阅读论文简单总结

Jinchao Zhang

Hebei University of Technology

2020.11.1

Contents

Proximal Newton-Type Methods

consider the following problem:

$$\min_{X \in \mathbb{R}^{n \times p}} f(X) + g(X) \tag{1}$$

f, g均为利普希茨连续凸函数，但是g一般不可微(非光滑)

重点

- 算法给出了全面的收敛性分析，也可以拓展到f为非凸函数，并给出相关的收敛性(本文没有给出)
- 相对一阶求解方法，具有收敛快、精度高、问题规模大，且对目标函数水平集的条件数不敏感
- 主要在于求解下降方向的子问题求解，文章中有精确求解和非精确求解,Hessian $H_k = \nabla^2 f(x_k)$,此时为牛顿法，取$\nabla^2 f(x_k)$的近似时，则为拟牛顿方法；当问题规模较大时，则选用Limited memory quasi-Newton method。
- 终止条件，自适应终止条件效果更好

Lee, Jason & Sun, Yuekai & Saunders, Michael. (2012). Proximal Newton-Type Methods for Minimizing Composite Functions. SIAM Journal on Optimization. 24. 10.1137/130921428

Existing Nonsmooth Optimization on Stiefel Manifolds

- Lai and Stanley Osher(2014)The splitting method for orthogonality constrained problem (SOC)
- Artiom Kovnatsky, Klaus Glashoff(2016)Manifold alternating direction method of multipliers (MADMM)
- Chen, Ji(2016)Proximal alternating minimization based on augmented Lagrangian method (PAMAL)
- Hosseini and Uschmajew (2017)Gradient sampling method and any limit point is a critical point.

Introduction

Definition 1 (Stiefel manifold)

(1) $S_{n,p}$ *is an embedded submanifold of* $\mathbb{R}^{n\times p}$*;*

(2) $p = 1$ *,*$S_{n,p}$ *reduces to the unit sphere* $S^{n-1} = \{x \in \mathbb{R}^n | \|x\|_2 = 1\}$*;*

(3) $dim(S_{n,p}) = np - \frac{1}{2}p(p+1)$*.*

Definition 2 (Tangent space)

The tangent space to Stiefel manifold $\mathcal{M}$ *at point* X *is denoted be* $T_X\mathcal{M}$*:*

$$T_X\mathcal{M} = \{Z \in \mathbb{R}^{n\times p} : X^T Z + Z^T X = 0\}$$

Definition 3 (Orthogonal projection)

The projection of Y onto the tangent space at $X \in S_{n,p}$ is given by:

$$Proj_{T_X\mathcal{M}} Y = (I_n - XX^T)Y + \frac{1}{2}X(X^TY - Y^TX) \tag{2}$$

Riemannian gradient of $f(X)$:$gradf(X) = Proj_{T_X\mathcal{M}} \nabla f(X)$.

Definition 4 (Retraction mapping)

A retraction on a differentiable manifold $\mathcal{M}$ is a smooth mapping Retr from the tangent bundle $T_X\mathcal{M}$ onto $\mathcal{M}$ satisfying the following two conditions:

(1) $Retr_X(0) = X, \forall X \in \mathcal{M}, 0 \in T_X\mathcal{M}$;

(2) $\forall X \in \mathcal{M}$

$$\lim_{T_X\mathcal{M} \ni \xi \to 0} \frac{\|Retr_X(\xi) - (X + \xi)\|_F}{\|\xi\|_F} = 0 \tag{3}$$

Introduction

1. QR decomposition:$\mathrm{Retr}_X^{\mathrm{QR}}(\xi) = \mathrm{qf}(X+\xi)$
2. exponential mapping:
   $\mathrm{Retr}_X^{\exp}(t\xi) = [X, Q] \exp\left(t\begin{bmatrix} -X^\top\xi & -R^\top \\ R & 0 \end{bmatrix}\right)\begin{bmatrix} I_r \\ 0 \end{bmatrix}$ where
   $QR = -(I_n - XX^T)\xi$ is the unique QR factorization;
3. polar decomposition: $\mathrm{Retr}_X^{\mathrm{polar}}(\xi) = (X+\xi)\left(I_r + \xi^\top\xi\right)^{-1/2}$
4. Cayley transformation: $\mathrm{Retr}_X^{\mathrm{cayley}}(\xi) = \left(I_n - \frac{1}{2}W(\xi)\right)^{-1}\left(I_n + \frac{1}{2}W(\xi)\right)X$
   where $W(\xi) = (I_n - \frac{1}{2}XX^T)\xi X^T - X\xi^T(I_n + \frac{1}{2}XX^T)$.

