北京大学
PEKING UNIVERSITY

A NEWTON-CG AUGMENTED LAGRANGIAN METHOD FOR SEMIDEFINITE PROGRAMMING*

XIN-YUAN ZHAO†, DEFENG SUN‡, AND KIM-CHUAN TOH§

彭秀辉（1401111549） 王琦少（1401111550）
王鹏昊 （1401214529）

目录

- 背景
- 本文的工作
  - 实用的共轭梯度（CG）算法
  - 半光滑Newton-CG算法
  - 半光滑Newton-CG增广拉格朗日算法
  - 收敛性分析
- 数值实验

背景

- 研究的问题：

利用增广拉格朗日方法解决如下SDP问题

$$\min\left\{b^{\mathrm{T}}y \mid \mathcal{A}^{*}y - C \succeq \mathbf{0}\right\},$$

对偶问题

$$\max\left\{\langle C, X\rangle \mid \mathcal{A}(X) = b, \quad X \succeq \mathbf{0}\right\}.$$

背景

- 增广拉格朗日方法

引入罚变量$\sigma > 0$，原问题的增广拉格朗日函数为

$$L_\sigma(y, X) = b^{\mathrm{T}} y + \frac{1}{2\sigma}(\|\Pi_{\mathcal{S}_+^n}(X - \sigma(\mathcal{A}^* y - C))\|^2 - \|X\|^2),\ (y, X) \in \Re^m \times \mathcal{S}^n,$$

给定$X^0 \in \mathcal{S}^n, \sigma_0 > 0, \rho > 1$，增广拉格朗日方法产生的序列$\{y^k\} \subset \mathcal{R}^m$和$\{X^k\} \subset \mathcal{S}^n$满足：

$$\begin{cases} y^{k+1} \approx \arg\min_{y \in \Re^m} L_{\sigma_k}(y, X^k), \\ X^{k+1} = \Pi_{\mathcal{S}_+^n}(X^k - \sigma_k(\mathcal{A}^* y^{k+1} - C)), \qquad k = 0, 1, 2, \ldots, \\ \sigma_{k+1} = \rho\sigma_k \text{ or } \sigma_{k+1} = \sigma_k. \end{cases}$$

本文的工作

- 算法

实用的共轭梯度（CG）算法

半光滑Newton-CG算法

半光滑Newton-CG增广拉格朗日算法

收敛性分析

实用的共轭梯度（CG）算法

- 考虑线性方程组

$$Ax = b\,,$$

其中$b \in \mathcal{R}^m$，而$A \in \mathcal{R}^{m\times m}$是对称正定矩阵.

- 实用的共轭梯度算法用到两个参数$i_{max} > 0$和罚函数$\eta \in (0, \|b\|)$.

ALGORITHM 1. A practical CG algorithm $(CG(\eta, i_{\max}))$.
Given $x^0 = 0$ and $r^0 = b$.
While $(\|r^i\| > \eta)$ or $(i < i_{\max})$
Step 1.1. $i = i + 1$
Step 1.2. If $i = 1$; $p^1 = r^0$; else; $\beta_i = \|r^{i-1}\|^2/\|r^{i-2}\|^2$, $p^i = r^{i-1} + \beta_i\, p^{i-1}$; end
Step 1.3. $\alpha_i = \|r^{i-1}\|^2/\langle p^i, Ap^i\rangle$
Step 1.4. $x^i = x^{i-1} + \alpha_i p^i$
Step 1.5. $r^i = r^{i-1} - \alpha_i A p^i$

- 算法得到的序列$\{x^i\}$上下界有估计式

$$\frac{1}{\lambda_{\max}(A)} \le \frac{\langle x^i, b\rangle}{\|b\|^2} \le \frac{1}{\lambda_{\min}(A)},$$

其中$\lambda_{max}(A)$和$\lambda_{min}(A)$分别是矩阵A的最大和最小特征值.

