机器学习中的优化算法

Lecture09: 复合优化算法–交替方向乘子法

张立平

清华大学数学科学系

办公室：理科楼#A302, Tel: 62798531

E-mail: lipingzhang@tsinghua.edu.cn

Contents and Acknowledgement

- 教材：**最优化：建模、算法与理论**

  http://bicmr.pku.edu.cn/ wenzw/bigdata2021.html

- 致谢：北京大学**文再文**教授

Outline of ADMM

- 交替方向乘子法
- Douglas-Rachford Splitting算法
- 常用变形和技巧
- 应用举例
- ADMM收敛性分析

ADMM针对的典型问题

▶典型优化问题形式:

$$\min_{x_1,x_2} \quad f_1(x_1)+f_2(x_2), \qquad \text{s.t.} \quad A_1x_1+A_2x_2=b. \tag{1}$$

- f_1, f_2是适当闭凸函数，但不要求是光滑的，$x_1\in\mathbb{R}^n, x_2\in\mathbb{R}^m$, $A_1\in\mathbb{R}^{p\times n}, A_2\in\mathbb{R}^{p\times m}, b\in\mathbb{R}^p$.
- 问题特点：目标函数可以分成彼此分离的两块，但是变量被线性约束结合在一起．常见的一些无约束和带约束的优化问题都可以表示成这一形式．

【例】可以分成两块的无约束优化问题:

$$\min_x \quad f_1(x)+f_2(x) \quad \Longleftrightarrow \quad \begin{cases} \min\limits_{x,z} & f_1(x)+f_2(z) \\ \text{s.t.} & x-z=0. \end{cases}$$

【例】 带线性变换的无约束优化问题:

$$\min_{x} \quad f_1(x) + f_2(Ax) \quad \Longleftrightarrow \quad \begin{cases} \min\limits_{x,z} & f_1(x) + f_2(z) \\ \text{s.t.} & Ax - z = 0. \end{cases}$$

【例】 凸集$C \subset \mathbb{R}^n$上的约束优化问题:

$$\begin{cases} \min\limits_{x} & f(x) \\ \text{s.t.} & Ax \in C \end{cases} \quad \Longleftrightarrow \quad \begin{cases} \min\limits_{x,z} & f(x) + I_C(z) \\ \text{s.t.} & Ax - z = 0. \end{cases}$$

【例】 全局一致性问题:

$$\min_{x} \quad \sum_{i=1}^{N} \phi_i(x) \quad \Longleftrightarrow \quad \begin{cases} \min\limits_{x_i,z} & \sum\limits_{i=1}^{N} \phi_i(x_i) \\ \text{s.t.} & x_i - z = 0,\ i = 1, 2, \cdots, N. \end{cases}$$

交替方向乘子法ADMM

▶典型优化问题(1)的增广拉格朗日函数:

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

▶ ALM迭代:

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

$$y^{k+1} = y^k + \tau\rho(A_1x_1^{k+1} + A_2x_2^{k+1} - b). \tag{4}$$

▶ ADMM的基本思路: ALM中(3)同时对x_1和x_2进行优化有时候比较困难，而固定一个变量求解关于另一个变量的极小问题可能比较简单，因此我们可

以考虑对x_1 和x_2 交替求极小.

$$x_1^{k+1} = \arg\min_{x_1} L_\rho(x_1, x_2^k, y^k), \tag{5}$$

$$x_2^{k+1} = \arg\min_{x_2} L_\rho(x_1^{k+1}, x_2, y^k), \tag{6}$$

$$y^{k+1} = y^k + \tau\rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \tag{7}$$

其中τ 为步长，通常取值于$\left(0, \frac{1+\sqrt{5}}{2}\right]$.

▶ ADMM的收敛准则: 根据最优性条件，若x_1^*, x_2^*为问题(1)的最优解，y^*为对应的拉格朗日乘子，则以下条件满足：

$$0 \in \partial_{x_1} L(x_1^*, x_2^*, y^*) = \partial f_1(x_1^*) + A_1^{\mathrm{T}} y^*, \tag{8a}$$

$$0 \in \partial_{x_2} L(x_1^*, x_2^*, y^*) = \partial f_2(x_2^*) + A_2^{\mathrm{T}} y^*, \tag{8b}$$

$$A_1 x_1^* + A_2 x_2^* = b, \tag{8c}$$

其中$L(x_1, x_2, y) = f_1(x_1) + f_2(x_2) + y^{\mathrm{T}}(A_1 x_1 + A_2 x_2 - b)$. 条件(8c)又称为原始可行性条件，条件(8a)和条件(8b)又称为对偶可行性条件.

ADMM单步迭代最优性条件

- 由x_2的更新

$$x_2^k = \arg\min_x \left\{ f_2(x) + \frac{\rho}{2} \left\| A_1 x_1^k + A_2 x - b + \frac{y^{k-1}}{\rho} \right\|^2 \right\},$$

$$\Longrightarrow \qquad 0 \in \partial f_2(x_2^k) + A_2^{\mathrm{T}}[y^{k-1} + \rho(A_1 x_1^k + A_2 x_2^k - b)]. \tag{9}$$

当$\tau = 1$时, 由(7)可知 $0 \in \partial f_2(x_2^k) + A_2^{\mathrm{T}} y^k$.

- 由x_1的更新

$$x_1^k = \arg\min_x \left\{ f_1(x) + \frac{\rho}{2} \| A_1 x + A_2 x_2^{k-1} - b + \frac{y^{k-1}}{\rho} \|^2 \right\},$$

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

当$\tau = 1$时, 由(7)可知

$$0 \in \partial f_1(x_1^k) + A_1^{\mathrm{T}}(y^k + A_2(x_2^{k-1} - x_2^k)). \tag{10}$$

对比(10)和条件(8a)可知, 多出来的项为$A_1^{\mathrm{T}} A_2(x_2^{k-1} - x_2^k)$. 因此要检测对偶可行性只需要检测残差:

$$s^k = A_1^{\mathrm{T}} A_2(x_2^{k-1} - x_2^k).$$

- 当x_2更新取到精确解且$\tau = 1$时, 判断ADMM 是否收敛只需要检测前述两个残差r^k, s^k是否充分小:

$$\begin{aligned} 0 &\approx \|r^k\| = \|A_1 x_1^k + A_2 x_2^k - b\| && \text{原始可行性}, \\ 0 &\approx \|s^k\| = \|A_1^{\mathrm{T}} A_2(x_2^{k-1} - x_2^k)\| && \text{对偶可行性}. \end{aligned} \tag{11}$$

► 典型优化问题(1)的ADMM的收敛准则:

$$\begin{aligned} 0 &\approx \|r^k\| = \|A_1 x_1^k + A_2 x_2^k - b\| && \text{原始可行} \\ 0 &\approx \|s^k\| = \|A_1^{\mathrm{T}} A_2(x_2^{k-1} - x_2^k)\| && \text{对偶可行} \end{aligned}$$

