BCD & ADMM 算法收敛理论

Lecture 11: ADMM 收敛性

罗自炎

北京交通大学数统学院

E-mail: zyluo@bjtu.edu.cn

参考资料

- 教材与参考文献：
  - 最优化：建模、算法与理论
  - Bolte et al., MP 2014
  - Fazel et al., SIMAX 2013
  - Rockafellar & Wets, Variational Analysis
- 致谢： 北京大学文再文教授; 清华大学张立平教授

回顾: 交替方向乘子法ADMM

- 优化问题形式:

$$\begin{aligned} \min_{x_1,x_2} \quad & f_1(x_1)+f_2(x_2), \\ \textbf{s.t.} \quad & A_1x_1+A_2x_2=b. \end{aligned} \tag{1}$$

- 增广拉格朗日函数:

$$\begin{aligned} L_\rho(x_1,x_2,y) =& f_1(x_1)+f_2(x_2)+y^{\mathrm{T}}(A_1x_1+A_2x_2-b) \\ & +\frac{\rho}{2}\|A_1x_1+A_2x_2-b\|_2^2. \end{aligned} \tag{2}$$

- **ADMM** 迭代:

$$x_1^{k+1} = \arg\min_{x_1} L_\rho(x_1,x_2^k,y^k), \tag{3}$$

$$x_2^{k+1} = \arg\min_{x_2} L_\rho(x_1^{k+1},x_2,y^k), \tag{4}$$

$$y^{k+1} = y^k + \tau\rho(A_1x_1^{k+1}+A_2x_2^{k+1}-b),\ \tau\in\left(0,\frac{1+\sqrt{5}}{2}\right) \tag{5}$$

回顾: 交替方向乘子法ADMM

- **基本假设:**

1. f_1, f_2 适当闭凸函数, ADMM 子问题(3),(4) 有唯一解;
2. 原问题(1)最优解集非空, 且Slater条件成立, i.e.,

$$\exists(\tilde{x}_1, \tilde{x}_2) \in \mathrm{int}(dom f_1 \times dom f_2) \text{ 使得 } A_1\tilde{x}_1 + A_2\tilde{x}_2 = b.$$

- **KKT点对(x_1^*, x_2^*, y^*):**

$$0 \in \partial_{x_1} L(x_1^*, x_2^*, y^*) = \partial f_1(x_1^*) + A_1^{\mathrm{T}} y^*, \tag{6a}$$

$$0 \in \partial_{x_2} L(x_1^*, x_2^*, y^*) = \partial f_2(x_2^*) + A_2^{\mathrm{T}} y^*, \tag{6b}$$

$$A_1 x_1^* + A_2 x_2^* = b, \tag{6c}$$

ADMM收敛性定理

Theorem 1. *若基本假设成立且A_1, A_2 列满秩, 取$\tau \in \left(0, \frac{1+\sqrt{5}}{2}\right)$, 则ADMM 算法产生的迭代点列$\{(x_1^k, x_2^k, y^k)\}$ 收敛到原问题(1)的一个KKT点对.*

分析：

- $(x_1^k - x_1^*, x_2^k - x_2^*, y^k - y^*) \to 0$
- $A_1(x_1^k - x_1^*) + A_2(x_2^k - x_2^*) \to 0$
- 构造$u^k \in \partial f_1(x_1^k), v^k \in \partial f_2(x_2^k)$, 使得$u^k \to -A_1^T y^*, v^k \to -A_2^T y^*$

ADMM收敛性分析

引入记号：

$$(e_1^k, e_2^k, e_y^k) \overset{\text{def}}{=\!=} (x_1^k, x_2^k, y^k) - (x_1^*, x_2^*, y^*),$$

$$\begin{aligned}
u^k &= -{A_1}^{\mathrm{T}}[y^k + (1-\tau)\rho(A_1 e_1^k + A_2 e_2^k) + \rho A_2(x_2^{k-1} - x_2^k)], \\
v^k &= -{A_2}^{\mathrm{T}}[y^k + (1-\tau)\rho(A_1 e_1^k + A_2 e_2^k)], \\
\Psi_k &= \frac{1}{\tau\rho}\|e_y^k\|^2 + \rho\|A_2 e_2^k\|^2, \\
\Phi_k &= \Psi_k + \max(1-\tau, 1-\tau^{-1})\rho\|A_1 e_1^k + A_2 e_2^k\|^2.
\end{aligned} \tag{7}$$