实用的共轭梯度（CG）算法

第i次迭代的误差 $e^i = x^* - x^i$

搜索方向是互相正交的 $\langle r^i, r^j \rangle = 0 \quad \text{for } j = 1, 2, \ldots, i-1,$

根据 $r^0 = b$, 和β_i; 的定义可以得到 $p^i = r^{i-1} + \beta_i p^{i-1}$

$$\langle p^1, b \rangle = \|r^0\|^2,$$

$$\langle p^i, b \rangle = \langle r^{i-1}, b \rangle + \beta_i \langle p^{i-1}, b \rangle = 0 + \prod_{j=2}^{i} \beta_j \langle p^1, b \rangle = \|r^{i-1}\|^2 \quad \forall i > 1.$$

已知定理 $\|e^{i-1}\|_A^2 = \|e^i\|_A^2 + \langle \alpha_i p^i, A(\alpha_i p^i) \rangle,$

根据算法第三步可以得到 $\alpha_i \|r^{i-1}\|^2 = \langle \alpha_i p^i, A(\alpha_i p^i) \rangle$

所以得到 $\alpha_i \|r^{i-1}\|^2 = \|e^{i-1}\|_A^2 - \|e^i\|_A^2.$

实用的共轭梯度（CG）算法

$$\|x\|_A := \sqrt{\langle x, Ax\rangle}. \quad x^i = x^{i-1} + \alpha_i p^i$$

利用上面的公式可以得到

$$\langle x^i, b\rangle = \langle x^{i-1}, b\rangle + \alpha_i\langle p^i, b\rangle = \langle x^0, b\rangle + \sum_{j=1}^{i} \alpha_j\langle p^j, b\rangle = \sum_{j=1}^{i} \alpha_j \|r^{j-1}\|^2$$

$$= \sum_{j=1}^{i} \left[\|e^{j-1}\|_A^2 - \|e^j\|_A^2\right] = \|e^0\|_A^2 - \|e^i\|_A^2,$$

可以推出 $\langle x^i, b\rangle \geq \langle x^{i-1}, b\rangle, \ i = 1, 2, \ldots, \bar{\imath}.$

$$\frac{1}{\lambda_{\max}(A)} \leq \alpha_1 = \frac{\langle x^1, b\rangle}{\|b\|^2} \leq \frac{\langle x^i, b\rangle}{\|b\|^2}.$$

实用的共轭梯度（CG）算法

因为 $e^0 = x^* - x^0 = A^{-1}b,$

$1 \le i \le \bar{i},$

$$\frac{\langle x^i, b\rangle}{\|b\|^2} \le \frac{\|e^0\|_A^2}{\|b\|^2} = \frac{\|A^{-1}b\|_A^2}{\|b\|^2} \le \frac{1}{\lambda_{\min}(A)}.$$

综上所述

$$\frac{1}{\lambda_{\max}(A)} \le \frac{\langle x^i, b\rangle}{\|b\|^2} \le \frac{1}{\lambda_{\min}(A)},$$

半光滑Newton-CG方法

为了应用增广的拉格朗日方法求解原问题,对于固定的$X \in S^n$和$\sigma > 0$考虑如下的凸内问题

$$\min\left\{\varphi(y) := L_\sigma(y, X) \mid y \in \Re^m\right\}.$$

其中$\varphi(\cdot)$是连续可微的凸函数,有

$$\nabla\varphi(y) = b - \mathcal{A}\Pi_{\mathcal{S}_+^n}(X - \sigma(\mathcal{A}^* y - C)) = 0$$

因为$\prod_{\mathcal{S}_+^n}(\cdot)$是不光滑的,所以$\nabla\varphi(y)$是不连续可微的.但$\prod_{\mathcal{S}_+^n}(\cdot)$是强半光滑的,所以可以用半光滑的Newton-CG方法求解上述非线性方程组.

半光滑Newton-CG方法

- Hessian矩阵

$\Pi_{\mathcal{S}_+^n}(\cdot)$是Lipschitz连续的且模为1. φ在y的广义Hessian矩阵定义为

$$\partial^2\varphi(y) := \partial(\nabla\varphi)(y),$$

其中$\partial(\nabla\varphi)(y)$是$\nabla\varphi$在y的Clarke广义雅克比矩阵.由于$\partial^2\varphi(y)$表示困难,所以定义$\hat{\partial}^2\varphi(y)$代替$\partial^2\varphi(y)$

$$\hat{\partial}^2\varphi(y)| := \sigma\mathcal{A}\,\partial\Pi_{\mathcal{S}_+^n}(X - \sigma(\mathcal{A}^*y - C))\mathcal{A}^*.$$

并且，对于$d \in \mathcal{R}^m$

$$\partial^2\varphi(y)d \subseteq \hat{\partial}^2\varphi(y)d,$$

上式表示,如果$\hat{\partial}^2\varphi(y)$的每个元素是正定的,则$\partial^2\varphi(y)$也是.因此，$\hat{\partial}^2\varphi(y)$代替$\partial^2\varphi(y)$是可行的.

