ISSN 1749-3889 (print), 1749-3897 (online)
International Journal of Nonlinear Science
Vol.31(2021) No.1,pp.73-80

A Hybird Conjugate Gradient Method on Stiefel Manifold

Maoya Wu *
School of Mathematical Sciences, Jiangsu University, Zhenjiang, 212013, P. R. China
(Received 1 March 2021, accepted 12 March 2021)

Abstract: In this paper, we make a convex combination with Fletcher-Reeves and Dai-Yuan conjugate gradient methods for solving optimization problems on Stiefel manifold. The directions generated by the method are always descent for the objective function. The property is independent of the line search used. Under the non-monotone line search condition, the global convergence of the method is proved. The numerical experiments show the efficiency of the proposed methods.

Keywords: Hybird conjugate gradient method; Fletcher-Reeves and Dai-Yuan conjugate gradient methods; Stiefel manifold; Global convergence

1 Introduction

Optimization problem constrained on the Stiefel manifold has attracted lots of research attentions, such as scientific engineering computing, data mining, machine learning, graphclustering and so on. Many methods in Euclidean space like Fletcher-Reeves(FR) method[1], Dai-Yuan(DY) method[4] and Polak-Ribi‘ere-Polyak(PRP) method[2][3], canbe promoted to solve the optimization problem in manifold. However, since these methods cannot practice well in Stiefel manifold, retractions and vector transports[6] are proposed to help us construct some hybird conjugate gradient methods which are more effectively. Zhu[9] has numerically proposed some vector transports combine with the retraction which constructed by Cayley transformation on Stiefel manifolds. Sakai[13] proposed a hybird conjugate gradient method which combines the Fletcher-Reeves method with the Polak-Ribi‘ere-Polyak method, and prove that under the strong Wolfe conditions, his method can globally convergence. Another methods proposed by Wen and Yin[11] canbe seen as a special case of retraction, which derived a equivalent formula with the Cayley transform, and the main computional complexity is to calculate the inverse of a $2p \times 2p$ matrix. By decomposing each feasible point into the range space of X and the null space of X^T, Jiang and Dai[13] proposed a unified framework, which reduce the main computational complexity to calculate the inverse of a $p \times p$ matrix.

In this paper, we proposed a hybird conjugate gradient method which is sufficient descent on Stiefel manifold. Different from the methods proposed by Sakai[13] and inspired by Zhang[7], we make a combination with modified Fletcher-Reeves(FR) conjugate gradient method and modified Dai-Yuan(DY) conjugate gradient method. Under mild conditions, we proved that use similar retractions and vector transports in Zhu[9], our method is globally convergent under the non-monotone line search. In numerical experiments, we use our methods to solve the linear eigenvalue problem with two situstions, the results shows that always guarantee its convergence.

The rest of this paper is organized as follows. In Section 2, we briefly review the modified conjugate gradient method, the property of Stiefel manifold and introduce a vector transport. In Section 3, we describe the hybrid FR and DY conjugate gradient method and provide its convergence analysis. In Section 4, we test the efficiency and effectiveness of the proposed method. Finally, its the conclusion of our paper.

*Corresponding author. *E-mail address*: 2281976557@qq.com

IJNS.2021.03.15

2 Preliminaries on hybird conjugate gradient methods

In this section, we first give a brief introduction of modified conjugate gradient methods on Euclidean space, then present some properties of Stiefel manifold.

