带正交约束条件的优化问题算法及其应用

报告人：张金超

河北工业大学理学院

指导教师： 刘新为 教授

2020. 11. 20

目录

研究背景

正交约束优化问题一般形式

$$\begin{aligned} \min_{X\in\mathbb{R}^{n\times p}} \quad & f(X)+g(X) \\ \text{s.t.} \quad & X^\top X = I_p \quad (p \ll n) \end{aligned} \tag{1}$$

- $f:\mathbb{R}^{n\times p}\mapsto\mathbb{R}$,连续可微
- $g=0$或一般为非光滑凸函数
- Stiefel 流行: $S_{n,p}=\{X\in\mathbb{R}^{n\times p}|X^TX=I_p\}$

难点和挑战:

- 非凸约束
- NP-难(特殊的f)
- 保持可行(正交化)

正交约束优化的应用

- 最小p个特征值/奇异值计算：

$$
\begin{aligned}
&\min_{X\in\mathbb{R}^{n\times p}} \operatorname{tr}(X^\top AX) &\qquad& \min_{U\in\mathbb{R}^{m\times p}} \operatorname{tr}(U^\top AA^\top U) \\
&s.t.\quad X^\top X = I_p && s.t.\ U^\top U = I_p
\end{aligned} \tag{2}
$$

- 稀疏主成分分析：
  - 观测矩阵: $A\in\mathbb{R}^{n\times m}$

$$
\begin{aligned}
&\min_{X\in\mathbb{R}^{n\times p}} -tr(X^TA^TAX)+\mu\|X\|_1 \\
&s.t\quad X^TX = I_p
\end{aligned} \tag{3}
$$

- 稀疏盲去卷积问题：

$$
\begin{aligned}
&\min_{a,x} \|y-a\circledast x\|+\mu\|x\|_1 \\
&s.t\quad \|a\|_2 = 1
\end{aligned} \tag{4}
$$

其中$y\in\mathbb{R}^m$ 和$\circledast$ 表示某种卷积.

正交约束优化的应用(续)

- Bose-Einstein凝聚：

$$\begin{aligned}\min_{\phi\in\mathcal{S}} &\ E(\phi) \\ s.t. &\quad X^\top X = I_p\end{aligned} \xrightarrow[\text{离散模型}]{} \begin{aligned}\min_{x\in\mathbb{R}^n} f(x) := \frac{1}{2}x^\top Ax + \frac{\beta}{2}\sum_{i=1}^n x_i^4 \\ \text{s. t.} \quad \|x\|_2 = 1\end{aligned} \tag{5}$$

- 电子结构计算：

- 变分问题的压缩模型
- 低秩相关矩阵
- 联合对角化问题
- ……

研究背景-(动机)

黎曼流行优化 来点基本介绍

$$\begin{aligned} &\min_{X\in\mathbb{R}^{n\times p}} f(X) \\ &\text{s.t. } X \in \mathscr{S}_{n,p} \end{aligned} \tag{6}$$

研究现状

$$\text{核心思想}\begin{cases}\text{测地线类}\\ \text{投影类}\\ \text{不可行方法}\end{cases}$$

测地线类：

- 1998：*Edelmana,Ariasta*开辟了正交约束优化问题的研究领域，给出了测地线的具体表示:

$$\mathscr{R}_X^{\text{geoe}}(tD) = [X,Q]\exp\left(t\begin{bmatrix} -X^\top D & -R^\top \\ R & 0 \end{bmatrix}\right)\begin{bmatrix} I_p \\ 0 \end{bmatrix} \tag{7}$$

其中$exp(X)$表示矩阵指数函数

- 2005：*Nishimoriy,Akahos*基于Cayley变换提出了一种拟测地线更新公式.

$$\mathscr{R}_X^{\text{qgeo}}(tD) = \left(I + \tfrac{t}{2}A\right)^{-1}\left(I - \tfrac{t}{2}A\right)X \tag{8}$$

投影类:

- 2002: *Manton* 提出了Stiefel流行上的标准投影法,

$$\mathscr{R}_X^{\mathrm{pj}}(tD) = \mathscr{P}_{S_{n,p}}(X - tD) \tag{9}$$

其中$D \in \mathscr{T}_X\mathscr{S}_{n,p}$,等价于$X - tD$的奇异值分解.

