\partial} F(X) = \left\{ \xi \in T_X\mathcal{M} : \langle \xi, V \rangle \le F^{\circ}(X,V)\ \forall V \in T_X\mathcal{M} \right\}$

由于 F 是正则函数，本文有 $\hat{\partial} F(X) = \mathrm{Proj}_{T_X\mathcal{M}}\left(\partial F(X)\right)$，其中 $\mathrm{Proj}_{T_X\mathcal{M}}$ 是正交投影。进一步我们可以得出问题(1)的一阶最优性条件，即 $0 \in grad\ f(X) + \mathrm{Proj}_{T_X\mathcal{M}}\left(\partial g(X)\right)$，其中 $\partial(\bullet)$ 表示为欧式空间中的次微分。

3. 可变度量的惯性邻近梯度算法

本节中，我们提出 Stiefel 流形约束优化问题可变度量的惯性邻近梯度算法。

为了满足约束，本文的算法通过更新迭代

$$X_{k+1} = \mathrm{Retr}_{X_k}\left(\alpha_k d_k\right) \tag{2}$$

来保持可行性，其中 α_k 是由 Armijo 条件决定的，d_k 是搜索方向。如同文献[13]中所示，搜索方向 d_k 是由如下限制在切空间中的邻近子问题所决定

$$d_k = \arg\min_{d \in T_{X_k}\mathcal{M}} \left\langle grad\ f(X_k), d \right\rangle + \frac{L}{2}\|d\|_F^2 + g(X_k + d), \tag{3}$$

其中 L 为可微函数 f 的梯度 Lipschitz 常数。

具体地，本文类似于无约束优化中的情况，首先给出如下的长和短 BB 步长：

$$\alpha_{k+1}^{BB1} = \arg\min_{\alpha \in \mathbb{R}} \left\| \alpha^{-1} S_k - Y_k \right\|_F = \frac{tr\left(S_k^{\mathrm{T}} S_k\right)}{\left| tr\left(S_k^{\mathrm{T}} Y_k\right) \right|},$$

和

$$\alpha_{k+1}^{BB2} = \arg\min_{\alpha \in \mathbb{R}} \left\| S_k - \alpha Y_k \right\|_F = \frac{tr\left(S_k^{\mathrm{T}} Y_k\right)}{\left| tr\left(Y_k^{\mathrm{T}} Y_k\right) \right|}$$

其中 $S_k = X_{k+1} - X_k$，$Y_k = d_{k+1} - d_k$。这里下降方向 d_{k+1} 和 d_k 分别是限制在不同切空间中邻近子问题的解。为了更好的获取函数 f 的 Hessian 信息，采用如下在第 k 次迭代中计算得出的度量 $U^k = Diag\left(u^k\right)$，$u^k = \left[u_1^k, \cdots, u_n^k\right] \in \mathbb{R}^n$，其中

$$u_i^k = \begin{cases} \dfrac{1}{\alpha_k^{BB1}} & \dfrac{S_k^i Y_k^i + \upsilon u_i^{k-1}}{\left(S_k^i\right)^2 + \upsilon} < \dfrac{1}{\alpha_k^{BB1}} \\ \dfrac{1}{\alpha_k^{BB2}} & \dfrac{S_k^i Y_k^i + \upsilon u_i^{k-1}}{\left(S_k^i\right)^2 + \upsilon} > \dfrac{1}{\alpha_k^{BB2}} \\ \dfrac{S_k^i Y_k^i + \upsilon u_i^{k-1}}{\left(S_k^i\right)^2 + \upsilon} & \text{否则} \end{cases} \tag{4}$$

此时，邻近子问题(3)就变为

$$d_k = \underset{d \in T_{X_k}\mathcal{M}}{\arg\min} \left\langle grad\ f\left(X_k\right), d\right\rangle + \frac{L}{2}\left\langle d, U^k d\right\rangle + g\left(X_k + d\right) \tag{5}$$

该子问题仍采用半光滑牛顿法求解。接下来，我们给出求解问题(1)的可变度量惯性邻近梯度算法。

算法 1. 可变度量惯性邻近梯度法

VM-AManPG 算法：

步 0：初始化 $X_0 \in \mathcal{M}$，线搜索参数 δ、$\sigma \in (0,1)$，超参数 $\upsilon > 0$，保护步中正整数 M，$t_0 = 1$。