Douglas-Rachford Splitting (DRS) 算法

DRS算法是一类非常重要的算子分裂算法, 用于求解复合优化问题

$$\min \quad \psi(x) = f(x) + h(x)$$

其中g, h 是闭凸函数. DRS算法的迭代格式是:

$$\begin{aligned} x^{k+1} &= \mathrm{prox}_{th}(z^k) \\ y^{k+1} &= \mathrm{prox}_{tf}(2x^{k+1} - z^k) \\ z^{k+1} &= z^k + y^{k+1} - x^{k+1} \end{aligned} \tag{12}$$

- 引入DRS算法的等价变形: 按y, z, x的顺序更新

$$\begin{aligned} y^{k+1} &= \mathrm{prox}_{tf}(2x^k - z^k) \\ z^{k+1} &= z^k + y^{k+1} - x^k \\ x^{k+1} &= \mathrm{prox}_{th}(z^{k+1}) \end{aligned}$$

- 引入辅助变量$w^k = z^k - x^k$, 消去z^k, z^{k+1}得等价的DRS算法的迭代格式:

$$\begin{aligned} y^{k+1} &= \text{prox}_{tf}(x^k - w^k) \\ x^{k+1} &= \text{prox}_{th}(w^k + y^{k+1}) \\ w^{k+1} &= w^k + y^{k+1} - x^{k+1} \end{aligned} \tag{13}$$

- DRS算法还可以写成关于z^k的不动点迭代: $z^{k+1} = T(z^k)$,

$$T(z) = z + \text{prox}_{tf}(2\text{prox}_{th}(z) - z) - \text{prox}_{th}(z).$$

  – 优势: 去掉变量x^k, y^k; 利用不动点迭代的收敛性、压缩算子; 可推广成加速算法; T的不动点与$\min \psi(x)$的最小值点密切相关.

  – $z = T(z) \quad \Rightarrow \quad \text{prox}_{th}(z) \in \arg\min \psi(x)$
  令$x = \text{prox}_{th}(z)$, 由$z = T(z)$知, $\text{prox}_{tf}(2x - z) = x = \text{prox}_{th}(z)$
  $\Rightarrow \quad x - z \in t\partial f(x), \quad z - x \in t\partial h(x) \quad \Rightarrow \quad \text{prox}_{th}(z) \in \arg\min \psi(x)$

  – 设$x \in \arg\min \psi(x)$, 则存在$u \in t\partial f(x) \cap (-t\partial h(x))$使得$x - u$是$T$的不动点.
  $u \in t\partial f(x) \cap (-t\partial h(x)) \Rightarrow x = \text{prox}_{tf}(x + u), \; x = \text{prox}_{th}(x - u)$

▶Douglas-Rachford迭代映射

$$T(z) = \ z + \mathrm{prox}_{tf}(2\mathrm{prox}_{th}(z) - z) - \mathrm{prox}_{th}(z)$$
$$G(z) = \ z - T(z) = \mathrm{prox}_{th}(z) - \mathrm{prox}_{tf}(2\mathrm{prox}_{th}(z) - z)$$

- T 是固定非扩张的 (firmly non-expansive)

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

① 令$x = \mathrm{prox}_{th}(z), \hat{x} = \mathrm{prox}_{th}(\hat{z})$, 则有

$$(x - \hat{x})^T (z - \hat{z}) \geq \|x - \hat{x}\|_2^2.$$

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

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

③ 令$a = x - \hat{x}, b = y - \hat{y}, c = z - \hat{z}$, 则由$T(z) = z + y - x$
和$T(\hat{z}) = \hat{z} + \hat{y} - \hat{x}$ 及①②知,

$$a^T c \geq \|a\|_2^2, \quad (2a - c)^T b \geq \|b\|_2^2, \quad T(z) - T(\hat{z}) = b + c - a.$$

于是,

$$
\begin{aligned}
&(T(z)-T(\hat{z}))^T(z-\hat{z}) \\
=&\|T(z)-T(\hat{z})\|_2^2+(b+c-a)^T(a-b) \\
=&\|T(z)-T(\hat{z})\|_2^2+(a^Tc-\|a\|_2^2)+\left((2a-c)^Tb-\|b\|_2^2\right) \\
\geq&\|T(z)-T(\hat{z})\|_2^2.
\end{aligned}
$$

- G也是固定非扩张的

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

- 添加松弛项来加快收敛速度: $1 < \rho < 2$超松弛; $0 < \rho < 1$欠松弛

$$z^{k+1} = z^k + \rho(T(z^k) - z^k)$$

- DRS的松弛形式1:

$$\begin{cases} x^{k+1} = \mathrm{prox}_{th}(z^k) \\ y^{k+1} = \mathrm{prox}_{tf}(2x^{k+1} - z^k) \\ z^{k+1} = z^k + \rho(y^{k+1} - x^{k+1}) \end{cases}$$

- DRS的松弛形式2:

$$\begin{cases} y^{k+1} = \mathrm{prox}_{tf}(x^k - w^k) \\ x^{k+1} = \mathrm{prox}_{th}((1-\rho)x^k + \rho y^{k+1} + w^k) \\ w^{k+1} = w^k + \rho y^{k+1} + (1-\rho)x^k - x^{k+1} \end{cases}$$

DRS与ADMM的关系

$$\text{(P)} \qquad \min \quad f_1(x_1) + f_2(x_2) \quad \textbf{s.t.} \quad A_1x_1 + A_2x_2 = b$$

$$\text{(D)} \qquad \max \quad -b^T z - f_1^*(-A_1^T z) - f_2^*(-A_2^T z) \quad \Leftrightarrow$$

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

▶ 对(P)用ADMM等价于对(D)用DRS

Theorem 1. 令$w^1 = -tA_2x_2^0$, *则对(D)用DRS算法*(13)*等价于对(P)用ADMM.*

- 由$y^{k+1} = \operatorname{prox}_{tf}(x^k - w^k)$知

$$0 \in tb - tA_1\partial f_1^*(-A_1^{\mathrm{T}} y^{k+1}) - x^k + w^k + y^{k+1},$$

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

$$y^{k+1} = x^k - w^k + t(A_1x_1^k - b), \tag{14}$$

因此,

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

- 由$x^{k+1} = \mathrm{prox}_{th}(w^k + y^{k+1})$知

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

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

$$x^{k+1} = x^k + t(A_1x_1^k + A_2x_2^k - b), \tag{15}$$

于是

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

- 由$w^{k+1} = w^k + y^{k+1} - x^{k+1}$及(14)和(15)知$w^{k+1} = -tA_2x_2^k$.
- 定义(P)的增广拉格朗日函数

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

令$z^k = x^{k+1}$, 则有

$$
\begin{aligned}
x_1^k &= \arg\min_{x_1} L_t(x_1, x_2^{(k-1)}, z^{(k-1)}) \\
&= \arg\min_{x_1} \left(f_1(x_1) + (z^{k-1})^T A_1 x_1 + \frac{t}{2} \|A_1 x_1 + A_2 x_2^{k-1} - b\|_2^2 \right) \\
x_2^k &= \arg\min_{x_2} L_t(x_1^k, x_2, z^{k-1}) \\
&= \arg\min_{x_2} \left(f_2(x_2) + (z^{k-1})^T A_2 x_2 + \frac{t}{2} \|A_1 x_1^k + A_2 x_2 - b\|_2^2 \right) \\
z^k &= z^{k-1} + t(A_1 x_1^k + A_2 x_2^k - b)
\end{aligned}
$$

ADMM中的常用技巧: 线性化

▶线性化技巧使用近似点项对子问题目标函数进行二次近似, 使得子问题有显式解.