2.1 Modified conjugate gradient methods

Let $f : R^n \to R$ be a conntinuously differentiable function. Consider the following unconstrained optimization problem:

$$\min_{x \in R^n} f(x).$$

The iterative process of the conjugate gradient method is given by

$$x_{k+1} = x_k + \alpha_k d_k,$$

where d_k is the search direction defined by

$$d_k = -g_k + \beta_k d_{k-1} \ .$$

Here we use g_k to denote $\nabla f(x_k)$. Different set on β_k refers to different conjugate gradient method. Well-known conjugate gradient methods include the Fletcher-Reeves(FR) method[1], the Polak-Ribiere-Polyak(PRP) method[2] and the Dai-Yuan(DY) method[3] etc. The parameter β_k of FR and DY methods which we used in this paper specified as

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \quad and \quad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}},$$

where $y_{k-1} = g_k - g_{k-1}$. Zhang[7] proposed a modified FR gradient method in the following way:

$$d_k = -\theta_k g_k + \beta_k d_{k-1}, \tag{1}$$

where

$$\theta_k = \frac{d_{k-1}^T y_{k-1}}{\|g_k\|^2} \quad and \quad \beta_k = \beta_k^{FR}.$$

Inspired by the modified FR conjugate gradient method, we define another modified FR and DY conjugate gradient methods, which directions defined by (1) with the following parameters (θ_k, β_k)

$$\theta_k = -\frac{d_{k-1}^T (g_k - \omega g_{k-1})}{d_{k-1}^T g_{k-1}} \quad and \quad \beta_k = \beta_k^{FR},$$

$$\theta_k = \frac{d_{k-1}^T (g_k + \omega y_{k-1})}{d_{k-1}^T y_{k-1}} \quad and \quad \beta_k = \beta_k^{DY},$$

the modified FR and DY conjugate gradient methods also provide descent direction if $\omega > 0$. Actually,

$$d_k^T g_k = -\omega \|g_k\|^2, \quad \forall k \geq 1.$$

2.2 The projection of the gradient and vector transport

Retractions and vector transports on Remannian manifolds have been systematacially summarized by Zhu[9]. Stiefel manifold is a special case of Remmannian manifolds, which denoted by

$$St(p, n) = \{X \in R^{p \times n} : X^T X = I\}.$$

The tangent space[8] $T_X St(n,p)$ of the Stiefel manifold can be given by

$$T_X St(n,p) = \left\{ Y \in {}^{n \times p} : X^T Y + Y^T X = 0 \right\}.$$

By comparing the common inner products used in $T_X St(n,p)$, Zhu [10] has given the difinition of generalized inner product, which defined by

$$\langle Y, Z\rangle_{\gamma,X} = trace\left\langle Y^T(I-\gamma XX^T)Z\right\rangle, \quad \forall Y, Z \in T_X St(n,p) \tag{2}$$

where $\gamma \in [0,1)$ is a parameter. For any differentiable function $F(X)$, we denote the gradient of $F(X)$ as $ggradF(X)$, under the above inner product, $ggradF(X)$ canbe expressed as

$$ggradF(X) = aX(X^TG - G^TX) + (I - XX^T)G,$$

where $G = (\frac{\partial F(X)}{\partial X_{ij}}) \in R^{n\times p}$, ∂X_{ij} is the (i,j) entries of X, and $a = \frac{1}{2(1-\gamma)} \geq \frac{1}{2}$.

If $a = \frac{1}{2}$, we can achieve the projection of the gradient gg_k and denote $gg_K \triangleq ggrad F(X_k)$, then

$$gg_K \triangleq ggrad F(X) = G - X\frac{X^TG + (X^TG)^T}{2}.$$

In order to employ our hybird conjugate gradient method on the Stiefel manifold, we use well-known Cayley transformation [5] to help us transport d_{k-1} from $T_{X_{k-1}}(n,p)$ into $T_{X_k}(n,p)$.

$$R_X(tY) = (I + \frac{t}{2}A_Y)^{-1}(I - \frac{t}{2}A_Y)X \quad \forall Y \in T_X St(n,p). \tag{3}$$

where $A_Y = (P_XY)X^T - X(P_XY)^T$ and $P_X = I - \frac{1}{2}XX^T$.

The aboved formula of $R_X(tY)$ is defined as a retraction on the Stiefel manifold. Similar properties of retraction were generalized by Zhu[9].