步 1：$Y_0 = X_0$，$Z_0 = X_0$；

步 2：**for** $k = 0, \cdots$ **do**

步 3： **if** mod(k, M)=0 **then**

步 4：   调用**算法 2：** $\left[Z_{k+M}, X_k, Y_k, t_k\right] = \mathrm{Alg}\,2\left(Z_{k+M}, X_k, Y_k, t_k, F\left(X_k\right)\right)$；

步 5： **end if**

步 6：通过半光滑牛顿法求解子问题(5)得出下降方向 d_k；

步 7：步通过式(2)更新 X_{k+1}；

步 8：$t_{k+1} = \dfrac{\sqrt{4t_k^2 + 1} + 1}{2}$；

步 9：计算 $Y_{k+1} = \mathrm{Retr}_{X_{k+1}}\left(\dfrac{1 - t_k}{t_{k+1}} \mathrm{Retr}_{X_{k+1}}^{-1}\left(X_k\right)\right)$；

步 10：**end for**

注 1. 其中 VM-AManPG 算法步骤 9 中的收缩映射 $\mathrm{Retr}_{(\cdot)}$ 本文考虑使用极分解，$\mathrm{Retr}_X^{polar}(\xi) = (X + \xi)\left(I_p + \xi^{\mathrm{T}}\xi\right)^{-1/2}$，且该收缩的逆是存在的。众所周知，惯性邻近梯度法是非单调的，基于文献[15]，本文同样采用重启策略使得算法 VM-AManPG 是满足下降性的。

算法 2. 算法 1.的保护步

VM-AManPG 算法：

步 0：输入 $\left(Z_k, X_k, Y_k, t_k, F\left(X_k\right)\right)$；

步 1：计算子问题(5)得到下降方向 d_{Z_k}；

步 2：令 $\alpha = 1$；

步 3：　　**while** $F\left(\mathrm{Retr}_{Z_k}\left(\alpha d_{Z_k}\right)\right) > F\left(Z_k\right) - \sigma\alpha\left\|d_{Z_k}\right\|_F^2$ **do**

步 4：　　　$\alpha = \delta\alpha$；

步 5：　　**end while**

步 6：若 $F\left(\mathrm{Retr}_{Z_k}\left(\alpha d_{Z_k}\right)\right) < F\left(X_k\right)$ 则

步 7：　$X_k = \mathrm{Retr}_{Z_k}\left(\alpha d_{Z_k}\right)$；$Y_k = \mathrm{Retr}_{Z_k}\left(\alpha d_{Z_k}\right)$，$t_k = 1$；

步 8：否则 X_k，Y_k，t_k 若保持不变；

步 9：输出 $Z_{k+M} = X_k$。

4. 收敛性分析

本节中分析了 **VM-AManPG** 算法的收敛性质。在正式给出结论之前，本文需要做出如下的假设。

假设 1. 函数 f 是可微的，且其梯度 ∇f 是 Lipschitz 连续的有常数 L。函数 g 是凸非光滑的，并且是 Lipschitz 连续的。

假设 2. 函数 F 是强制的，即当 $\|X\|_F \to \infty$ 时 $F(X) \to +\infty$。

假设 3. 存在两个正常数 $0 < \gamma < \overline{\gamma}$，使得对角矩阵 U 在任意点 X_k 处的特征值满足在 $\gamma, \overline{\gamma}$ 范围内。

引理 1： 序列 $\{Z_k\},\{d_k\}$ 分别是由算法 VM-AManPG 生成的迭代序列，当**假设 1** 成立时，存在常数 $\overline{\alpha}$ 使得对任意的 $0 < \alpha \le \min(1, \overline{\alpha})$，满足

$$F\left(\mathrm{Retr}_{Z_k}\left(\alpha d_k\right)\right) - F\left(Z_k\right) \le -\frac{\alpha L}{2}\left\|d_k\right\|_F^2$$

上述引理说明算法 1 是有定义的。事实上，算法 1 中序列 $\{Z_k\}$ 的下标为 $k+M$。当本文中所采用的对角 BB 步长矩阵 U^k 为单位阵时，子问题(5)退化为文献[13]中所求解的子问题，此时，文献[13]中的结论在本文中都成立。进一步**假设 3** 成立的条件下，无论 U^k 是否为单位阵，上述引理都是成立的。