ADMM收敛性分析

Lemma 1. *假设*$\{(x_1^k, x_2^k, y^k)\}$*为ADMM*(3)-(5) *产生一个迭代序列, 则*$\forall k \geq 1$*有*

$$u^k \in \partial f_1(x_1^k), \quad v^k \in \partial f_2(x_2^k), \tag{8}$$

和

$$\begin{aligned} \Phi_k - \Phi_{k+1} \geq & \min(\tau, 1+\tau-\tau^2)\rho\|A_2(x_2^k - x_2^{k+1})\|^2 \\ & + \min(1, 1+\tau^{-1}-\tau)\rho\|A_1 e_1^{k+1} + A_2 e_2^{k+1}\|^2. \end{aligned} \tag{9}$$

注：只有当$\tau \in \left(0, \frac{1+\sqrt{5}}{2}\right)$时, (9) 中不等号右侧的项才为非负.

Lemma 1 证明

Proof. 先证明次梯度部分(8).

- 先证$u^{k+1} \in \partial f_1(x_1^{k+1})$： 对$x_1^{k+1}$, 由子问题(3)的最优性条件可知,

$$0 \in \partial f_1(x_1^{k+1}) + A_1^{\mathrm{T}} y^k + \rho A_1^{\mathrm{T}}(A_1 x_1^{k+1} + A_2 x_2^k - b).$$

将$y^k = y^{k+1} - \tau\rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b)$代入上式,

$$-A_1^{\mathrm{T}}\left(y^{k+1} + (1-\tau)\rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) + \rho A_2(x_2^k - x_2^{k+1})\right) \in \partial f_1(x_1^{k+1}).$$

根据u^{k+1}的定义以及$b = A_1 x_1^* + A_2 x_2^*$ 可知：$u^{k+1} \in \partial f_1(x_1^{k+1})$.

- 再证$v^{k+1} \in \partial f_2(x_2^{k+1})$： 对$x_2^{k+1}$, 由子问题(4)的最优性条件可知,

$$0 \in \partial f_2(x_2^{k+1}) + A_2^{\mathrm{T}} y^k + \rho A_2^{\mathrm{T}}(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b),$$

又由(5) $y^k = y^{k+1} - \tau\rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b)$ 知

$$-A_2^{\mathrm{T}}\left(y^{k+1} + (1-\tau)\rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b)\right) \in \partial f_2(x_2^{k+1}).$$

根据v^{k+1}的定义以及$b = A_1 x_1^* + A_2 x_2^*$, $v^{k+1} \in \partial f_2(x_2^{k+1})$. 故(8)得证.

再证序列$\{\Phi_k\}$下降性(9). 根据(6)和(8),

$$u^{k+1} \in \partial f_1(x_1^{k+1}), \quad -A_1^{\mathrm{T}} y^* \in \partial f_1(x_1^*),$$

$$v^{k+1} \in \partial f_2(x_2^{k+1}), \quad -A_2^{\mathrm{T}} y^* \in \partial f_2(x_2^*).$$

根据f_1, f_2凸, 其次微分算子为单调算子, 从而

$$\left\langle u^{k+1} + A_1^{\mathrm{T}} y^*, x_1^{k+1} - x_1^* \right\rangle \geq 0,$$

$$\left\langle v^{k+1} + A_2^{\mathrm{T}} y^*, x_2^{k+1} - x_2^* \right\rangle \geq 0.$$

再结合u^{k+1}, v^{k+1}的定义, 得到

$$\left\langle -e_y^{k+1} - (1-\tau)\rho(A_1 e_1^{k+1} + A_2 e_2^{k+1}) - \rho A_2(x_2^k - x_2^{k+1}), A_1 e_1^{k+1} \right\rangle \geq 0,$$

$$\left\langle -e_y^{k+1} - (1-\tau)\rho(A_1 e_1^{k+1} + A_2 e_2^{k+1}), A_2 e_2^{k+1} \right\rangle \geq 0.$$

