

# 交替方向乘子法(ADMM)

文再文

北京大学北京国际数学研究中心

教材《最优化：建模、算法与理论》配套电子教案

<http://bicmr.pku.edu.cn/~wenzw/optbook.html>

致谢：本教案由**华奕轩**协助准备

- 1 交替方向乘子法
- 2 常见变形和技巧
- 3 应用举例
- 4 Douglas-Rachford splitting 算法
- 5 ADMM的收敛性分析

# 典型问题形式

考虑如下凸问题：

$$\begin{aligned} \min_{x_1, x_2} \quad & f_1(x_1) + f_2(x_2), \\ \text{s.t.} \quad & A_1 x_1 + A_2 x_2 = b, \end{aligned} \tag{1}$$

- $f_1, f_2$  是适当的闭凸函数，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。

## 问题形式举例

- 可以分成两块的非约束优化问题

$$\min_x f_1(x) + f_2(x).$$

引入一个新的变量 $z$ 并令 $x = z$ , 将问题转化为

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

- 带线性变换的非约束优化问题

$$\min_x f_1(x) + f_2(Ax).$$

可以引入一个新的变量 $z$ , 令 $z = Ax$ , 则问题变为

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

## 问题形式举例

- 凸集  $C \subset \mathbb{R}^n$  上的约束优化问题

$$\min_x f(x),$$

$$\text{s.t. } Ax \in C,$$

$I_C(z)$  是集合  $C$  的示性函数, 引入约束  $z = Ax$ , 那么问题转化为

$$\min_{x,z} f(x) + I_C(z),$$

$$\text{s.t. } Ax - z = 0.$$

- 全局一致性问题

$$\min_x \sum_{i=1}^N \phi_i(x).$$

令  $x = z$ , 并将  $x$  复制  $N$  份, 分别为  $x_i$ , 那么问题转化为

$$\min_{x_i, z} \sum_{i=1}^N \phi_i(x_i),$$

$$\text{s.t. } x_i - z = 0, \quad i = 1, 2, \dots, N.$$

# 增广拉格朗日函数法

- 首先写出问题(1)的增广拉格朗日函数

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

其中  $\rho > 0$  是二次罚项的系数.

- 常见的求解带约束问题的增广拉格朗日函数法为如下更新:

$$(x_1^{k+1}, x_2^{k+1}) = \underset{x_1, x_2}{\operatorname{argmin}} L_{\rho}(x_1, x_2, y^k), \quad (3)$$

$$y^{k+1} = y^k + \tau \rho (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \quad (4)$$

其中  $\tau$  为步长.

# 交替方向乘子法

## Alternating direction method of multipliers, ADMM

- 交替方向乘子法的基本思路: 第一步迭代(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
- 其迭代格式可以总结如下:

$$x_1^{k+1} = \underset{x_1}{\operatorname{argmin}} L_\rho(x_1, x_2^k, y^k), \quad (5)$$

$$x_2^{k+1} = \underset{x_2}{\operatorname{argmin}} L_\rho(x_1^{k+1}, x_2, y^k), \quad (6)$$

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

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

## 原问题最优性条件

- 因为 $f_1, f_2$ 均为闭凸函数，约束为线性约束，所以当Slater条件成立时，可以使用凸优化问题的KKT条件来作为交替方向乘子法的收敛准则。问题(1)的拉格朗日函数为

$$L(x_1, x_2, y) = f_1(x_1) + f_2(x_2) + y^T(A_1x_1 + A_2x_2 - b).$$

- 根据最优性条件定理，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：

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

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

$$A_1 x_1^* + A_2 x_2^* = b. \quad (8c)$$

在这里条件(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 $x_2$ 的更新步骤

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

根据最优性条件不难推出

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

当 $\tau = 1$ 时, 根据(7)可知上式方括号中的表达式就是 $y^k$ , 最终有

$$0 \in \partial f_2(x_2^k) + A_2^T y^k,$$

- 由 $x_1$ 的更新公式

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

假设子问题能精确求解, 根据最优性条件

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

# ADMM单步迭代最优性条件

- 根据ADMM的第三式(7)取 $\tau = 1$ 有

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

对比条件(8a)可知多出来的项为 $A_1^T A_2(x_2^{k-1} - x_2^k)$ 。因此要检测对偶可行性只需要检测残差

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

- 综上当 $x_2$ 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：

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

# 提纲

- 1 交替方向乘子法
- 2 常见变形和技巧**
- 3 应用举例
- 4 Douglas-Rachford splitting 算法
- 5 ADMM的收敛性分析

# 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.
- 不失一般性, 我们考虑第一个子问题, 即

$$\min_{x_1} f_1(x_1) + \frac{\rho}{2} \|A_1 x_1 - v^k\|^2, \quad (12)$$

其中  $v^k = b - A_2 x_2^k - \frac{1}{\rho} y^k$ .

- 当子问题目标函数可微时, 线性化将问题(12)变为

$$x_1^{k+1} = \operatorname{argmin}_{x_1} \left\{ (\nabla f_1(x_1^k) + \rho A_1^T (A_1 x_1^k - v^k))^T x_1 + \frac{1}{2\eta_k} \|x_1 - x^k\|_2^2 \right\},$$

其中  $\eta_k$  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, 即

$$x_1^{k+1} = \operatorname{argmin}_{x_1} \left\{ f(x_1) + \rho (A_1^T (A_1 x_1^k - v^k))^T x_1 + \frac{1}{2\eta_k} \|x_1 - x^k\|_2^2 \right\},$$

这等价于做一步近似点梯度步.

## 缓存分解

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

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

- 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  $C^T C + \rho A_1^T A_1$  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  $\rho$  发生更新时，就要重新进行分解。特别地，当  $C^T C + \rho A_1^T A_1$  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。

## 优化转移

- 有时候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 $A_1^T A_1$ ，此时子问题(12)替换为

$$\begin{aligned} x_1^{k+1} = \operatorname{argmin}_{x_1} & \left\{ f_1(x_1) + \frac{\rho}{2} \|A_1 x_1 - v^k\|_2^2 \right. \\ & \left. + \frac{\rho}{2} (x_1 - x^k)^T (D - A_1^T A_1) (x_1 - x^k) \right\}. \end{aligned}$$

这种方法也称为优化转移.

- 通过选取合适的 $D$ ，当计算 $\operatorname{argmin}_{x_1} \left\{ f_1(x_1) + \frac{\rho}{2} x_1^T D x_1 \right\}$ 明显比计算 $\operatorname{argmin}_{x_1} \left\{ f_1(x_1) + \frac{\rho}{2} x_1^T A_1^T A_1 x_1 \right\}$ 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 $D = \frac{\eta_k}{\rho} I$ 时，优化转移等价于做单步的近似点梯度步.

## 二次罚项系数的动态调节

- 原始可行性和对偶可行性分别用 $\|r^k\|$ 和 $\|s^k\|$ 度量.
- 求解过程中二次罚项系数 $\rho$ 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, 则会有相反的效果. 这样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 $\rho$ 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. 一个简单有效的方式是令

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

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. 如果发现 $\|r^k\|$ 或 $\|s^k\|$ 下降过慢就应该相应增大或减小二次罚项系数 $\rho^k$ .

- 在(6)式与(7)式中,  $A_1x_1^{k+1}$ 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.



# 多块问题的ADMM

- 考虑有多块变量的情形

$$\begin{aligned} \min_{x_1, x_2, \dots, x_N} \quad & f_1(x_1) + f_2(x_2) + \dots + f_N(x_N), \\ \text{s.t.} \quad & A_1 x_1 + A_2 x_2 + \dots + A_N x_N = b. \end{aligned} \tag{13}$$

这里 $f_i(x_i)$ 是闭凸函数,  $x_i \in \mathbb{R}^{n_i}$ ,  $A_i \in \mathbb{R}^{m \times n_i}$ .