定理 1： 在**假设 1**、**2** 和**假设 3** 成立的条件下，序列 $\{Z_k\}$ 是由**算法 1** 产生的迭代序列。令 Z_* 是序列 $\{Z_k\}$ 的任意聚点，则有

$$0 \in \mathrm{Proj}_{T_Z\mathcal{M}}\, \partial F\left(Z_*\right)$$

即序列 $\{Z_k\}$ 的任意聚点是问题(1)的一个稳定点。

证明：在**假设 1** 和**假设 2** 条件下，由于 Stiefel 流形是紧集，则次水平集 $\Omega_{X_0} = \left\{X \in \mathcal{M} \middle| F(X) \le F(X_0)\right\}$ 是紧的，由于假设函数 F 是强制的，所以次水平集 Ω_{X_0} 是有界的。进一步由于 F 是连续函数，则 F 在次水平集 Ω_{X_0} 是有界的。

当下降方向 $d_k = 0$ 时，根据子问题(5)的最优性条件有

$$0 \in \mathit{grad}\, f\left(Z_k\right) + \frac{1}{t}U^k d_k + \mathrm{Proj}_{T_Z\mathcal{M}}\, \partial g\left(Z_k + d_k\right) \tag{6}$$

成立，这恰好是原问题(1)的一阶必要性条件。接下来我们就说明序列 $\{d_k\}$ 是满足 $\lim_{k\to\infty}\|d_k\|_F = 0$。

根据算法 2 中步骤 3 的 Armijo 条件，有 $F(Z_{k+1}) < F(Z_k)$，即函数序列 $\{F(Z_k)\}$ 是单调下降的，进一步，由于函数 F 是有下界的，则序列 $\{F(Z_k)\}$ 的极限存在，有

$$\lim_{k\to\infty} F\left(Z_k\right) - F\left(\mathrm{Retr}_{Z_k}\left(\alpha d_k\right)\right) = 0$$

结合**引理 1**，则有 $\lim_{k\to\infty}\|d_k\|_F=0$ 成立。

由子问题(5)的最优性条件(6)，在式(6)两边分别加上 $grad\ f\left(Z_k+d_k\right)$，可得

$$\begin{aligned}&grad\ f\left(Z_k+d_k\right)-grad\ f\left(Z_k\right)-\frac{1}{t}U^k d_k\\&\in \mathrm{Proj}_{T_Z\mathcal{M}}\,\partial g\left(Z_k+d_k\right)+grad\ f\left(Z_k+d_k\right)=\mathrm{Proj}_{T_Z\mathcal{M}}\,\partial F\left(Z_k+d_k\right)\end{aligned}\tag{7}$$

且有

$$\begin{aligned}&\left\|grad\ f\left(Z_k\right)-grad\ f\left(Z_k+d_k\right)+\frac{1}{t}U^k d_k\right\|_F\\&=\left\|\mathrm{Proj}_{T_Z\mathcal{M}}\,\nabla f\left(Z_k\right)-\mathrm{Proj}_{T_Z\mathcal{M}}\,\nabla f\left(Z_k+d_k\right)+\frac{1}{t}U^k d_k\right\|_F\\&\le\left(L+\frac{1}{t}\right)\|d_k\|_{U^k}\to 0\end{aligned}\tag{8}$$

当 $k\to\infty$ 时，其中 $\mathrm{Proj}_{T_Z\mathcal{M}}(\bullet)$ 为光滑映射，满足范数不等式。

令 $\left\{Z_{k_i}\right\}$ 为 $\left\{Z_k\right\}$ 的子序列且收敛到 Z_*。接下来我们证明 $0\in\mathrm{Proj}_{T_Z\mathcal{M}}\,\partial F\left(Z_*\right)$。根据式(7)，存在序列 $\left\{\tau_k\right\}\in N_{Z_k}\mathcal{M}$，其中 $N_{Z_k}\mathcal{M}$ 为流形 $\mathcal{M}$ 在点 Z_k 处的法空间，于是有

$$grad\ f\left(Z_k+d_k\right)-grad\ f\left(Z_k\right)-\frac{1}{t}U^k d_k+\tau_k\in\partial F\left(Z_k+d_k\right)$$

进一步结合当 $i\to\infty$ 时，$\left\{Z_{k_i}\right\}\to Z_*$，则有

$$grad\ f\left(Z_{k_i}+d_{k_i}\right)-grad\ f\left(Z_{k_i}\right)-\frac{1}{t}U^{k_i}d_{k_i}+\tau_{k_i}\in\partial F\left(Z_{k_i}+d_{k_i}\right)$$

