章节 7.5 交替方向乘子法

问题形式

考虑如下可分的凸问题：

$$\begin{aligned} \min_{x_1,x_2} \quad & f_1(x_1)+f_2(x_2), \\ \text{s.t.} \quad & A_1x_1+A_2x_2=b, \end{aligned} \tag{171}$$

- f_1，f_2 是适当的闭凸函数，不需要是光滑的，$x_1 \in \mathbb{R}^n, x_2 \in \mathbb{R}^m$，$A_1 \in \mathbb{R}^{p\times n}, A_2 \in \mathbb{R}^{p\times m}$，$b \in \mathbb{R}^p$.
- 可分：目标函数可以分成两个变量独占的函数，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

- 考虑如下问题

$$\min_{x} \quad f_1(x) + f_2(x).$$

若 f_1, f_2 都是非光滑函数，之前学过的算法，例如近似点梯度法，则无法处理。引入一个新的变量 z 并令 $x = z$，将问题转化为 (171)的形式：

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

- 带线性变换的无约束优化问题

$$\min_{x} \quad f_1(x) + f_2(Ax).$$

可以引入一个新的变量 z，令 $z = Ax$，则问题变为

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

同样转化为 (171)的形式。

例

全局一致性问题

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

令 $x = z$，并将 x 复制 N 份，分别为 x_i，那么问题转化为

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

- 首先写出问题(171)的增广拉格朗日函数

$$
\begin{aligned}
L_\rho(x_1, x_2, y) =& f_1(x_1) + f_2(x_2) + y^{\mathrm{T}}(A_1x_1 + A_2x_2 - b) \\
&+ \frac{\rho}{2}\|A_1x_1 + A_2x_2 - b\|_2^2,
\end{aligned} \tag{172}
$$

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 (ALM) 为如下更新：

$$
(x_1^{k+1}, x_2^{k+1}) = \operatorname*{argmin}_{x_1, x_2} L_\rho(x_1, x_2, y^k), \tag{173}
$$

$$
y^{k+1} = y^k + \tau\rho(A_1x_1^{k+1} + A_2x_2^{k+1} - b), \tag{174}
$$

其中 τ 为步长.
ALM 的缺点：子问题(173)不易求解。

英文名：Alternating direction method of multipliers, 简称 ADMM

- 交替方向乘子法的基本思路: 第一步迭代(173)同时对 x_1 和 x_2 进行求解有时候比较困难，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：

$$x_1^{k+1} = \operatorname*{argmin}_{x_1} L_\rho(x_1, x_2^k, y^k), \tag{175}$$

$$x_2^{k+1} = \operatorname*{argmin}_{x_2} L_\rho(x_1^{k+1}, x_2, y^k), \tag{176}$$

$$y^{k+1} = y^k + \tau\rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \tag{177}$$

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

对于基追踪问题. 设 $A\in\mathbb{R}^{m\times n}(m\leqslant n)$, $b\in\mathbb{R}^m$, $x\in\mathbb{R}^n$, 基追踪问题被描述为

$$\min_{x\in\mathbb{R}^n} \quad \|x\|_1, \quad \text{s.t.} \quad Ax=b. \tag{178}$$

ALM 迭代更新格式为

$$\begin{cases} x^{k+1} = \underset{x\in\mathbb{R}^n}{\arg\min}\left\{\|x\|_1+\frac{\sigma}{2}\left\|Ax-b+\frac{\lambda^k}{\sigma}\right\|_2^2\right\}, \\ \lambda^{k+1} = \lambda^k+\sigma\left(Ax^{k+1}-b\right). \end{cases} \tag{179}$$

引入 $y=x$, 问题变为

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} \quad \|y\|_1, \quad \text{s.t.} \quad Ax=b, \quad x=y. \tag{180}$$

ADMM 迭代更新格式为

$$\begin{cases} x^{k+1} = \underset{x\in\mathbb{R}^n}{\arg\min}\left\{\langle\lambda_{1,k}, Ax-b\rangle+\frac{\sigma}{2}\|Ax-b\|_2^2+\langle\lambda_{2,k}, x-y_k\rangle+\frac{\sigma}{2}\|x-y_k\|_2^2\right\}, \\ y^{k+1} = \underset{y\in\mathbb{R}^m}{\arg\min}\left\{\|y\|_1+\langle\lambda_{2,k}, y-x_{k+1}\rangle+\frac{\sigma}{2}\|y-x_{k+1}\|_2^2\right\}, \\ \lambda_{1,k+1} = \lambda_{1,k}+\sigma\left(Ax^{k+1}-b\right), \\ \lambda_{2,k+1} = \lambda_{2,k}+\sigma\left(y_{k+1}-x_{k+1}\right) \end{cases} \tag{181}$$

- 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

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

对于 x_k 的子问题，有如下方式减少每个迭代步的计算量:

- 因为 $\rho > 0$，所以 $A^{\mathrm{T}}A + \rho I$ 总是可逆的．若使用固定的罚因子 ρ，我们使用例如 Cholesky 分解得到 $A^{\mathrm{T}}A + \rho I$ 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（即 $m \ll n$），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 $O(n^3)$
- 可以利用 SMW 公式减少矩阵求逆计算量:

$$(A^{\mathrm{T}}A + \rho I_n)^{-1} = \rho^{-1}I - \rho^{-1}A^{\mathrm{T}}(\rho I_m + AA^{\mathrm{T}})^{-1}A$$

应用举例：Fused 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:

$$\min_x \quad \mu\|Dx\|_1 + \frac{1}{2}\|Ax - b\|^2. \tag{182}$$

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

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

且 $A = I$ 时，广义 LASSO 问题为

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

这个问题就是图像去噪问题模型.