半光滑Newton-CG方法

- 替代Hessian矩阵中元素的表示方法

为了应用半光滑的Newton-CG方法，需要$\hat{\partial}^2\varphi(y)$中的一个元素，将对称矩阵$X-\sigma(\mathcal{A}^*y-C)$正交对角化

$$X-\sigma(\mathcal{A}^*y-C)=Q\,\Gamma_y\,Q^{\mathrm{T}},$$

其中Γ_y是对角矩阵且对角元是$X-\sigma(\mathcal{A}^*y-C)$的特征值$\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n$.
定义指标集合

$$\alpha:=\{i\mid\lambda_i>0\},\quad\beta:=\{i\mid\lambda_i=0\},\quad\text{and}\quad\gamma:=\{i\mid\lambda_i<0\}.$$

则可得到$V_y^0\in\hat{\partial}^2\varphi(y)$

$$V_y^0 d:=\sigma\mathcal{A}[\,Q(\Omega\circ(\,Q^{\mathrm{T}}(\mathcal{A}^*d)\,Q))\,Q^{\mathrm{T}}],\quad d\in\Re^m.$$

其中“∘”为两个矩阵的Hadamard积.

$$\Omega=\begin{bmatrix}E_{\alpha\alpha} & \nu_{\alpha\bar{\alpha}}\\ \nu_{\alpha\bar{\alpha}}^{\mathrm{T}} & 0\end{bmatrix},\quad \nu_{ij}:=\frac{\lambda_i}{\lambda_i-\lambda_j},\; i\in\alpha, j\in\bar{\alpha},$$

其中$\bar{\alpha}=\{1,\cdots,n\}\backslash\alpha, E^{\alpha\alpha}\in\mathcal{S}^{|\alpha|}$是元素为1的矩阵.

半光滑Newton-CG方法

- 替代Hessian矩阵中元素的正定性

文中命题3.2说明了替代Hessian矩阵中元素的正定性等价于对应对偶问题的约束非退化条件成立,又等价于原问题非退化条件.因此,在满足约束非退化条件的情况下，Hessian矩阵正定,半光滑Newton-CG方法求得极小.

PROPOSITION 3.2. *Suppose that the problem (56) satisfies condition (7). Let* $(\hat{y}, \widehat{Z}) \in \Re^m \times \mathcal{S}^n$ *be a pair that satisfies the KKT conditions (57), and let* P *be an orthogonal matrix such that* $\widehat{Z}$ *and* $\widehat{Z} - (X - \sigma(\mathcal{A}^* \hat{y} - C))$ *have the spectral decomposition as in (14). Then the following conditions are equivalent:*

(i) *The constraint nondegenerate condition*

$$\mathcal{A} \operatorname{lin}(\mathcal{T}_{\mathcal{S}^n_+}(\widehat{Z})) = \Re^m$$

holds at $\widehat{Z}$, *where* $\operatorname{lin}(\mathcal{T}_{\mathcal{S}^n_+}(\widehat{Z}))$ *denotes the lineality space of* $\mathcal{T}_{\mathcal{S}^n_+}(\widehat{Z})$, *i.e.,*

$$\operatorname{lin}(\mathcal{T}_{\mathcal{S}^n_+}(\widehat{Z})) = \{B \in \mathcal{S}^n \mid [P_\beta \; P_\gamma]^T B \, [P_\beta \; P_\gamma] = 0\}.$$

(ii) *Every* $V_{\hat{y}} \in \hat{\partial}^2 \varphi(\hat{y})$ *is symmetric and positive definite.*

(iii) $V^0_{\hat{y}} \in \hat{\partial}^2 \varphi(\hat{y})$ *is symmetric and positive definite.*

半光滑Newton-CG方法

- 算法

Algorithm 2. A semismooth Newton-CG algorithm ($NCG(y^0, X, \sigma)$).

Step 0. Given $\mu \in (0, 1/2)$, $\bar{\eta} \in (0, 1)$, $\tau \in (0, 1]$, $\tau_1, \tau_2 \in (0, 1)$, and $\delta \in (0, 1)$.

Step 1. For $j = 0, 1, 2, \ldots$

Step 1.1. Given a maximum number of CG iterations $n_j > 0$, compute

$$\eta_j := \min(\bar{\eta}, \|\nabla\varphi(y^j)\|^{1+\tau}).$$

Apply the practical CG algorithm, Algorithm 1 ($CG(\eta_j, n_j)$), to find an approximation solution d^j to

$$(V_j + \varepsilon_j I)\, d = -\nabla\varphi(y^j),$$

where $V_j \in \hat{\partial}^2\varphi(y^j)$ is defined in (54) and $\varepsilon_j := \tau_1 \min\{\tau_2, \|\nabla\varphi(y^j)\|\}$.

