# 第七章 复合优化算法

修贤超

https://xianchaoxiu.github.io

# 目录

# 典型问题形式

- 考虑如下凸问题

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

- $f_1$，$f_2$ 是适当的闭凸函数，但不要求是光滑的
- 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

# 问题形式举例

- **例 7.13** 可以分成两块的无约束优化问题

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

引入一个新的变量 $z$ 并令 $x = z$，将问题转化为

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

- **例 7.14** 带线性变换的无约束优化问题

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

引入一个新的变量 $z$，令 $z = Ax$，则问题变为

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

- 例 7.15　凸集$C \subset \mathbb{R}^n$ 上的约束优化问题

$$
\begin{aligned}
\min_{x} \quad & f(x) \\
\text{s.t.} \quad & Ax \in C
\end{aligned}
$$

引入约束 $z = Ax$，那么问题转化为

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

- 例 7.16 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

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

# 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

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

- 增广拉格朗日函数法为如下更新

$$
\begin{aligned}
(x_1^{k+1}, x_2^{k+1}) &= \arg\min_{x_1, x_2} L_\rho(x_1, x_2, y^k) \\
y^{k+1} &= y^k + \tau\rho(A_1x_1^{k+1} + A_2x_2^{k+1} - b)
\end{aligned}
$$

# 交替方向乘子法

- Alternating direction method of multipliers, ADMM
- 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
- 其迭代格式可以总结如下

$$
\begin{aligned}
x_1^{k+1} &=_{x_1} L_\rho(x_1, x_2^k, y^k) \\
x_2^{k+1} &= \arg\min_{x_2} L_\rho(x_1^{k+1}, x_2, y^k) \\
y^{k+1} &= y^k + \tau\rho(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b)
\end{aligned}
$$

# 原问题最优性条件

- 问题(1)的拉格朗日函数为

$$L(x_1, x_2, y) = f_1(x_1) + f_2(x_2) + y^\top (A_1 x_1 + A_2 x_2 - b)$$

- 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

$$0 \in \partial_{x_1} L(x_1^*, x_2^*, y^*) = \partial f_1(x_1^*) + A_1^\top y^* \tag{2a}$$

$$0 \in \partial_{x_2} L(x_1^*, x_2^*, y^*) = \partial f_2(x_2^*) + A_2^\top y^* \tag{2b}$$

$$A_1 x_1^* + A_2 x_2^* = b \tag{2c}$$

- 条件(2c)又称为原始可行性条件，条件(2a)和条件(2b)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

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

- 根据最优性条件推出

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

- 当 $\tau = 1$ 时知

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

# ADMM 单步迭代最优性条件

- 由$x_1$ 的更新公式

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

- 假设子问题能精确求解，根据最优性条件

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

- 当 $\tau = 1$ 时知

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

# ADMM 单步迭代最优性条件

- 对比条件(2a)可知多出来的项为 $A_1^\top A_2(x_2^{k-1}-x_2^k)$, 因此要检测对偶可行性只需要检测残差

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

- 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

$$\begin{aligned}
0 &\approx \|r^k\| = \|A_1x_1^k + A_2x_2^k - b\| \quad (\text{原始可行性}) \\
0 &\approx \|s^k\| = \|A_1^\top A_2(x_2^{k-1}-x_2^k)\| \quad (\text{对偶可行性})
\end{aligned}$$

# 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
- 考虑第一个子问题

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

- 当子问题目标函数可微时，线性化为

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

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

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

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 $f_1(x_1) = \frac{1}{2}\|Cx_1 - d\|_2^2$，那么针对 $x_1$ 的更新等价于求解线性方程组

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

- 虽然子问题有显式解，但是每步求解的复杂度仍然比较高
- 首先对 $C^\top C + \rho A_1^\top A_1$ 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 $C^\top C + \rho A_1^\top A_1$ 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

- 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 $A_1^\top A_1$，此时子问题可替换为

$$x_1^{k+1} = \arg\min_{x_1} \left\{ f_1(x_1) + \frac{\rho}{2}\|A_1 x_1 - v^k\|_2^2 + \frac{\rho}{2}(x_1 - x^k)^\top (D - A_1^\top A_1)(x_1 - x^k) \right\}.$$