Fused LASSO 问题

- 通过引入约束 $Dx = z$:

$$\begin{aligned} \min_{x,z} \quad & \frac{1}{2}\|Ax-b\|^2 + \mu\|z\|_1, \\ \text{s.t.} \quad & Dx - z = 0, \end{aligned} \tag{183}$$

- 引入乘子 y，其增广拉格朗日函数为

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

- 此问题的 x 迭代是求解方程组

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

而 z 迭代依然通过 ℓ_1 范数的邻近算子.

- 因此交替方向乘子法所产生的迭代为

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

- 对于全变差去噪问题，$A^{\mathrm{T}}A + \rho D^{\mathrm{T}}D$ 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 $O(n\log n)$.

$$b = Kx_t + w$$

- 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。

图像去模糊的实例

我们有如下图像添加噪声和去噪声例子：求解下面问题，以恢复出带噪声/模糊的图片，

$$\min_{x} \frac{1}{2}\|Kx - b\|^2 + \|Dx\|_1.$$

- 1024 × 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

矩阵分离问题, 或称鲁棒主成分分析（RPCA）

在经典的 PCA 中，数据被分解为几个主成分，这些主成分捕获了数据中的主要变异性。然而，当数据中存在离群点或异常值时，PCA 的性能可能会大大下降，因为它试图捕获所有数据点的变异性，包括异常值。

RPCA 解决了这个问题。它将数据矩阵分解为两部分：一个低秩矩阵和一个稀疏矩阵。低秩矩阵捕获数据的主要结构，而稀疏矩阵则包含异常值或离群点。通过这种方式，RPCA 能够在保持数据主要结构的同时，有效地处理异常值。

- 其数学模型如下：

$$\begin{aligned} \min_{X,S} \quad & \|X\|_* + \mu\|S\|_1, \\ \text{s.t.} \quad & X + S = M, \end{aligned} \tag{184}$$

其中 $\|\cdot\|_1$ 与 $\|\cdot\|_*$ 分别表示矩阵 ℓ_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

$$L_\rho(X,S,Y) = \|X\|_* + \mu\|S\|_1 + \langle Y, X+S-M\rangle + \frac{\rho}{2}\|X+S-M\|_F^2. \tag{185}$$

对于 X 子问题，

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

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 的约化奇异值分解.

- 对于 S 子问题，

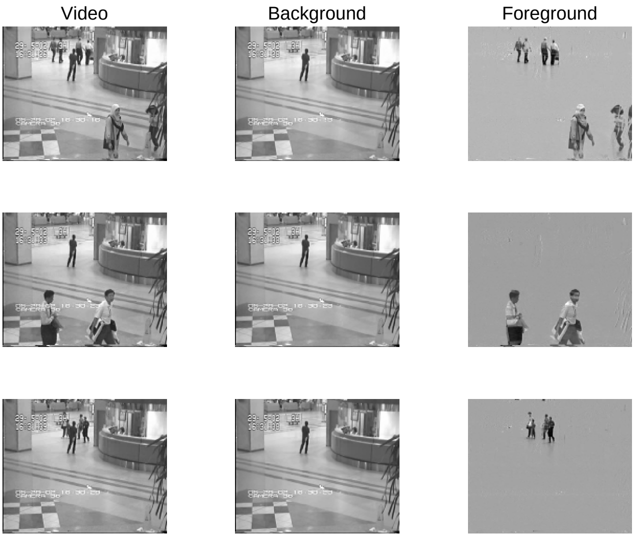
$$\begin{aligned} S^{k+1} &= \underset{S}{\operatorname{argmin}}\ L_\rho(X^{k+1}, S, Y^k) \\ &= \underset{S}{\operatorname{argmin}} \left\{ \mu\|S\|_1 + \frac{\rho}{2}\left\| X^{k+1} + S - M + \frac{Y^k}{\rho} \right\|_F^2 \right\} \\ &= \operatorname{prox}_{(\mu/\rho)\|\cdot\|_1}\left(M - X^{k+1} -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

- 那么交替方向乘子法的迭代格式为

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

例：矩阵分解

通过 RPCA，将视频图片分为背景（低秩）和前景（稀疏）两个部分。



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.

定理 12

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.