由于目标函数 F 在紧集 Ω_{X_0} 上是连续的，则存在常数 $G>0$ 使得 $\max_{Z\in\Omega_{X_0}}\max_{\omega\in\partial F(Z)}\|\omega\|_F<G$，于是有 $\left\|\tau_{k_i}\right\|_F<G$，存在收敛子列 $\left\{\tau_{k_{i_j}}\right\}$ 且极限点为 τ_*。则当 $j\to\infty$ 时，有

$$grad\ f\left(Z_{k_{i_j}}+d_{k_{i_j}}\right)-grad\ f\left(Z_{k_{i_j}}\right)-\frac{1}{t}U^{k_{i_j}}d_{k_{i_j}}+\tau_{k_{i_j}}\to\tau_*$$

以及 $Z_{k_{i_j}}+d_{k_{i_j}}\to Z_*$。因此有 $0\in\partial F\left(Z_*\right)$，进一步根据投影 $\mathrm{Proj}_{N_Z\mathcal{M}}$ 是光滑的，有

$$\mathrm{Proj}_{N_{Z_{k_{i_j}}}\mathcal{M}}\,\tau_{k_{i_j}}\to\mathrm{Proj}_{N_{Z_*}\mathcal{M}}\,\tau_*$$

所以，我们可以得出 $0\in\mathrm{Proj}_{T_{Z_*}}\,\partial F\left(Z_*\right)$。

5. 数值实验

本节中，我们将在数据分析领域中的稀疏主成分分析问题上验证算法的有效性。所有的数值实验均是通过运行环境为 64bit Ubuntu platform，CPU (Intel Core i5-5200U)为 2.20GHz 的 Matlab R2019a 实现。

稀疏主成分分析(SPCA)已经成为一种强大的数据分析技术，通过识别数据中的本地空间结构和消除不同时间尺度之间的歧义，提供对低秩结构的改进描述。该问题在文献[1]中经过离散化，可以转化为如下的模型：

$$\begin{aligned}&\min -\mathrm{Tr}\left(X^{\mathrm{T}}A^{\mathrm{T}}AX\right)+\mu\|X\|_1\\&\text{s.t.}\quad X\in S_{n,p},\end{aligned}$$

其中矩阵 $A\in\mathbb{R}^{m\times n}$，$\|X\|_1=\sum_{ij}|X_{ij}|$，$\mu>0$ 是正则化参数，可以控制上述模型的稀疏性。

本文通过与最新的算法 ManPG-Ad [13]，AManPG [15]，ARPG [16]进行比较来验证所提出算法的有效性。在数值实验中，算法 VM-AManPG 的参数设置如下：上述算法的终止条件均设置为 $\|d_k/t\|_F^2\le\varepsilon:=10^{-8}np$。超参数 $\upsilon=2$，线搜索参数 $\delta=0.5$，$\sigma=0.001$，常数 $L=4\|A\|_F^2$，参数 $M=5$。另外，算法 ManPG-Ada，AManPG，ARPG 的相关参数均保持为文献中的默认值。

在数值试验中，通过两类不同的数据测试算法。第一类是采取随机数据，即矩阵 A 和变量 X 都是随机生成的，对于数据矩阵 A，其生成方式为：首先产生随机矩阵 $A=\mathrm{randn}(m,n)$，且使该矩阵的列均值为 0，最后使其列的欧氏范数等于 1。在所有的测试中均令 $m=50$。第二类是采用文献[15]中真实的 DNA 数据集，其变量维数 $n=24589$，$m=113$，$p=4$。

本文首先基于随机生成的数据，通过选取不同的变量维数 n，p 和正则化参数 μ 来测试算法 VM-AManPG 的性能。分别从表 1、表 2、表 3 中我们可以发现，VM-AManPG 无论在 CPU 时间上还是在迭代次数上都明显的优于算法 ManPG-Ada、AManPG 和 ARPG。数值结果表明，VM-AManPG 具有明显的加速效果，并且说明了对角 BB 步长策略的有效性。

Table 1. Comparison on SPCA model with varies n with $p=5$ and regular parameter $\mu=0.5$
表 1. 在 SPCA 模型上不同变量维数 n，其中维数 $p=5$，正则化参数 $\mu=0.5$ 地比较

n	ManPG-Ada		AManPG		ARPG		VM-AManPG	
	iter	cpu	iter	cpu	iter	cpu	iter	cpu
512	203	0.47	77	0.25	76	0.43	**57**	**0.11**
1024	241	0.67	87	0.40	84	0.59	**58**	**0.18**
2156	299	0.82	87	0.49	89	0.66	**58**	**0.20**
3500	258	0.64	102	0.67	92	0.75	**59**	**0.28**
6000	345	1.37	117	0.83	104	1.46	**65**	**0.46**
8000	410	2.14	122	1.51	112	1.84	**68**	**0.62**