Step 1.2. Set $\alpha_j = \delta^{m_j}$, where m_j is the first nonnegative integer m for which

$$\varphi(y^j + \delta^m d^j) \le \varphi(y^j) + \mu\delta^m \langle\nabla\varphi(y^j), d^j\rangle.$$

Step 1.3. Set $y^{j+1} = y^j + \alpha_j\, d^j$.

半光滑Newton-CG方法的收敛性

- 收敛性分析

算法中使用的是线性搜索方法.搜索方向是用CG算法解线性方程组得到的.满足CG算法序列上下界的估计式,由此得到命题3.3.

PROPOSITION 3.3. *For every $j \geq 0$, the search direction d^j generated in Step 1.2 of Algorithm 2 satisfies*

$$\frac{1}{\lambda_{\max}(\widetilde{V_j})|} \leq \frac{\langle -\nabla\varphi(y^j), d^j\rangle}{\|\nabla\varphi(y^j)\|^2} \leq \frac{1}{\lambda_{\min}(\widetilde{V_j})},$$

where $\widetilde{V}_j := V_j + \varepsilon_j I$ and $\lambda_{\max}(\widetilde{V}_j)$ and $\lambda_{\min}(\widetilde{V}_j)$ are the largest and smallest eigenvalues of $\widetilde{V}_j$, respectively.

由于$\tilde{V}_j$是正定矩阵，故d^j与$\nabla\varphi(y^j)$方向相反，故搜索方向始终是下降方向.作者引用文献说明了水平集$\mathcal{L} := \{y \in \mathcal{R}^m | \varphi(y) \leq \varphi(y^0)\}$是一个闭凸集.因此序列$\{y_j\}$有界.令$\hat{y}$是序列$\{y_j\}$的任一聚点,由命题3.3和$\nabla\varphi(\cdot)$的Lipschitz连续性,容易得到$\nabla\varphi(\hat{y}) = 0$.根据$\varphi(\cdot)$的凸性，得到$\hat{y}$是最优解.即下面的收敛性定理.

THEOREM 3.4. *Suppose that problem (56) satisfies condition (7). Then Algorithm 2 is well defined and any accumulation point $\hat{y}$ of $\{y^j\}$ generated by Algorithm 2 is an optimal solution to the inner problem (46).*

半光滑Newton-CG方法

- 收敛速率

先给出收敛速率定理

THEOREM 3.5. *Assume that problem (56) satisfies condition (7). Let $\hat{y}$ be an accumulation point of the infinite sequence $\{y^j\}$ generated by Algorithm 2 for solving the inner problem (46). Suppose that at each step $j \geq 0$, when the practical CG algorithm, Algorithm 1, terminates, the tolerance η_j is achieved (e.g., when $n_j = m+1$); i.e.,*

$$\|\nabla\varphi(y^j) + (V_j + \varepsilon_j I)\, d^j\| \leq \eta_j .$$

Assume that the constraint nondegenerate condition (58) holds at $\widehat{Z} := \Pi_{\mathcal{S}_+^n}(X - \sigma(\mathcal{A}^*\hat{y} - C))$. Then the whole sequence $\{y^j\}$ converges to $\hat{y}$ and

$$\|y^{j+1} - \hat{y}\| = O(\|y^j - \hat{y}\|^{1+\tau}).$$

由上面的定理可以看到,半光滑Newton-CG算法的收敛速率为$(1+\tau)$阶的,

当$\tau = 1$时,对应二阶收敛.为了节约计算时间,实际选择$\tau = 0.1\sim0.2$,仍能保证

算法是超线性收敛的.

半光滑Newton-CG方法

- 收敛速率定理证明

Since $\Pi_{\mathcal{S}^n_+}(\cdot)$ is strongly semismooth [36], it holds that for all j sufficiently large,