- 考虑子问题(5):

$$\min_{x_1} \quad f_1(x_1) + \frac{\rho}{2}\|A_1x_1 - v^k\|^2, \quad v^k = b - A_2x_2^k - \frac{1}{\rho}y^k. \tag{16}$$

- 当子问题目标函数可微时, 线性化将问题(16)变为

$$x_1^{k+1} = \arg\min_{x_1} \left\{ \left(\nabla f_1(x_1^k) + \rho A_1^{\mathrm{T}}(A_1x_1^k - v^k)\right)^{\mathrm{T}} x_1 + \frac{1}{2\eta_k}\|x_1 - x_1^k\|_2^2 \right\},$$

其中η_k是步长参数. 这等价于做一步梯度下降.

- 当目标函数不可微时, 可以考虑只将二次项线性化:

$$x_1^{k+1} = \arg\min_{x_1} \left\{ f_1(x_1) + \rho\left(A_1^{\mathrm{T}}(A_1x_1^k - v^k)\right)^{\mathrm{T}} x_1 + \frac{1}{2\eta_k}\|x_1 - x_1^k\|_2^2 \right\}.$$

这等价于做一步近似点梯度步. 若$f_1(x_1)$是可微函数与不可微函数的和时, 将可微部分线性化.

ADMM中的常用技巧: 缓存分解

► 缓存分解

- 如果目标函数中含二次函数, 例如$f_1(x_1) = \frac{1}{2}\|Cx_1 - d\|_2^2$, 那么针对$x_1$的更新(5)等价于求解线性方程组

$$(C^{\mathrm{T}}C + \rho A_1^{\mathrm{T}} A_1)x_1 = C^{\mathrm{T}}d + \rho A_1^{\mathrm{T}} v^k.$$

- 虽然子问题有显式解, 但是每步求解的复杂度仍然比较高，这时候可以考虑用**缓存分解**的方法. 首先对$C^{\mathrm{T}}C + \rho A_1^{\mathrm{T}} A_1$进行Cholesky 分解并缓存分解的结果, 在每步迭代中只需要求解简单的三角形方程组.

- 当ρ发生更新时, 就要重新进行分解. 特别地, 当$C^{\mathrm{T}}C + \rho A_1^{\mathrm{T}} A_1$具有特殊结构(一部分容易求逆，另一部分是低秩的情形)时，可以用SMW公式来求逆. $(A + BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}$

ADMM中的常用技巧: 优化转移

▶ 优化转移就是为了方便求解子问题, 可以用一个性质好的矩阵D近似二次项$A_1^{\mathrm{T}}A_1$, 此时子问题(16)替换为

$$\begin{aligned} x_1^{k+1} &= \arg\min_{x_1}\left\{f_1(x_1)+\frac{\rho}{2}\|A_1x_1-v^k\|_2^2\right. \\ &\quad \left.+\frac{\rho}{2}(x_1-x^k)^{\mathrm{T}}(D-A_1^{\mathrm{T}}A_1)(x_1-x_1^k)\right\}. \end{aligned}$$

- 通过选取合适的D, 当计算

$$\arg\min_{x_1}\left\{f_1(x_1)+\frac{\rho}{2}x_1^{\mathrm{T}}Dx_1\right\}$$

明显比计算

$$\arg\min_{x_1}\{f_1(x_1)+\frac{\rho}{2}x_1^{\mathrm{T}}A_1^{\mathrm{T}}A_1x_1\}$$

要容易时, 优化转移可以极大地简化子问题的计算. 特别地, 当$D=\dfrac{\eta_k}{\rho}I$时, 优化转移等价于做单步的近似点梯度步.

ADMM中的常用技巧: 二次罚项系数的动态调节

▶ 动态调节二次罚项系数在交替方向乘子法的实际应用中是一个非常重要的数值技巧.

- 由(11)知, 求解过程中二次罚项系数ρ太大会导致原始可行性$\|r^k\|$下降很快, 但是对偶可行性$\|s^k\|$下降很慢; 二次罚项系数太小, 则效果相反. 这都会导致收敛比较慢或得到的解的可行性很差. 一个自然的想法是在每次迭代时动态调节惩罚系数ρ的大小, 从而使得原始可行性和对偶可行性能够以比较一致的速度下降到零.

- 简单有效的动态调节二次罚项系数的方式:

$$\rho^{k+1} = \begin{cases} \gamma_p \rho^k, & \|r^k\| > \mu \|s^k\|, \\ \rho^k/\gamma_d & \|s^k\| > \mu \|r^k\|, \\ \rho^k, & \text{otherwise}, \end{cases}$$

其中$\mu > 1, \gamma_p > 1, \gamma_d > 1$是参数, 常见的选择为$\mu = 10, \gamma_p = \gamma_d = 2$. 在迭代过程中将原始可行性$\|r^k\|$和对偶可行性$\|s^k\|$保持在彼此的$\mu$倍内.

选择合适的惩罚项$\|\cdot\|_P \quad (P = F^T F)$

ADMM中的常用技巧: 超松弛

考虑求解优化问题(1)的ADMM迭代格式:

$$x_1^{k+1} = \arg\min_{x_1} L_\rho(x_1, x_2^k, y^k), \tag{17}$$

$$x_2^{k+1} = \arg\min_{x_2} L_\rho(x_1^{k+1}, x_2, y^k), \tag{18}$$

$$y^{k+1} = y^k + \tau\rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b). \tag{19}$$

- 在(18)与(19)中, $A_1 x_1^{k+1}$可以被替换为

  $$\alpha_k A_1 x_1^{k+1} + (1-\alpha_k)(A_2 x_2^k - b),$$

  其中$\alpha_k \in (0,2)$是一个松弛参数.

- 当$\alpha_k > 1$时, 这种技巧称为超松弛. 当$\alpha_k < 1$时, 这种技巧称为欠松弛. 实验表明$\alpha_k \in [1.5, 1.8]$的超松弛可以提高收敛速度.

ADMM应用

LASSO 问题 (合适的拆分: $x = z$)

$$\min \quad \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|^2 \quad \Longleftrightarrow \quad \begin{cases} \min\limits_{x,z} & \frac{1}{2}\|Ax - b\|^2 + \mu\|z\|_1, \\ \text{s.t.} & x = z. \end{cases}$$

- 交替方向乘子法迭代格式为

$$\begin{aligned} x^{k+1} &= \arg\min_x \left\{ \frac{1}{2}\|Ax - b\|^2 + \frac{\rho}{2}\|x - z^k + y^k/\rho\|_2^2 \right\}, \\ &= (A^{\mathrm{T}}A + \rho I)^{-1}(A^{\mathrm{T}}b + \rho z^k - y^k), \\ z^{k+1} &= \arg\min_z \left\{ \mu\|z\|_1 + \frac{\rho}{2}\|x^{k+1} - z + y^k/\rho\|^2 \right\}, \\ &= \operatorname{prox}_{(\mu/\rho)\|\cdot\|_1}\left(x^{k+1} + y^k/\rho\right), \\ y^{k+1} &= y^k + \tau\rho(x^{k+1} - z^{k+1}). \end{aligned}$$

- 因为$\rho > 0$, 所以$A^{\mathrm{T}}A + \rho I$总是可逆的．x迭代本质上是计算一个岭回归问题(ℓ_2范数平方正则化的最小二乘问题). 而对z的更新为ℓ_1范数的邻近算子, 同样有显式解．在求解x迭代时，若使用固定的罚因子ρ, 我们可以缓存矩阵$A^{\mathrm{T}}A + \rho I$的初始分解, 从而减小后续迭代中的计算量.