- 同样写出增广拉格朗日函数 $L_\rho(x_1, x_2, \dots, x_N, y)$ , 相应的多块ADMM迭代格式为

$$\begin{aligned} x_1^{k+1} &= \operatorname{argmin}_x L_\rho(x, x_2^k, \dots, x_N^k, y^k), \\ x_2^{k+1} &= \operatorname{argmin}_x L_\rho(x_1^{k+1}, x, \dots, x_N^k, y^k), \\ &\dots\dots\dots \\ x_N^{k+1} &= \operatorname{argmin}_x L_\rho(x_1^{k+1}, x_2^{k+1}, \dots, x, y^k), \\ y^{k+1} &= y^k + \tau \rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} + \dots + A_N x_N^{k+1} - b), \end{aligned}$$

其中 $\tau \in (0, (\sqrt{5} + 1)/2)$ 为步长参数.

# 提纲

- 1 交替方向乘子法
- 2 常见变形和技巧
- 3 应用举例
- 4 Douglas-Rachford splitting 算法
- 5 ADMM的收敛性分析

# LASSO 问题的Primal 形式

- LASSO 问题

$$\min \quad \mu \|x\|_1 + \frac{1}{2} \|Ax - b\|^2.$$

转换为标准问题形式：

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

- 交替方向乘子法迭代格式为

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

# LASSO 问题的Primal 形式

- 注意，因为  $\rho > 0$ ，所以  $A^T A + \rho I$  总是可逆的。  $x$  迭代本质上是计算一个岭回归问题（ $\ell_2$  范数平方正则化的最小二乘问题）；而对  $z$  的更新为  $\ell_1$  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  $\rho$ ，我们可以缓存矩阵  $A^T A + \rho I$  的初始分解，从而减小后续迭代中的计算量。
- 需要注意的是，在 LASSO 问题中，矩阵  $A \in \mathbb{R}^{m \times n}$  通常有较多的列（即  $m \ll n$ ），因此  $A^T A \in \mathbb{R}^{n \times n}$  是一个低秩矩阵，二次罚项的作用就是将  $A^T A$  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 公式则可进一步降低每次迭代的运算量）

# LASSO 问题的对偶形式

- 考虑LASSO 问题的对偶问题

$$\begin{aligned} \min \quad & b^T y + \frac{1}{2} \|y\|^2, \\ \text{s.t.} \quad & \|A^T y\|_\infty \leq \mu. \end{aligned} \quad (14)$$

- 引入约束  $A^T y + z = 0$ , 可以得到如下等价问题:

$$\begin{aligned} \min \quad & \underbrace{b^T y + \frac{1}{2} \|y\|^2}_{f(y)} + \underbrace{I_{\|z\|_\infty \leq \mu}(z)}_{h(z)}, \\ \text{s.t.} \quad & A^T y + z = 0. \end{aligned} \quad (15)$$

- 对约束  $A^T y + z = 0$  引入乘子  $x$ , 对偶问题的增广拉格朗日函数为

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

# LASSO 问题的对偶形式

- 当固定  $y, x$  时, 对  $z$  的更新即向无穷范数球  $\{z \mid \|z\|_\infty \leq \mu\}$  做欧几里得投影, 即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

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

- 因此得到 ADMM 迭代格式为

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

- 虽然 ADMM 应用于对偶问题也需要求解一个线性方程组, 但由于 LASSO 问题的特殊性 ( $m \ll n$ ), 求解  $y$  更新的线性方程组需要的计算量是  $O(m^3)$ , 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.

# 广义LASSO 问题

- 对许多问题 $x$ 本身不稀疏，但在某种变换下是稀疏的：

$$\min_x \mu \|Fx\|_1 + \frac{1}{2} \|Ax - b\|^2. \quad (16)$$

- 一个重要的例子是当 $F \in \mathbb{R}^{(n-1) \times n}$  是一阶差分矩阵

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

且 $A = I$ 时，广义LASSO问题为

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

这个问题就是图像去噪问题的TV模型；当 $A = I$ 且 $F$ 是二阶差分矩阵时，问题(16)被称为一范数趋势滤波。

# 广义LASSO 问题

- 通过引入约束  $Fx = z$ :

$$\begin{aligned} \min_{x,z} \quad & \frac{1}{2} \|Ax - b\|^2 + \mu \|z\|_1, \\ \text{s.t.} \quad & Fx - z = 0, \end{aligned} \tag{17}$$

- 引入乘子  $y$ , 其增广拉格朗日函数为

$$L_\rho(x, z, y) = \frac{1}{2} \|Ax - b\|^2 + \mu \|z\|_1 + y^T (Fx - z) + \frac{\rho}{2} \|Fx - z\|^2.$$

- 此问题的  $x$  迭代是求解方程组

$$(A^T A + \rho F^T F)x = A^T b + \rho F^T \left( z^k - \frac{y^k}{\rho} \right),$$

而  $z$  迭代依然通过  $\ell_1$  范数的邻近算子.



- 因此交替方向乘子法所产生的迭代为

$$x^{k+1} = (A^T A + \rho F^T F)^{-1} \left( A^T b + \rho F^T \left( z^k - \frac{y^k}{\rho} \right) \right),$$

$$z^{k+1} = \text{prox}_{(\mu/\rho)\|\cdot\|_1} \left( Fx^{k+1} + \frac{y^k}{\rho} \right),$$

$$y^{k+1} = y^k + \tau \rho (Fx^{k+1} - z^{k+1}).$$

- 对于全变差去噪问题,  $A^T A + \rho F^T F$  是三对角矩阵, 所以此时  $x$  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  $\mathcal{O}(n \log n)$ ; 对于一范数趋势滤波问题,  $A^T A + \rho F^T F$  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

# 半定规划问题

考虑

$$\begin{aligned} \min_{X \in S^n} \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \langle A^{(i)}, X \rangle = b_i, \quad i = 1, \dots, m, \\ & X \succeq 0 \end{aligned}$$

对偶问题为

$$(D) \quad \begin{cases} \min_{y \in \mathbb{R}^m, S \in S^n} & -b^\top y \\ \text{s.t.} & \mathcal{A}^*(y) + S = C, \quad S \succeq 0, \end{cases}$$

增广拉格朗日函数为

$$\mathcal{L}_\mu(X, y, S) = -b^\top y + \langle X, \mathcal{A}^*(y) + S - C \rangle + \frac{1}{2\mu} \|\mathcal{A}^*(y) + S - C\|_F^2.$$

# ADMM求解半定规划问题

ADMM迭代格式为

$$\begin{aligned}y^{k+1} &:= \arg \min_{y \in \mathbb{R}^m} \mathcal{L}_\mu(X^k, y, S^k), \\&= -(\mathcal{A}\mathcal{A}^*)^{-1} (\mu(\mathcal{A}(X^k) - b) + \mathcal{A}(S^k - C)) \\S^{k+1} &:= \arg \min_{S \in S^n} \mathcal{L}_\mu(X^k, y^{k+1}, S), \quad S \succeq 0, \\X^{k+1} &:= X^k + \frac{\mathcal{A}^*(y^{k+1}) + S^{k+1} - C}{\mu}.\end{aligned}$$

● 关于 $S$ 的子问题:

$$\min_{S \in S^n} \|S - V^{k+1}\|_F^2, \quad S \succeq 0,$$

其中 $V^{k+1} := V(S^k, X^k) = C - \mathcal{A}^*(y(S^k, X^k)) - \mu X^k$ .