$$\begin{aligned}
\|y^j + d^j - \hat{y}\| &= \|y^j + (V_j + \varepsilon_j I)^{-1}((\nabla\varphi(y^j) + (V_j + \varepsilon_j I)\, d^j) - \nabla\varphi(y^j)) - \hat{y}\| \\
&\leq \|y^j - \hat{y} - (V_j + \varepsilon_j I)^{-1}\nabla\varphi(y^j)\| + \|(V_j + \varepsilon_j I)^{-1}\|\, \|\nabla\varphi(y^j) + (V_j + \varepsilon_j I)\, d^j\| \\
&\leq \|(V_j + \varepsilon_j I)^{-1}\|(\|\nabla\varphi(y^j) - \nabla\varphi(\hat{y}) - V_j(y^j - \hat{y})\| + \varepsilon_j\|y^j - \hat{y}\| + \eta_j) \\
&\leq O(\|\mathcal{A}\|\|\Pi_{\mathcal{S}^n_+}(X - \sigma(\mathcal{A}^* y^j - C)) - \Pi_{\mathcal{S}^n_+}(X - \sigma(\mathcal{A}^*\hat{y} - C)) - W_j(\sigma\mathcal{A}^*(y^j - \hat{y}))\|) \\
&\quad + O(\tau_1\|\nabla\varphi(y^j)\|\|y^j - \hat{y}\| + \|\nabla\varphi(y^j)\|^{1+\tau}) \\
&\leq O(\|\sigma\mathcal{A}^*(y^j - \hat{y})\|^2) + O(\tau_1\, \|\nabla\varphi(y^j) - \nabla\varphi(\hat{y})\|\|y^j - \hat{y}\| + \|\nabla\varphi(y^j) - \nabla\varphi(\hat{y})\|^{1+\tau}) \\
&\leq O(\|y^j - \hat{y}\|^2) + O(\tau_1\sigma\|\mathcal{A}\|\|\mathcal{A}^*\|\|y^j - \hat{y}\|^2 + (\sigma\|\mathcal{A}\|\|\mathcal{A}^*\|\|y^j - \hat{y}\|)^{1+\tau}) \\
&= O(\|y^j - \hat{y}\|^{1+\tau}),
\end{aligned}$$

半光滑Newton-CG方法

- 收敛速率定理证明

① $\|y^j + d^j - \hat{y}\| = O\left(\|y^j - \hat{y}\|^{1+\tau}\right) \Longrightarrow \quad y^j - \hat{y} = -d^j + O\left(\|d^j\|^{1+\tau}\right)$

② $\langle \nabla\varphi(y^j), d^j \rangle + \langle d^j, (V_j + \varepsilon_j I)\, d^j \rangle$

$$\leq \eta_j \|d^j\| \;\leq\; \|\nabla\varphi(y^j)\|^{1+\tau}\|d^j\| \;=\; \|\nabla\varphi(y^j) - \nabla\varphi(\hat{y})\|^{1+\tau}\|d^j\|$$

$$\leq \sigma\|d^j\|\|\mathcal{A}\|\|\mathcal{A}^*\|\|y^j - \hat{y}\|^{1+\tau}$$

$$\leq O(\|d^j\|^{2+\tau}),$$

结合上面两个式子,以及$\|(V_j+\varepsilon_j I)^{-1}\|$是一致有界,可以得出存在$\hat{\delta} > 0$使

$$-\langle \nabla\varphi(y^j), d^j \rangle \geq \hat{\delta}\|d^j\|^2 \qquad \forall\ j \text{ sufficiently large.}$$

半光滑Newton-CG方法

- 收敛速率定理证明

由参考文献,对于$\mu \in (0, 1/2)$,存在j_0,当迭代次数$j \geq j_0$时,恒有

$$\varphi(y^j + d^j) \leq \varphi(y^j) + \mu\langle\nabla\varphi(y^j), d^j\rangle,$$

即算法中$\alpha_j = 0$，即

$$y^{j+1} = y^j + d^j.$$

前面已经得到$\|y^j + d^j - \hat{y}\| = o\left(\|y^j - \hat{y}\|^{1+\tau}\right)$,故

$$\|y^{j+1} - \hat{y}\| = O(\|y^j - \hat{y}\|^{1+\tau})$$

得证.

Newton-CG的增广拉格朗日方法

- SDPNAL算法是给定一个非负的单调不减序列以及初始点后，产生满足以下条件的序列点 (y^k, X^k) 直找到最优解。

$$
\begin{cases}
y^{k+1} \approx \arg\min_{y \in \Re^m} L_{\sigma_k}(y, X^k), \\
X^{k+1} = \Pi_{\mathcal{S}_+^n}(X^k - \sigma_k(\mathcal{A}^* y^{k+1} - C)), \qquad k = 0, 1, 2, \ldots, \\
\sigma_{k+1} = \rho\sigma_k \text{ or } \sigma_{k+1} = \sigma_k.
\end{cases}
$$

Newton-CG的增广拉格朗日方法

- SDPNAL算法的终止准则

对于任意的 $k \geq 0,\ \varphi_k(\cdot) \equiv L_{\sigma_k}(\cdot, X^k)$. ，因为求解内问题没有用到精确解，所以将用到下面的终止准则作为SPDNAL算法的终止准则。