- 在LASSO 问题中, 矩阵$A \in \mathbb{R}^{m\times n}$通常有较多的列$m \ll n$, 因此$A^{\mathrm{T}}A \in \mathbb{R}^{n\times n}$是一个低秩矩阵，二次罚项的作用就是将$A^{\mathrm{T}}A$增加了一个正定项．该ADMM 主要运算量来自更新$x$变量时求解线性方程组，复杂度为$O(n^3)$. 若使用缓存分解技术或SMW 公式:

$$(A^{\mathrm{T}}A + \rho I_n)^{-1} = \frac{1}{\rho}\left[I_n - A^{\mathrm{T}}(AA^{\mathrm{T}} + \rho I_m)^{-1}A\right],$$

  则可进一步降低每次迭代的运算量, 复杂度为$O(m^3)$.

■ **考虑LASSO 问题的对偶问题**:

$$\begin{aligned} \min \quad & b^{\mathrm{T}}y + \frac{1}{2}\|y\|^2, \\ \text{s.t.} \quad & \|A^{\mathrm{T}}y\|_\infty \le \mu. \end{aligned} \tag{20}$$

- 引入约束$A^{\mathrm{T}}y + z = 0$, 则

$$(20) \quad \Longleftrightarrow \quad \left\{ \begin{aligned} \min \quad & \underbrace{b^{\mathrm{T}}y + \frac{1}{2}\|y\|^2}_{f(y)} + \underbrace{I_{\|z\|_\infty \le \mu}(z)}_{h(z)}, \\ \text{s.t.} \quad & A^{\mathrm{T}}y + z = 0. \end{aligned} \right. \tag{21}$$

- 对偶问题(21)的增广拉格朗日函数为

$$L_\rho(y, z, x) = b^{\mathrm{T}}y + \frac{1}{2}\|y\|^2 + I_{\|z\|_\infty \le \mu}(z) - x^{\mathrm{T}}(A^{\mathrm{T}}y + z) + \frac{\rho}{2}\|A^{\mathrm{T}}y + z\|^2.$$

- 当固定y, x时, 对z的更新即向无穷范数球$\{z \| \|z\|_\infty \le \mu\}$做欧几里得投影, 将每个分量截断在区间$[-\mu, \mu]$中. 当固定$z, x$时, 对$y$的更新即求解线性方程组

$$(I + \rho A A^{\mathrm{T}})y = A(x^k - \rho z^{k+1}) - b. \tag{22}$$

- LASSO 问题的对偶问题(21)的ADMM 迭代格式为

$$
\begin{aligned}
z^{k+1} &= \mathcal{P}_{\|z\|_\infty \le \mu}\left(x^k/\rho - A^{\mathrm{T}} y^k\right), \\
y^{k+1} &= (I + \rho A A^{\mathrm{T}})^{-1}\Big(A(x^k - \rho z^{k+1}) - b\Big), \\
x^{k+1} &= x^k - \tau\rho(A^{\mathrm{T}} y^{k+1} + z^{k+1}).
\end{aligned}
$$

  - 虽然ADMM 应用于对偶问题也需要求解一个线性方程组(22), 但由于LASSO 问题的特殊性$m \ll n$, 求解y更新的线性方程组(22)需要的计算量是$O(m^3)$, 使用缓存分解技巧后可进一步降低至$O(m^2)$, 这大大小于针对原始问题的ADMM.

- $m = 512; n = 1024; A = \text{randn}(m, n); u = \text{sprandn}(n, 1, r); b = A * u;$
取$\mu = 10^{-3}$, 用ADMM求解LASSO原问题和对偶问题: 取$\tau = 1.618$, 原问题和对偶问题分别取$\rho = 0.01, 100.$

终止条件设为: $|f(x^k) - f(x^{k-1})| < 10^{-8}$ 和最大迭代次数2000. 对原问题还添加终止条件$\|x^k - z^k\| < 10^{-10}$; 对对偶问题$\|A^T y^k + z^k\| < 10^{-10}$.

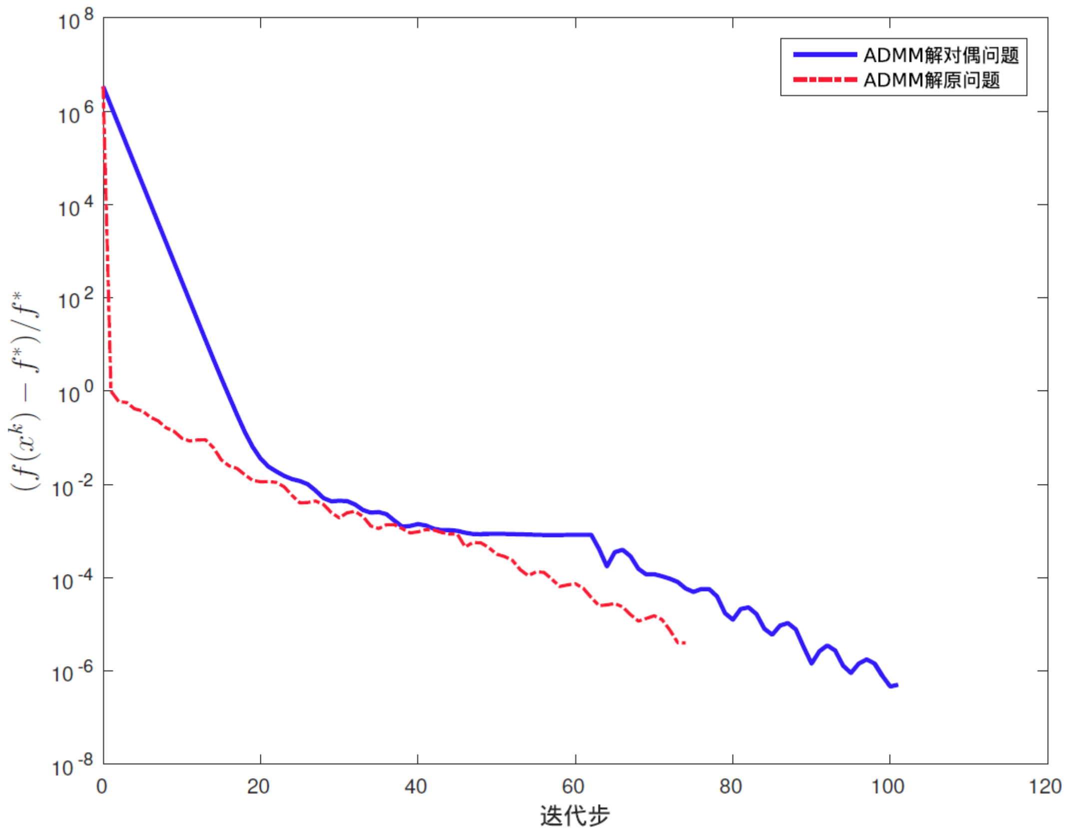

Figure 1: ADMM 求解LASSO 问题: 原问题所需迭代次数少, 对偶问题每一次迭代所需时间短.