- 考虑有多块变量的情形

$$\begin{aligned} \min_{x_1,x_2,\cdots,x_N} \quad & f_1(x_1)+f_2(x_2)+\cdots+f_N(x_N), \\ \text{s.t.} \quad & A_1x_1+A_2x_2+\cdots+A_Nx_N=b. \end{aligned} \tag{186}$$

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 $L_\rho(x_1,x_2,\cdots,x_N,y)$，相应的多块 ADMM 迭代格式为

$$\begin{aligned} x_1^{k+1} &= \operatorname*{argmin}_x L_\rho(x,x_2^k,\cdots,x_N^k,y^k), \\ x_2^{k+1} &= \operatorname*{argmin}_x L_\rho(x_1^{k+1},x,\cdots,x_N^k,y^k), \\ &\cdots\cdots\cdots\cdots \\ x_N^{k+1} &= \operatorname*{argmin}_x L_\rho(x_1^{k+1},x_2^{k+1},\cdots,x,y^k), \\ y^{k+1} &= y^k+\tau\rho(A_1x_1^{k+1}+A_2x_2^{k+1}+\cdots+A_Nx_N^{k+1}-b), \end{aligned}$$

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

考虑最优化问题

$$\begin{aligned} \min \quad & 0, \\ \text{s.t.} \quad & A_1x_1 + A_2x_2 + A_3x_3 = 0, \end{aligned} \tag{187}$$

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(187)只有零解．此时容易计算出最优解对应的乘子为 $y = (0, 0, 0)^{\mathrm{T}}$．

(187)的增广拉格朗日函数为

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

- 当固定 x_2, x_3, y 时，对 x_1 求最小可推出

$$A_1^{\mathrm{T}} y + \rho A_1^{\mathrm{T}}(A_1x_1 + A_2x_2 + A_3x_3) = 0,$$

整理可得

$$x_1 = -\frac{1}{\|A_1\|^2}\left(A_1^{\mathrm{T}}\left(\frac{y}{\rho} + A_2x_2 + A_3x_3\right)\right).$$

可类似地计算 x_2, x_3 的表达式

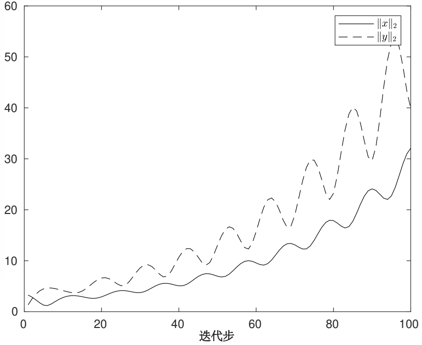
- 因此多块交替方向乘子法的迭代格式可以写为

$$\begin{aligned}
x_1^{k+1} &= -\frac{1}{\|A_1\|^2}A_1^{\mathrm{T}}\left(\frac{y^k}{\rho} + A_2x_2^k + A_3x_3^k\right),\\
x_2^{k+1} &= -\frac{1}{\|A_2\|^2}A_2^{\mathrm{T}}\left(\frac{y^k}{\rho} + A_1x_1^{k+1} + A_3x_3^k\right),\\
x_3^{k+1} &= -\frac{1}{\|A_3\|^2}A_3^{\mathrm{T}}\left(\frac{y^k}{\rho} + A_1x_1^{k+1} + A_2x_2^{k+1}\right),\\
y^{k+1} &= y^k + \rho(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1}).
\end{aligned} \tag{188}$$

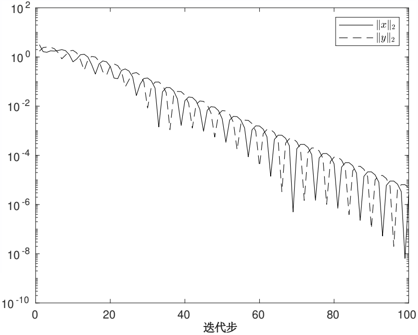
- 自变量初值初值选为 $(1,1,1)$，乘子选为 $(0,0,0)$. 选取 A 为

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

- 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 $\widetilde{A}$，ADMM 不收敛。



(b) 系数矩阵为 $\widehat{A}$

1. 给出 ADMM 求解线性规划标准形式，以及其对偶问题的迭代形式，要求写出子问题的求解公式。
2. 考虑如下问题
$$\min \sum_{i=1}^{n} f_i(x_i), \quad \text{s.t.} \quad x_1 = x_2 = \ldots = x_n.$$
其中，f_i 均为闭凸函数，且其近似点映射有显式解。写出多块 ADMM 求解该问题的迭代形式, 要求每个 x_i 的子问题均有显式解。
3. 考虑如下问题：给定对称正定矩阵 $C \in \mathbb{R}^{n\times n}$，$\lambda > 0$ 为给定的实数，求解
$$\min_X \mathrm{Tr}(CX) - \log\det X + \lambda\|X\|_1,$$
这里，$\|X\|_1 = \sum_{i=1}^n \sum_{j=1}^n |X_{ij}|$。使用 ADMM 求解该问题，并给出子问题的显式解。(提示：该问题为凸问题，默认 X 的定义域为对称正定矩阵集合，即无需添加正定矩阵约束，直接考虑无约束问题，$\log\det X$ 的梯度为 X^{-1}）.