ADMM算法的线性收敛性

1301110046 吴昌晶
1301110049 何　迪

2014年12月3日

On the Linear Convergence of the Alternating Direction Method of Multipliers

问题

目标函数：可分非光滑凸函数；

限制条件：线性等式约束；

$$\begin{aligned} \min \quad & f(x) = f_1(x_1) + f_2(x_2) + \cdots + f_K(x_K) \\ s.t. \quad & Ex = E_1x_1 + E_2x_2 + \cdots + E_Kx_K = q \\ & x_k \in X_k, \quad k = 1, 2, \ldots, K \end{aligned}$$

对于不等式约束$Ex \geq q$，可以引入一个变量$x_{K+1} \geq 0$，改写成等式约束

$$Ex - x_{K+1} = q$$

并且在目标函数中增加$f_{K+1}(x_{K+1}) = i_{R_+^m}(x_{K+1})$
其中

$$i_{R_+^m}(x_{K+1}) = \begin{cases} 0, & x_{K+1} \geq 0; \\ \infty, & otherwise. \end{cases}$$

增广拉格朗日函数:

$$L(x;y) = f(x) + \langle y, q - Ex \rangle + \frac{\rho}{2}||q - Ex||^2$$

增广对偶函数:

$$d(y) = \min_x f(x) + \langle y, q - Ex \rangle + \frac{\rho}{2}||q - Ex||^2$$

对偶问题:

$$\max_y = d(y)$$

ADMM算法

用Gauss-Seidel迭代更新$\{x_k\}$，用dual ascent方法更新$\{y_k\}$，得到ADMM算法：

$$x_k^{r+1} = \arg\min_{x_k \in X_k} L(x_1^{r+1}, \ldots, x_{k-1}^{r+1}, x_k, x_{k+1}^r, \ldots, x_K^r; y^r),$$

$$y^{r+1} = y^r + \alpha(q - Ex^{r+1}) = y^r + \alpha\left(q - \sum_{k=1}^{K} E_k x_k^{r+1}\right),$$

收敛性

- 结论：在一定条件下，$\{x^r\}, \{y^r\}, \{f(x^r)\}$线性收敛至$x^{opt}, y^{opt}, p^*$.

- 线性收敛：称$\{x^r\}$线性收敛至x^∞, 当且仅当$\left||x^r - x^\infty|\right| \le c\mu^r, \forall r$,其中$0 < \mu < 1$.
- 线性收敛的误差呈指数阶下降。例： 对f优化，当f强凸时，梯度下降法也是线性收敛的（见课程讲义）。
- 数列的线性收敛等价于其一阶差分列线性收敛至0。

$$\left||x^{r+1} - x^r|\right| \le \left||x^{r+1} - x^\infty|\right| + \left||x^r - x^\infty|\right| \le 2c\mu^r$$

$$\left||x^r - x^\infty|\right| \le \sum_{k=r}^{\infty} \left||x^{k+1} - x^k|\right| \le \sum_{k=r}^{\infty} c\mu^k = \frac{c}{1-\mu}\mu^r$$

基本假设

(a)原始问题和对偶问题的全局极值可达，可行域非空。

(b)$f = f_1(x_1) + \cdots + f_K(x_K), f_k$可分解为

$$f_k(x_k) = g_k(A_k x_k) + h_k(x_k)$$

其中g_k, h_k都是凸的连续函数。

(c) g_k强凸且导数Lipschitz连续。

(d) h_k的epigraph是polyhedron，或者h_k是x_k的1-模与2-模的组合。

(e)限制条件$Ex = q$中每个E_k列满秩。

(f)可行域X为polyhedron，且**X**紧或迭代序列$\{(\boldsymbol{x}_r, \boldsymbol{y}_r)\}$有界。

证明思路

- 设$\Delta_d^r = d^* - d(y^r), \Delta_p^r = L(x^{r+1}; y^r) - d(y^r)$.这里$d(y^r)$是第r步迭代时对偶函数值（并未解出）。由$d^* = \max_y d(y)$及$d(y^r) = \min_x L(x; y^r)$可见$\Delta_d^r, \Delta_p^r \geq 0$.
- 往证$\{\Delta_d^r + \Delta_p^r\}$线性收敛，且$\{x^r\}$的一阶差分列可以被$\{\Delta_d^r + \Delta_p^r\}$的一阶差分列控制。$\{y^r\}$与$\{f(x^r)\}$亦可类似地被控制住。

- 记$X(y^r) = \{x | x = \arg\min_x L(x; y^r)\}, \bar{x}^r = \arg\min_{x \in X(y^r)} ||x - x^r||$.
- 引理1：$\Delta_d^r - \Delta_d^{r-1} \le -\alpha(Ex^r - q)^T(E\bar{x}^r - q)$
- 证：

$$\begin{aligned}
\Delta_d^r - \Delta_d^{r-1} &= [d^* - d(y^r)] - [d^* - d(y^{r-1})] \\
&= d(y^{r-1}) - d(y^r) \\
&= L(\bar{x}^{r-1}; y^{r-1}) - L(\bar{x}^r; y^r) \\
&= [L(\bar{x}^r; y^{r-1}) - L(\bar{x}^r; y^r)] + [L(\bar{x}^{r-1}; y^{r-1}) - L(\bar{x}^r; y^{r-1})] \\
&= (y^{r-1} - y^r)^T(q - E\bar{x}^r) + [L(\bar{x}^{r-1}; y^{r-1}) - L(\bar{x}^r; y^{r-1})] \\
&= -\alpha(Ex^r - q)^T(E\bar{x}^r - q) + [L(\bar{x}^{r-1}; y^{r-1}) - L(\bar{x}^r; y^{r-1})] \\
&\le -\alpha(Ex^r - q)^T(E\bar{x}^r - q), \quad \forall\, r \ge 1,
\end{aligned}$$