上面两式相加得到:

$$\left\langle e_y^{k+1}, -(A_1 e_1^{k+1} + A_2 e_2^{k+1}) \right\rangle - (1-\tau)\rho \|A_1 e_1^{k+1} + A_2 e_2^{k+1}\|^2$$

$$+\rho\langle -A_2(x_2^k - x_2^{k+1}), A_1 e_1^{k+1} \rangle \geq 0$$

注意到恒等式

$$A_1e_1^{k+1}+A_2e_2^{k+1} = A_1x_1^{k+1}+A_2x_2^{k+1}-b = (\tau\rho)^{-1}(y^{k+1}-y^k) = (\tau\rho)^{-1}(e_y^{k+1}-e_y^k), \tag{10}$$

我们有

$$\begin{aligned} &\frac{1}{\tau\rho}\left\langle e_y^{k+1}, e_y^k - e_y^{k+1}\right\rangle - (1-\tau)\rho\|A_1x_1^{k+1}+A_2x_2^{k+1}-b\|^2 \\ &+\rho\left\langle A_2(x_2^{k+1}-x_2^k), A_1x_1^{k+1}+A_2x_2^{k+1}-b\right\rangle \\ &-\rho\left\langle A_2(x_2^{k+1}-x_2^k), A_2e_2^{k+1}\right\rangle \geq 0. \end{aligned} \tag{11}$$

注意到不等式(11)和(9) 主要的差别在下面这一项上:

$$\rho\left\langle A_2(x_2^{k+1}-x_2^k), A_1x_1^{k+1}+A_2x_2^{k+1}-b\right\rangle.$$

接下来, 我们估计这一项的上界. 引入新符号

$$\nu^{k+1} := y^{k+1}+(1-\tau)\rho(A_1x_1^{k+1}+A_2x_2^{k+1}-b),$$

$$M^{k+1} := (1-\tau)\rho\left\langle A_2(x_2^{k+1}-x_2^k), A_1x_1^k+A_2x_2^k-b\right\rangle,$$

则$-A_2^{\mathrm{T}}\nu^{k+1} \in \partial f_2(x_2^{k+1}), -A_2^{\mathrm{T}}\nu^k \in \partial f_2(x_2^k)$. 因$f_2$凸, 故

$$\left\langle -A_2^{\mathrm{T}}(\nu^{k+1}-\nu^k), x_2^{k+1}-x_2^k \right\rangle \geq 0.$$

即:

$$\left\langle \nu^{k+1}-\nu^k, A_2(x_2^{k+1}-x_2^k) \right\rangle \leq 0.$$

从而有

$$\begin{aligned}
&\rho\left\langle A_2(x_2^{k+1}-x_2^k), A_1x_1^{k+1}+A_2x_2^{k+1}-b \right\rangle \\
=&(1-\tau)\rho\left\langle A_2(x_2^{k+1}-x_2^k), A_1x_1^{k+1}+A_2x_2^{k+1}-b \right\rangle \\
&+\left\langle A_2(x_2^{k+1}-x_2^k), y^{k+1}-y^k \right\rangle \\
=&M^{k+1}+\left\langle \nu^{k+1}-\nu^k, A_2(x_2^{k+1}-x_2^k) \right\rangle \leq M^{k+1}.
\end{aligned}$$

因此, 不等式(11)可以放缩成

$$\begin{aligned}
&\frac{1}{\tau\rho}\left\langle e_y^{k+1}, e_y^k-e_y^{k+1} \right\rangle-(1-\tau)\rho\|A_1x_1^{k+1}+A_2x_2^{k+1}-b\|^2 \\
&+M^{k+1}-\rho\left\langle A_2(x_2^{k+1}-x_2^k), A_2e_2^{k+1} \right\rangle \geq 0.
\end{aligned}$$

利用恒等式

$$\langle a, b\rangle = \frac{1}{2}(\|a+b\|^2 - \|a\|^2 - \|b\|^2),$$

可得: $2\left\langle e_y^{k+1}, e_y^k - e_y^{k+1}\right\rangle = \|e_y^k\|^2 - \|e_y^{k+1}\|^2 - \|e_y^k - e_y^{k+1}\|^2$

$$= \|e_y^k\|^2 - \|e_y^{k+1}\|^2 - (\tau\rho)^2\|A_1x_1^{k+1} + A_2x_2^{k+1} - b\|^2$$