- 2009: *Absil* 基于QR分解和极分解,提出了两类投影收缩方法.

$$\begin{aligned} \mathscr{R}_X^{\mathrm{qr}}(tD) &= \mathrm{qr}(X - tD) \\ \mathscr{R}_X^{\mathrm{pd}}(tD) &= (X - tD)\left(I_p + t^2 D^\top D\right)^{-1/2} \end{aligned} \tag{10}$$

$qr(X)$表示矩阵X进行QR分解之后的Q矩阵.

研究现状

已有的算法:

♣ 流行优化算法

- 最速下降法:[]
- 共轭梯度法:[]
- 牛顿法:[]
- 拟牛顿法:[]
- 信赖域算法:[]
- 测地线方法:[]
- 拟测地线、Cayley:[]

♣ 切空间搜索

- 直接投影法:[]
- 保约束可行法:[]

♣ 其他相关研究

- 算子分裂算法,ADMM, 邻近梯度算法:[]
- 随机算法:[]

♣ Absil-Mahony-Sepulchre, Optimization algorithms on matrix manifolds, Princeton University Press, 2008

研究现状(非光滑目标函数:$f(X)+g(X)$)

- 1998：*Ferreira and Oliveira*研究了对于在黎曼流行上凸函数求极小的次梯度方法收敛性.

- 2015：*Hosseini*基于Kurdyka-Lojasiewicz不等式性质研究了次梯度下降方法的收敛性

- 2017：*Hosseini and Uschmajew*提出了黎曼梯度采样算法，证明每一个迭代点的聚点都是Clarke 稳定点.

需要次梯度信息建立二次规划来找到一个下降方向

$$\hat{g} \longleftarrow \min_{g \in conv(W)} \|g\| \tag{11}$$

其中$conv(W)$是集合$W=\{G_j, j=1,\dots,J\}$的凸包，G_j是在当前点X周围可微点的黎曼梯度

研究现状

- 2014：*R. Lai and S. Osher*针对正交约束优化问题提出了算子分裂算法. 正交约束优化问题(1)的等价形式为：

$$\min F(Y) \qquad \text{s.t. } Y = X, X \in S_{n,p} \tag{12}$$

增广朗格朗日函数:

$$L_\beta(X,Y;\Lambda) := F(Y) - \langle \Lambda, Y - X \rangle + \frac{\beta}{2}\|Y - X\|_F^2 \tag{13}$$

主要迭代形式为结构:

$$\begin{aligned} Y^{k+1} &:= \arg\min_Y L_\beta\left(Y, X^k; \Lambda^k\right) \\ X^{k+1} &:= \arg\min_X L_\beta\left(Y^{k+1}, X; \Lambda^k\right) \text{ s.t. } X \in S_{n,p} \\ \Lambda^{k+1} &:= \Lambda^k - \beta(Y^{k+1} - X^{k+1}) \end{aligned} \tag{14}$$

- 2016：*A. Kovnatsky*提出了MADMM(manifold ADMM)算法对于问题：

$$\min_{X,Z} f(X)+g(Z) \qquad \text{s.t. } Z=AX, X\in S_{n,p} \tag{15}$$

增广朗格朗日函数:

$$L_\beta(X,Z;\Lambda) := f(X)+g(Z)-\langle \Lambda, Y-X\rangle + \frac{\beta}{2}\|Z-AX\|_F^2 \tag{16}$$

MADMM迭代形式为结构:

$$\begin{aligned} X^{k+1} &:= \arg\min_X L_\beta\left(X,Z^k;\Lambda^k\right) \text{ s.t. } X\in S_{n,p} \\ Z^{k+1} &:= \arg\min_X L_\beta\left(X^{k+1},Z;\Lambda^k\right) \\ \Lambda^{k+1} &:= \Lambda^k-\beta(Z^{k+1}-AX^{k+1}) \end{aligned} \tag{17}$$