显示解为

$$S^{k+1} := V_{\dagger}^{k+1} := Q_{\dagger} \Sigma_{+} Q_{\dagger}^{\top}$$

$$\text{其中 } V^{k+1} = Q \Sigma Q^{\top} = \begin{pmatrix} Q_{\dagger} & Q_{\ddagger} \end{pmatrix} \begin{pmatrix} \Sigma_{+} & 0 \\ 0 & \Sigma_{-} \end{pmatrix} \begin{pmatrix} Q_{\dagger}^{\top} \\ Q_{\ddagger}^{\top} \end{pmatrix}$$

# ADMM求解半定规划问题

更新拉格朗日乘子 $X^{k+1}$

- 更新格式:

$$X^{k+1} := X^k + \frac{\mathcal{A}^*(y^{k+1}) + S^{k+1} - C}{\mu}$$

- 等价形式:

$$X^{k+1} = \frac{1}{\mu}(S^{k+1} - V^{k+1}) = \frac{1}{\mu}V_{\ddagger}^{k+1},$$

其中 $V_{\ddagger}^{k+1} := -Q_{\ddagger}\Sigma - Q_{\ddagger}$ .

- 注意到 $X^{k+1}$ 也是如下优化问题的最优解

$$\min_{X \in S^n} \quad \|\mu X + V^{k+1}\|_F^2, \quad X \succeq 0.$$

# 稀疏逆协方差矩阵估计

- 该问题的基本形式是

$$\min_X \langle S, X \rangle - \ln \det X + \mu \|X\|_1, \quad (18)$$

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。

- 目标函数由光滑项和非光滑项组成，因此引入约束 $X = Z$ 将问题的两部分分离：

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

引入乘子 $U$ 作用在约束 $X - Z = 0$ 上，可得增广拉格朗日函数为

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

# 稀疏逆协方差矩阵估计

- 首先, 固定 $Z^k, U^k$ , 则 $X$ 子问题是凸光滑问题, 对 $X$ 求矩阵导数并令其为零,

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

这是一个关于 $X$ 的矩阵方程, 可以求出满足上述矩阵方程的唯一正定的 $X$ 为

$$X^{k+1} = Q \text{Diag}(x_1, x_2, \dots, x_n) Q^T,$$

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.

- 固定 $X^{k+1}, U^k$ , 则 $Z$ 的更新为矩阵 $\ell_1$  范数的邻近算子.
- 最后是常规的乘子更新.

# 矩阵分离问题

- 考虑矩阵分离问题：

$$\begin{aligned} \min_{X, S} \quad & \|X\|_* + \mu \|S\|_1, \\ \text{s.t.} \quad & X + S = M, \end{aligned} \tag{19}$$

其中  $\|\cdot\|_1$  与  $\|\cdot\|_*$  分别表示矩阵  $\ell_1$  范数与核范数.

- 引入乘子  $Y$  作用在约束  $X + S = M$  上，我们可以得到此问题的增广拉格朗日函数

$$L_\rho(X, S, Y) = \|X\|_* + \mu \|S\|_1 + \langle Y, X + S - M \rangle + \frac{\rho}{2} \|X + S - M\|_F^2. \tag{20}$$

# 矩阵分离问题

- 对于 $X$ 子问题,

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

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 $U \operatorname{Diag}(\sigma(A)) V^T$ 为 $A$ 的约化奇异值分解.



# 矩阵分离问题

- 对于  $S$  子问题,

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

- 那么交替方向乘子法的迭代格式为

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

# 图像去噪模型

$$b = Kx_t + w$$

- $x_t$  为未知图像
- $b$  为观察到的图像，模糊且有噪声； $w$  为噪声
- $N \times N$  的像素点按列储存为长为  $N^2$  的向量

## 模糊矩阵 $K$

- 表示一个2维的卷积，是有空间不动点的扩散函数
- 满足周期边界条件，有循环块(circulant blocks)
- 可对角化，即存在酉的2维离散傅立叶变换矩阵  $W$ ，使得

$$K = W^H \mathbf{diag}(\lambda) W.$$

系数矩阵为  $I + K^T K$  的线性方程组可在  $O(N^2 \log N)$  的时间内求解。

# 全变差去模糊方法

$$\begin{aligned} \min \quad & \|Kx - b\|_1 + \gamma \|Dx\|_{tv} \\ \text{s.t.} \quad & 0 \leq x \leq 1 \end{aligned}$$

目标函数的第二项称为全变差罚函数

- $Dx$  是离散化的水平和垂直的方向导数

$$\begin{pmatrix} I \otimes D_1 \\ D_1 \otimes I \end{pmatrix}, \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 & 0 & 1 \\ 1 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix}$$

- $\|\cdot\|_{tv}$  是欧式距离的和:  $\|(u, v)\|_{tv} = \sum_{i=1}^n \sqrt{u_i^2 + v_i^2}$

# ADMM求解图像去模糊问题

考虑对原问题进行拆分:

$$\min \|u\|_1 + \gamma \|v\|_{tv}, \quad \text{s.t. } u = Kx - b, \quad v = Dx, \quad y = x, \quad 0 \leq y \leq 1$$

**ADMM** 算法要求:

- 将 $\|u\|_1$  和 $\|v\|_{tv}$ 的proximal算子以及在集合 $C$ 上的投影这三部分分离
- 在 $O(N^2 \log N)$ 的时间内求解系数矩阵为 $I + K^T K + D^T D$ 的线性方程组

# 图像去模糊的实例

- $1024 \times 1024$  的图像，满足周期边界条件
- 高斯模糊
- 椒盐噪声 (salt-and-pepper noise)：50% 的像素点被随机替换为0/1



original



noisy/blurred



restored

# 全局一致性优化问题

$$\begin{aligned} \min_{x_i, z} \quad & \sum_{i=1}^N \phi_i(x_i), \\ \text{s.t.} \quad & x_i - z = 0, \quad i = 1, 2, \dots, N. \end{aligned}$$

- 增广拉格朗日函数为

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

- 固定 $z^k, y_i^k$ , 更新 $x_i$ 的公式为

$$x_i^{k+1} = \operatorname{argmin}_x \left\{ \phi_i(x) + \frac{\rho}{2} \|x - z^k + y_i^k / \rho\|^2 \right\}. \quad (21)$$

- 注意：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 $x_i$ 可以看成是一个整体. 相应地, 所有乘子 $y_i$ 也可以看成是一个整体.

# 全局一致性优化问题

- 迭代式(21)的具体计算依赖于 $\phi_i$ 的形式，在一般情况下更新 $x_i$ 的表达式为

$$x_i^{k+1} = \text{prox}_{\phi_i/\rho} (z^k - y_i^k/\rho).$$

- 固定 $x_i^{k+1}, y_i^k$ ，问题关于 $z$ 是二次函数，因此可以直接写出显式解：

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

- 综上，该问题的交替方向乘子法迭代格式为

$$x_i^{k+1} = \text{prox}_{\phi_i/\rho} (z^k - y_i^k/\rho), \quad i = 1, 2, \dots, N,$$

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

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

# 分布式ADMM I

形式一：考虑 $f$ 是不可分(inseparable)函数， $g$ 是可分(separable)函数

$$\min_{\mathbf{x}, \mathbf{z}} \sum_{l=1}^L (f_l(\mathbf{x}) + g_l(\mathbf{z}_l)), \text{ s.t. } \mathbf{A}\mathbf{x} + \mathbf{z} = \mathbf{b}$$

- 将 $\mathbf{x}$ 复制 $L$ 份:  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L$
- 拆分矩阵和向量

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_L \end{bmatrix}, \mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_L \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_L \end{bmatrix}$$

将 $\mathbf{A}\mathbf{x} + \mathbf{z} = \mathbf{0}$ 转化为

$$\mathbf{A}_l \mathbf{x}_l + \mathbf{z}_l = \mathbf{b}_l, \mathbf{x}_l - \mathbf{x} = \mathbf{0}, l = 1, \dots, L.$$