Proximal gradient method on the Stiefel manifold

Optimization with Structure:

$$\begin{aligned} &\min_{X\in\mathbb{R}^{n\times p}} F(X) = f(X) + h(X) \\ &s.t \quad X^T X = I_p, \quad p \ll n \end{aligned} \tag{4}$$

Assumption:

(1) f is smooth and possibly nonconvex,and its gradient ∇f is Lipschitz continuous with Lipschitz constant L.

(2) h is convex,possibly nonsmooth ,and Lipschitz continuous with constant L_h

[1]Chen S , Ma S , So M C , et al. Proximal gradient method for nonsmooth optimization over the stiefel manifold[J]. SIAM Journal on Optimization, 2020, 30(1):210-239.

Proximal gradient method on the Stiefel manifold

In the Euclidean setting generates the iterates as follows:

$$X_{k+1} = \arg\min_{Y} f(X_k) + \langle \nabla f(x_k), Y - X_k \rangle + \frac{1}{2t}\| Y - X_k \|_F^2 + h(Y). \tag{5}$$

In order to deal with the manifold constraint,then:

$$\begin{aligned} &(1)\, V_k = \arg\min_{V} \langle gradf(x_k), V \rangle + \frac{1}{2t}\| V \|_F^2 + h(X_k + V) \\ &\qquad s.t. \quad V \in T_{X_k}\mathcal{M} \\ &(2) X_{k+1} = Retr_{X_k(\alpha V_k)} \end{aligned} \tag{6}$$

Where $t > 0$ is the stepsize, we can interpret $X_k + V$ as the sum of X_k and V in the ambient Euclidean space $\mathbb{R}^{n \times p}$, as $\mathcal{M}$ is an embedded submanifold of $\mathbb{R}^{n \times p}$.

Proximal gradient method on the Stiefel manifold

Following the definition of $gradf$, we have:

$$\langle gradf(x_k), V\rangle = \langle \nabla f(x_k), V\rangle \quad \forall V \in T_{X_k\mathcal{M}}$$

proximal gradient step:

$$\begin{aligned} V_k &:= \arg\min_V \langle \nabla f(x_k), V\rangle + \tfrac{1}{2t}\|V\|_F^2 + h(X_k + V) \\ &s.t. \quad V \in T_{X_k}\mathcal{M} \end{aligned} \tag{7}$$

The tangent space: $T_X\mathcal{M} = \{V | V^T X + X^T V = 0\}$

Define the linear operator $\mathcal{A}_k := V^T X_k + X_k^T V$

Regularized semismooth Newton method for subproblem

Rewrite subproblem:How to solve?

$$\begin{aligned} V_k := \arg\min_{V} & \langle \nabla f(x_k), V \rangle + \tfrac{1}{2t}\| V \|_F^2 + h(X_k + V) \\ s.t. \quad & \mathcal{A}_k(V) = 0. \end{aligned} \tag{8}$$

Consider the composite convex optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x) + h(x) \tag{9}$$

Where f, h are real-valued convex functions,in general h is a regularization function.

Contribution

- Bridge the gap between first-order and second-order type methods for composite convex programs;
- Propose an adaptive semi-smooth Newton method and establish its convergence to global optimality

[2] X. Xiao, Y. Li, Z. Wen, and L. Zhang, A regularized semi-smooth Newton method with projection steps for composite convex programs, J. Sci. Comput., 76 (2018), pp. 364–389.