Combine the aboved retraction with Cayley transform, we can achieve the following vector transport [6] formula

$$T_{tY}(D) = (I + \frac{t}{2}A_Y)^{-1}(I - \frac{t}{2}A_Y)D \in T_X St(n,p) \quad \forall\, Y, D \in T_X St(n,p), \tag{4}$$

where T_{tY} is a mapping from $T_X St(n,p)$ into $T_{R_X(tY)}St(n,p)$ on the Stiefel manifold. Under the aboved inner product (2), (3) defines an isometric vector transport. Similar result can be found in [9].

To avoid time-consuming matrix inverse $(I + \frac{t}{2}A_Y)^{-1}$ in (4), by using Sherman-Morrison Woodbury formula, we can rewrite $T_{tY}(D)$ as

$$T_{tY}(D) = D - tU(I + \frac{t}{2}V^TU)^{-1}V^TD,$$

where $U = [P_XD \quad X] \in R^{n\times 2p}$ and $V = [X \quad -P_XD] \in R^{n\times 2p}$.

3 Hybrid FR and DY conjugate gradient method

From now on, we consider a hybrid conjugate gradient method which make a convex combination of FR and DY conjugate gradient method for optimization problems. We focus on the Stiefel manifold

$$\min_{X\in St(n,p)} \quad F(X), \tag{5}$$

where $F : R^{n\times p} \to R$ is continuously differentiable.

3.1 Nonmontone Conjugate Gradient Method on Stiefel Mainfold

For optimization problem (5), we update the current iterate X_k by

$$X_{k+1} = R_{X_k}(\alpha_k A_k) = (I_n + \frac{\alpha_k}{2}A_k)^{-1}(I_n - \frac{\alpha_k}{2}A_k)X_k,$$

where α_k is the step length which determined by some line search and

$$A_k = (P_{X_k}d_k)X_k^T - X_k(P_{X_k}d_k)^T \quad A_k \in T_X St(n,p),$$

d_k is the search direction combined , which defined by

$$d_k = \begin{cases} -\omega gg_k & if \quad k=0, \\ -[\varphi_k\theta_k^{FR}+(1-\varphi_k)\theta_k^{DY}]gg_k + [\varphi_k\beta_k^{FR}+(1-\varphi_k)\beta_k^{DY}]T_{\alpha_{k-1}d_{k-1}}(d_{k-1}) & if \quad k>0, \end{cases} \tag{6}$$

where $\omega \geq 1$ and parameters as followed:

$$\beta_k^{DY} = \frac{\|gg_k\|^2}{\left\langle T_{\alpha_{k-1}d_{k-1}}(d_{k-1}), gg_k \right\rangle - \langle d_{k-1}, gg_{k-1}\rangle} \quad and \quad \beta_k^{FR} = \frac{\|gg_k\|^2}{\omega\|gg_{k-1}\|^2};$$

$$\theta_k^{DY} = \frac{(1+\omega)\left\langle T_{\alpha_{k-1}d_{k-1}}(d_{k-1}), gg_k \right\rangle - \omega\left\langle d_{k-1,}gg_{k-1}\right\rangle}{\left\langle T_{\alpha_{k-1}d_{k-1}}(d_{k-1}), gg_k \right\rangle - \left\langle d_{k-1,}gg_{k-1}\right\rangle} \quad and \quad \theta_k^{FR} = -\frac{\left\langle T_{\alpha_{k-1}d_{k-1}}(d_{k-1}), gg_k \right\rangle - \omega\left\langle d_{k-1,}gg_{k-1}\right\rangle}{\left\langle d_{k-1,}g_{k-1}\right\rangle}.$$

Obviously, the search direction generated by the hybird conjugate gradient is a sufficient descent direction, since

$$\langle d_k, gg_k\rangle = -\omega\|gg_k\|^2, \forall k \geq 1 \quad \text{and} \quad d_k \in T_{X_k}St(p,n). \tag{7}$$