■ **广义LASSO 问题**指x本身不稀疏, 但在某种变换下是稀疏的:

$$\min_{x} \quad \mu\|Fx\|_1 + \frac{1}{2}\|Ax-b\|^2. \tag{23}$$

- 当$F \in \mathbb{R}^{(n-1)\times n}$是一阶差分矩阵

$$F_{ij} = \begin{cases} 1, & j = i+1, \\ -1, & j = i, \\ 0, & \text{otherwise}, \end{cases}$$

且$A = I$时, 广义LASSO问题(23)为图像去噪问题的TV 模型:

$$\min_{x} \quad \frac{1}{2}\|x-b\|^2 + \mu\sum_{i=1}^{n-1}|x_{i+1}-x_i|.$$

当$A = I$且F是二阶差分矩阵时, 问题(23)被称为一范数趋势滤波.

► 广义LASSO问题(23)等价于

$$\begin{aligned} \min_{x,z} \quad & \frac{1}{2}\|Ax-b\|^2 + \mu\|z\|_1 \\ \text{s.t.} \quad & Fx - z = 0, \end{aligned} \tag{24}$$

- 对广义LASSO问题(24)的ADMM迭代为

$$\begin{aligned} x^{k+1} &= (A^{\mathrm{T}}A + \rho F^{\mathrm{T}}F)^{-1}\left(A^{\mathrm{T}}b + \rho F^{\mathrm{T}}\left(z^k - \frac{y^k}{\rho}\right)\right), \\ z^{k+1} &= \mathrm{prox}_{(\mu/\rho)\|\cdot\|_1}\left(Fx^{k+1} + \frac{y^k}{\rho}\right), \\ y^{k+1} &= y^k + \tau\rho(Fx^{k+1} - z^{k+1}). \end{aligned}$$

- 注: 对于全变差去噪问题, $A^{\mathrm{T}}A + \rho F^{\mathrm{T}}F$是三对角矩阵, 此时$x$迭代可以在$\mathcal{O}(n)$的时间复杂度内解决；对于图像去模糊问题，$A$ 是卷积算子，则利用傅里叶变换可将求解方程组的复杂度降低至$O(n\log n)$；对于一范数趋势滤波问题，$A^{\mathrm{T}}A + \rho F^{\mathrm{T}}F$是五对角矩阵，所以$x$迭代仍可以在$\mathcal{O}(n)$的时间复杂度内解决. GOLUB GH, VAN LOAN CF, *Matrix computations* 2012

■ 稀疏逆协方差矩阵估计问题:

$$\min_{X} \quad \langle S, X\rangle - \ln\det X + \mu\|X\|_1, \tag{25}$$

其中S是已知的对称矩阵, 通常由样本协方差矩阵得到. 变量$X \in \mathcal{S}^n_{++}$, $\|\cdot\|_1$定义为矩阵所有元素绝对值的和.

- (25)的目标函数由光滑项和非光滑项组成, 将问题的两部分分离:

$$\begin{aligned} \min \quad & \underbrace{\langle S, X\rangle - \ln\det X}_{f(X)} + \underbrace{\mu\|Z\|_1}_{h(Z)}, \\ \text{s.t.} \quad & X = Z. \end{aligned}$$

- 其增广拉格朗日函数为

$$L_\rho(X, Z, U) = \langle S, X\rangle - \ln\det X + \mu\|Z\|_1 + \langle U, X - Z\rangle + \frac{\rho}{2}\|X - Z\|_F^2.$$

▶ ADMM迭代:

- 对X的更新, 固定Z^k, U^k, 关于X的子问题是凸的, 故由最优性条件

$$S - X^{-1} + U^k + \rho(X - Z^k) = 0,$$

$$\Longrightarrow \quad X^{k+1} = Q\mathrm{Diag}(x_1, x_2, \cdots, x_n)Q^{\mathrm{T}},$$

其中Q包含矩阵$S - \rho Z^k + U^k$的所有特征向量, x_i的表达式为

$$x_i = \frac{-d_i + \sqrt{d_i^2 + 4\rho}}{2\rho},$$

d_i为矩阵$S-\rho Z^k+U^k$的第i个特征值.

- 对Z的更新, 固定X^{k+1}, U^k, 关于Z的更新为矩阵ℓ_1 范数的邻近算子.
- 乘子更新: $U^{k+1}=U^k+\tau\rho(X^{k+1}-Z^{k+1})$.

■ 矩阵分离问题:

$$\begin{aligned} \min_{X,S} \quad & \|X\|_*+\mu\|S\|_1 \\ \textbf{s.t.} \quad & X+S=M, \end{aligned} \tag{26}$$

其中$\|\cdot\|_1$与$\|\cdot\|_*$分别表示矩阵ℓ_1范数与核范数.

- 问题(26)的增广拉格朗日函数

$$L_\rho(X,S,Y)=\|X\|_*+\mu\|S\|_1+\langle Y, X+S-M\rangle+\frac{\rho}{2}\|X+S-M\|_F^2.$$

- 对X的更新

$$
\begin{aligned}
X^{k+1} &= \arg\min_X \ L_\rho(X, S^k, Y^k) \\
&= \arg\min_X \left\{ \|X\|_* + \frac{\rho}{2} \left\| X + S^k - M + \frac{Y^k}{\rho} \right\|_F^2 \right\} \\
&= \arg\min_X \left\{ \frac{1}{\rho}\|X\|_* + \frac{1}{2} \left\| X + S^k - M + \frac{Y^k}{\rho} \right\|_F^2 \right\} \\
&= U \mathrm{Diag}\Big(\mathrm{prox}_{(1/\rho)\|\cdot\|_1}(\sigma(A)) \Big) V^{\mathrm{T}},
\end{aligned}
$$

其中$A = M - S^k - \frac{Y^k}{\rho}$, $\sigma(A)$为A的所有非零奇异值构成的向量并且$U\mathrm{Diag}(\sigma(A))V^{\mathrm{T}}$为$A$的奇异值分解.