Regularized semismooth Newton method for subproblem

Definition 5 (Clarke's generalized Jacobian)

Let $F : \mathcal{O} \to \mathcal{R}^m$ be locally Lipschitz continuous at $x \in \mathcal{O}$. The B-subdifferential of F at x is defined by

$$\partial_B F(x) := \{\lim_{k\to\infty} F^{'}(x^k) | x^k \in D_F, x^k \to x\} \tag{10}$$

Then set $\partial F(x) = conv(\partial_B(F(x)))$ is called Clarke's generalized Jacobian.

Definition 6 (semi-smooth)

Let $F : \mathcal{O} \to \mathcal{R}^m$ be locally Lipschitz continuous at $x \in \mathcal{O}$. if
(a) F is directionally differentiable at x;
(b) for any $d \in \mathcal{O}$ and $J \in \partial F(x+d)$,

$$\|F(x+d) - F(x) - Jd\|_2 = o(\|d\|_2) \quad as \ d \to 0$$

if F is semi-smooth and for any $d \in \mathcal{O}$ and $J \in \partial F(x+d)$,

$$\|F(x+d) - F(x) - Jd\|_2 = O(\|d\|_2^2) \quad as \ d \to 0$$

F is said to be strongly semi-smooth.

Regularized semismooth Newton method for subproblem

The examples of semi-smooth functions:

(1) smooth functions, all convex functions (thus norm), and the piecewise differentiable functions.

(2) $\|\cdot\|_p, p \in [1, \infty]$ is strong semi-smooth.

Lemma 1

For a Lipschitz continuous mapping $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$,F is monotone if and only if each element of $\partial_B F(x)$ is positive semidefinite for any $x \in \mathbb{R}^n$

Consider the ℓ_1-regularized optimization problem:

$$\min h(x) + \mu\|x\|_1$$

Where $h(x)$ is convex continuously differentiable,$f(x) = \mu\|x\|_1$; Corresponding to the FBS method is

$$F(x) = x - prox_{tf}(x - t\nabla h(x)) = 0$$

Regularized semismooth Newton method

$J(x) \in \partial_B F(x)$

$$J(x) = I - M(x)(I - t\partial^2 h(x)) \tag{11}$$

$M(x) \in \partial prox_{tf}(x - t\nabla h(x))$ and $\partial^2 h(x)$ is generalized Hessian matrix of $h(x)$. shrinkage operator:

$$(prox_{tf}(x))_i = sign(x_i)max(|x_i| - \mu t, 0)$$

Take $M(x)$ is a diagonal matrix

$$M_{ii}(z) = \begin{cases} 1, & |(z)_i| > t \\ 0, & \text{otherwise} \end{cases}$$

Then we have:

$$(J(x) + \mu_k I)d = -F^k \tag{12}$$

Regularized semismooth Newton method

For

$$F(z) = 0$$

$$(J(z) + \mu_k I)d = -F^k \tag{13}$$

Where $F^k = F(z^k), \mu_k = \lambda_k \|F^k\|_2$,$\lambda_k > 0$ is a regularization parameter. define

$$r^k := (J_k + \lambda_k \|F(z^k)\|_2 I)d + F(z^k)$$

(1) Inexactly solve linear system(16) seek d^k satisfy:

$$\|r^k\|_2 \leq \tau \min\{1, \lambda_k \|F(z^k)\|_2 \|d^k\|_2\} \tag{14}$$

Then obtain a trial point

$$u^k = z^k + d^k$$

(2) IF $\|F(u^k)\|_2$ sufficiently decreased,take a Newton step,i.e.,

Regularized semismooth Newton method

When $\|F(u^k)\|_2 \leq \nu\xi_k$; $z^{k+1} = u^k, \xi_{k+1} = \|F(u^k)\|_2, \lambda_{k+1} = \lambda_k$ [Newton step]
(3) Otherwise take a safeguard step,define

$$\rho_k = -\frac{\langle F(u^k), d^k\rangle}{\|d^k\|_2^2}$$

Let $0 < \eta_1 < \eta_2 < 1, 1 < \gamma_1 < \gamma_2$ If $\rho_k \geq \eta_1$, the iteration is said to be successful, for a successful iteration, we take a hyperplane projection step when the residual of the projection step is non-increasing and take a fixed-point iteration when it is increasing Otherwise unsuccessful,i.e.,