For the step length α_k of our algorithm, we can set t_k as its first choice, let

$$t_k = \begin{cases} \frac{\alpha_{k-1}\|d_{k-1}\|^2}{\left|trace\left((gg_k-gg_{k-1})^T d_{k-1}\right)\right|} & if \quad trace\left((gg_k-gg_{k-1})^T d_{k-1}\right) \neq 0, \\ 1 & otherwise. \end{cases} \tag{8}$$

Abviously, t_k is a special case of Barzilai-Borwein [12] step size on the Stiefel manifold. In Algorithm1, we require the final step length α_k satisfies the following non-monotone line search condition

$$F\left(R_{X_k}(\alpha_k A_k)\right) \leq \mu_k - \delta_1\omega\alpha_k\|gg_k\|^2 - \delta_2\alpha_k^2\|d_k\|^2, \tag{9}$$

where μ_{k+1} is taken to be the convex combination of μ_k and $F(X_{k+1})$.

Given the formula of descent direction and the related parameters, the algorithm of this method is presented as follows.

Algorithm 1 :HFRDY method-nonmontone hybird FR and DY conjugate gradient methods on Stiefel manifold

Input: $\omega > 0,\ \ \rho, \delta_1, \delta_2, \eta \in (0,1),\ X_0 \in St(p,n),\ \xi_0 = 1, \mu_0 = F(X_0)$.

Output: $\{X_k\}_k \geq 0$.

While $\|gg_k\| \geq \varepsilon$ do
Set $\theta_k = \varphi_k\theta_k^{FR} + (1-\varphi_k)\theta_k^{DY};\quad \beta_k = \varphi_k\beta_k^{FR} + (1-\varphi_k)\beta_k^{DY}$.
Compute d_k by (6);
Compute t_k by (9);
Deteimine $\alpha_k = \max\{t_k\rho^j, j = 0,1,2,\cdots\}$ satisfying non-monotone line search condition (8);
Update X_{k+1} by $X_{k+1} = R_{X_k}(\alpha_k A_k)$;
Update $\xi_{k+1} \leftarrow \eta\xi_k + 1$ and $\mu_{k+1} \leftarrow (\eta\xi_k\mu_k + F(X_{k+1}))/\xi_{k+1}$;
Update $k \leftarrow k+1$;
Return $\{X_k\}_k \geq 0$.

3.2 Sufficient Descent Properties of Hybird Conjugate Gradient Methods

To analyze the convergence theorem, the following assumption is imposed to the objective function F.

Assumption 1 *F is continuously differentiable with Lipschitz gradient, i.e., there exists a constant $l_f > 0$ such that*

$$\left\|\frac{\partial F}{\partial X}(X) - \frac{\partial F}{\partial X}(Y)\right\| \leq l_f\,\|Y - X\|\,.$$

Similar to Lemma 4.4 in [5], we can achieve the following conclusion.

Lemma 2 *For all $k \geq 0$ and $0 \leq \alpha \leq t_k$,there exist $l_1 > 0$ and $l_2 > 0$, such that the update scheme follows*

$$\|R_{X_k}(\alpha A_k) - R_{X_k}(0)\| \leq l_1 \alpha \|d_k\| \quad and \left\| R'_{X_k}(\alpha A_k) - R'_{X_k}(0) \right\| \leq l_2 \alpha \|d_k\|^2.$$

Lemma 3 *If $\alpha_k = t_k \rho^{j_k}$ satisfies the line search condition, then*

$$\alpha_k > \frac{2\rho\omega(1-\delta_1)\theta\|g_k\|^2}{(c_1 + 2\delta_2)\|d_k\|^2}, \tag{10}$$

where $c_1 = \bar{c}\, l_2 + c_1 l_f l_1$. Moreover, when $\delta_1 < 1$,

$$\lim_{k\to\infty} \alpha_k \|g_k\|^2 = 0 \quad and \quad \lim_{k\to\infty} \alpha_k \|d_k\| = 0. \tag{11}$$

It follows from Lemma 2 and 3 that the following Zoutendijk condition holds:

$$\sum_{k=1}^{\infty} \frac{\|gg_k\|^4}{\|d_k\|^2} < +\infty.$$

Theorem 4 *Suppose Assumption* 1 *holds and Algorithm does not terminate in finitely many iterations. Then the sequence X_k generated by Algorithm converges in the sense that*

$$\lim \inf_{k\to\infty} \|gg_k\| = 0,$$

which means there exists at least one accumulation point satiesfies the first optimality condition[4].