(A) $\varphi_k(y^{k+1}) - \inf \varphi_k \leq \epsilon_k^2/2\sigma_k,\ \epsilon_k \geq 0,\ \sum_{k=0}^{\infty} \epsilon_k < \infty.$

(B) $\varphi_k(y^{k+1}) - \inf \varphi_k \leq (\delta_k^2/2\sigma_k)\|X^{k+1} - X^k\|^2,\ \delta_k \geq 0,\ \sum_{k=0}^{\infty} \delta_k < \infty.$

(B′) $\|\nabla \varphi_k(y^{k+1})\| \leq (\delta_k^{'}/\sigma_k)\|X^{k+1} - X^k\|,\ 0 \leq \delta_k^{'} \to 0.$

- SDPNAL算法

Step 0. Given $(y^0, X^0) \in \Re^m \times \mathcal{S}_+^n$, $\sigma_0 > 0$, a threshold $\bar{\sigma} \geq \sigma_0 > 0$, and $\rho > 1$.

Step 1. For $k = 0, 1, 2, \ldots$

Step 1.1. Starting with y^k as the initial point, apply Algorithm 2 to $\varphi_k(\cdot)$ to find $y^{k+1} = \mathrm{NCG}(y^k, X^k, \sigma_k)$ and $X^{k+1} = \Pi_{\mathcal{S}_+^n}(X^k - \sigma_k(\mathcal{A}^* y^{k+1} - C))$ satisfying (A), (B), or (B′).

Step 1.2. If $\sigma_k \leq \bar{\sigma}$, $\sigma_{k+1} = \rho\,\sigma_k$ or $\sigma_{k+1} = \sigma_k$.

Newton-CG的增广拉格朗日方法

$$\mathcal{A}^* z^0 - C \succ \mathbf{0}, \tag{69}$$

$$\begin{cases} \mathcal{A} : \mathcal{S}^n \to \Re^m \ \textit{is onto}, \\ \exists X_0 \in \mathcal{S}^n_+ \ \textit{such that}\ \mathcal{A}(X_0) = b,\ X_0 \succ \mathbf{0}, \end{cases} \tag{7}$$

$$\mathcal{A}^* \Re^m + \operatorname{conv}\left(\bigcup_{y \in \mathcal{M}(\overline{X})} \left(\mathcal{T}_{\mathcal{S}^n_+}(\mathcal{A}^* y - C) \cap \overline{X}^{\perp} \right) \right) = \mathcal{S}^n, \tag{23}$$

$$\mathcal{A}^* \Re^m + \operatorname{lin}(\mathcal{T}_{\mathcal{S}^n_+}(\mathcal{A}^* \bar{y} - C)) = \mathcal{S}^n. \tag{36}$$

$$\mathcal{A} \operatorname{lin}(\mathcal{T}_{\mathcal{S}^n_+}(\widehat{Z})) = \Re^m \tag{58}$$

Newton-CG的增广拉格朗日方法

SDPNAL算法的收敛性分析

THEOREM 4.2. *Let Algorithm 3 be executed with stopping criteria* (A) *and* (B). *Assume that* (*D*) *satisfies condition* (69) *and* (*P*) *satisfies condition* (7). *If the extended strict primal-dual constraint qualification* (23) *holds at* $\overline{X}$, *where* $\overline{X}$ *is an optimal solution to* (*P*), *then the generated sequence* $\{X^k\} \subset \mathcal{S}^n_+$ *is bounded and* $\{X^k\}$ *converges to the unique solution* $\overline{X}$ *with* $\max(P) = \min(D)$, *and*

$$\|X^{k+1} - \overline{X}\| \leq \theta_k \|X^k - \overline{X}\| \quad \forall\ k \text{ sufficiently large},$$

where

$$\theta_k = \left[a_g(a_g^2 + \sigma_k^2)^{-1/2} + \delta_k\right](1 - \delta_k)^{-1} \to \theta_\infty = a_g(a_g^2 + \sigma_\infty^2)^{-1/2} < 1 \quad \text{as } \sigma_k \to \sigma_\infty\,,$$

and a_g *is a Lipschitz constant of* $-(\partial g)^{-1}$ *at the origin (cf. Proposition* 2.1*). The conclusions of Theorem* 4.1 *about* $\{y^k\}$ *are valid.*