$$z^{k+1} = \begin{cases} v^k, \text{ if } \rho_k \geq \eta_1 \text{ and } \left\|F\left(u^k\right)\right\|_2 \leq \|F(z^k)\|_2 & [\text{ projection step }] \\ w^k, \text{ if } \rho_k \geq \eta_1 \text{ and } \left\|F\left(u^k\right)\right\|_2 \geq |F(z^k)\|_2 & [\text{ fixed -point step }] \\ z^k, \text{ otherwise} & [\text{unsuccessful iteration}] \end{cases} \tag{15}$$

$v^k = z^k - \frac{\langle F(u^k), z^k - u^k\rangle}{\|F(u^k)\|_2^2} F(u^k), w^k = z^k - \beta F(z^k), \beta \in (0, \frac{1}{\alpha})$
update the parameter $\xi_{k+1} = \xi_k$

$$\lambda_{k+1} \in \begin{cases} (\lambda_0, \lambda_k), & \text{if } \rho_k \geq \eta_2 \\ [\lambda_k, \gamma_1\lambda_k], & \text{if } \eta_1 \leq \rho_k < \eta_2 \\ (\gamma_1\lambda_k, \gamma_2\lambda_k], & \text{otherwise} \end{cases} \tag{16}$$

An Adaptive Semi-smooth Newton method (ASSN)

Algorithm 1 (The regularized semi-smooth Newton method)

1: **Input:** $0 < \eta_1 < \eta_2 < 1, 1 < \gamma_1 < \gamma_2$ and $0 < \tau < 1, \nu < 1$.

2: Choose z^0 and $\epsilon > 0$.Set $\xi_0 = \|F(z^0)\|_2$

3:**While** $\|F(z^k)\|_2 > \epsilon$ **do**

4: **Step 1:** Select $J_k \in \partial_B F(z^k)$.

5: **Step 2:** Solve the linear system (16) approximately such that d^k satisfies (17)

6: Compute $u^k = z^k + d^k$ and ratio ρ_k

7: **Step 3:**If $\|F(u^k)\|_2 \le \nu\xi_k$,update $z^{k+1}, \xi_{k+1}, \lambda_{k+1}$ according to [Newton step],Otherwise

8: set them according to (18) and (19), respectively;

9: Set $k = k + 1$;

10:**Output:** z^k.

Proximal gradient method on the Stiefel manifold

Lagrangian function:

$$\mathcal{L}(V, \Lambda) = \langle \nabla f(x_k), V \rangle + \frac{1}{2t} \| V \|_F^2 + h(X_k + V) - \langle \mathcal{A}_k(V), \Lambda \rangle \tag{17}$$

KKT systerm:

$$\begin{aligned} 0 &\in \partial_V \mathcal{L}(V, \Lambda); \\ \mathcal{A}_k(V) &= 0. \end{aligned} \tag{18}$$

$V(\Lambda) = prox_{th}(B(\Lambda)) - X_k$ with $B(\Lambda) = X_k - t(\nabla f(X_k) - \mathcal{A}_k^*(\Lambda))$, Where $\mathcal{A}_k^*$ denotes the adjoint opterator of $\mathcal{A}_k$.

Regularized semismooth Newton method for subproblem

$$\begin{aligned} E(\Lambda) \equiv \mathcal{A}_k(V(\Lambda)) = V(\Lambda)^T X_k + X_k^T V(\Lambda) = 0 \\ vec(E(\Lambda)) = (X_k^T \otimes I_p) K_{nr} vec(V(\Lambda)) + (I_r \otimes X_k^T) vec(V(\Lambda)) \end{aligned} \quad (20)$$

We can show operator E is monotone and Lipschitz continuous.