利用恒等式

$$\langle a, b\rangle = \frac{1}{2}(\|a\|^2 + \|b\|^2 - \|a-b\|^2),$$

得

到$2\left\langle A_2(x_2^{k+1} - x_2^k), A_2e_2^{k+1}\right\rangle = \|A_2(x_2^{k+1} - x_2^k)\|^2 + \|A_2e_2^{k+1}\|^2 - \|A_2e_2^k\|^2.$

从而

$$\begin{aligned}&\frac{1}{\tau\rho}(\|e_y^k\|^2 - \|e_y^{k+1}\|^2) - (2-\tau)\rho\|A_1x_1^{k+1} + A_2x_2^{k+1} - b\|^2 \\ &+ 2M^{k+1} - \rho\|A_2(x_2^{k+1} - x_2^k)\|^2 - \rho\|A_2e_2^{k+1}\|^2 + \rho\|A_2e_2^k\|^2 \geq 0.\end{aligned} \tag{12}$$

除了M^{k+1}中的项, (12)中的其他项均在不等式(9) 中出现.

由于M^{k+1}的符号和τ的取法有关, 下面针对τ的两种取法进行讨论.

情形一: $\tau \in (0,1]$: 此时$1-\tau \geq 0$, 根据基本不等式

$$
\begin{aligned}
&2\left\langle A_2(x_2^{k+1}-x_2^k), A_1x_1^k + A_2x_2^k - b\right\rangle \\
\leq &\|A_2(x_2^{k+1}-x_2^k)\|^2 + \|A_1x_1^k + A_2x_2^k - b\|^2.
\end{aligned}
$$

代入(12)得

$$
\begin{aligned}
&\frac{1}{\tau\rho}\|e_y^k\|^2 + \rho\|A_2e_2^k\|^2 + (1-\tau)\rho\|A_1e_1^k + A_2e_2^k\|^2 \\
&-\left[\frac{1}{\tau\rho}\|e_y^{k+1}\|^2 + \rho\|A_2e_2^{k+1}\|^2 + (1-\tau)\rho\|A_1e_1^{k+1} + A_2e_2^{k+1}\|^2\right] \\
\geq &\rho\|A_1x_1^{k+1} + A_2x_2^{k+1} - b\|^2 + \tau\rho\|A_2(x_2^{k+1}-x_2^k)\|^2.
\end{aligned} \tag{13}
$$

情形二： $\tau > 1$: 此时$1-\tau<0$, 根据基本不等式

$$
\begin{aligned}
& -2\left\langle A_2(x_2^{k+1}-x_2^k), A_1x_1^k+A_2x_2^k-b\right\rangle \\
\le & \tau\|A_2(x_2^{k+1}-x_2^k)\|^2+\frac{1}{\tau}\|A_1x_1^k+A_2x_2^k-b\|^2.
\end{aligned}
$$

同样代入不等式(12)可以得到

$$
\begin{aligned}
& \frac{1}{\tau\rho}\|e_y^k\|^2+\rho\|A_2e_2^k\|^2+\left(1-\frac{1}{\tau}\right)\rho\|A_1e_1^k+A_2e_2^k\|^2 \\
& -\left[\frac{1}{\tau\rho}\|e_y^{k+1}\|^2+\rho\|A_2e_2^{k+1}\|^2+\left(1-\frac{1}{\tau}\right)\rho\|A_1e_1^{k+1}+A_2e_2^{k+1}\|^2\right] \\
\ge & \left(1+\frac{1}{\tau}-\tau\right)\rho\|A_1x_1^{k+1}+A_2x_2^{k+1}-b\|^2 \\
& +(1+\tau-\tau^2)\rho\|A_2(x_2^{k+1}-x_2^k)\|^2.
\end{aligned}
\tag{14}
$$

综合两种情形可得$\Phi_k-\Phi_{k+1}$ 满足(9). $\square$

ADMM 收敛性定理证明

Theorem 1 若基本假设成立且A_1, A_2 列满秩, 取$\tau \in \left(0, \frac{1+\sqrt{5}}{2}\right)$, 则ADMM 算法产生的迭代点列$\{(x_1^k, x_2^k, y^k)\}$ 收敛到原问题(1)的一个KKT点对.