Moreover, if the stopping criterion (B′) *is also used and the constraint nondegenerate conditions* (36) *and* (58) *hold at* $\bar{y}$ *and* $\overline{X}$, *respectively, then, in addition to the above conclusions, the sequence* $\{y^k\} \to \bar{y}$, *where* $\bar{y}$ *is the unique optimal solution to* (*D*), *and one has*

$$\|y^{k+1} - \bar{y}\| \leq \theta'_k \|X^{k+1} - X^k\| \quad \forall\ k \text{ sufficiently large},$$

where $\theta'_k = a_l(1 + \delta'_k)/\sigma_k \to \delta_\infty = a_l/\sigma_\infty$ *and* a_l *is a Lipschitz constant of* T_l^{-1} *at the origin.*

定理4.2的第一部分证明是由文献32,33结合命题2得出的，结合极大单调算子的逆算子是Lipschitz连续性和文献7得出了定理的第二部分证明。

Newton-CG的增广拉格朗日方法

$$\begin{cases} y^{k+1} \approx \arg\min_{y\in\Re^m} L_{\sigma_k}(y, X^k), \\ X^{k+1} = \Pi_{\mathcal{S}_+^n}(X^k - \sigma_k(\mathcal{A}^* y^{k+1} - C)), \qquad k = 0, 1, 2, \ldots, \\ \sigma_{k+1} = \rho\sigma_k \text{ or } \sigma_{k+1} = \sigma_k. \end{cases}$$

对上式添加 $\frac{1}{2\sigma_k}\|y - y^k\|^2$ to $L_{\sigma_k}(y, X^k)$，则 $L_{\sigma_k}(y, X^k) + \frac{1}{2\sigma_k}\|y - y^k\|^2$ 是一个强凸函数。

等价于如下的迭代：

$$\begin{cases} y^{k+1} \approx \arg\min_{y\in\Re^m} \left\{L_{\sigma_k}(y, X^k) + \frac{1}{2\sigma_k}\|y - y^k\|^2\right\}, \\ X^{k+1} = \Pi_{\mathcal{S}_+^n}(X^k - \sigma_k(\mathcal{A}^* y^{k+1} - C)), \\ \sigma_{k+1} = \rho\sigma_k \quad \text{or} \quad \sigma_{k+1} = \sigma_k. \end{cases}$$

收敛性分析与上面相同。

数值试验分析

- 应用半光滑的Newton-CG的增广拉格朗日方法（SPDNAL）求解各种SDP问题时，计算原始问题与对偶问题的不可行性与最优性约束如下表示：

$$R_D = \frac{\|C + S - \mathcal{A}^* y\|}{1 + \|C\|}, \quad R_P = \frac{\|b - \mathcal{A}(X)\|}{1 + \|b\|}, \quad \text{gap} = \frac{b^T y - \langle C, X\rangle}{1 + |b^T y| + |\langle C, X\rangle|},$$

where $S = (\Pi_{\mathcal{S}^n_+}(W) - W)/\sigma$ with $W = X - \sigma(\mathcal{A}^* y - C)$

- 数值试验中，SPDNAL算法的终止准则为：

$$\max\{R_D, R_P\} \le 10^{-6}.$$

choose the initial iterate $y^0 = 0$, $X^0 = 0$, and $\sigma_0 = 10$.

数值试验分析—random sparse

TABLE 1
Results for the SDPNAL algorithm on the random sparse SDP problems considered in [19].

Problem	$m \mid n_s; n_l$	it| itsub| pcg	$\langle C, X \rangle$	$b^T y$	R_P| R_D| gap	Time
Rn6m50p3	50000 | 600;	10| 50| 58.2	-3.86413091 2	-3.86353173 2	2.8-7| 8.5-7| -7.7-5	7:53
Rn6m60p3	60000 | 600;	9| 47| 48.3	6.41737682 2	6.41803361 2	5.0-7| 8.7-7| -5.1-5	7:00
Rn7m50p3	50000 | 700;	12| 52| 31.6	3.13203609 2	3.13240876 2	7.4-7| 5.4-7| -5.9-5	6:18
Rn7m70p3	70000 | 700;	10| 48| 41.6	-3.69557843 2	-3.69479811 2	2.4-7| 8.7-7| **-1.1-4**	8:48
Rn8m70p3	70000 | 800;	11| 51| 33.3	2.33139641 3	2.33149302 3	1.8-7| 9.9-7| -2.1-5	9:37
Rn8m100p3	100000 | 800;	10| 52| 55.8	2.25928848 3	2.25937157 3	1.3-7| 7.3-7| -1.8-5	18:49

TABLE 2
Results obtained by the boundary-point method in [19] on the random sparse SDP problems considered therein. The parameter σ_0 is set to 0.1, which gives better timings than the default initial value of 1.

