BCD & ADMM 算法收敛理论

Lecture 9: DR分裂算法与ADMM($\tau = 1$)

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

Douglas-Rachford Splitting 算法

- **复合优化模型:**

$$\min_x f(x) := g(x) + h(x) \tag{1}$$

其中g, h 是适当闭凸函数.

- **最优性分析:**

$$
\begin{aligned}
& 0 \in \partial g(x) + \partial h(x) \\
\Longleftrightarrow\quad & 0 \in \partial tg(x) + \partial th(x) \\
\Longleftrightarrow\quad & x - z \in \partial tg(x) \;\;\&\;\; z - x \in \partial th(x) \\
\Longleftrightarrow\quad & (2x - z) - x \in \partial tg(x) \;\;\&\;\; z - x \in \partial th(x) \\
\Longleftrightarrow\quad & x = \mathrm{prox}_{tg}(2x - z) \;\;\&\;\; x = \mathrm{prox}_{th}(z) \\
\Longleftrightarrow\quad & F(z) := z + \mathrm{prox}_{tg}(2\mathrm{prox}_{th}(z) - z) - \mathrm{prox}_{th}(z) = z \;\;\&\;\; x = \mathrm{prox}_{th}(z)
\end{aligned}
$$

Douglas-Rachford Splitting 算法

- **不动点方程:**

$$F(z) := z + \text{prox}_{tg}(2\text{prox}_{th}(z) - z) - \text{prox}_{th}(z) = z \qquad (2)$$

- **不动点迭代:**

$$z^k = F(z^{k-1})$$

即:

$$\begin{cases} x^k = \text{prox}_{th}(z^{k-1}), \\ y^k = \text{prox}_{tg}(2x^k - z^{k-1}), \\ z^k = z^{k-1} + y^k - x^k, \\ x^{k+1} = \text{prox}_{th}(z^k) \\ \vdots \end{cases}$$

- 适当闭凸函数f关键性质: $u = \text{prox}_f(x) \iff x - u \in \partial f(u)$.

等价形式

- 从y的更新开始

$$y^+ = \mathrm{prox}_{tg}(2x - z); \quad z^+ = z + y^+ - x; \quad x^+ = \mathrm{prox}_{th}(z^+)$$

- 交换z和x的更新顺序

$$y^+ = \mathrm{prox}_{tg}(2x-z); \quad x^+ = \mathrm{prox}_{th}(z+y^+-x); \quad z^+ = z+y^+-x$$

- 作变量替换$w = z - x$

DR 迭代的等价形式: 从任意初始点$x^0 \in \mathbf{dom}\, h, w^0 \in t\partial h(x^0)$开始

$$\begin{aligned} y^+ &= \mathrm{prox}_{tg}(x - w) \\ x^+ &= \mathrm{prox}_{th}(y^+ + w) \\ w^+ &= w + y^+ - x^+ \end{aligned}$$

Douglas-Rachford迭代的松弛形式

- 不动点迭代的松弛形式

$$z^+ = z + \rho(F(z) - z)$$

若$1 < \rho < 2$, 称为**超松弛**; 若$0 < \rho < 1$称为**次松弛**.

- **DR迭代的松弛形式1**

$$\begin{cases} x^+ = \mathrm{prox}_{th}(z), \\ y^+ = \mathrm{prox}_{tg}(2x^+ - z), \\ z^+ = z + y^+ - x^+, \end{cases} \xrightarrow{relax} \begin{cases} x^+ = \mathrm{prox}_{th}(z) \\ y^+ = \mathrm{prox}_{tg}(2x^+ - z) \\ z^+ = z + \rho(y^+ - x^+) \end{cases}$$

- 不动点迭代的松弛形式

$$z^+ = z + \rho(F(z) - z)$$

- **DR迭代的松弛形式2**

$$\begin{cases} y^+ = \mathrm{prox}_{tg}(x - w) \\ x^+ = \mathrm{prox}_{th}(y^+ + w) \\ w^+ = w + y^+ - x^+ \end{cases} \xrightarrow[F(z)-z=y^+-x]{y^+ + w = z^+} \begin{array}{l} y^+ = \mathrm{prox}_{tg}(x - w) \\ x^+ = \mathrm{prox}_{th}((1-\rho)x + \rho y^+ + w) \\ w^+ = w + \rho y^+ + (1-\rho)x - x^+ \end{array}$$