Proof. 由Lemma 1中$\{\Phi_k\}$ 序列下降性可知$\{\Phi_k\}$ 是有界序列, 结合Φ_k的定义式

$$\Phi_k = \frac{1}{\tau\rho}\|e_y^k\|^2 + \rho\|A_2 e_2^k\|^2 + \max\left(1-\tau, 1-\tau^{-1}\right)\rho\|A_1 e_1^k + A_2 e_2^k\|^2$$

可知

$$\|e_y^k\|, \quad \|A_2 e_2^k\|, \quad \|A_1 e_1^k + A_2 e_2^k\|$$

均有界. 根据三角不等式

$$\|A_1 e_1^k\| \leq \|A_1 e_1^k + A_2 e_2^k\| + \|A_2 e_2^k\|,$$

可推出$\{\|A_1 e_1^k\|\}$也是有界序列.

注意到$A_1^{\mathrm{T}}A_1 \succ 0, A_2^{\mathrm{T}}A_2 \succ 0$, 即$\lambda_{\min}(A_1^TA_1) > 0, \lambda_{\min}(A_2^TA_2) > 0$. 结合

$$\lambda_{\min}(A_1^TA_1)\|e_1^k\|^2 \leq \|A_1e_1^k\|^2, \ \ \lambda_{\min}(A_2^TA_2)\|e_2^k\|^2 \leq \|A_2e_2^k\|^2$$

可得$\{(x_1^k - x_1^*, x_2^k - x_2^*)\}$ 有界. 结合$\|e_y^k\| = \|y^k - y^*\|$ 有界, 可知

$\{(x_1^k, x_2^k, y^k)\}$是有界序列.

考虑到$\{\Phi_k\}$ 序列下降性以及$\{\Phi_k\}$ 有界, 可知:

$$\sum_{k=0}^{\infty}\|A_1e_1^k + A_2e_2^k\|^2 < \infty, \quad \sum_{k=0}^{\infty}\|A_2(x_2^{k+1} - x_2^k)\|^2 < \infty$$

这表明

$$\|A_1e_1^k + A_2e_2^k\| = \|A_1x_1^k + A_2x_2^k - b\| \to 0, \quad \|A_2(x_2^{k+1} - x_2^k)\| \to 0. \tag{15}$$

由于$\{(x_1^k, x_2^k, y^k)\}$是有界序列, 因此存在一个收敛子列，设

$$(x_1^{k_j}, x_2^{k_j}, y^{k_j}) \to (x_1^\infty, x_2^\infty, y^\infty).$$

由$\{u^k\}$与$\{v^k\}$的定义式(7)

$$u^k = -{A_1}^{\mathrm{T}}[y^k + (1-\tau)\rho(A_1 e_1^k + A_2 e_2^k) + \rho A_2(x_2^{k-1} - x_2^k)],$$

$$v^k = -{A_2}^{\mathrm{T}}[y^k + (1-\tau)\rho(A_1 e_1^k + A_2 e_2^k)],$$

以及证明得到的可行性结论(15)

$$\|A_1 e_1^k + A_2 e_2^k\| = \|A_1 x_1^k + A_2 x_2^k - b\| \to 0, \quad \|A_2(x_2^{k+1} - x_2^k)\| \to 0.$$

可得$\{u^k\}$与$\{v^k\}$相应的子列也收敛. 从而

$$u^\infty \stackrel{\text{def}}{=\!=} \lim_{j\to\infty} u^{k_j} = -A_1^{\mathrm{T}} y^\infty, \quad v^\infty = \lim_{j\to\infty} v^{k_j} = -A_2^{\mathrm{T}} y^\infty. \tag{16}$$

由(8)知, 对任意的$k \geq 1$，有$u^k \in \partial f_1(x_1^k)$, $v^k \in \partial f_2(x_2^k)$. 由次梯度映射的图像是闭集可知

$$-A_1^T y^\infty \in \partial f_1(x_1^\infty), \quad -A_2^T y^\infty \in \partial f_2(x_2^\infty).$$

由(15)可知

$$\lim_{j\to\infty}\|A_1x_1^{k_j}+A_2x_2^{k_j}-b\|=\|A_1x_1^{\infty}+A_2x_2^{\infty}-b\|=0.$$

这表明$(x_1^{\infty},x_2^{\infty},y^{\infty})$是问题(1)的一个KKT对.