  这种方法也称为优化转移

- 通过选取合适的 $D$，当计算 $\arg\min_{x_1} \left\{ f_1(x_1) + \frac{\rho}{2} x_1^\top D x_1 \right\}$ 明显比计算 $\arg\min_{x_1} \{ f_1(x_1) + \frac{\rho}{2} x_1^\top A_1^\top A_1 x_1 \}$ 要容易时，优化转移简化子问题的计算

- 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.
- 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

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

- 常见的选择为 $\mu = 10, \gamma_p = \gamma_d = 2$

# 多块问题的 ADMM

- **考虑有多块变量的情形**

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

- **多块 ADMM 迭代格式为**

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

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

# 应用举例: LASSO 问题

- 考虑 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

$$\begin{aligned} x^{k+1} &= \arg\min_x \left\{ \frac{1}{2}\|Ax - b\|^2 + \frac{\rho}{2}\|x - z^k + y^k/\rho\|_2^2 \right\} \\ &= (A^\top A + \rho I)^{-1}(A^\top b + \rho z^k - y^k) \end{aligned} \tag{3}$$

# 应用举例: LASSO 问题

- 交替方向乘子法迭代格式为

$$
\begin{aligned}
z^{k+1} &= \arg\min_z \left\{ \mu\|z\|_1 + \frac{\rho}{2}\|x^{k+1} - z + y^k/\rho\|^2 \right\} \\
&= \mathrm{prox}_{(\mu/\rho)\|\cdot\|_1}\left(x^{k+1} + y^k/\rho\right) \\
y^{k+1} &= y^k + \tau\rho(x^{k+1} - z^{k+1})
\end{aligned}
$$

- 在求解 $x$ 迭代时，可以使用固定的罚因子 $\rho$，缓存矩阵 $A^\top A + \rho I$ 的初始分解
- 主要运算量来自更新 $x$ 变量时求解线性方程组，复杂度为 $O(n^3)$

# 应用举例: LASSO 问题

- 考虑 LASSO 问题的对偶问题

$$\begin{array}{ll} \min & b^\top y + \frac{1}{2}\|y\|^2 \\ \text{s.t.} & \|A^\top y\|_\infty \le \mu \end{array}$$

- 引入约束 $A^\top y + z = 0$，可以得到如下等价问题

$$\begin{array}{ll} \min & \underbrace{b^\top y + \dfrac{1}{2}\|y\|^2}_{f(y)} + \underbrace{I_{\|z\|_\infty \le \mu}(z)}_{h(z)} \\ \text{s.t.} & A^\top y + z = 0 \end{array}$$

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

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

# 应用举例: LASSO 问题

- 当固定 $y, x$ 时，对 $z$ 的更新即向无穷范数球 $\{z \mid \|z\|_\infty \le \mu\}$ 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

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

- ADMM 迭代格式为

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

- 由于 $m \ll n$，求解 $y$ 更新的线性方程组需要的计算量是 $O(m^3)$

# 应用举例: 矩阵分离问题

- 考虑矩阵分离问题

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

- 引入乘子 $Y$ 作用在约束 $X + S = M$ 上，得到增广拉格朗日函数

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

# 应用举例: 矩阵分离问题

- 对于$X$ 子问题

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

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

# 应用举例: 矩阵分离问题

- 对于$S$ 子问题

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

- 交替方向乘子法的迭代格式为

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

# 应用举例: 全局一致性优化问题

- 考虑全局一致性优化问题

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

- 增广拉格朗日函数为

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

- 固定 $z^k, y_i^k$，更新 $x_i$ 的公式为

$$
x_i^{k+1} = \arg\min_x \left\{ \phi_i(x) + \frac{\rho}{2} \|x - z^k + y_i^k/\rho\|^2 \right\}
$$

# 应用举例：全局一致性优化问题

- 在一般情况下更新 $x_i$ 的表达式为

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

- 固定 $x_i^{k+1}, y_i^k$，关于 $z$ 可以直接写出显式解

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

- 交替方向乘子法迭代格式为

$$\begin{aligned}
x_i^{k+1} &= \operatorname{prox}_{\phi_i/\rho}(z^k - y_i^k/\rho),\ \ i = 1, 2, \cdots, N \\
z^{k+1} &= \frac{1}{N}\sum_{i=1}^{N}(x_i^{k+1} + y_i^k/\rho) \\
y_i^{k+1} &= {y_i}^k + \tau\rho({x_i}^{k+1} - z^{k+1}),\ \ i = 1, 2, \cdots, N
\end{aligned}$$

# *Q&A*

*Thank you!*

感谢您的聆听和反馈