# 分布式ADMM I

模型一：

$$\begin{aligned} \min_{\mathbf{x}, \{\mathbf{x}_l\}, \mathbf{z}} \quad & \sum_{l=1}^L (f_l(\mathbf{x}_l) + g_l(\mathbf{z}_l)) \\ \text{s.t.} \quad & \mathbf{A}_l \mathbf{x}_l + \mathbf{z}_l = \mathbf{b}_l, \mathbf{x}_l - \mathbf{x} = \mathbf{0}, l = 1, \dots, L. \end{aligned}$$

- $\mathbf{x}_l$  是  $\mathbf{x}$  的备份
- $\mathbf{z}_l$  是  $\mathbf{z}$  的一部分
- 将  $\{\mathbf{x}_l\}, \mathbf{z}, \mathbf{x}$  分为两块
  - $\{\mathbf{x}_l\}$ : 给定  $\mathbf{z}$  和  $\mathbf{x}$ ,  $\mathbf{x}_l$  的更新是可分的
  - $(\mathbf{z}, \mathbf{x})$ : 给定  $\{\mathbf{x}_l\}$ ,  $\mathbf{z}_l$  和  $\mathbf{x}$  的更新也是可分的  
因此可以使用标准的2块的ADMM算法
- 还可以在目标函数中加入简单的正则项  $h(\mathbf{x})$

# 分布式ADMM I

考虑用MPI建立 $L$ 个计算节点。

- $\mathbf{A}_l$  是仅储存在节点 $l$  上的局部数据
- $\mathbf{x}_l, \mathbf{z}_l$  是局部变量,  $\mathbf{x}_l$  仅在节点 $l$ 上 储存和更新
- $\mathbf{x}$  是全局变量, 由MPI指派和计算
- $\mathbf{y}_l, \bar{\mathbf{y}}_l$  分别是 $\mathbf{A}_l \mathbf{x}_l + \mathbf{z}_l = \mathbf{b}_l$  和  $\mathbf{x}_l - \mathbf{x} = \mathbf{0}$  对应的拉格朗日乘子, 仅在节点 $l$ 上 储存和更新

每次迭代中,

- 每个节点 $l$  各自使用数据 $\mathbf{A}_l$  计算 $\mathbf{x}_l^{k+1}$
- 每个节点 $l$  各自计算 $\mathbf{z}_l^{k+1}$ , 并准备 $\mathbf{P}_l = (\dots)$
- MPI 收集 $\mathbf{P}_l$  并将他们的平均值 $\mathbf{x}^{k+1}$  分发到各节点上
- 每个节点 $l$  各自计算乘子 $\mathbf{y}_l^{k+1}, \bar{\mathbf{y}}_l^{k+1}$

## 分布式ADMM II

形式二：考虑 $f$ 和 $g$ 均是可分的函数

$$\min \sum_{j=1}^N f_j(\mathbf{x}_j) + \sum_{i=1}^M g_i(\mathbf{z}_i), \text{ s.t. } \mathbf{A}\mathbf{x} + \mathbf{z} = \mathbf{b},$$

其中

$$\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N), \mathbf{z} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M).$$

将 $\mathbf{A}$ 分解为

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1N} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2N} \\ & & \cdots & \\ \mathbf{A}_{M1} & \mathbf{A}_{M2} & \cdots & \mathbf{A}_{MN} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_M \end{bmatrix}.$$

我们可以得到类似的模型

$$\min \sum_{j=1}^N f_j(\mathbf{x}_j) + \sum_{i=1}^M g_i(\mathbf{z}_i), \text{ s.t. } \sum_{j=1}^N \mathbf{A}_{ij}\mathbf{x}_j + \mathbf{z}_i = \mathbf{b}_i, i = 1, \dots, M.$$

# 分布式ADMM II

注意到 $\mathbf{A}_{ij}\mathbf{x}_j'$ 在约束中是耦合的，因此令

$$\mathbf{p}_{ij} = \mathbf{A}_{ij}\mathbf{x}_j,$$

模型二：

$$\min \sum_{j=1}^N f_j(\mathbf{x}_j) + \sum_{i=1}^M g_i(\mathbf{z}_i), \text{ s.t. } \begin{cases} \sum_{j=1}^N \mathbf{p}_{ij} + \mathbf{z}_i = \mathbf{b}_i, \forall i, \\ \mathbf{p}_{ij} - \mathbf{A}_{ij}\mathbf{x}_j = 0, \forall i, j. \end{cases}$$

ADMM迭代

- 交替更新 $\{\mathbf{p}_{ij}\}$  和  $(\{\mathbf{x}_j\}, \{\mathbf{z}_i\})$
- 关于 $\mathbf{p}_{ij}$  的子问题有闭形式解
- 关于 $(\{\mathbf{x}_j\}, \{\mathbf{z}_i\})$ 的子问题对 $\mathbf{x}_j$  和 $\mathbf{z}_i$ 是可分的
  - $\mathbf{x}_j$ 的更新需要 $f_j$  和  $\mathbf{A}_{1j}^T\mathbf{A}_{1j}, \dots, \mathbf{A}_{Mj}^T\mathbf{A}_{Mj}$ ;
  - $\mathbf{z}_i$ 的更新需要 $g_i$ .

思考：如何进一步将 $f_j$ 和 $\mathbf{A}_{1j}^T\mathbf{A}_{1j}, \dots, \mathbf{A}_{Mj}^T\mathbf{A}_{Mj}$  去耦合？

# 分布式ADMM III

对每个 $\mathbf{x}_j$ ，作 $M$ 个独立的备份 $\mathbf{x}_{1j}, \mathbf{x}_{2j}, \dots, \mathbf{x}_{Mj}$ .

模型三：

$$\min \sum_{j=1}^N f_j(\mathbf{x}_j) + \sum_{i=1}^M g_i(\mathbf{z}_i), \quad \text{s.t.} \quad \begin{aligned} \sum_{j=1}^N \mathbf{p}_{ij} + \mathbf{z}_i &= \mathbf{b}_i, & \forall i, \\ \mathbf{p}_{ij} - \mathbf{A}_{ij}\mathbf{x}_{ij} &= \mathbf{0}, & \forall i, j, \\ \mathbf{x}_j - \mathbf{x}_{ij} &= \mathbf{0}, & \forall i, j. \end{aligned}$$

## ADMM

- 交替更新 $(\{\mathbf{x}_j\}, \{\mathbf{p}_{ij}\})$  和  $(\{\mathbf{x}_j\}, \{\mathbf{z}_i\})$
- 关于 $(\{\mathbf{x}_j\}, \{\mathbf{p}_{ij}\})$ 的子问题是可分的
  - $\mathbf{x}_j$ 的更新只需要 $f_j$ ，即只需计算 $\text{prox}_{f_j}$
  - $\mathbf{p}_{ij}$ 的更新有闭形式解
- 关于 $(\{\mathbf{x}_{ij}\}, \{\mathbf{z}_i\})$ 的子问题也是可分的
  - $\mathbf{x}_{ij}$ 的更新需要 $(\alpha I + \beta \mathbf{A}_{ij}^T \mathbf{A}_{ij})$ ;
  - $\mathbf{y}_i$ 的更新只需要 $g_i$ ，即只需计算 $\text{prox}_{g_i}$ .

