6.4 交替方向乘子法

SMaLL

[1] 中国石油大学（华东）
SMaLL 课题组
small.sem.upc.edu.cn
liangxijunsd@163.com

2023

交替方向乘子法 (ADMM)

1. 对偶上升法 Dual Ascent
2. 乘子法
3. 交替方向乘子法
4. 收敛性
5. 例子

对偶问题

考虑凸函数等式约束优化问题

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array} \tag{1}$$

- 拉格朗日函数: $L(x, y) = f(x) + y^T(Ax - b)$
- 对偶函数: $g(y) = \inf_x L(x, y)$
- 对偶问题: maximize $g(y)$
- 原始

  $x^\star = \text{argmin}_x \, L(x, y^\star)$

  $y^\star = \text{argmax}_y \, g(y)$

对偶上升法

- 对于对偶问题: $y^{k+1} = y^k + \alpha^k \nabla g\left(y^k\right)$
- $\tilde{x} = \operatorname{argmin}_x L\left(x, y^k\right) \rightarrow \nabla g\left(y^k\right) = A\tilde{x} - b$
- 对偶上升

$$
\begin{aligned}
x^{k+1} &:= \operatorname{argmin}_x L\left(x, y^k\right) && // x\text{ -minimization} \\
y^{k+1} &:= y^k + \alpha^k\left(Ax^{k+1} - b\right) && // \text{ dual update}
\end{aligned} \tag{2}
$$

使用大量有力的假设

对偶分解

假设 f 是可分离的:

$$f(x)=f_1\left(x_1\right)+\cdots+f_N\left(x_N\right), x=\left(x_1, \ldots, x_N\right) \tag{3}$$

$\rightarrow$ L 在 x 中是可分离的:

$$\begin{aligned} & L(x, y)=L_1\left(x_1, y\right)+\cdots+L_N\left(x_N, y\right)-y^T b, \\ & L_i\left(x_i, y\right)=f_i\left(x_i\right)+y^T A_i x_i \end{aligned} \tag{4}$$

$\rightarrow$ 在对偶上升法中，最小化 x 分裂为 N 个最小化

$$x_i^{k+1}:=\underset{x_i}{\operatorname{argmin}} L_i\left(x_i, y^k\right) \tag{5}$$

并且可以同时进行

对偶分解

对偶分解

$$\begin{aligned} x_i^{k+1} &:= \underset{x_i}{\operatorname{argmin}} L_i\left(x_i, y^k\right), \quad i = 1, \ldots, N \\ y^{k+1} &:= y^k + \alpha^k \left(\sum_{i=1}^{N} A_i x_i^{k+1} - b\right) \end{aligned} \tag{6}$$

伴随许多假设 → 乘子法

交替方向乘子法 (ADMM)

1. 对偶上升法 Dual Ascent
2. 乘子法
3. 交替方向乘子法
4. 收敛性
5. 例子

乘子法

一种具有良好鲁棒性的对偶上升方法

- 使用增广拉格朗日函数, $\rho > 0$

$$L_\rho(x, y) = f(x) + y^T(Ax - b) + (\rho/2)\|Ax - b\|_2^2 \tag{7}$$

- 乘子法

$$\begin{aligned} x^{k+1} &:= \underset{x}{\operatorname{argmin}} L_\rho\left(x, y^k\right) \\ y^{k+1} &:= y^k + \rho\left(Ax^{k+1} - b\right) \end{aligned} \tag{8}$$

二次惩罚破坏了更新后 x 的，因此无法对其进行分解.

交替方向乘子法 (ADMM)

交替方向乘子法

- 对偶上升法: 支持对偶分解
- 乘子法: 良好的鲁棒性

$\rightarrow$ 交替方向乘子法 (ADMM)

交替方向乘子法

ADMM 算法主要用于求解下列形式的优化问题 (其中 f, g 是凸函数)

$$\begin{array}{ll} \text{minimize} & f(x)+g(z) \\ \text{subject to} & Ax+Bz=c \end{array} \tag{9}$$

- 其中 $x \in \mathbf{R}^n$ 且 $z \in \mathbf{R}^m$
- $A \in \mathbf{R}^{p\times n}$, $B \in \mathbf{R}^{p\times m}$, 且 $c \in \mathbf{R}^p$

交替方向乘子法

考虑一般线性等式约束凸优化问题

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \end{array} \tag{10}$$

- $x \in \mathbf{R}^n$
- $A \in \mathbf{R}^{m \times n}$
- $f : \mathbf{R}^n \to \mathbf{R}$ 是凸的

交替方向乘子法

变量 $x \rightarrow$ 被分成了两个部分, x 和 z

$$\begin{array}{ll} \text{minimize} & f(x) + g(z) \\ \text{subject to} & Ax + Bz = c \end{array} \tag{11}$$