- 对S的更新

$$\begin{aligned} S^{k+1} &= \arg\min_S \ L_\rho(X^{k+1}, S, Y^k) \\ &= \arg\min_S \left\{ \mu\|S\|_1 + \frac{\rho}{2} \left\| X^{k+1} + S - M + \frac{Y^k}{\rho} \right\|_F^2 \right\} \\ &= \mathrm{prox}_{(\mu/\rho)\|\cdot\|_1} \left(M - X^{k+1} - \frac{Y^k}{\rho} \right). \end{aligned}$$

▶ 矩阵分离问题(26)的ADMM迭代:

$$\begin{aligned} X^{k+1} &= U\mathrm{Diag}\left(\mathrm{prox}_{(1/\rho)\|\cdot\|_1}(\sigma(A)) \right) V^{\mathrm{T}}, \\ S^{k+1} &= \mathrm{prox}_{(\mu/\rho)\|\cdot\|_1} \left(M - L^{k+1} - \frac{Y^k}{\rho} \right), \\ Y^{k+1} &= Y^k + \tau\rho(X^{k+1} + S^{k+1} - M). \end{aligned}$$

全局一致性优化问题:

$$\min_{x_i,z} \sum_{i=1}^{N} \phi_i(x_i) \quad \text{s.t.} \;\; x_i - z = 0, \; i = 1, 2, \cdots, N.$$

- 增广拉格朗日函数

$$L_\rho(x_1, \cdots, x_N, z, y_1, \cdots, y_N) = \sum_{i=1}^{N} \phi_i(x_i) + \sum_{i=1}^{N} y_i^{\mathrm{T}}(x_i - z) + \frac{\rho}{2} \sum_{i=1}^{N} \|x_i - z\|^2.$$

注: 虽然表面上看增广拉格朗日函数有$(N+1)$个变量块, 但本质上还是两个变量块. 这是因为在更新某个x_i时并没有利用其他x_i, 所有x_i可以看成一个整体. 相应地, 所有乘子y_i也可以看成一个整体.

▶ 全局一致性优化问题的ADMM迭代:

$$x_i^{k+1} = \mathrm{prox}_{\phi_i/\rho}\left(z^k - y_i^k/\rho\right), \; i = 1, 2, \cdots, N,$$

$$z^{k+1} = \frac{1}{N} \sum_{i=1}^{N} \left(x_i^{k+1} + y_i^k/\rho\right),$$

$$y_i^{k+1} = {y_i}^k + \tau\rho({x_i}^{k+1} - z^{k+1}), \; i = 1, 2, \cdots, N.$$

■ 混合全变差和ℓ_1正则 $\min_x$ **TV**$(x) + \alpha\|Wx\|_1$, **s.t.** $\|Rx - b\|_2 \le \sigma$

- 新模型:

$$\begin{aligned}\min_x \ & \sum_{i=1}^{N}\|z_i\|_2 + \alpha\|Wx\|_1 + \mathcal{I}(\|y\|_2 \le \sigma)\\ \text{s.t. } & z_i = D_i x,\ \forall i = 1, \cdots, N\\ & y = Rx - b.\end{aligned}$$

- 将变量分为两块: x 和$(y, \{z_i\})$, 约束化成:

$$\begin{pmatrix} R \\ D_1 \\ \vdots \\ D_N \end{pmatrix} x - \begin{pmatrix} y \\ z_1 \\ \vdots \\ z_N \end{pmatrix} = \begin{pmatrix} b \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

- x的子问题不易求解, 使用线性化二次项技巧简化proximal算子,进行非精确更新

非凸约束优化问题的ADMM

考虑如下约束优化问题:

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}} & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in \mathcal{S} \end{array} \quad \Leftrightarrow \quad \begin{array}{ll} \min\limits_{\boldsymbol{x}} & f(\boldsymbol{x}) + \mathbb{I}_{\mathcal{S}}(\boldsymbol{z}) \\ \text{s.t.} & \boldsymbol{x} - \boldsymbol{z} = \mathbf{0} \end{array}$$

其中f是闭凸函数, $\mathcal{S}$是非凸集合. 其交替方向乘子法迭代格式:

$$\begin{cases} \boldsymbol{x}^{k+1} = \arg\min\limits_{\boldsymbol{x}} \left(f(\boldsymbol{x}) + \frac{\rho}{2} \|\boldsymbol{x} - \boldsymbol{z}^k + \boldsymbol{u}^k\|_2^2 \right), \\ \boldsymbol{z}^{k+1} = \mathcal{P}_{\mathcal{S}}(\boldsymbol{x}^{k+1} + \boldsymbol{u}^k), \quad \mathcal{P}_{\mathcal{S}}(\boldsymbol{z}) \text{ 是将}\boldsymbol{z}\text{投影到集合}\mathcal{S}, \text{难!} \\ \boldsymbol{u}^{k+1} = \boldsymbol{u}^k + \tau\rho(\boldsymbol{x}^{k+1} - \boldsymbol{z}^{k+1}). \end{cases}$$

▶ 特殊非凸集合$\mathcal{S}$

- 基数: $\mathcal{S} = \{\boldsymbol{x} | \|\boldsymbol{x}\|_0 \le c\}$, 计算向量$\boldsymbol{v}$到$\mathcal{S}$中的投影就是保留$\boldsymbol{v}$分量中绝对值从大到小排列的前$c$ 个, 其余分量变成0. 假设$\boldsymbol{v}$ 的各个分量满

足$|v_{i_1}| \geq |v_{i_2}| \geq \cdots \geq |v_{i_n}|$, 则投影算子

$$(\mathcal{P}_{\mathcal{S}}(\boldsymbol{v}))_i = \begin{cases} v_i, & i \in \{i_1, \ldots, i_c\}, \\ 0, & \textbf{otherwise}. \end{cases}$$

- 低秩投影: $\mathcal{S} = \{X \in \mathbb{R}^{m\times n} | \mathrm{rank}(X) \leq r\}$, 计算矩阵$X$到$\mathcal{S}$中的投影就是对$X$做截断奇异值分解. 设$X$的奇异值分解为

$$X = \sum_i^{\min\{m,n\}} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T,$$

其中$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{m,n\}}$为奇异值, $\boldsymbol{u}_i, \boldsymbol{v}_i$为对应的左右奇异向量, 则投影算子

$$\mathcal{P}_{\mathcal{S}}(X) = \sum_{i=1}^r \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T.$$

- 布尔约束: $\mathcal{S} = \{\boldsymbol{x} | x_i \in \{0, 1\}\}$, 则$\mathcal{P}_{\mathcal{S}}(\boldsymbol{v})$就是把每个分量$v_i$变为0和1 中离它更近的数.

■ **非负矩阵分解和补全:** 已知一个非负矩阵的部分元素,求其非负分解. 它可以看作非负矩阵分解问题和低秩矩阵补全问题的结合, 是一个非常重要的统计学习方法.

- 已知从一个非负、秩为r 的矩阵$M \in \mathbb{R}^{m\times n}$中采样的部分元素$M_{ij}$, $(i,j) \in \Omega \subset \{1,2,\ldots,m\} \times \{1,2,\ldots,n\}$, 目标是要找到非负矩阵$X \in \mathbb{R}^{m\times p}, Y \in \mathbb{R}^{p\times n}$使得$\|M - XY\|_F^2$极小. 则非负矩阵分解和补全问题:

$$\begin{array}{ll} \min\limits_{X,Y} & \|P \odot (XY - M)\|_F^2 \\ \text{s.t.} & X \geq 0, \ \ Y \geq 0. \end{array} \qquad P_{ij} = \begin{cases} 1, & (i,j) \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$$

这个问题是非凸的.