Table 2. Comparison on SPCA model with varies p with $n=5000$ and regular parameter $\mu=0.5$
表 2. 在 SPCA 模型上不同变量维数 p，其中维数 $n=5000$，正则化参数 $\mu=0.5$ 地比较

p	ManPG-Ada		AManPG		ARPG		VM-AManPG	
	iter	cpu	iter	cpu	iter	cpu	iter	cpu
2	163	0.25	77	0.21	60	0.32	**38**	**0.19**
4	179	0.42	92	0.34	76	0.76	**54**	**0.20**
6	336	1.42	112	0.80	100	0.94	**78**	**0.41**
8	1453	6.57	262	1.38	239	3.12	**178**	**1.01**
10	1767	9.69	342	2.28	299	3.67	**218**	**1.48**

Table 3. Comparison on SPCA model with varies regular parameter μ with $n = 3000$ and $p = 5$
表 3. 在 SPCA 模型上不同正则化参数 μ，其中维数 $n = 3000$，$p = 5$ 地比较

μ	ManPG-Ada		AManPG		ARPG		VM-AManPG	
	iter	cpu	iter	cpu	iter	cpu	iter	cpu
0.5	329	1.18	107	0.74	104	1.11	**68**	**0.41**
0.75	241	0.83	97	0.50	89	0.97	**58**	**0.35**
1	172	0.77	87	0.43	69	0.79	**48**	**0.31**
1.5	114	0.41	77	0.49	62	0.92	**57**	**0.39**
2	130	0.50	75	0.48	59	0.95	**47**	**0.36**
2.5	117	0.45	72	0.44	64	0.86	**49**	**0.33**

接下来我们采用真实 DNA 数据集，通过选取不同的正则化参数 μ 评估算法性能。数值结果如表 4 所示，我们所提出的算法与已有的算法相比较，满足相同终止条件所需的 CPU 时间和迭代次数都具有明显的优势。

Table 4. Comparison on SPCA model with varies regular parameter μ, DNA data $n = 24589$, $m = 113$, $p = 4$
表 4. 在 SPCA 模型上不同正则化参数 μ，DNA 数据 $n = 24589$，$m = 113$，$p = 4$ 地比较

μ	ManPG-Ada		AManPG		ARPG		VM-AManPG	
	iter	cpu	iter	cpu	iter	cpu	iter	cpu
0.5	5089	119.88	342	11.88	315	13.34	**268**	**8.28**
1	4281	103.35	287	9.06	284	12.11	**217**	**9.06**
1.5	2640	63.43	282	9.58	310	14.05	**108**	**4.16**
2	1964	47.70	382	13.38	349	15.61	**248**	**10.88**
2.5	1645	42.61	287	12.40	259	13.73	**218**	**9.16**

6. 总结

本文针对一类在 Stiefel 流形上的非凸非光滑优化问题，提出了一种可变度量的惯性邻近梯度算法。通过将欧氏空间中的对角 Barzilai-Borwein 类步长策略推广到 Stiefel 流形上可以进一步加速算法收敛，在一定的假设条件下，可以证明算法全局收敛到原问题的一个稳定点。数值实验表明，无论是在随机数据还是在真实数据集中的数值结果，我们所提出的算法在求解效率上都具有一定的优势。

参考文献

[1] Erichson, N.B., Zheng, P., Manohar, K., Brunton, S.L., Kutz, J.N. and Aravkin, A.Y. (2020) Sparse Principal Component Analysis via Variable Projection. *SIAM Journal on Applied Mathematics*, **80**, 977-1002. https://doi.org/10.1137/18M1211350

[2] Barekat, F. (2014) On the Consistency of Compressed Modes for Variational Problems Associated with the Schrödinger Operator. *SIAM Journal on Mathematical Analysis*, **46**, 3568-3577. https://doi.org/10.1137/130942747

[3] Zhang, Y., Lau, Y., Kuo, H.-W., Cheung, S., Pasupathy, A. and Wright, J. (2017) On the Global Geometry Ofsphere-Constrained Sparse Blind Deconvolution. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 4381-4389. https://doi.org/10.1109/CVPR.2017.466 https://par.nsf.gov/biblio/10203504