因此上述分析中的(x_1^*,x_2^*,y^*)均可替换为$(x_1^{\infty},x_2^{\infty},y^{\infty})$.

注意到$\{\Phi_k\}$是单调下降的，且对子列$\{\Phi_{k_j}\}$有

$$\begin{aligned}&\lim_{j\to\infty}\Phi_{k_j}\\&=\lim_{j\to\infty}\left(\frac{1}{\tau\rho}\|e_y^{k_j}\|^2+\rho\|A_2e_2^{k_j}\|^2+\max\left\{1-\tau,1-\frac{1}{\tau}\right\}\rho\|A_1e_1^{k_j}+A_2e_2^{k_j}\|^2\right)\\&=0.\end{aligned}$$

这说明 $\|e_y^k\|\to 0,\quad \|A_2e_2^k\|\to 0,\quad \|A_1e_1^k+A_2e_2^k\|\to 0.$

进一步有

$$0\le\limsup_{k\to\infty}\|A_1e_1^k\|\le\lim_{k\to\infty}\left(\|A_2e_2^k\|+\|A_1e_1^k+A_2e_2^k\|\right)=0.$$

注意到$A_1^{\mathrm T}A_1\succ 0$, $A_2^{\mathrm T}A_2\succ 0$, 故全序列$\{(x_1^k,x_2^k,y^k)\}$收敛. $\square$

多块ADMM收敛性反例

- 考虑最优化问题[2]

$$\begin{aligned} \min \quad & 0, \\ \text{s.t.} \quad & A_1x_1 + A_2x_2 + A_3x_3 = 0, \end{aligned} \tag{17}$$

其中非零向量$A_i \in \mathbb{R}^3$, $i \in [3]; x_i \in \mathbb{R}, i \in [3]$. 若$A_1, A_2, A_3$ 线性无关, 则问题(17) 只有零解. 此时对应的乘子为$y = 0 \in \mathbb{R}^3$.
【$\nabla_x L(x, y) = A^T y = 0 \Rightarrow y = 0$】

- 增广拉格朗日函数为

$$L_\rho(x, y) = 0 + y^{\mathrm{T}}(A_1x_1 + A_2x_2 + A_3x_3) + \frac{\rho}{2}\|A_1x_1 + A_2x_2 + A_3x_3\|^2.$$

[2] C. Chen et al., The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent, Math Programming 155(1-2) (2016) 57-79

多块ADMM收敛性反例

- 当固定x_2, x_3, y时，对x_1求最小可推出

$$A_1^{\mathrm{T}} y + \rho A_1^{\mathrm{T}}(A_1 x_1 + A_2 x_2 + A_3 x_3) = 0,$$

整理可得

$$x_1 = -\frac{1}{\|A_1\|^2}\left(A_1^{\mathrm{T}}\left(\frac{y}{\rho} + A_2 x_2 + A_3 x_3\right)\right).$$

可类似地计算x_2, x_3的表达式

多块ADMM收敛性反例

- 求解该问题的多块交替方向乘子法的迭代格式可以写为

$$
\begin{aligned}
x_1^{k+1} &= -\frac{1}{\|A_1\|^2} A_1^{\mathrm{T}} \left(\frac{y^k}{\rho} + A_2 x_2^k + A_3 x_3^k \right), \\
x_2^{k+1} &= -\frac{1}{\|A_2\|^2} A_2^{\mathrm{T}} \left(\frac{y^k}{\rho} + A_1 x_1^{k+1} + A_3 x_3^k \right), \\
x_3^{k+1} &= -\frac{1}{\|A_3\|^2} A_3^{\mathrm{T}} \left(\frac{y^k}{\rho} + A_1 x_1^{k+1} + A_2 x_2^{k+1} \right), \\
y^{k+1} &= y^k + \rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1}).
\end{aligned}
\tag{18}
$$