- 为了利用交替方向乘子法的优势, 考虑其等价形式：

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

- 增广拉格朗日函数

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

- 交替方向乘子法

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

求解可得

$$
\begin{aligned}
X^{k+1} &= \left(Z^k (Y^k)^T + \alpha U^k - \Lambda^k\right)\left(Y^k (Y^k)^T + \alpha I\right)^{-1} \\
Y^{k+1} &= \left(X^{k+1} (X^{k+1})^T + \beta I\right)^{-1}\left((X^{k+1})^T Z^k + \beta V^k - \Pi^k\right) \\
Z^{k+1} &= X^{k+1} Y^{k+1} + P \odot (M - X^{k+1} Y^{k+1}) \\
U^{k+1} &= \mathcal{P}_+\left(X^{k+1} + \frac{\Lambda^k}{\alpha}\right) \\
V^{k+1} &= \mathcal{P}_+\left(Y^{k+1} + \frac{\Pi^k}{\beta}\right) \\
\Lambda^{k+1} &= \Lambda^k + \tau\alpha(X^{k+1} - U^{k+1}), \\
\Pi^{k+1} &= \Pi^k + \tau\beta(Y^{k+1} - V^{k+1}).
\end{aligned}
$$

ADMM收敛性分析

- **必要假设:** 问题(1)的解集非空, $f_1(x), f_2(x)$ 均为闭凸函数, 且每个ADMM迭代子问题存在唯一解.

- **记号:** 由(1)的解集非空可设(x_1^*, x_2^*, y^*) 是KKT 对

$$-A_1^{\mathrm{T}} y^* \in \partial f_1(x_1^*), \quad -A_2^{\mathrm{T}} y^* \in \partial f_2(x_2^*), \quad A_1 x_1^* + A_2 x_2^* = b. \tag{27}$$

目的是证明ADMM 迭代序列$\{(x_1^k, x_2^k, y^k)\}$收敛到问题(1)的一个KKT 对, 于是定义

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

和

$$\begin{aligned} u^k &= -A_1{}^{\mathrm{T}}[y^k + (1-\tau)\rho(A_1 e_1^k + A_2 e_2^k) + \rho A_2(x_2^{k-1} - x_2^k)], \\ v^k &= -A_2{}^{\mathrm{T}}[y^k + (1-\tau)\rho(A_1 e_1^k + A_2 e_2^k)], \\ \Psi_k &= \frac{1}{\tau\rho}\|e_y^k\|^2 + \rho\|A_2 e_2^k\|^2, \\ \Phi_k &= \Psi_k + \max(1-\tau, 1-\tau^{-1})\rho\|A_1 e_1^k + A_2 e_2^k\|^2. \end{aligned} \tag{28}$$

Lemma 1. *假设$\{(x_1^k, x_2^k, y^k)\}$为ADMM(5)-(7) 产生一个迭代序列, 则对任意的$k \geq 1$有*

$$u^k \in \partial f_1(x_1^k), \quad v^k \in \partial f_2(x_2^k), \tag{29}$$

和

$$\begin{aligned} \Phi_k - \Phi_{k+1} \geq & \min(\tau, 1+\tau-\tau^2)\rho\|A_2(x_2^k - x_2^{k+1})\|^2 \\ & + \min(1, 1+\tau^{-1}-\tau)\rho\|A_1 e_1^{k+1} + A_2 e_2^{k+1}\|^2. \end{aligned} \tag{30}$$

注：只有当$\tau \in \left(0, \frac{1+\sqrt{5}}{2}\right)$时, (30) 中不等号右侧的项才为非负.

Proof. 先证(29). 对x_1^{k+1}, 由(5)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的定义和$b = A_1x_1^* + A_2x_2^*$, $u^k \in \partial f_1(x_1^k)$. 对$x_2^{k+1}$, 由(6)知

$$0 \in \partial f_2(x_2^{k+1}) + A_2^{\mathrm{T}} y^k + \rho A_2^{\mathrm{T}}(A_1x_1^{k+1} + A_2x_2^{k+1} - b),$$

又由(7)知

$$-A_2^{\mathrm{T}}\Big(y^{k+1} + (1-\tau)\rho(A_1x_1^{k+1} + A_2x_2^{k+1} - b)\Big) \in \partial f_2(x_2^{k+1}).$$

根据v^k的定义和$b = A_1x_1^* + A_2x_2^*$, $v^k \in \partial f_2(x_2^k)$. 故(29)得证.

现证(30). 根据(27)和(29),

$$u^{k+1} \in \partial f_1(x_1^{k+1}), \quad -A_1^{\mathrm{T}} y^* \in \partial f_1(x_1^*),$$
$$v^{k+1} \in \partial f_2(x_2^{k+1}), \quad -A_2^{\mathrm{T}} y^* \in \partial f_2(x_2^*).$$

根据f_1, f_2凸,

$$\left\langle u^{k+1} + A_1^{\mathrm{T}} y^*, x_1^{k+1} - x_1^* \right\rangle \geq 0,$$
$$\left\langle v^{k+1} + A_2^{\mathrm{T}} y^*, x_2^{k+1} - x_2^* \right\rangle \geq 0.$$

再结合u^{k+1}, v^{k+1}的定义, 并注意到恒等式

$$A_1x_1^{k+1} + A_2x_2^{k+1} - b = (\tau\rho)^{-1}(y^{k+1} - y^k) = (\tau\rho)^{-1}(e_y^{k+1} - e_y^k), \quad (31)$$

我们有

$$
\begin{aligned}
&\frac{1}{\tau\rho}\left\langle e_y^{k+1}, e_y^k - e_y^{k+1}\right\rangle - (1-\tau)\rho\|A_1e_1^{k+1} + A_2e_2^{k+1}\|^2 \\
&+ \rho\left\langle A_2(x_2^{k+1} - x_2^k), A_1e_1^{k+1} + A_2e_2^{k+1}\right\rangle \\
&- \rho\left\langle A_2(x_2^{k+1} - x_2^k), A_2e_2^{k+1}\right\rangle \geq 0.
\end{aligned} \tag{32}
$$

注意到不等式(32)和(30) 主要的差别在下面这一项上:

$$
\rho\left\langle A_2(x_2^{k+1} - x_2^k), A_1e_1^{k+1} + A_2e_2^{k+1}\right\rangle.
$$

接下来, 我们估计这一项的上界. 引入新符号

$$
\nu^{k+1} := y^{k+1} + (1-\tau)\rho(A_1e_1^{k+1} + A_2e_2^{k+1}),
$$

$$
M^{k+1} := (1-\tau)\rho\left\langle A_2(x_2^{k+1} - x_2^k), A_1e_1^k + A_2e_2^k\right\rangle,
$$