# 去中心化ADMM

对 $\mathbf{x}$ 进行局部备份得到 $\mathbf{x}_i$ ，不使用约束

$$\mathbf{x}_i - \mathbf{x} = \mathbf{0}, i = 1, \dots, M,$$

而是考虑图 $\mathcal{G} = (\mathcal{V}, \varepsilon)$  其中 $\mathcal{V}$ 为顶点集， $\varepsilon$ 为边集。



根据图 $\mathcal{G}$ 建立约束

$$\begin{aligned} & \mathbf{x}_i - \mathbf{x}_j = \mathbf{0}, \quad \forall (i, j) \in \varepsilon, \quad \text{or} \\ & \mathbf{x}_i - \mathbf{z}_{ij} = \mathbf{0}, \mathbf{x}_j - \mathbf{z}_{ij} = \mathbf{0} \quad \forall (i, j) \in \varepsilon, \quad \text{or} \\ & \text{mean}\{\mathbf{x}_j : (i, j) \in \varepsilon\} - \mathbf{x}_i = \mathbf{0}, \quad \forall i \in \mathcal{V}. \end{aligned}$$

# 去中心化ADMM

- 去中心化ADMM在互相连接的网络上运行
- 没有控制和分发数据的中心
- 应用：
  - wireless sensor networks
  - collaborative learning
- 去中心化ADMM交替进行如下步骤
  - 每个节点进行局部计算
  - 相邻节点交流或传播信息
- 因为数据没有共享或储存于某个中心，因此数据安全性得到保证
- 算法的收敛速度受如下条件影响
  - 问题的性质，如目标函数的凸性，问题的条件数等
  - 图的规模，连通性，谱性质等

# 浅变量图模型的选择

## V. Chandrasekaran, P. Parrilo, A. Willsky

包含正则项的极大似然模型

$$\min_{R, S, L} \langle R, \hat{\Sigma}_X \rangle - \log \det(R) + \alpha \|S\|_1 + \beta \text{Tr}(L), \text{ s.t. } R = S - L, R \succ 0, L \succeq 0,$$

其中  $X$  为观察到的变量,  $\Sigma_X^{-1} \approx R = S - L$ ,  $S$  稀疏,  $L$  低秩. 目标函数的前两项来自对数极大似然函数

$$l(K; \Sigma) = \log \det(K) - \text{tr}(K\Sigma).$$

引入指示函数

$$\mathcal{I}(L \succeq 0) := \begin{cases} 0, & \text{if } L \succeq 0 \\ +\infty, & \text{otherwise.} \end{cases}$$

模型可以改写为3项的ADMM算法

$$\min_{R, S, L} \langle R, \hat{\Sigma}_X \rangle - \log \det(R) + \alpha \|S\|_1 + \beta \text{Tr}(L) + \mathcal{I}(L \succeq 0), \text{ s.t. } R - S + L = 0.$$



# 稳定主成分追踪(PCP)

模型一：

$$\begin{aligned} \min_{L,S,Z} \quad & \|L\|_* + \rho \|S\|_1 \\ \text{s.t.} \quad & L + S + Z = M \\ & \|Z\|_F \leq \sigma, \end{aligned}$$

注： $M$  = 低秩矩阵+ 稀疏矩阵+ 噪声.

在图像处理中还要加入非负约束 $L \geq 0$ .

模型二：

$$\begin{aligned} \min_{L,S,Z,K} \quad & \|L\|_* + \rho \|S\|_1 + \mathcal{I}(\|Z\|_F \leq \sigma) + \mathcal{I}(K \geq 0) \\ \text{s.t.} \quad & L + S + Z = M \\ & L - K = 0. \end{aligned}$$

约束可以整合为

$$\begin{pmatrix} I & I \\ I & 0 \end{pmatrix} \begin{pmatrix} L \\ S \end{pmatrix} + \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} Z \\ K \end{pmatrix} = \begin{pmatrix} M \\ 0 \end{pmatrix}.$$

## 混合全变差和 $l_1$ 正则

原始模型：

$$\min_x TV(x) + \alpha \|Wx\|_1, \text{ s.t. } \|Rx - b\|_2 \leq \sigma.$$

新模型：

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

将变量分为两块： $x$  和  $(y, \{z_i\})$

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

$x$  的子问题不易求解。

# 将变量去耦合的方法

- 使用线性化技巧简化proximal算子，进行非精确更新
- 引入新变量作为桥梁，类似分布式ADMM

例如考虑如下问题

$$\min_{\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}} (f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2)) + g(\mathbf{y}), \text{ s.t. } (\mathbf{A}_1 \mathbf{x}_1 + \mathbf{A}_2 \mathbf{x}_2) + \mathbf{B} \mathbf{y} = \mathbf{b}.$$

在ADMM迭代中 $(\mathbf{x}_1, \mathbf{x}_2)$ 的子问题中 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 是耦合的，但是进行线性化后 $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 的子问题是独立的。

$$\min_{\mathbf{x}_1, \mathbf{x}_2} (f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2)) + \left\langle \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}, \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \right\rangle + \frac{1}{2t} \left\| \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} - \begin{bmatrix} \mathbf{x}_1^k \\ \mathbf{x}_2^k \end{bmatrix} \right\|_2^2.$$

## 非凸约束问题

考虑如下约束优化问题：

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}), \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{S}, \end{aligned}$$

其中 $f$ 是凸的，但是 $\mathcal{S}$ 是非凸的。可以将上述问题改写为：

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

交替方向乘子法产生如下迭代：

$$\begin{aligned} \mathbf{x}^{k+1} &= \operatorname{argmin}_{\mathbf{x}} (f(\mathbf{x}) + (\rho/2)\|\mathbf{x} - \mathbf{z}^k + \mathbf{u}^k\|_2^2), \\ \mathbf{z}^{k+1} &= \Pi_{\mathcal{S}}(\mathbf{x}^{k+1} + \mathbf{u}^k), \\ \mathbf{u}^{k+1} &= \mathbf{u}^k + (\mathbf{x}^{k+1} - \mathbf{z}^{k+1}) \end{aligned}$$

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

# 非凸约束问题

一般来说，这种投影很难计算，但是在下面列出的这些特殊情形中可以精确求解。

- 基数：如果  $\mathcal{S} = \{\mathbf{x} | \text{card}(\mathbf{x}) \leq c\}$ ，其中  $\text{card}(\mathbf{v})$  表示非零元素的数目，那么  $\Pi_{\mathcal{S}}(\mathbf{v})$  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：

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{Ax} - \mathbf{b}\|_2^2, \\ \text{s.t.} \quad & \text{card}(\mathbf{x}) \leq c. \end{aligned}$$

- 秩：如果  $\mathcal{S}$  是秩为  $c$  的矩阵的集合，那么  $\text{card}(\mathbf{V})$  可以通过对  $\mathbf{V}$  做奇异值分解， $\mathbf{V} = \sum_i \sigma_i \mathbf{u}_i \mathbf{u}_i^T$ ，然后保留前  $c$  大的奇异值及奇异向量，即  $\Pi_{\mathcal{S}}(\mathbf{V}) = \sum_{i=1}^c \sigma_i \mathbf{u}_i \mathbf{u}_i^T$ 。
- 布尔约束：如果  $\mathcal{S} = \{\mathbf{x} | x_i \in \{0, 1\}\}$ ，那么  $\Pi_{\mathcal{S}}(\mathbf{v})$  就是简单地把每个元素变为 0 和 1 中离它更近的数。

# 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：

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

其中， $\Omega$ 表示矩阵 $\mathbf{M}$ 中的已知元素的下标集合， $\mathcal{P}_{\Omega}(\mathbf{A})$ 表示得到一个新的矩阵 $\mathbf{A}'$ ，其下标在集合 $\Omega$ 中的所对应的元素等于矩阵 $\mathbf{A}$ 的对应元素，其下标不在集合 $\Omega$ 中的所对应的元素为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：