Proof. If $\lim \inf_{k\to\infty} \alpha_k > 0$, from Lemma 3, it holds that $\lim \inf_{k\to\infty} \|gg_k\| = 0$.

If $\lim \inf_{k\to\infty} \alpha_k = 0$. Suppose

$$\lim \inf_{k\to\infty} \alpha_k \neq 0,$$

then there exiats a constant $\zeta > 0$ such that $\|gg_k\| \geq \zeta, \quad \forall k \geq 0$.

We get from (6) and (7) that

$$\|d_k\|^2 = [\varphi_k \beta_k^{FR} + (1-\varphi_k)\beta_k^{DY}]^2 \|d_{k-1}\|^2 + 2\omega[\varphi_k \theta_k^{FR} + (1-\varphi_k)\theta_k^{DY}] \|gg_k\|^2$$
$$-[\varphi_k \theta_k^{FR} + (1-\varphi_k)\theta_k^{DY}]^2 \|gg_k\|^2 .$$

Dividing both sides of this equality by $\|gg_k\|^4$, combing with $\omega \geq 1$, we have

$$\begin{aligned}\frac{\|d_k\|^2}{\|gg_k\|^4} &= [\varphi_k \beta_k^{FR} + (1-\varphi_k)\beta_k^{DY}]^2 \frac{\|d_{k-1}\|^2}{\|gg_k\|^4} + \frac{2\omega[\varphi_k \theta_k^{FR} + (1-\varphi_k)\theta_k^{DY}]}{\|gg_k\|^2} - \frac{[\varphi_k \theta_k^{FR} + (1-\varphi_k)\theta_k^{DY}]^2}{\|gg_k\|^2} \\ &\leq [\varphi_k \beta_k^{FR} + (1-\varphi_k)\beta_k^{DY}]^2 \frac{\|d_{k-1}\|^2}{\|gg_k\|^4} + \frac{\omega^2}{\|gg_k\|^2}.\end{aligned} \tag{12}$$

Define $\theta_k = \frac{\langle T_{\alpha_{k-1} d_{k-1}}(d_{k-1}), gg_k \rangle - \langle d_{k-1}, gg_{k-1}\rangle - \omega\|gg_{k-1}\|^2}{\|gg_{k-1}\|^2}, \quad c_k = \frac{\varphi_k}{\omega} + \frac{1-\varphi_k}{\omega+\theta_k}$, then

$$\begin{aligned}\varphi_k \beta_k^{FR} + (1-\varphi_k)\beta_k^{DY} &= \frac{\varphi_k}{\omega} \frac{\|gg_k\|^2}{\|gg_{k-1}\|^2} + \frac{(1-\varphi_k)}{\omega+\theta_k} \frac{\|gg_k\|^2}{\|gg_{k-1}\|^2} \frac{(\omega+\theta)\|gg_{k-1}\|^2}{\langle T_{\alpha_{k-1} d_{k-1}}(d_{k-1}), gg_k \rangle - \langle d_{k-1}, gg_{k-1}\rangle} \\ &= c_k \frac{\|gg_k\|^2}{\|gg_{k-1}\|^2}.\end{aligned}$$

Suppose that $\theta_k \geq 0$, then $0 \leq c_k = \frac{\varphi_k}{\omega} + \frac{1-\varphi_k}{\omega+\theta_k} \leq \frac{1}{\omega} \leq 1$.