最优解可以由下式表示:

$$p^{\star} = \inf\{f(x) + g(z) \mid Ax + Bz = c\}. \tag{12}$$

它的增广拉格朗日函数如下

$$L_{\rho}(x, z, y) = f(x) + g(z) + y^{T}(Ax + Bz - c) + (\rho/2)\|Ax + Bz - c\|_2^2 \tag{13}$$

交替方向乘子法

如果我们联合最小化两个变量 x,z, 归约为乘子法

$$\begin{aligned}\left(x^{k+1}, z^{k+1}\right) &:= \underset{x,z}{\operatorname{argmin}} L_{\rho}\left(x, z, y^{k}\right) \\ y^{k+1} &:= y^{k} + \rho\left(Ax^{k+1} + Bz^{k+1} - c\right).\end{aligned} \tag{14}$$

其中第一步关于两个原始变量联合最小化增广拉格朗日函数

在 ADMM 中, x 的更新步骤与 z 的更新步骤次序将会**调换**.

交替方向乘子法

ADMM 由迭代组成

$$
\begin{aligned}
x^{k+1} &:= \underset{x}{\operatorname{argmin}} L_\rho\left(x, z^k, y^k\right), \quad // \; x\text{ -minimization step} \\
z^{k+1} &:= \underset{z}{\operatorname{argmin}} L_\rho\left(x^{k+1}, z, y^k\right), \quad // \; z\text{ -minimization step} \\
y^{k+1} &:= y^k + \rho\left(Ax^{k+1} + Bz^{k+1} - c\right), \quad // \text{ dual variable update}
\end{aligned} \tag{15}
$$

其中 $\rho > 0$. ADMM 类似于对偶上升法和乘子法.

标度变换形式

- 定义残差 $r = Ax + Bz - c$,

$$\begin{aligned} y^T r + (\rho/2)\|r\|_2^2 &= (\rho/2)\|r + (1/\rho)y\|_2^2 - (1/2\rho)\|y\|_2^2 \\ &= (\rho/2)\|r + u\|_2^2 - (\rho/2)\|u\|_2^2, \end{aligned} \tag{16}$$

  其中 $u = (1/\rho)y$ 为标度变换.

- ADMM (标度对偶形式):

$$\begin{aligned} x^{k+1} &:= \operatorname*{argmin}_x \left(f(x) + (\rho/2) \|Ax + Bz^k - c + u^k\|_2^2 \right) \\ z^{k+1} &:= \operatorname*{argmin}_z \left(g(z) + (\rho/2) \|Ax^{k+1} + Bz - c + u^k\|_2^2 \right) \\ u^{k+1} &:= u^k + Ax^{k+1} + Bz^{k+1} - c \end{aligned} \tag{17}$$

交替方向乘子法 (ADMM)

收敛性

假设 1.. 扩展实值函数 $f: \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ 以及 $g: \mathbf{R}^m \to \mathbf{R} \cup \{+\infty\}$ 是闭的、正常的、凸的.

函数 f 满足假设 1 $\leftrightarrow$

$$\mathbf{epi} f = \{(x, t) \in \mathbf{R}^n \times \mathbf{R} \mid f(x) \leq t\} \tag{18}$$

$\Rightarrow$

- 意味着 x 的更新和 z 的更新中产生的子问题是可解的.
- 允许 f 和 g 不可微，并且函数值可取为 $+\infty$

收敛性

Assumption. [假设 2.] 拉格朗日函数 L_0 有一个鞍点.

具体地说，存在 (可能不是唯一的) 点对 $(x^\star, z^\star, y^\star)$, 对所有的 x, z, y 下式成立

$$L_0\left(x^\star, z^\star, y\right) \leq L_0\left(x^\star, z^\star, y^\star\right) \leq L_0\left(x, z, y^\star\right)$$

连续性

在假设 1 和假设 2 下，ADMM 算法具有以下性质:

- 残差收敛. $r^k \to 0$, $k \to \infty$,
  即迭代点列逼近可行解.
- 目标函数收敛. $f(x^k) + g(z^k) \to p^\star$ as $k \to \infty$,
  即迭代的目标函数逼近最优值.
- 对偶变量收敛. $y^k \to y^\star$,$k \to \infty$,
  其中 $y^\star$ 是对偶最优解.