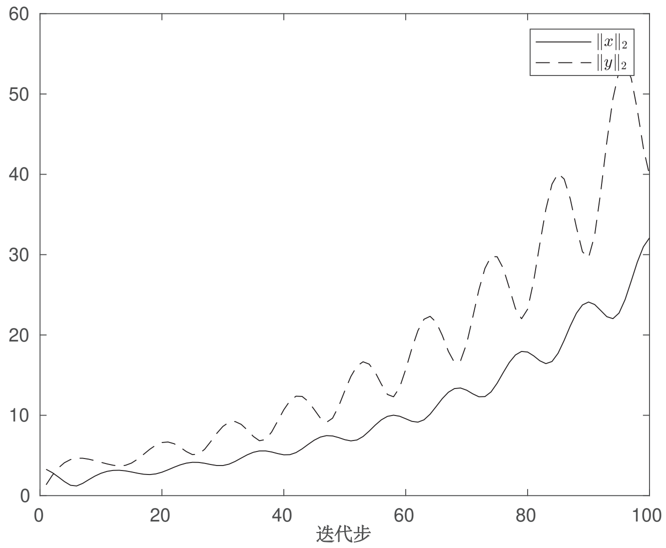
多块ADMM收敛性反例

- 自变量初值初值选为$(1,1,1)$，乘子选为$(0,0,0)$. 选取A为

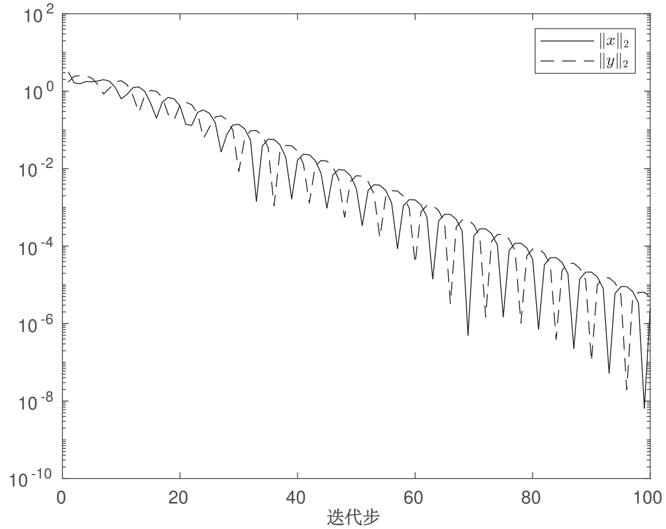
$$\widetilde{A}=\begin{bmatrix}1 & 1 & 2\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix} \quad \text{或} \quad \widehat{A}=\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 2\\ 1 & 2 & 2\end{bmatrix}.$$

多块ADMM收敛性反例

下图记录了在不同A下x和y的ℓ_2范数随迭代的变化过程.



(a) 系数矩阵为$\widetilde{A}$



(b) 系数矩阵为$\widehat{A}$

Figure 1: 选取不同A时的数值结果

显然： $A=\tilde{A}$时, 发散; $A=\hat{A}$ 时, 收敛, 且$\|x\|, \|y^k\|$ 并非单调下降, 而是在下降大趋势下有规律地振荡【原因分析见文献[2]】.

非凸优化的ADMM举例:非凸约束问题

考虑如下约束优化问题：

$$\begin{aligned} \min_x \quad & f(x), \\ s.t. \quad & x \in \mathcal{S}, \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \min_x \quad & f(x) + \mathbb{I}_{\mathcal{S}}(z), \\ s.t. \quad & x - z = 0, \end{aligned}$$

其中f是凸的，但是$\mathcal{S}$是非凸的. 交替方向乘子法产生如下迭代：

$$\begin{aligned} x^{k+1} &= \arg\min_x \left(f(x) + (\rho/2)\|x - z^k + u^k\|_2^2 \right), \\ z^{k+1} &= \Pi_{\mathcal{S}}(x^{k+1} + u^k), \\ u^{k+1} &= u^k + (x^{k+1} - z^{k+1}) \end{aligned}$$

其中, $\Pi_{\mathcal{S}}(z)$是将z投影到集合$\mathcal{S}$中. 因为f是凸的，所以上述x-极小化步是凸问题, 但是z-极小化步是向一个非凸集合的投影.