[4] Tang, J. and Liu, H. (2012) Unsupervised Feature Selection for Linked Social Mediadata. *Proceedings of the* 18*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Beijing, 12-16 August 2012, 904-912. https://doi.org/10.1145/2339530.2339673
https://dl.acm.org/doi/10.1145/2339530.2339673

[5] Absil, P.-A. and Hosseini, S. (2019) A Collection of Nonsmooth Riemannian Optimization Problems. In: Hosseini, S., Mordukhovich, B. and Uschmajew, A., Eds., *Nonsmooth Optimization and Its Applications, International Series of Numerical Mathematics*, Vol. 170, Birkhäuser, Cham, 1-15. https://doi.org/10.1007/978-3-030-11370-4_1

[6] Hu J, Jiang, B., Liu, X. and Wen, Z.W. (2016) A Note on Semidenite Programming Relaxations for Polynomialoptimization over a Single Sphere. *Science China Mathematics*, **59**, 1543-1560.
https://doi.org/10.1007/s11425-016-0301-5

[7] Li, X., Chen, S., Deng, Z., Qu, Q., Zhu, Z. and So, A.M.-C. (2021) Weakly Convex Optimization over Stiefel Manifold Using Riemannian Subgradient-Type Methods. *SIAM Journal on Optimization*, **31**, 1605-1634.
https://doi.org/10.1137/20M1321000
https://epubs.siam.org/doi/10.1137/20M1321000

[8] 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
https://epubs.siam.org/doi/abs/10.1137/16M1069298

[9] Lai, R. and Osher, S. (2014) A Splitting Method for Orthogonality Constrained Problems. *Journal of Scientific Computing*, **58**, 431-449. https://doi.org/10.1007/s10915-013-9740-x

[10] Chen, W., Ji, H. and You, Y. (2016) An Augmented Lagrangian Method for l_1-Regularized Optimization Problems with Orthogonality Constraints. *SIAM Journal on Scientific Computing*, **38**, B570-B592.
https://doi.org/10.1137/140988875
https://epubs.siam.org/doi/abs/10.1137/140988875

[11] Ferreira, O.P. and Oliveira, P.R. (2002) Proximal Point Algorithm on Riemannian Manifold. *Optimization*, **51**, 257-270.
https://doi.org/10.1080/02331930290019413

[12] Bento, G.C., Cruz Neto, J.X. and Oliveira, P.R. (2016) A New Approach to the Proximal Pointmethod: Convergence on General Riemannian Manifolds. *Journal of Optimization Theory and Applications*, **168**, 743-755.
https://doi.org/10.1007/s10957-015-0861-2

[13] Chen, S., Ma, S., Man-Cho So, A. and Zhang, T. (2020) Proximal Gradient Method for Nonsmooth Optimization over the Stiefel Manifold. *SIAM Journal on Optimization*, **30**, 210-239. https://doi.org/10.1137/18M122457X
https://epubs.siam.org/doi/abs/10.1137/18M122457X

[14] Xiao, X., Li, Y., Wen, Z. and Zhang, L. (2018) A Regularized Semi-Smooth Newton Method with Projection Steps for Composite Convex Programs. *Journal of Scientific Computing*, **76**, 364-389.
https://doi.org/10.1007/s10915-017-0624-3

[15] Huang, W. and Wei, K. (2022) An Extension of Fast Iterative Shrinkage-Thresholding Algorithm to Riemannian Optimization for Sparse Principal Component Analysis. *Numerical Linear Algebra with Applications*, **29**, Article No. e2409.
https://doi.org/10.1002/nla.2409

[16] Huang, W. and Wei, K. (2021) Riemannian Proximal Gradient Methods. *Mathematical Programming*, 1-43.
https://doi.org/10.1007/s10107-021-01632-3

[17] Bonettini, S., Porta, F. and Ruggiero, V. (2016) A Variable Metric Forward-Backward Method with Extrapolation. *SIAM Journal on Scientific Computing*, **38**, A2558-A2584. https://doi.org/10.1137/15M1025098
https://epubs.siam.org/doi/10.1137/15M1025098

[18] Park, Y., Dhar, S., Boyd, S. and Shah, M. (2020) Variable Metric Proximal Gradient Method with Diagonal Barzilai-Borwein Stepsize. 2020 *IEEE International Conference on Acoustics*, *Speech and Signal Processing*, Barcelona, 4-8 May 2020, 3597-3601. https://doi.org/10.1109/ICASSP40776.2020.9054193
https://ieeexplore.ieee.org/document/9054193