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

## 非负矩阵分解和补全

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

$$\mathbf{X}^{k+1} = \operatorname{argmin}_{\mathbf{X}} L_{\alpha,\beta}(\mathbf{X}, \mathbf{Y}^k, \mathbf{Z}^k, \mathbf{U}^k, \mathbf{V}^k, \mathbf{\Lambda}^k, \mathbf{\Pi}^k),$$

$$\mathbf{Y}^{k+1} = \operatorname{argmin}_{\mathbf{Y}} L_{\alpha,\beta}(\mathbf{X}^{k+1}, \mathbf{Y}, \mathbf{Z}^k, \mathbf{U}^k, \mathbf{V}^k, \mathbf{\Lambda}^k, \mathbf{\Pi}^k),$$

$$\mathbf{Z}^{k+1} = \operatorname{argmin}_{\mathcal{P}_{\Omega}(\mathbf{Z}-\mathbf{M})=0} L_{\alpha,\beta}(\mathbf{X}^{k+1}, \mathbf{Y}^{k+1}, \mathbf{Z}, \mathbf{U}^k, \mathbf{V}^k, \mathbf{\Lambda}^k, \mathbf{\Pi}^k),$$

$$\mathbf{U}^{k+1} = \operatorname{argmin}_{\mathbf{U} \geq 0} L_{\alpha,\beta}(\mathbf{X}^{k+1}, \mathbf{Y}^{k+1}, \mathbf{Z}^{k+1}, \mathbf{U}, \mathbf{V}^k, \mathbf{\Lambda}^k, \mathbf{\Pi}^k),$$

$$\mathbf{V}^{k+1} = \operatorname{argmin}_{\mathbf{V} \geq 0} L_{\alpha,\beta}(\mathbf{X}^{k+1}, \mathbf{Y}^{k+1}, \mathbf{Z}^{k+1}, \mathbf{U}^{k+1}, \mathbf{V}, \mathbf{\Lambda}^k, \mathbf{\Pi}^k),$$

$$\mathbf{\Lambda}^{k+1} = \mathbf{\Lambda}^k + \tau\alpha(\mathbf{X}^{k+1} - \mathbf{U}^{k+1}),$$

$$\mathbf{\Pi}^{k+1} = \mathbf{\Pi}^k + \tau\beta(\mathbf{Y}^{k+1} - \mathbf{V}^{k+1}).$$

- 1 交替方向乘子法
- 2 常见变形和技巧
- 3 应用举例
- 4 Douglas-Rachford splitting 算法**
- 5 ADMM的收敛性分析



# Douglas-Rachford splitting 算法

考虑复合优化问题

$$\min f(x) = g(x) + h(x)$$

其中  $g, h$  是闭凸函数.

**Douglas-Rachford 迭代:** 从任意初始点  $z^0$  开始,

$$x^{(k)} = \text{prox}_{th}(z^{k-1})$$

$$y^{(k)} = \text{prox}_{tg}(2x^k - z^{k-1})$$

$$z^{(k)} = z^{k-1} + y^k - x^k$$

- $t$  为正常数
- 通常用于  $g, h$  的 proximal 算子计算代价较小的场景
- 在较弱的条件下(如极小点存在), 迭代点列  $x^k$  收敛

# 等价形式

- 从 $y$ 的更新开始

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

- 交换 $z$ 和 $x$ 的更新顺序

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

- 作变量替换 $w = z - x$

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

$$y^+ = \text{prox}_{tg}(x - w)$$

$$x^+ = \text{prox}_{th}(y^+ + w)$$

$$w^+ = w + y^+ - x^+$$

# DR迭代和不动点迭代

Douglas-Rachford迭代可以写成不动点迭代的形式

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

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

DR迭代和不动点迭代的关系

- 若  $z$  是不动点, 则  $x = \text{prox}_{th}(z)$  是目标函数的极小点

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

- 若  $x$  是目标函数的极小点,  $u \in t\partial g(x) \cap -t\partial h(x)$ ,  
则  $x - u = F(x - u)$

# Douglas-Rachford迭代的松弛形式

- 不动点迭代的松弛形式

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

$1 < \rho < 2$  称为超松弛,  $0 < \rho < 1$  称为次松弛

- DR迭代的松弛形式一

$$x^+ = \text{prox}_{th}(z)$$

$$y^+ = \text{prox}_{tg}(2x^+ - z)$$

$$z^+ = z + \rho(y^+ - x^+)$$

- DR迭代的松弛形式二

$$y^+ = \text{prox}_{tg}(x - w)$$

$$x^+ = \text{prox}_{th}((1 - \rho)x + \rho y^+ + w)$$

$$w^+ = w + \rho y^+ + (1 - \rho)x - x^+$$

## 凸的原始问题

$$\begin{array}{ll}\min & f_1(x_1) + f_2(x_2) \\ \text{s.t.} & A_1x_1 + A_2x_2 = b\end{array}$$

## 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：

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

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

- 第一式的最优性条件为

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

上式等价于存在  $x_1^k \in \partial f_1^*(-A_1^T y^{k+1})$ , 使得

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

这就是如下更新的最优性条件.

$$x_1^k = \underset{x_1}{\operatorname{argmin}} \left\{ f_1(x_1) + (x^k)^T (A_1 x_1 - b) + \frac{t}{2} \|A_1 x_1 - b - w^k/t\|_2^2 \right\}.$$

- 类似地，第二式的最优性条件为

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

其等价于存在  $x_2^k \in \partial f_2^*(-A_2^T x^{k+1})$ ，使得

$$x^{k+1} = x^k + t(A_1 x_1^k + A_2 x_2^k - b).$$

进一步等价于如下更新的最优性条件.

$$x_2^k = \operatorname{argmin}_{x_2} \left\{ f_2(x_2) + (x^k)^T (A_2 x_2) + \frac{t}{2} \|A_1 x_1^k + A_2 x_2 - b\|_2^2 \right\}.$$

- 由第一、第二式的等价形式整理可得  $w^{k+1} = -tA_2 x_2^k$ .

# DR迭代与ADMM

定义增广拉格朗日函数

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

① 关于 $x_1$ 极小化增广拉格朗日函数

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

② 关于 $x_2$ 极小化增广拉格朗日函数

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

③ 对偶变量更新 $z^{(k)} = z^{(k-1)} + t(A_1 x_1^{(k)} + A_2 x_2^{(k)} - b)$



# 非扩张函数

## 定义

函数 $f$ 称为非扩张的(*non-expansive*), 若

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

函数 $f$ 称为固定非扩张的(*firmly non-expansive*), 若

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

定理:  $\text{prox}_h$  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).

- 设 $u = \text{prox}_h(x), v = \text{prox}_h(y)$ , 由proximal算子的性质

$$x - u \in \partial h(u), y - v \in \partial h(v) \Rightarrow (x - u - y + v)^{\top} (u - v) \geq 0$$

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

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

# Douglas-Rachford迭代映射

定义迭代映射 $F, G$ 如下

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

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

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

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

- 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  $\text{prox}_{th}$  和  $\text{prox}_{tg}$  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  $f^* = \inf_x (g(x) + h(x))$  有限且可取到
- $\rho_k \in [\rho_{\min}, \rho_{\max}]$ , 其中  $0 < \rho_{\min} < \rho_{\max} < 2$

结论

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

## Proof.

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

$$\begin{aligned}\|z^+ - z^*\|_2^2 - \|z - z^*\|_2^2 &= 2(z^+ - z)^T(z - z^*) + \|z^+ - z\|_2^2 \\ &= -2\rho G(z)^T(z - z^*) + \rho^2\|G(z)\|_2^2 \\ &\leq -\rho(2 - \rho)\|G(z)\|_2^2 \\ &\leq -M\|G(z)\|_2^2\end{aligned}\tag{22}$$

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

● (22) 可以推出

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

● 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 $\lim_{k \rightarrow \infty} \|z^{(k)} - z^*\|_2$  存在

## Proof.

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

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

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

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

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

均存在, 则由  $z^{(k)}$  的收敛性知

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



# 提纲

- 1 交替方向乘子法
- 2 常见变形和技巧
- 3 应用举例
- 4 Douglas-Rachford splitting 算法
- 5 ADMM的收敛性分析**

# 假设

我们先引入一些必要的假设.