Problem	$m \mid n_s; n_l$	it	$\langle C, X \rangle$	$b^T y$	R_P| R_D| gap	Time
Rn6m50p3	50000 | 600;	142	-3.86413897 2	-3.86413511 2	9.9-7| 5.7-8| -5.0-7	1:21
Rn6m60p3	60000 | 600;	137	6.41736718 2	6.41736746 2	9.9-7| 3.0-8| -2.2-8	2:09
Rn7m50p3	50000 | 700;	165	3.13202583 2	3.13205602 2	9.9-7| 1.1-7| -4.8-6	2:07
Rn7m70p3	70000 | 700;	136	-3.69558765 2	-3.69558700 2	9.9-7| 4.2-8| -8.9-8	2:10
Rn8m70p3	70000 | 800;	158	2.33139551 3	2.33139759 3	9.9-7| 8.3-8| -4.5-7	2:54
Rn8m100p3	100000 | 800;	135	2.25928693 3	2.25928781 3	9.4-7| 2.9-8| -1.9-7	4:16

数值试验分析—FAPs

TABLE 3
Results for the SDPNAL algorithm on the FAPs.

Problem	m | $n_s; n_l$	it| itsub| pcg	$\langle C, X\rangle$	$b^T y$	R_P| R_D| gap	Time
fap09	15225 | 174; 14025	22| 120| 38.4	1.07978114 1	1.07978423 1	8.9-7| 9.6-7| -1.4-6	41
fap10	14479 | 183; 13754	23| 140| 57.4	9.67044948-3	9.74974306-3	1.5-7| 9.3-7| -7.8-5	1:18
fap11	24292 | 252; 23275	25| 148| 69.0	2.97000004-2	2.98373492-2	7.7-7| 6.0-7| **-1.3-4**	3:21
fap12	26462 | 369; 24410	25| 169| 81.3	2.73251961-1	2.73410714-1	6.0-7| 7.8-7| **-1.0-4**	9:07
fap25	322924 | 2118; 311044	24| 211| 84.8	1.28761356 1	1.28789892 1	3.2-6| 5.0-7| **-1.1-4**	10:53:22
fap36	1154467 | 4110; 1112293	17| 197| 87.4	6.98561787 1	6.98596286 1	7.7-7| 6.7-7| -2.5-5	65:25:07

TABLE 4
Results obtained by the boundary-point method in [19] *on the FAPs. The parameter* σ_0 *is set to* 1 *(better than* 0.1*).*

Problem	m | $n_s; n_l$	it	$\langle C, X\rangle$	$b^T y$	R_P| R_D| gap	Time
fap09	15225 | 174; 14025	2000	1.07978251 1	1.07982902 1	9.2-7| 9.8-6| -2.1-5	59
fap10	14479 | 183; 13754	2000	1.70252739-2	2.38972400-2	1.1-5| **1.1-4**| **-6.6-3**	1:25
fap11	24292 | 252; 23275	2000	4.22711513-2	5.94650102-2	8.8-6| **1.4-4**| **-1.6-2**	2:31
fap12	26462 | 369; 24410	2000	2.93446247-1	3.26163363-1	6.0-6| **1.5-4**| **-2.0-2**	4:37
fap25	322924 | 2118; 311044	2000	1.31895665 1	1.35910952 1	4.8-6| **2.0-4**| **-1.4-2**	8:04:00
fap36	1154467 | 4110; 1112293	2000	7.03339309 1	7.09606078 1	3.9-6| **1.4-4**| **-4.4-3**	46:59:28
fap09	15225 | 174; 14025	300	1.08257732 1	1.09208378 1	**1.7-4**| **7.2-4**| **-4.2-3**	09
fap10	14479 | 183; 13754	300	5.54148690-2	9.98476591-2	8.3-5| **6.9-4**| **-3.8-2**	12
fap11	24292 | 252; 23275	300	1.33930656-1	1.82368305-1	**2.4-4**| **7.9-4**| **-3.7-2**	22
fap12	26462 | 369; 24410	300	4.11473718-1	5.69735906-1	**1.2-4**| **8.4-4**| **-8.0-2**	41
fap25	322924 | 2118; 311044	300	1.47010392 1	1.69017693 1	**1.1-4**| **1.2-3**| **-6.8-2**	1:10:36
fap36	1154467 | 4110; 1112293	300	7.28509749 1	7.67389918 1	8.6-5| **8.9-4**| **-2.6-2**	6:53:36