$$\mathcal{G}(vec)(\Lambda) = 2t(K_{rr} + I_{r^2})(I_r \otimes X_k^T)\mathcal{J}(y)|_{y=vec(B(\Lambda))}(I_r \otimes X_k)$$

$$\begin{aligned} \mathcal{G}(\bar{vec})(\Lambda) &= tU_r^+ \mathcal{G}(vec(\Lambda))U_r & (21) \\ &= 4tU_r^+(I_r \otimes X_k^T)\mathcal{J}(y)|_{y=vec(B(\Lambda))}(I_r \otimes X_k)U_r & (22) \end{aligned}$$

$\mathcal{J}$ is the generalized Jacobian of $prox_{th}(y)$.$\mathcal{G}(\bar{vec})(\Lambda)$ is positive semidefinite,because E is positive semidefinite.Now we have

$$E(\lambda) = 0$$

$$(\mathcal{G}(\bar{vec})(\Lambda_k) + \eta I)d = -\bar{vec}(E(\Lambda_k)) \quad (23)$$

and then apply Regularized semismooth Newton method.

[3]J. R. Magnus and H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley Series in Probability and Mathematical Statistics, 1988.

Algorithm

Algorithm 2 (Manifold proximal gradient method (ManPG))

1: **Input**： initial point $X_0 \in \mathcal{M}$,$\gamma \in (0,1)$,stepsize $t > 0$.

2: **for** $k = 0, 1...$**do**

3: obtain V_k by solving the subproblem

4: set $\alpha = 1$

5: **while** $F(Retr_{X_k}(\alpha V_k)) > F(X_k) - \frac{\alpha \| V_k \|_F^2}{2t}$ **do**

6： $\alpha = \lambda\alpha$

7： **end while**

8： set $X_{k+1} = Retr_{X_k}(\alpha V_k)$

9: **end for**

Algorithm

Algorithm 3 (ManPG-Ada)

Input: initial point $X_0 \in \mathcal{M}$,$\gamma \in (0,1)$,$\tau > 1$ and Lipschitz constant L.

set $t = 1/L$

for $k = 0, 1$...**do**

  obtain V_k by solving the subproblem

  set $\alpha = 1$ and linesearchflag=0

  while $F(Retr_{X_k}(\alpha V_k)) > F(X_k) - \frac{\alpha \| V_k \|_F^2}{2t}$ **do**

    $\alpha = \lambda\alpha$

    linesearchflag=1

  end while

  set $X_{k+1} = Retr_{X_k}(\alpha V_k)$

  if linesearchflag=1 **then**

    $t = \tau t$

  else

    $t = \max\{1/L, \tau/t\}$

  end if

end for

Global convergence and iteration complexity

Lemma 2

Given the iterate X_k,let

$$g(V) := \langle \nabla f(X_k), V \rangle + \frac{1}{2t}\|V\|_F^2 + h(X_k + V), V \in T_{X_k}\mathcal{M} \tag{24}$$

denote the objective function in subproblem.Then the following holds for any $\alpha \in [0,1]$:

$$g(\alpha V_k) - g(0) \leq \frac{(\alpha-2)\alpha}{2t}\|V_k\|_F^2. \tag{25}$$

Global convergence and iteration complexity

Definition 7

A point $X \in \mathcal{M}$ is called a stationary point of problem (1.1) if it satisfies the first-order necessary condition;i.e.,$0 \in \mathrm{grad}\, f(X) + \mathrm{Proj}_{\mathrm{T}_X\mathcal{M}}(\partial h(X))$

Lemma 3

For $\forall\ t > 0$,$\exists$ *constant* $\bar{\alpha} > 0$ *such that for any* $0 < \alpha \le \min\{1, \bar{\alpha}\}$,*the condition in step 5 of Algorithm 1 is satisfied, and the sequence* $\{X_k\}$ *generated by Algorithm 1 satisfies*

$$F(X_{k+1}) - F(X_k) \le -\frac{\alpha}{2t}\|V_k\|_F^2 \tag{26}$$

Lemma 4

If $V_k = 0$,*then* X_k *is a stationary point of original problem.*

Proof:Optimality conditions of the subproblem

$$\begin{aligned} &0 \in \frac{1}{t}V_k + gradf(X_k) + Proj_{T_{X_k}\mathcal{M}}\partial h(X_k + V_k),\ V_k \in T_{X_k}\mathcal{M} \\ &V_k = 0, 0 \in gradf(X_k) + Proj_{T_{X_k}\mathcal{M}}\partial h(X_k) \end{aligned} \tag{27}$$

Contributions

Definition 8 ($\epsilon - sationaritypoint$)

$X_k \in \mathcal{M}$ is an $\epsilon-$ stationarity point of original problem if the solution V_k to subproblem with $t = \frac{1}{L}$ satisfies $\| V_k \|_F \leq \frac{\epsilon}{L}$.