On the other hand, if $\theta_k < 0$, and $\omega + \theta_k > 0$, then

$$c_k - 1 = \frac{(\theta_k - \omega)\varphi_k + \omega(1 - \omega - \theta_k)}{\omega(\omega + \theta_k)} \quad and \quad c_k + 1 = \frac{(\theta_k - \omega)\varphi_k + \omega(1 + \omega + \theta_k)}{\omega(\omega + \theta_k)}.$$

It is easy to see that ,if

$$\omega + \theta_k > 0, \quad \varphi_k \in [\max\{0, \frac{\omega(1 - \omega - \theta_k)}{\omega - \theta_k}\}, \min\{1, \frac{\omega(1 + \omega + \theta_k)}{\omega - \theta_k}\}].$$

$$\omega + \theta_k < 0, \quad \varphi_k \in [\max\{0, \frac{\omega(1 + \omega + \theta_k)}{\omega - \theta_k}\}, \min\{1, \frac{\omega(1 - \omega - \theta_k)}{\omega - \theta_k}\}].$$

Then, $|c_k| \leq 1, \quad [\varphi_k \beta_k^{FR} + (1 - \varphi_k)\beta_k^{DY}]^2 \leq \frac{\|gg_k\|^4}{\|gg_{k-1}\|^4}$.
From (6) and (7),(12), we can get

$$\begin{aligned}
\frac{\|d_k\|^2}{\|gg_k\|^4} &\leq [\varphi_k \beta_k^{FR} + (1 - \varphi_k)\beta_k^{DY}]^2 \frac{\|d_{k-1}\|^2}{\|gg_k\|^4} + \frac{\omega^2}{\|gg_k\|^2} \\
&\leq \frac{\|d_{k-1}\|^2}{\|gg_{k-1}\|^4} + \frac{\omega^2}{\|gg_k\|^2} \leq \frac{\|d_0\|^2}{\|gg_0\|^4} + \sum_{k=1}^{k} \frac{\omega^2}{\|gg_i\|^2} \\
&\leq \frac{\omega^2(k+1)}{\zeta^2}.
\end{aligned}$$

The last inequalities implies

$$\sum_{k=1}^{\infty} \frac{\|gg_k\|^4}{\|d_k\|^2} \geq +\infty.$$

which contradicts Zoutandijk condition. ■

Then, combine Theorem 4 with Zoutendijk condition, the hybird conjugate gradient convergence globally on the Stiefel manifold.

4 Numerical Experiments

In this part, we mainly use the hybird conjugate gradient method to solve linear eigenvalue problem on Stiefel manifold and analysis its efficiency under different conditions.

The formula of linear eigenvalue problem as followed:

$$\min_{X \in R^{n \times p}} \quad tr(X^T AX) \quad s.t.\ X^T X = I,$$

where A is a symmetric matrix.

The corresponding objective function and gradient of this problem are as following :

$$f(X) = -tr(X^T AX) \quad \text{and} \quad G = \frac{\partial f(X)}{\partial X} = -2AX.$$

In numerical experiments, we consider two situations of our experiment. First, we fix the size of this problem, where $n = 1000$ and $p = 5$, and take values of parameter ω which appeared in Algorithm 1 as $0.1, 0.5, 1, 1.5, 2$. Second, after choosing the most suitable value of ω, we take different sizes of the linear eigenvalue problem, where $n = 500, 1000, 2000, 3000, 4000$ and $p = 5$. The results of the aboved conditions are presented in the following Table 1 and Table 2.

In these tables, "FV" means the number of the objective function, "NG" means the norm of the gradient and "NGc" means $\|\nabla FX)\|_F = \left\|\frac{\partial}{\partial X} L(X, G^T X)\right\|_F$, "itr" means the number of iterations, "feasi" means the feasibilty of the problem. Table 1 shows that, when $\omega = 0.1, 0.5$, the Algorithm1 cannot convergence. Therefore, the range of parameter ω must be $\omega \geq 1$, and we find that when $\omega = 1.5$, its the best one among these values. After we fix the value of the Parameter ω, we see that with the sizes of the linear eigenvalue problem increase, the calculation amount of Algorithm1 become more complex.