交替方向乘子法 (ADMM)

凸优化约束

考虑 ADMM 对于一般问题

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{C} \end{array} \tag{19}$$

ADMM 的迭代格式: 以 g 作为 $\mathcal{C}$ 的指标

$$\begin{array}{ll} \text{minimize} & f(x) + g(z) \\ \text{subject to} & x - z = 0 \end{array} \tag{20}$$

算法:

$$\begin{aligned} x^{k+1} &:= \underset{x}{\operatorname{argmin}} \left(f(x) + (\rho/2) \left\| x - z^k + u^k \right\|_2^2 \right) \\ z^{k+1} &:= \Pi_{\mathcal{C}} \left(x^{k+1} + u^k \right) \\ u^{k+1} &:= u^k + x^{k+1} - z^{k+1} \end{aligned} \tag{21}$$

Lasso

Lasso 问题

$$\text{minimize } (1/2)\|Ax-b\|_2^2+\lambda\|x\|_1 \tag{22}$$

ADMM 的迭代格式:

$$\begin{array}{ll}\text{minimize} & f(x)+g(z)\\ \text{subject to} & x-z=0\end{array} \tag{23}$$

- $f(x)=(1/2)\|Ax-b\|_2^2$
- $g(z)=\lambda\|z\|_1$

ADMM:

$$\begin{aligned}x^{k+1} &:= \left(A^TA+\rho I\right)^{-1}\left(A^Tb+\rho\left(z^k-u^k\right)\right)\\ z^{k+1} &:= S_{\lambda/\rho}\left(x^{k+1}+u^k\right)\\ u^{k+1} &:= u^k+x^{k+1}-z^{k+1}.\end{aligned} \tag{24}$$

Bi-convex 问题

另一个使用精确 ADMM 更新的重要问题是双凸规划 (bi-convex),

$$\begin{array}{ll} \text{minimize} & F(x, z) \\ \text{subject to} & G(x, z) = 0 \end{array} \tag{25}$$

- $F : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ 是 bi-convex 的

当 F 在 x 和 z 中是可分离的，G 关于 x 和 z 中是共同仿射时,
$\Rightarrow$ 模型就简化为标准的 ADMM 可求解的模型的形式.

Bi-convex 问题

ADMM 的迭代格式:

$$
\begin{aligned}
x^{k+1} &:= \underset{x}{\operatorname{argmin}} \left(F(x, z^k) + (\rho/2) \left\| G(x, z^k) + u^k \right\|_2^2 \right) \\
z^{k+1} &:= \operatorname{argmin}_z \left(F(x^{k+1}, z) + (\rho/2) \left\| G(x^{k+1}, z) + u^k \right\|_2^2 \right) \\
u^{k+1} &:= u^k + G(x^{k+1}, z^{k+1})
\end{aligned}
\tag{26}
$$

由于 x 和 z 的更新仅涉及到凸优化问题，因此易于处理.

当 $G = 0$ 时, $\Rightarrow$ ADMM 退化为交替极小化算法，这也是 Bi-convex 极小化的标准方法.

非负矩阵分解问题

考虑如下非负矩阵分解问题:

$$\begin{array}{ll} \text{minimize} & (1/2)\|VW - C\|_F^2 \\ \text{subject to} & V_{ij} \geq 0, \quad W_{ij} \geq 0, \end{array} \tag{27}$$

其中 $V \in \mathbf{R}^{p\times r}$, $W \in \mathbf{R}^{r\times q}$, 且数据 $C \in \mathbf{R}^{p\times q}$

- 目标函数是双仿射的
- 但没有等式约束

$\Rightarrow$ 因此 ADMM 成为非负矩阵分解的标准方法，交替地在

- 最小化 V, 在 W 固定时
- 最小化 W, 在 V 固定时

非负矩阵分解问题

我们还可以引入一个新变量，将目标中的双线性项移到约束中:

$$
\begin{array}{ll}
\text{minimize} & (1/2)\|X - C\|_F^2 + I_+(V) + I_+(W) \\
\text{subject to} & X - VW = 0
\end{array} \tag{28}
$$

其中变量为 X, V, W,

I_+ 是元素非负矩阵的指示函数.

非负矩阵分解问题

因为 (X, V) 代替的是 x, W 代替 z ,

$$
\begin{aligned}
\left(X^{k+1}, V^{k+1}\right) &:= \underset{X, V \geq 0}{\operatorname{argmin}} \left(\|X - C\|_F^2 + (\rho/2) \left\|X - VW^k + U^k\right\|_F^2 \right) \\
W^{k+1} &:= \underset{W \geq 0}{\operatorname{argmin}} \left\|X^{k+1} - V^{k+1} W + U^k\right\|_F^2 \\
U^{k+1} &:= U^k + X^{k+1} - V^{k+1} W^{k+1}
\end{aligned} \tag{29}
$$

Step 1 分离 X 和 V, 其中 X 和 V 可以通过求解二次规划得到,

Step 2 计算 W , 可以通过并行的求解二次规划来计算矩阵 W 的每一列.

习题:

运用 ADMM 求解 lasso 问题

参考文献和进一步阅读

Extensions and/or analyses:

- Stephen P. Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein: Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Found. Trends Mach. Learn. 3(1): 1-122 (2011)

参考文献和进一步阅读

Helpful lecture notes/books:

- Ryan Tibshirani. Convex Optimization. `http://www.stat.cmu.edu/~ryantibs/convexopt/`
- E. Candes, Lecture notes for Math 301, Stanford University, Winter 2010 – 2011