DR迭代与ADMM

凸的原始问题

$$\begin{aligned} \min \quad & f_1(x_1) + f_2(x_2) \\ \text{s.t.} \quad & A_1 x_1 + A_2 x_2 = b \end{aligned}$$

对偶问题

$$\max \quad -b^T z - f_1^*(-A_1^T z) - f_2^*(-A_2^T z)$$

对对偶问题应用Douglas-Rachford 迭代：

$$\min_z \underbrace{b^T z + f_1^*(-A_1^T z)}_{g(z)} + \underbrace{f_2^*(-A_2^T z)}_{h(z)}$$

- $\partial g(z) = b - A_1 \partial f_1^*(-A_1^T z); \quad \partial h(z) = -A_2 \partial f_2^*(-A_2^T z)$

DR迭代与ADMM

DR 迭代：

$$
\begin{aligned}
y^{k+1} &= \mathrm{prox}_{tg}(x^k - w^k), \\
x^{k+1} &= \mathrm{prox}_{th}(w^k + y^{k+1}), \\
w^{k+1} &= w^k + y^{k+1} - x^{k+1}.
\end{aligned} \tag{3}
$$

- 第一式的最优性条件为

$$
0 \in t(b - A_1 \partial f_1^*(-A_1^{\mathrm{T}} y^{k+1})) + y^{k+1} - (x^k - w^k),
$$

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

$$
y^{k+1} = x^k - w^k + t(A_1 x_1^k - b). \tag{4}
$$

利用闭凸函数f关键性质: $x \in \partial f(y) \Longleftrightarrow y \in \partial f^*(x)$, 结合(4)得

$$
x_1^k \in \arg\min_{x_1} \left\{ f_1(x_1) + (x^k)^{\mathrm{T}}(A_1 x_1 - b) + \frac{t}{2}\|A_1 x_1 - b - w^k/t\|_2^2 \right\}. \tag{5}
$$

DR迭代与ADMM

- 类似地，第二式的最优性条件为

$$0 \in tA_2\partial f_2^*(-A_2^{\mathrm{T}}x^{k+1}) + w^k + y^{k+1} - x^{k+1},$$

即存在$x_2^k \in \partial f_2^*(-A_2^{\mathrm{T}}x^{k+1})$，使得

$$x^{k+1} = x^k + t(A_1x_1^k + A_2x_2^k - b). \tag{6}$$

由$x_2^k \in \partial f_2^*(-A_2^{\mathrm{T}}x^{k+1})$等价于$-A_2^{\mathrm{T}}x^{k+1} \in \partial f_2(x_2^k)$,结合(6)可得

$$x_2^k \in \arg\min_{x_2}\left\{f_2(x_2) + (x^k)^{\mathrm{T}}(A_2x_2) + \frac{t}{2}\|A_1x_1^k + A_2x_2 - b\|_2^2\right\}. \tag{7}$$

- 由(3)(4)(6)可得

从而式(5):

$$w^{k+1} = -tA_2x_2^k. \tag{8}$$

$$x_1^k \in \arg\min_{x_1}\left\{f_1(x_1) + (x^k)^{\mathrm{T}}(A_1x_1 - b) + \frac{t}{2}\|A_1x_1 - b + A_2x_2^{k-1}\|_2^2\right\}.$$

DR迭代与ADMM

原问题的增广拉格朗日函数:

$$L_t(x_1,x_2,z)=f_1(x_1)+f_2(x_2)+z^T(A_1x_1+A_2x_2-b)+\frac{t}{2}\|A_1x_1+A_2x_2-b\|_2^2$$

- 整理公式(5),(7),(8)得到:

1. $x_1^{(k)}$的更新为极小化原问题的增广拉格朗日函数:

$$\begin{aligned} x_1^{(k)} &\in \arg\min_{x_1}\left\{f_1(x_1)+(z^{(k-1)})^TA_1x_1+\frac{t}{2}\|A_1x_1+A_2x_2^{(k-1)}-b\|_2^2\right\} \\ &= \arg\min_{x_1} L_t(x_1,x_2^{(k-1)},z^{(k-1)}) \end{aligned}$$

2. $x_2^{(k)}$的更新为极小化原问题的增广拉格朗日函数:

$$\begin{aligned} x_2^{(k)} &\in \arg\min_{x_2}\left(f_2(x_2)+(z^{(k-1)})^TA_2x_2+\frac{t}{2}\|A_1x_1^{(k)}+A_2x_2-b\|_2^2\right) \\ &= \arg\min_{x_2} L_t(x_1^{(k)},x_2,z^{(k-1)}) \end{aligned}$$

3. 对偶变量更新: $z^{(k)}=z^{(k-1)}+t(A_1x_1^{(k)}+A_2x_2^{(k)}-b)\Rightarrow$ ADMM $(\tau=1)$

Dual DRS $\Longleftrightarrow$ ADMM with $\tau = 1$

非扩张函数

Definition 1. *称函数f 为非扩张的(non-expansive), 若*

$$\|x-y\|^2 \geq \|f(x)-f(y)\|^2;$$

称函数f为固定非扩张的(firmly non-expansive), 若

$$(f(x)-f(y))^\top (x-y) \geq \|f(x)-f(y)\|^2.$$

Theorem 1. prox_h *是固定非扩张的,进而也是非扩张的,因而是Lipschitz连续的(常数为1).*

证明: 设$u = \mathrm{prox}_h(x), v = \mathrm{prox}_h(y)$，由邻近点算子的性质可得

$$x - u \in \partial h(u), y - v \in \partial h(v)$$

由∂h的单调性可得

$$(x - u - y + v)^\top (u - v) \geq 0.$$

由柯西不等式$\|u - v\|^2 \leq (x - y)^\top (u - v) \leq \|x - y\|\|u - v\|$, 因此

$$\|\mathrm{prox}_h(x) - \mathrm{prox}_h(y)\|_2 \leq \|x - y\|_2.$$

Douglas-Rachford迭代映射

定义迭代映射F, G如下

$$\begin{aligned} F(z) =& \ z + \mathrm{prox}_{tg}(2\mathrm{prox}_{th}(z) - z) - \mathrm{prox}_{th}(z) \\ G(z) =& \ z - F(z) = \mathrm{prox}_{th}(z) - \mathrm{prox}_{tg}(2\mathrm{prox}_{th}(z) - z) \end{aligned}$$

- F 是固定非扩张的

$$(F(z) - F(\hat{z}))^T(z - \hat{z}) \geq \|F(z) - F(\hat{z})\|_2^2 \quad \forall z, \hat{z}$$

- 进而G 也是固定非扩张的

$$\begin{aligned} & (G(z) - G(\hat{z}))^T(z - \hat{z}) \\ = & \ \|G(z) - G(\hat{z})\|_2^2 + (F(z) - F(\hat{z}))^T(z - \hat{z}) - \|F(z) - F(\hat{z})\|_2^2 \\ \geq & \ \|G(z) - G(\hat{z})\|_2^2 \end{aligned}$$

F固定非扩张性的证明

- 设$x = \mathrm{prox}_{th}(z), \hat{x} = \mathrm{prox}_{th}(\hat{z})$, 以及

$$y = \mathrm{prox}_{tg}(2x - z), \quad \hat{y} = \mathrm{prox}_{tg}(2\hat{x} - \hat{z})$$