- 2016：*W. Chen*提出了增广拉格朗日法与邻近交替极小结合的PAMAL方法：

$$\min_{X,Q,P} f(P)+g(Q) \qquad \text{s.t. } Q=X,\ P=X,\ X\in S_{n,p} \tag{18}$$

$$\begin{aligned} L_\beta(X,Q,P;\Lambda_1,\Lambda_2) := & f(P)+g(Q)-\langle\Lambda_1,Q-X\rangle-\langle\Lambda_2,P-X\rangle \\ & +\frac{\beta}{2}\|Q-X\|_F^2+\frac{\beta}{2}\|P-X\|_F^2 \end{aligned} \tag{19}$$

增广拉格朗日法求解形式为:

$$\begin{aligned} \left(X^{k+1},Q^{k+1},P^{k+1}\right) & := \arg\min_{X,Q,P} L_\beta\left(X,Q,P;\Lambda_1^k,\Lambda_2^k\right) \qquad \text{s.t. } X\in St_{n,p} \\ \Lambda_1^{k+1} & := \Lambda_1^k-\beta\left(Q^{k+1}-X^{k+1}\right) \\ \Lambda_2^{k+1} & := \Lambda_2^k-\beta\left(P^{k+1}-X^{k+1}\right) \end{aligned} \tag{20}$$

拉格朗日子问题仍然求解困难,结合邻近交替极小策略非精确求解该子问题！

研究现状

- 2017：*H. Zhu, D. Chu*针对广义正交约束优化问题，基于增广拉格朗日算法和PALM[] 提出了EPALMAL 算法，与PAMAl主要区别在于，增广拉格朗日子问题非精确求解是基于PALM同时也证明了算法产生迭代点序列的极限点都是KKT点.

- 2020：*S. Chen and S. Ma*等提出了收缩类邻近梯度法ManPG求解Stiefel流行上的非光滑问题.

$$
\begin{aligned}
&(1)V_k = \arg\min_{V} \langle gradf(x_k), V\rangle + \frac{1}{2t}\|V\|_F^2 + h(X_k + V) \\
&\qquad s.t. \quad V \in T_{X_k} S_{n,p} \\
&(2)X_{k+1} = Retr_{X_k(\alpha V_k)}
\end{aligned}
\tag{21}
$$

核心:其子问题(1)转换为求解非线性方程组,利用半光滑牛顿法求解[],以此达到更快的收敛速度

研究现状

Algorithm 1 (Manifold proximal gradient method (ManPG))

Input: initial point $X_0 \in \mathcal{M}, \gamma \in (0,1)$, stepsize $t > 0$.
for $k = 0, 1 \ldots$ **do**
  obtain V_k by semi-smooth newton solving the subproblem;
  set $\alpha = 1$
  while $F(Retr_{X_k}(\alpha V_k)) > F(X_k) - \frac{\alpha \|V_k\|_F^2}{2t}$ **do**
    $\alpha = \lambda \alpha$
  end while
  set $X_{k+1} = Retr_{X_k}(\alpha V_k)$
end for

研究现状

- 2020：*W. Huang*在ManPG的基础上,引入Nesterov加速技术,提出了加速版Acc-ManPG.

Algorithm 2 (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 for safeguard;
  $t_0 = 1, y_0 = x_0, z_0 = x_0$;
for $k = 0, 1 \ldots$ **do**
  obtain V_k by solving the subproblem
  if $mod(k,N) = 0$ **then** Invoke safeguard every N iterations
    Invoke Algorithm 3:$[z_{k+N}, x_k, y_k, t_k] = Alg5(z_k, x_k, y_k, t_k, F(x_k))$;
  end if
  Compute $\eta_{y_k} = \arg\min_{\eta \in T_{y_k}\mathscr{M}} \langle \nabla f(y_k), \eta \rangle + \frac{1}{2\mu} \|\eta\|_F^2 + g(y_k + \eta)$;
  $x_{k+1} = R_{y_k}(\eta_{y_k})$;
  $t_{k+1} = \frac{1+\sqrt{4t_k^2+1}}{2}$
  Compute $y_{k+1} = R_{x_{k+1}}(\frac{1-t_k}{t_{k+1}} R_{x_{k+1}}^{-1}(x_k))$;
end for

研究现状

Algorithm 3 (Safeguard for Alg4)

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

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

3: Set $\alpha = 1$

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

5: $\alpha = \nu \alpha$;

5:**end while**

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

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

8:**else**

9: x_k, y_k, t_k keep unchanged;

10:**end if**

11: $z_{k+N} = x_k$;

研究现状

- 2020：*N. Xiao, X. Liu, Y. Yuan*对于非光滑项为$\ell_{2,1}$范数的正交约束优化问题，提出了基于精确罚函数模型子问题具有显示表达式的非精确邻近梯度算法PenCPG，并且建立了全局收敛性.