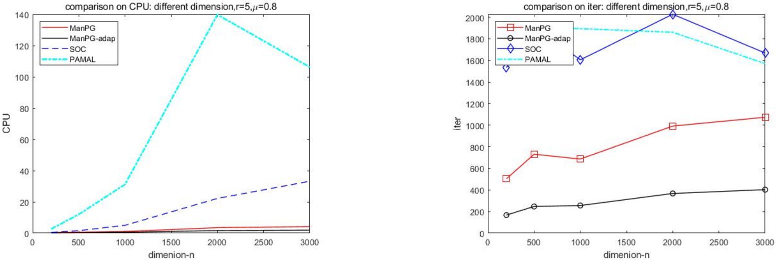
Theorem 1

Under Assumption 1.1, every limit point of the sequence $\{X_k\}$ generated by Algorithm 2 is a stationary point of original problem.Moreover, Algorithm 3 with $t = 1/L$ will return an $\epsilon - stationarypoint$ of original problem in at most $\lceil \frac{2L(F(X_0)-F^*)}{\gamma \bar{\alpha} \epsilon^2} \rceil$ iterations, where $\bar{\alpha}$ is defined in Lemma 5.2 and F^* is the optimal value of original problem.

Numerical experiments

Problem:

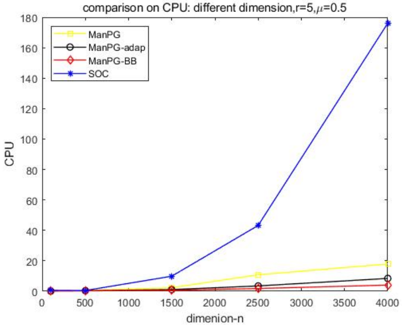
$$\min_{X \in \mathbb{R}^{n \times p}} -tr(X^T A^T A X^T) + \mu \|X\|_1 \qquad (28)$$
$$s.t \quad X^T X = I_r$$



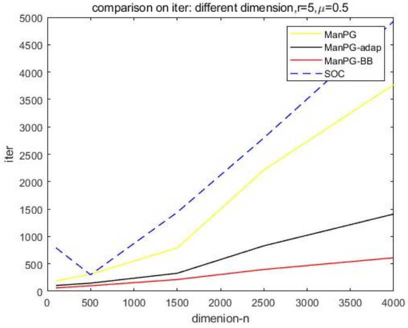
(a) CPU (b) Iteration

Figure 1: Comparison on SPCA problem with different n.

Numerical experiments:Non-monotone line search with BB steps



(a) CPU



(b) Iteration

Figure 2: Comparison on SPCA problem with different n.

Accelerated variant riemannian proximal gradient

Algorithm 4 (AccManPG)

Input： Lipschitz constant L on ∇f, $\mu \in (0, 1/L]$ in the proximal mapping,line
search parameter $\sigma \in (0, 1)$, shrinking parameter in line search $\nu \in (0, 1)$, positive integer N f
$t_0 = 1, y_0 = x_0, z_0 = x_0$;
for $k = 0, 1...$**do**
obtain V_k by solving the subproblem
if $mod(k, N) = 0$ **then** Invoke safeguard every N iterations
Invoke Algorithm 5:$[z_{k+N}, x_k, y_k, t_k] = Alg5(z_k, x_k, y_k, t_k, F(x_k))$;
end if
Compute $\eta_{y_k} = \arg\min_{\eta \in T_{y_k}\mathcal{M}} \langle \nabla f(y_k), \eta \rangle + \frac{1}{2\mu} \|\eta\|_F^2 + g(y_k + \eta)$;
$x_{k+1} = R_{y_k}(\eta_{y_k})$;
$t_{k+1} = \frac{\sqrt{4t_k^2+1}}{2}$
Compute $y_{k+1} = R_{x_{k+1}}(\frac{1-t_k}{t_{k+1}} R_{x_{k+1}}^{-1}(x_k))$;
end for

[4]Huang W , Wei K . Extending FISTA to riemannian optimization for sparse PCA. arXiv:1909.05485 .