- $f_1(x), f_2(x)$  均为闭凸函数, 且每个ADMM 迭代子问题存在唯一解;
- 原始问题的解集非空, 且Slater 条件满足.

注: 假设给出的条件是很基本的.

- $f_1$  和  $f_2$  的凸性保证了要求解的问题是凸问题, 每个子问题存在唯一解是为了保证迭代的良定义
- 在Slater 条件满足的情况下, 原始问题的KKT 对和最优解是对应的, 因此可以很方便地使用KKT 条件来讨论收敛性.



## 记号

由于原始问题解集非空，不妨设 $(x_1^*, x_2^*, y^*)$ 是KKT对，即满足条件

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

我们最终的目的是证明ADMM迭代序列 $\{(x_1^k, x_2^k, y^k)\}$ 收敛到原始问题的一个KKT对，因此引入如下记号来表示当前迭代点和KKT对的误差：

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

我们进一步引入如下辅助变量来简化之后的证明：

$$\begin{aligned} u^k &= -A_1^T [y^k + (1 - \tau)\rho(A_1 e_1^k + A_2 e_2^k) + \rho A_2 (x_2^{k-1} - x_2^k)], \\ v^k &= -A_2^T [y^k + (1 - \tau)\rho(A_1 e_1^k + A_2 e_2^k)], \\ \Psi_k &= \frac{1}{\tau\rho} \|e_y^k\|^2 + \rho \|A_2 e_2^k\|^2, \\ \Phi_k &= \Psi_k + \max(1 - \tau, 1 - \tau^{-1}) \rho \|A_1 e_1^k + A_2 e_2^k\|^2. \end{aligned} \tag{23}$$

# ADMM的收敛性分析

## 引理

假设 $\{(x_1^k, x_2^k, y^k)\}$ 为交替方向乘子法产生一个迭代序列, 那么, 对任意的 $k \geq 1$ 有

$$u^k \in \partial f_1(x_1^k), v^k \in \partial f_2(x_2^k), \quad (24)$$

$$\begin{aligned} \Phi_k - \Phi_{k+1} \geq & \min(\tau, 1 + \tau - \tau^2) \rho \|A_2(x_2^k - x_2^{k+1})\|^2 \\ & + \min(1, 1 + \tau^{-1} - \tau) \rho \|A_1 e_1^{k+1} + A_2 e_2^{k+1}\|^2. \end{aligned} \quad (25)$$

注: 只有当 $\tau \in \left(0, \frac{1+\sqrt{5}}{2}\right)$ 时, (25) 式中不等号右侧的项才为非负.

# ADMM的收敛性分析

## Proof.

先证明(24)式的两个结论. 根据交替方向乘子法的迭代过程, 对 $x_1^{k+1}$ 我们有

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

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, 消去 $y^k$ 就有

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

根据 $u^k$ 的定义自然有 $u^k \in \partial f_1(x_1^k)$  (注意代回 $b = A_1 x_1^* + A_2 x_2^*$ ). 类似地, 对 $x_2^{k+1}$ 我们有

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

同样利用 $y^k$ 的表达式消去 $y^k$ , 得到

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

# ADMM的收敛性分析

## Proof.

根据 $v^k$ 的定义自然有 $v^k \in \partial f_2(x_2^k)$ .

接下来证明不等式(25). 首先根据 $(x_1^*, x_2^*, y^*)$ 的最优性条件以及关系式(24),

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

根据凸函数的单调性,

$$\begin{aligned}\langle u^{k+1} + A_1^T y^*, x_1^{k+1} - x_1^* \rangle &\geq 0, \\ \langle v^{k+1} + A_2^T y^*, x_2^{k+1} - x_2^* \rangle &\geq 0.\end{aligned}$$

将上述两个不等式相加, 结合 $u^{k+1}, v^{k+1}$ 的定义, 并注意到恒等式

$$A_1 x_1^{k+1} + A_2 x_2^{k+1} - b = (\tau\rho)^{-1}(y^{k+1} - y^k) = (\tau\rho)^{-1}(e_y^{k+1} - e_y^k), \quad (26)$$

# ADMM的收敛性分析

## Proof.

我们可以得到

$$\begin{aligned} & \frac{1}{\tau\rho} \langle e_y^{k+1}, e_y^k - e_y^{k+1} \rangle - (1 - \tau)\rho \|A_1 x_1^{k+1} + A_2 x_2^{k+1} - b\|^2 \\ & + \rho \langle A_2(x_2^{k+1} - x_2^k), A_1 x_1^{k+1} + A_2 x_2^{k+1} - b \rangle \\ & - \rho \langle A_2(x_2^{k+1} - x_2^k), A_2 e_2^{k+1} \rangle \geq 0. \end{aligned} \tag{27}$$

不等式(27)的形式和不等式(25)还有一定差异，主要的差别就在

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

这一项上。我们接下来估计这一项的上界。

# ADMM的收敛性分析

## Proof.

引入新符号

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

则  $-A_2^T\nu^{k+1} \in \partial f_2(x_2^{k+1})$  以及  $-A_2^T\nu^k \in \partial f_2(x_2^k)$ . 再利用单调性知

$$\langle -A_2^T(\nu^{k+1} - \nu^k), x_2^{k+1} - x_2^k \rangle \geq 0. \quad (28)$$

根据这些不等式关系我们最终得到

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

# ADMM的收敛性分析

## Proof.

估计完这一项之后，不等式(27)可以放缩成

$$\begin{aligned} & \frac{1}{\tau\rho} \langle e_y^{k+1}, e_y^k - e_y^{k+1} \rangle - (1 - \tau)\rho \|A_1 x_1^{k+1} + A_2 x_2^{k+1} - b\|^2 \\ & + M^{k+1} - \rho \langle A_2(x_2^{k+1} - x_2^k), A_2 e_2^{k+1} \rangle \geq 0. \end{aligned}$$

上式中含有内积项，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

$$\begin{aligned} & \frac{1}{\tau\rho} (\|e_y^k\|^2 - \|e_y^{k+1}\|^2) - (2 - \tau)\rho \|A_1 x_1^{k+1} + A_2 x_2^{k+1} - b\|^2 \\ & + 2M^{k+1} - \rho \|A_2(x_2^{k+1} - x_2^k)\|^2 - \rho \|A_2 e_2^{k+1}\|^2 + \rho \|A_2 e_2^k\|^2 \geq 0. \end{aligned} \tag{29}$$

# ADMM的收敛性分析

## Proof.

此时除了 $M^{k+1}$ 中的项, (29)中的其他项均在不等式(25)中出现. 由于 $M^{k+1}$ 的符号和 $\tau$ 的取法有关, 下面我们针对 $\tau$ 的两种取法进行讨论.