Table 1: Different sizes of parameter ω for Algorithm1

ω	0.1	0.5	1	1.5	2
FV	-2.516292e+03	-2.520052e+03	-9.7425556e+03	-9.723070e+03	-9.695318e+03
NG	2.24e+03	2.23e+03	3.54e-02	1.28e-02	1.70e-02
NGc	2.24e+03	2.23e+03	3.54e-02	1.28e-02	1.70e-02
itr	2	2	139.2	137.8	189
Time	0.0168052	0.019159	0.1714356	0.1802832	0.242674667
feasi	1.791273e-15	1.828956e-15	3.966999e-15	9.197988e-15	1.984286e-14

Table 2: Different sizes of the linear eigenvalue problem

n	500	1000	2000	3000	4000
FV	-4.750444e+03	-9.725608e+03	-1.9640047e+4	-2.958271e+04	-3.955398e+04
NG	4.218619e-03	8.754726e-03	4.5880502e-02	0.154919793	0.215752572
NGc	4.218619e-03	8.754726e-03	4.5880502e-02	0.154919793	0.215752572
itr	113	151	175.5	210.5	215.5
Time	0.0952655	0.191996167	1.187099	2.951548833	5.3789635
feasi	3.387220e-15	3.832722e-15	4.692564e-15	5.465049E-15	5.489136e-15

5 Conclusions

In this paper, by introducing an vector transport, we propose an hybird FR and DY conjugate gradient methods for optimization problems constrained on the Stiefel manifold and proved that the deirections are always descent under the non-momotone line search. The global convergence is given under the Zoutendijk condition. In numerical experiments we indicate that the proposed methods are effcient. The hybird conjugate gradient method performs well both on theories and experiments for solving the Stiefel optimization problems.

Acknowledgments

The authors would like to thank our editor and the reviewers who went far beyond the call of duty with many very valuable suggestions that enhanced the readability and relevance of this paper.

References

[1] R. Fletcher and C. M. Reeves. Function minimization by conjugate gradients. *The computer journal*, 7(1964)(2):149–154.

[2] E. Polak and G. Ribiere. Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 3(1969)(R1):35–43.

[3] B. T. Polyak. The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics*, 9(1969)(4):94–112.

[4] Y.-H. Dai and Y. Yuan. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on optimization*, 10(1999)(1):177–182.

[5] B. Jiang and Y.-H. Dai. A framework of constraint preserving update schemes for optimization on stiefel manifold. *Mathematical Programming*, 153(2015)(2):535–575.

[6] P.-A. Absil, R. Mahony and R. Sepulchre. Optimization algorithms on matrix manifolds. Princeton University Press. 2009.

[7] L. Zhang, W. Zhou and D. Li. Global convergence of a modified fletcher–reeves conjugate gradient method with armijo-type line search. *Numerische mathematik*, 104(2006)(4):561–572.

[8] A. Edelman, T. A. Arias and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(1998)(2):303–353.

[9] X. Zhu. A riemannian conjugate gradient method for optimization on the stiefel manifold. *Computational optimization and Applications*, 67(2017)(1):73–110.

[10] H. Zhu et al. Multiple graphs clustering by gradient flow method. *Journal of the Franklin Institute*, 355(2018)(4):1819–1845.

[11] Z. Wen and W. Yin. A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, 142(2013)(1):397–434.

[12] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA journal of numerical analysis*, 8(1988)(1):141–148.

[13] S. Hiroyuki and I. Hideaki. Sufficient Descent Riemannian Conjugate Gradient Method. *arXiv preprint arXiv*. 2009.

[14] B. Jiang and Y.-H. Dai. A framework of constraint preserving update schemes for optimization on stiefel manifold. *Mathematical Programming*, 153(2015)(2):535–575.