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

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

从而有

$$
\begin{aligned}
&\rho\left\langle A_2(x_2^{k+1}-x_2^k), A_1e_1^{k+1}+A_2e_2^{k+1}\right\rangle \\
=\ &(1-\tau)\rho\left\langle A_2(x_2^{k+1}-x_2^k), A_1e_1^{k+1}+A_2e_2^{k+1}\right\rangle \\
&+\left\langle A_2(x_2^{k+1}-x_2^k), y^{k+1}-y^k\right\rangle \\
=\ &M^{k+1}+\left\langle \nu^{k+1}-\nu^k, A_2(x_2^{k+1}-x_2^k)\right\rangle \le M^{k+1}.
\end{aligned}
$$

因此, 不等式(32)可以放缩成

$$
\begin{aligned}
&\frac{1}{\tau\rho}\left\langle e_y^{k+1}, e_y^k-e_y^{k+1}\right\rangle-(1-\tau)\rho\|A_1e_1^{k+1}+A_2e_2^{k+1}\|^2 \\
&+M^{k+1}-\rho\left\langle A_2(x_2^{k+1}-x_2^k), A_2e_2^{k+1}\right\rangle \ge 0.
\end{aligned}
$$

利用恒等式

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

进一步得到

$$\begin{aligned}&\frac{1}{\tau\rho}(\|e_y^k\|^2-\|e_y^{k+1}\|^2)-(2-\tau)\rho\|A_1e_1^{k+1}+A_2e_2^{k+1}\|^2\\&+2M^{k+1}-\rho\|A_2(x_2^{k+1}-x_2^k)\|^2-\rho\|A_2e_2^{k+1}\|^2+\rho\|A_2e_2^k\|^2\ge 0.\end{aligned}\tag{33}$$

除了M^{k+1}中的项, (33)中的其他项均在不等式(30) 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$(1-\tau)\rho\ge 0$, 根据基本不等式, 可得

$$\begin{aligned}2M^{k+1}=&2(1-\tau)\rho\left\langle A_2(x_2^{k+1}-x_2^k),A_1e_1^k+A_2e_2^k\right\rangle\\\le&(1-\tau)\rho\left(\|A_2(x_2^{k+1}-x_2^k)\|^2+\|A_1e_1^k+A_2e_2^k\|^2\right).\end{aligned}$$

代入(33)得

$$\begin{aligned}&\frac{1}{\tau\rho}\|e_y^k\|^2+\rho\|A_2e_2^k\|^2+(1-\tau)\rho\|A_1e_1^k+A_2e_2^k\|^2\\&-\left[\frac{1}{\tau\rho}\|e_y^{k+1}\|^2+\rho\|A_2e_2^{k+1}\|^2+(1-\tau)\rho\|A_1e_1^{k+1}+A_2e_2^{k+1}\|^2\right]\\\ge&\rho\|A_1e_1^{k+1}+A_2e_2^{k+1}\|^2+\tau\rho\|A_2(x_2^{k+1}-x_2^k)\|^2.\end{aligned}\tag{34}$$

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

$$
\begin{aligned}
&-2\left\langle A_2(x_2^{k+1}-x_2^k), A_1e_1^k + A_2e_2^k \right\rangle \\
\le&\tau\|A_2(x_2^{k+1}-x_2^k)\|^2 + \frac{1}{\tau}\|A_1e_1^k + A_2e_2^k\|^2,
\end{aligned}
$$

可得

$$
-2M^{k+1} \ge (1-\tau)\rho\left(\tau\|A_2(x_2^{k+1}-x_2^k)\|^2 + \frac{1}{\tau}\|A_1e_1^k + A_2e_2^k\|^2\right).
$$

同样代入不等式(33)可以得到

$$
\begin{aligned}
&\frac{1}{\tau\rho}\|e_y^k\|^2 + \rho\|A_2e_2^k\|^2 + \left(1-\frac{1}{\tau}\right)\rho\|A_1e_1^k + A_2e_2^k\|^2 \\
&-\left[\frac{1}{\tau\rho}\|e_y^{k+1}\|^2 + \rho\|A_2e_2^{k+1}\|^2 + \left(1-\frac{1}{\tau}\right)\rho\|A_1e_1^{k+1} + A_2e_2^{k+1}\|^2\right] \\
\ge&\left(1+\frac{1}{\tau}-\tau\right)\rho\|A_1e_1^{k+1} + A_2e_2^{k+1}\|^2 \\
&+(1+\tau-\tau^2)\rho\|A_2(x_2^{k+1}-x_2^k)\|^2.
\end{aligned} \tag{35}
$$

□

Theorem 2. *设必要假设成立且*A_1, A_2*列满秩．如果*$\tau \in \left(0, \dfrac{1+\sqrt{5}}{2}\right)$*, 则序列*$\{(x_1^k, x_2^k, y^k)\}$*收敛到问题(1)的一个KKT对, 也就是最优解.*

Proof. 引理1表明$\{\Phi_k\}$是有界序列, 故由Φ_k的定义(28)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．根据不等式

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

可推出$\{\|A_1 e_1^k\|\}$也是有界序列．注意到$A_1^{\mathrm{T}} A_1 \succ 0, A_2^{\mathrm{T}} A_2 \succ 0$，因此$\{(x_1^k, x_2^k, y^k)\}$是有界序列. 而且,

$$\sum_{k=0}^{\infty} \|A_1 e_1^k + A_2 e_2^k\|^2, \quad \sum_{k=0}^{\infty} \|A_2(x_2^{k+1} - x_2^k)\|^2$$

都是收敛的, 这表明

$$\|A_1 e_1^k + A_2 e_2^k\| = \|A_1 x_1^k + A_2 x_2^k - b\| \to 0, \quad \|A_2(x_2^{k+1} - x_2^k)\| \to 0. \tag{36}$$

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(28)及(36)可得$\{u^k\}$与$\{v^k\}$相应的子列也收敛. 记

$$u^\infty \overset{\text{def}}{=\!=} \lim_{j\to\infty} u^{k_j} = -A_1^{\mathrm{T}} y^\infty, \quad v^\infty = \lim_{j\to\infty} v^{k_j} = -A_2^{\mathrm{T}} y^\infty. \tag{37}$$

由(29)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

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

由(36)可知

$$\lim_{j\to\infty} \|A_1 x_1^{k_j} + A_2 x_2^{k_j} - b\| = \|A_1 x_1^\infty + A_2 x_2^\infty - b\| = 0.$$

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

$$\begin{aligned}
&\lim_{j\to\infty} \Phi_{k_j} \\
&= \lim_{j\to\infty} \left(\frac{1}{\tau\rho}\|e_y^{k_j}\|^2 + \rho\|A_2 e_2^{k_j}\|^2 + \max\left\{1-\tau, 1-\frac{1}{\tau}\right\}\rho\|A_1 e_1^{k_j} + A_2 e_2^{k_j}\|^2 \right) \\
&= 0.
\end{aligned}$$

这说明

$$\|e_y^k\| \to 0, \quad \|A_2 e_2^k\| \to 0, \quad \|A_1 e_1^k + A_2 e_2^k\| \to 0.$$

进一步有

$$0 \le \limsup_{k\to\infty} \|A_1 e_1^k\| \le \lim_{k\to\infty} \left(\|A_2 e_2^k\| + \|A_1 e_1^k + A_2 e_2^k\| \right) = 0.$$

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. $\square$