非凸优化的ADMM举例:非凸约束问题

可以计算的非凸集上的投影：

- 稀疏集：如果$\mathcal{S} = \{x | \|x\|_0 \leq c\}$，其中$\|\cdot\|_0$ 表示非零元素的数目，那么$\Pi_{\mathcal{S}}(v)$保持前c大的元素不变，其他元素变为0。

  例如回归选择（也叫特征选择）问题：

$$\begin{aligned} \min_x \quad & \|Ax - b\|_2^2, \\ s.t. \quad & \|x\|_0 \leq c. \end{aligned}$$

- 低秩集：如果$\mathcal{S} := \{X \in \mathbb{R}^{m\times n} : \mathrm{rank}(X) \leq c\}$, 对于矩阵$V = \sum_i \sigma_i u_i u_i^T$, $\sigma_1 \geq \cdots \geq \sigma_{\min m,n}$, 其在$\mathcal{S}$ 上的投影为: $\Pi_{\mathcal{S}}(V) = \sum_{i=1}^{c} \sigma_i u_i u_i^T$.

- 布尔约束集：$\mathcal{S} = \{x | x_i \in \{0, 1\}\}$，那么$\Pi_{\mathcal{S}}(v)$就是简单地把每个元素变为0 和1中离它更近的数。

非凸优化的ADMM举例:非负矩阵分解和补全

非负矩阵分解和补全问题可以写成如下形式:

$$\begin{aligned} \min_{X,Y} \quad & \|\mathcal{P}_\Omega(XY - M)\|_F^2, \\ s.t. \quad & X_{ij} \geq 0, Y_{ij} \geq 0, \forall i, j, \end{aligned}$$

其中，Ω表示矩阵M中的已知元素的下标集合，$\mathcal{P}_\Omega(A)$表示得到一个新的矩阵A'，其下标在集合Ω中的所对应的元素等于矩阵A的对应元素, 其他元素为0. 注意到，这个问题是非凸的。

为了利用交替方向乘子法的优势，我们考虑如下的等价形式:

$$\begin{aligned} \min_{U,V,X,Y,Z} \quad & \frac{1}{2}\|XY - Z\|_F^2, \\ s.t. \quad & X = U, Y = V, \\ & U \geq 0, V \geq 0, \\ & \mathcal{P}_\Omega(Z - M) = 0. \end{aligned}$$

非凸优化的ADMM举例:非负矩阵分解和补全

$$
\begin{aligned}
L_{\alpha,\beta}(X,Y,Z,U,V,\Lambda,\Pi) = {} & \frac{1}{2}\|XY-Z\|_F^2 + \Lambda \bullet (X-U) \\
& + \Pi \bullet (Y-V) + \frac{\alpha}{2}\|X-U\|_F^2 + \frac{\beta}{2}\|Y-V\|_F^2,
\end{aligned}
$$

$$
\begin{aligned}
X^{k+1} &= \arg\min_{X} \ L_{\alpha,\beta}(X,Y^k,Z^k,U^k,V^k,\Lambda^k,\Pi^k), \\
Y^{k+1} &= \arg\min_{Y} \ L_{\alpha,\beta}(X^{k+1},Y,Z^k,U^k,V^k,\Lambda^k,\Pi^k), \\
Z^{k+1} &= \arg\min_{\mathcal{P}_\Omega(Z-M)=0} \ L_{\alpha,\beta}(X^{k+1},Y^{k+1},Z,U^k,V^k,\Lambda^k,\Pi^k), \\
U^{k+1} &= \arg\min_{U\geq 0} \ L_{\alpha,\beta}(X^{k+1},Y^{k+1},Z^{k+1},U,V^k,\Lambda^k,\Pi^k), \\
V^{k+1} &= \arg\min_{V\geq 0} \ L_{\alpha,\beta}(X^{k+1},Y^{k+1},Z^{k+1},U^{k+1},V,\Lambda^k,\Pi^k), \\
\Lambda^{k+1} &= \Lambda^k + \tau\alpha(X^{k+1}-U^{k+1}), \\
\Pi^{k+1} &= \Pi^k + \tau\beta(Y^{k+1}-V^{k+1}).
\end{aligned}
$$

Thank You!