★ 证明正交约束条件下非光滑问题的拉格朗日乘子在任意一阶稳定点X^*具有显示表达式$\Lambda(X^*)$

$$\Lambda(X) := \Phi(X^T \nabla f(X)) + \sum_{i=1}^{n} \gamma_i S(X_{i\cdot}^T) \tag{22}$$

$\Phi(X) = \frac{X+X^T}{2}$,$Q \in \mathbb{S}^{p\times p}$,$S_Q : \mathbb{S}^p \mapsto \mathbb{S}^{p\times p}$定义为

$$S_Q(x) := \begin{cases} \frac{xQx^\top}{\|x\|_2}, & \text{if } x \neq 0 \\ \mathbf{0}_{p,p}, & \text{otherwise} \end{cases}$$

研究现状

★ 构造精确罚模型作为增广拉格朗日函数

$$\mathscr{L}(X,\Lambda)=f(X)+g(X)-\frac{1}{2}\left\langle\Lambda,X^\top X-I_p\right\rangle+\frac{\beta}{4}\left\|X^\top X-I_p\right\|_\mathrm{F}^2 \tag{23}$$

★ 对精确罚函数引入紧凸约束，保证其有界性；

★ 对$\frac{1}{2}\left\langle\Lambda(Y),Y^\top Y-I_p\right\rangle$做线性化近似，在当前迭代点$X^k$近似为：

$$\left\langle Y-X^k,X^k\Lambda\left(X^k\right)\right\rangle+\frac{1}{2}\left\langle\Lambda\left(X^k\right),{X^k}^\top X^k-I_p\right\rangle \tag{24}$$

研究现状

Algorithm 4 (PenCPG)

Input:

Initialization: $X_0, k := 0, t_0 = 1$.

while: certain stopping criterion is not reached **do**:

Step 1: Run the iteration: $G^k := \nabla f(X^k) - X^k \Lambda(X^k) + \beta X^k({X^k}^T X^k - I_p)$;

Step 2: Choose stepsize η^k by certain strategy;

Step 3: Compute $Y_{i.}^k = \begin{cases} \frac{(\|X_i^k - \eta^k G_i^k\|_2 - \gamma_i \eta^k)}{\|X_i^k - \eta^k G_i^k\|_2} (X_{i.}^k - \eta^k G_{i.}^k), & \text{when } \|X_{i.}^k - \eta^k G_{i.}^k\|_2 \geq \gamma_i \eta^k \\ 0, & \text{otherwise.} \end{cases}$

Step 4: if $\|Y^k\|_F > K$, **then**

$X^{k+1} = \frac{K}{\|Y^k\|_F} Y^k$;

else

$X^{k+1} = Y^k$;

end if

Step 5: $k = k + 1$;

Output: X^k .

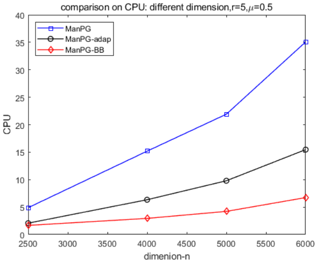
研究目标

- 在文献[*N*. *Xiao*, *X*. *Liu*, *Y*. *Yuan*,2020]所提出的精确罚框架中，基于作者对于精确罚函数构造的近似邻近映射所给出的显示解，计划借助Nesterov加速技术，实现加速版的PenCPG.

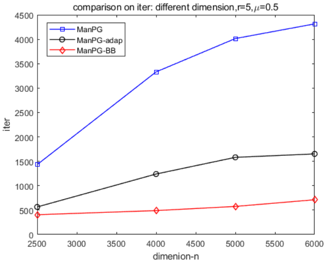
- 在文献[*N*. *Xiao*, *X*. *Liu*, *Y*. *Yuan*,2020]基础上，针对一般的非光滑项展开研究，此时拉格朗日乘子不一定会有显示解存在，则拟以精确罚函数为增广拉格朗日函数，拉格朗日乘子选择对偶上升步更新迭代，其增广朗格朗日子问题基于拟使用半光滑牛顿法进行求解.