Accelerated variant riemannian proximal gradient

Algorithm 5 (Safeguard for Alg4)

Input: $(z_k, x_k, y_k, t_k, F(x_k))$;

step 1: Compute $\eta_{z_k} = \underset{\eta \in T_{z_k}\mathcal{M}}{\arg\min} \langle \nabla f(z_k), \eta \rangle + \frac{1}{2\mu}\|\eta\|_F^2 + g(z_k + \eta)$;

Set $\alpha = 1$

while $F(R_{z_k}(\alpha\eta_{z_k})) > F(z_k) - \sigma\alpha\|\eta_{z_k}\|_F^2$ **do**

  $\alpha = \nu\alpha$;

end while

if $F(R_{z_k}(\alpha\eta_{z_k})) < F(x_k)$ **then**

  $x_k = R_{z_k}(\alpha\eta_{z_k}), y_k = R_{z_k}(\alpha\eta_{z_k}), t_k = 1$;

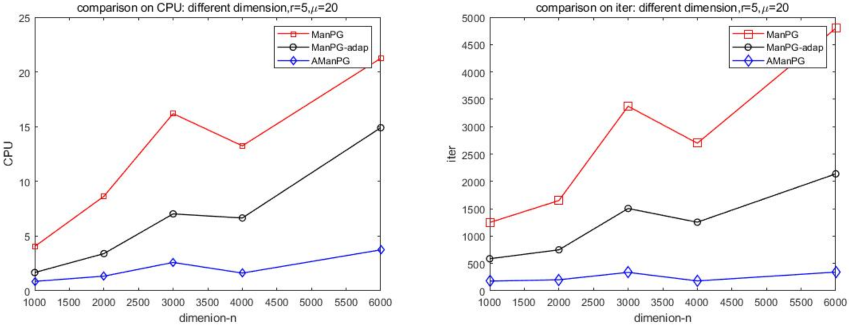
else

  x_k, y_k, t_k keep unchanged;

end if

  $z_{k+N} = x_k$;

Numerical experiments:AccManPG



(a) CPU (b) Iteration

Figure 3: Comparison on SPCA problem with different n.

References

Chen S , Ma S , So M C , et al. Proximal gradient method for nonsmooth optimization over the stiefel manifold[J]. SIAM Journal on Optimization, 2020, 30(1):210-239.

X. Xiao, Y. Li, Z. Wen, and L. Zhang, A regularized semi-smooth Newton method with projection steps for composite convex programs, J. Sci. Comput., 76 (2018), pp. 364–389.

Huang W , Wei K . Extending FISTA to riemannian optimization for sparse PCA. arXiv:1909.05485 .

Gutman Li X , Chen S , Deng Z , et al. Nonsmooth optimization over stiefel manifold: riemannian subgradient methods. arXiv:1911.05047v3.

Mannel, F., Rund, A. A hybrid semismooth quasi-Newton method for nonsmooth optimal control with PDEs.Optim Eng (2020).https://doi.org/10.1007/s11081-020-09523-w

References

Rongjie Lai and Stanley Osher. A splitting method for orthogonality constrained problems. Journal of Scientific Computing, 58(2):431–449, 2014.

Artiom Kovnatsky, Klaus Glashoff, and Michael M Bronstein. Madmm: a generic algorithm for non-smooth optimization on manifolds. In European Conference on Computer Vision, pages 680–696. Springer, 2016.

Weiqiang Chen, Hui Ji, and Yanfei You. An augmented lagrangian method for 1-regularized optimization problems with orthogonality constraints. SIAM Journal on Scientific Computing, 38(4):B570–B592, 2016.

Thank you!