- 代入$F(z) = z + y - x$ 和$F(\hat{z}) = \hat{z} + \hat{y} - \hat{x}$:

$$\begin{aligned}
&(F(z) - F(\hat{z}))^T(z - \hat{z}) \\
&\geq (z + y - x - \hat{z} - \hat{y} + \hat{x})^T(z - \hat{z}) - (x - \hat{x})^T(z - \hat{z}) + \|x - \hat{x}\|_2^2 \\
&= (y - \hat{y})^T(z - \hat{z}) + \|z - x - \hat{z} + \hat{x}\|_2^2 \\
&= (y - \hat{y})^T(2x - z - 2\hat{x} + \hat{z}) - \|y - \hat{y}\|_2^2 + \|F(z) - F(\hat{z})\|_2^2 \\
&\geq \|F(z) - F(\hat{z})\|_2^2
\end{aligned}$$

其中用到了prox_{th}和prox_{tg} 的固定非扩张性

$$(x - \hat{x})^T(z - \hat{z}) \geq \|x - \hat{x}\|_2^2, \quad (2x - z - 2\hat{x} + \hat{z})^T(y - \hat{y}) \geq \|y - \hat{y}\|_2^2$$

DR迭代的收敛性分析

不动点迭代格式:

$$\begin{aligned} z^{(k)} &= (1-\rho_k)z^{(k-1)} + \rho_k F(z^{(k-1)}) \\ &= z^{(k-1)} - \rho_k G(z^{(k-1)}) \end{aligned}$$

假设

- 最优值$f^* = \inf_x\{g(x)+h(x)\}$ 有限且可达到
- $\rho_k \in [\rho_{\min}, \rho_{\max}]$, 其中$0 < \rho_{\min} < \rho_{\max} < 2$

结论(收敛性)

- $z^{(k)}$ 收敛到F 的不动点z^*
- $x^{(k)} = \mathrm{prox}_{th}(z^{(k-1)})$ 收敛到原目标函数的极小点$x^* = \mathrm{prox}_{th}(z^*)$ (利用prox_{th} 的连续性)

证明

设z^* 为$F(z)$ 的任意不动点，即为$G(z)$的零点. 考虑第k步迭代，
记$z = z^{(k-1)}, \rho = \rho_k, z^+ = z^{(k)}$, 则

$$\begin{aligned} \|z^+ - z^*\|_2^2 - \|z - z^*\|_2^2 &= 2(z^+ - z)^T(z - z^*) + \|z^+ - z\|_2^2 \\ &= -2\rho G(z)^T(z - z^*) + \rho^2\|G(z)\|_2^2 \\ &\le -\rho(2-\rho)\|G(z)\|_2^2 \\ &\le -M\|G(z)\|_2^2 \end{aligned} \tag{9}$$

其中$M = \rho_{\min}(2 - \rho_{\max})$, 第三行用到了$G$的固定非扩张性.

- (9) 可以推出

$$M\sum_{k=0}^{\infty}\|G(z^{(k)})\|_2^2 \le \|z^{(0)} - z^*\|_2^2, \quad \|G(z^{(k)})\|_2 \to 0$$

- 并且$\|z^{(k)} - z^*\|_2$不增；进而$z^{(k)}$ 有界
- 因为$\|z^{(k)} - z^*\|_2$ 不增, 因此极限$\lim_{k\to\infty}\|z^{(k)} - z^*\|_2$ 存在

证明(续)

- 因为迭代点列$z^{(k)}$有界，因此它有收敛子列
- 设$\bar{z}_k$为收敛子列，极限为$\bar{z}$; 由G的连续性,

$$0 = \lim_{k\to\infty} G(\bar{z}_k) = G(\bar{z})$$

因此$\bar{z}$是G的零点，且极限$\lim_{k\to\infty} \|z^{(k)} - \bar{z}\|_2$存在

- 设存在两个子列分别收敛于$\bar{z}_1, \bar{z}_2$, 则

$$\lim_{k\to\infty} \|z^{(k_{j_1})} - \bar{z}_1\|_2, \quad \lim_{k\to\infty} \|z^{(k_{j_2})} - \bar{z}_2\|_2$$

均存在. 由$z^{(k)}$的收敛性知

$$\|\bar{z}_2 - \bar{z}_1\|_2 = \lim_{k\to\infty} \|z^{(k)} - \bar{z}_1\|_2 = \lim_{k\to\infty} \|z^{(k)} - \bar{z}_2\|_2 = 0$$

(全序列收敛)