情形一  $\tau \in (0, 1]$ , 此时 $M^{k+1} \geq 0$ , 根据基本不等式,

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

代入不等式(29)得到

$$\begin{aligned} & \frac{1}{\tau\rho} \|e_y^k\|^2 + \rho \|A_2e_2^k\|^2 + (1 - \tau)\rho \|A_1e_1^k + A_2e_2^k\|^2 \\ & - \left[ \frac{1}{\tau\rho} \|e_y^{k+1}\|^2 + \rho \|A_2e_2^{k+1}\|^2 + (1 - \tau)\rho \|A_1e_1^{k+1} + A_2e_2^{k+1}\|^2 \right] \quad (30) \\ & \geq \rho \|A_1x_1^{k+1} + A_2x_2^{k+1} - b\|^2 + \tau\rho \|A_2(x_2^{k+1} - x_2^k)\|^2. \end{aligned}$$



# ADMM的收敛性分析

## Proof.

情形二  $\tau > 1$ , 此时  $M^{k+1} < 0$ , 根据基本不等式,

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

同样代入不等式(29)可以得到

$$\begin{aligned} & \frac{1}{\tau\rho} \|e_y^k\|^2 + \rho \|A_2e_2^k\|^2 + \left(1 - \frac{1}{\tau}\right) \rho \|A_1e_1^k + A_2e_2^k\|^2 \\ & - \left[ \frac{1}{\tau\rho} \|e_y^{k+1}\|^2 + \rho \|A_2e_2^{k+1}\|^2 + \left(1 - \frac{1}{\tau}\right) \rho \|A_1e_1^{k+1} + A_2e_2^{k+1}\|^2 \right] \quad (31) \\ & \geq \left(1 + \frac{1}{\tau} - \tau\right) \rho \|A_1x_1^{k+1} + A_2x_2^{k+1} - b\|^2 \\ & + (1 + \tau - \tau^2) \rho \|A_2(x_2^{k+1} - x_2^k)\|^2. \end{aligned}$$

# ADMM的收敛性分析

## 定理

在假设的条件下，进一步假定 $A_1, A_2$ 列满秩。如果 $\tau \in \left(0, \frac{1+\sqrt{5}}{2}\right)$ ，则序列 $\{(x_1^k, x_2^k, y^k)\}$ 收敛到原始问题的一个KKT对。

## Proof.

前证引理表明 $\Phi_k$ 都是有界列，根据 $\Phi_k$ 的定义(23)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

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

可以进一步推出 $\{\|A_1 e_1^k\|\}$ 也是有界序列。注意到 $A_1^T A_1 \succ 0, A_2^T A_2 \succ 0$ ，因此以上有界性也等价于 $\{(x_1^k, x_2^k, y^k)\}$ 是有界序列。

# ADMM的收敛性分析

## Proof.

另一个直接结果是无穷级数

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

都是收敛的，这表明

$$\begin{aligned} \|A_1 e_1^k + A_2 e_2^k\| &= \|A_1 x_1^k + A_2 x_2^k - b\| \rightarrow 0, \\ \|A_2(x_2^{k+1} - x_2^k)\| &\rightarrow 0. \end{aligned} \tag{32}$$

利用这些结果我们就可以推导收敛性了。首先证明迭代点子列的收敛性。由于 $\{(x_1^k, x_2^k, y^k)\}$ 是有界序列，因此它存在一个收敛子列，设

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

由(23)式中的 $u^k, v^k$ 的定义及(32)式可得 $\{u^k\}$ 与 $\{v^k\}$ 相应的子列也收敛。

# ADMM的收敛性分析

## Proof.

记

$$u^\infty \stackrel{\text{def}}{=} \lim_{j \rightarrow \infty} u^{k_j} = -A_1^T y^\infty, \quad v^\infty = \lim_{j \rightarrow \infty} v^{k_j} = -A_2^T y^\infty. \quad (33)$$

从(24)式我们知道对于任意的 $k \geq 1$ , 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(32)的第一式可知

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

这表明 $(x_1^\infty, x_2^\infty, y^\infty)$ 是原始问题的一个KKT对. 因此上述分析中的 $(x_1^*, x_2^*, y^*)$ 均可替换为 $(x_1^\infty, x_2^\infty, y^\infty)$ .

# ADMM的收敛性分析

## Proof.

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

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

这说明

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

进一步有

$$0 \leq \limsup_{k \rightarrow \infty} \|A_1 e_1^k\| \leq \lim_{k \rightarrow \infty} (\|A_2 e_2^k\| + \|A_1 e_1^k + A_2 e_2^k\|) = 0.$$

注意到 $A_1^T A_1 \succ 0$ ,  $A_2^T A_2 \succ 0$ , 所以最终我们得到全序列收敛.



## 多块ADMM收敛性反例

- 考虑最优化问题

$$\begin{aligned} \min \quad & 0, \\ \text{s.t.} \quad & A_1x_1 + A_2x_2 + A_3x_3 = 0, \end{aligned} \tag{34}$$

其中  $A_i \in \mathbb{R}^3$ ,  $i = 1, 2, 3$  为三维空间中的非零向量,  $x_i \in \mathbb{R}$ ,  $i = 1, 2, 3$  是自变量. 该问题实际上就是求解三维空间中的线性方程组, 若  $A_1, A_2, A_3$  之间线性无关, 则问题(34) 只有零解. 此时容易计算出最优解对应的乘子为  $y = (0, 0, 0)^T$ .

- 增广拉格朗日函数为

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

## 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

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

整理可得

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

可类似地计算 $x_2, x_3$ 的表达式

- 因此多块交替方向乘子法的迭代格式可以写为

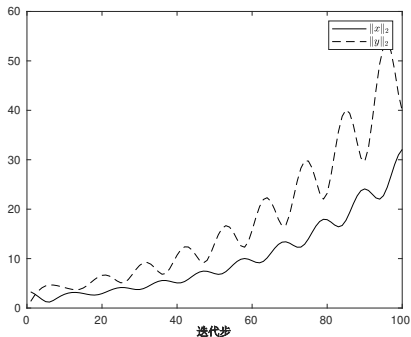
$$\begin{aligned} x_1^{k+1} &= -\frac{1}{\|A_1\|^2} A_1^T \left( \frac{y^k}{\rho} + A_2 x_2^k + A_3 x_3^k \right), \\ x_2^{k+1} &= -\frac{1}{\|A_2\|^2} A_2^T \left( \frac{y^k}{\rho} + A_1 x_1^{k+1} + A_3 x_3^k \right), \\ x_3^{k+1} &= -\frac{1}{\|A_3\|^2} A_3^T \left( \frac{y^k}{\rho} + A_1 x_1^{k+1} + A_2 x_2^{k+1} \right), \\ y^{k+1} &= y^k + \rho (A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1}). \end{aligned} \tag{35}$$

## 多块ADMM收敛性反例

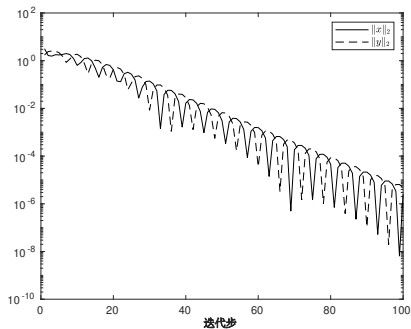
- 自变量初值初值选为 $(1, 1, 1)$ ，乘子选为 $(0, 0, 0)$ . 选取 $A$ 为

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

- 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 $\tilde{A}$



(b) 系数矩阵为 $\hat{A}$



# References

## Douglas-Rachford method, ADMM, Spingarn's method

- J. E. Spingarn, *Applications of the method of partial inverses to convex programming: decomposition*, Mathematical Programming (1985)
- J. Eckstein and D. Bertsekas, *On the Douglas-Rachford splitting method and the proximal algorithm for maximal monotone operators*, Mathematical Programming (1992)
- P.L. Combettes and J.-C. Pesquet, *A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery*, IEEE Journal of Selected Topics in Signal Processing (2007)
- S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers* (2010)
- N. Parikh, S. Boyd, *Block splitting for distributed optimization* (2013)

**image deblurring:** the example is taken from  
D. O'Connor and L. Vandenberghe, *Primal-dual decomposition by operator splitting and applications to image deblurring* (2014)