研究成果

- 基于文献[*S. Chen and S. Ma*, 2020]所提出的ManPG算法，将其中的回溯线搜索更换为结合交替BB步长的非单调线搜索技术.



(a) CPU



(b) Iteration

Figure: Comparison on SPCA problem with different n.

Algorithm 5 (PenCPG with extrapolation)

Input:

Initialization: $X_0, Y_0, k := 0, t_0 = 1.$

while: certain stopping criterion is not reached **do**:

Step 1: Run the iteration: $G^k := \nabla f(X^k) - X^k \Lambda(X^k) + \beta X^k (X^{k^T} X^k - I_p);$

Step 2: Choose stepsize η^k by certain strategy;

Step 3: Compute $Y_{i.}^k = \begin{cases} \frac{\left(\|X_i^k - \eta^k G_i^k\|_2 - \gamma_i \eta^k\right)}{\|X_i^k - \eta^k G_i^k\|_2} \left(X_{i.}^k - \eta^k G_{i.}^k\right), & \text{when } \|X_{i.}^k - \eta^k G_{i.}^k\|_2 \geq \gamma_i \eta^k \\ 0, & \text{otherwise.} \end{cases}$

Step 4: $t_k = \frac{1+\sqrt{4t_k^2+1}}{2};$

$Y^k := Y^k + \frac{t_k - 1}{t_{k+1}}(Y^k - Y^{k-1});$

Step 5: **if** $\|Y^k\|_F > K$

$X^{k+1} = \frac{K}{\|Y^k\|_F} Y^k;$

else

$X^{k+1} = Y^k;$

end if

Step 6: $k = k+1;$

Output: X^k .

研究计划

- 2020 年 11 月底之前完成引入外推技术的算法实验、论文基本框架,月底完成初稿.
- 2020 年 12 月底之前完成A semi-smooth newton augmented Lagrangian method for nonsmooth optimization over the Stiefel Manifold初稿.
- 2021 年 1 月后继续研读带正交约束条件优化问题的相关文献, 深入理解问题本质,以及针对非线性规划其他算法例如SQP算法探讨其对于正交约束优化问题的适用性等, 进一步拓展到带正交约束条件的随机优化问题进行尝试研究.

参考文献

- [1] Y. Nesterov, A method of solving a convex programming problem with convergence rate$O(1/k^2)$, Soviet Math. Dokl., 27 (1983), pp. 372 – 376.
- [2] Y .Nesterov, ”An approach to constructing optimal methods for minimization of smooth convex functions”. èkonom. i mat. metody 3(1988):509-517.
- [3] Tseng, P.. (2008). On accelerated proximal gradient methods for convex-concave optimization. submitted to SIAM Journal on Optimization.

参考文献

[4] Beck A , Teboulle M . A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences, 2009, 2(1):183-202.

[5] Chambolle, A., Dossal, C. On the convergence of the iterates of the "Fast Iterative Shrinkage/Thresholding Algorithm" . J Optim Theory Appl 166, 968 – 982 (2015).

[6] Kim D, Fessler JA. Another look at the FAST ITERATIVE SHRINKAGE/THRESHOLDING ALGORITHM (FISTA). SIAM J Optim. 2018;28(1):223-250. doi:10.1137/16M108940X.

参考文献

- [7] Mai V V, Johansson M. Anderson acceleration of proximal gradient methods.[J]. arXiv: Optimization and Control, 2019.
- [8] Bauschke H H, Bolte J, Teboulle M, et al. A descent lemma beyond lipschitz gradient continuity: First-Order Methods Revisited and Applications[J]. Mathematics of Operations Research, 2017, 42(2): 330-348.
- [9] Gutman D H, Pena J F. Convergence rates of proximal gradient methods via the convex conjugate[J]. SIAM Journal on Optimization, 2019, 29(1): 162-174.
- [10] Yao Q . Efficient inexact proximal gradient algorithm for nonconvex problems[J]. 2016.

谢谢!