- 引理2：$\Delta_p^r - \Delta_p^{r-1} \le \alpha\left||Ex^r - q|\right|^2 - \gamma\left||x^{r+1} - x^r|\right|^2 - \alpha(Ex^r - q)^T(E\bar{x}^r - q)$.
- 证：$\Delta_p^r - \Delta_p^{r-1} = \left(L(x^{r+1};y^r) - d(y^r)\right) - \left(L(x^r;y^{r-1}) - d(y^{r-1})\right)$

$$= L(x^{r+1};y^r) - L(x^r;y^{r-1}) + \Delta_d^r - \Delta_d^{r-1}$$

$$= L(x^{r+1};y^r) - L(x^r;y^r) + L(x^r;y^r) - L(x^r;y^{r-1}) + \Delta_d^r - \Delta_d^{r-1}$$

而$L(x^r;y^{r-1}) = f(x^r) + \langle y^{r-1}, q - Ex^r\rangle + \frac{\rho}{2}\left||Ex^r - q|\right|^2$,

$L(x^r;y^r) = f(x^r) + \langle y^{r-1} + \alpha(q - Ex^r), q - Ex^r\rangle + \dfrac{\rho}{2}\left||Ex^r - q|\right|^2$.

故$L(x^r;y^r) = L(x^r;y^{r-1}) + \alpha\left||Ex^r - q|\right|^2$.

只需再证第一项$L(x^{r+1};y^r) - L(x^r;y^r) \le -\gamma\left||x^{r+1} - x^r|\right|^2$.

- 回顾增广拉格朗日函数：

$$L(x;y) = \sum_{k=1}^{K}\left(f_k(x_k) + \langle y_k, q_k - E_k x_k\rangle\right) + \frac{\rho}{2}\left\|\sum_{k=1}^{K} E_k x_k - q\right\|^2$$

- E_k列满秩保证其关于x_k强凸，从而有：

$$L(x_k^r; y^r) \geq L\left(x_k^{r+1}; y^r\right) + \rho\lambda_{min}\left(E_k^T E_k\right)\left\|x_k^r - x_k^{r+1}\right\|^2$$

因此取$\gamma = \rho \min_k \lambda_{min}\left(E_k^T E_k\right)$,即有

$L(x^r; y^r) \geq L(x^{r+1}; y^r) + \gamma\left\|x^{r+1} - x^r\right\|^2$,即

$L(x^{r+1}; y^r) - L(x^r; y^r) \leq -\gamma\left\|x^{r+1} - x^r\right\|^2$.

- 结合引理1、2，有

$$\Delta_d^r + \Delta_p^r - \Delta_d^{r-1} - \Delta_p^{r-1} \le \alpha\left|\left|Ex^r - q\right|\right|^2 - \gamma\left|\left|x^{r+1} - x^r\right|\right|^2 - 2\alpha(Ex^r - q)^T(E\bar{x}^r - q)$$

$$= \alpha\left|\left|Ex^r - E\bar{x}^r\right|\right|^2 - \alpha\left|\left|E\bar{x}^r - q\right|\right|^2 - \gamma\left|\left|x^{r+1} - x^r\right|\right|^2$$

- 为把$x^r - \bar{x}^r$转化为$x^{r+1} - x^r$的形式，还需要如下引理：

引理3：在基本假设下，存在$\tau > 0$使得对**所有可行的**$\boldsymbol{x}$ 均有

$$dist\big(x, X(y)\big) \le \tau\left|\left|\tilde{\nabla}_x L(x; y)\right|\right|,$$

其中$\tilde{\nabla}$是proximal gradient算子，$\tilde{\nabla} f(x) := x - \mathrm{prox}_h(x - \nabla(f(x) - h(x)))$

引理4：只要g_k导数Lipschitz连续，那么存在$\sigma > 0$使得

$$\left|\left|\tilde{\nabla} L(x^r; y^r)\right|\right| \le \sigma\left|\left|x^{r+1} - x^r\right|\right|.$$

- 因此，$\Delta_d^r + \Delta_p^r - \Delta_d^{r-1} - \Delta_p^{r-1} \le \alpha\left||Ex^r - E\bar{x}^r|\right|^2 - \alpha\left||E\bar{x}^r - q|\right|^2 - \gamma\left||x^{r+1} - x^r|\right|^2$
  $\le \left(\alpha\left||E|\right|^2\tau^2\sigma^2 - \gamma\right)\left||x^{r+1} - x^r|\right|^2 - \alpha\left||E\bar{x}^r - q|\right|^2$

只要步长α充分小即可保证$\{\Delta_d^r + \Delta_p^r\}$单调下降。

为了证明收敛是线性，只要说明右端$\le -c(\Delta_d^r + \Delta_p^r), c > 0$, 从而

$$(1 + c)\left(\Delta_d^r + \Delta_p^r\right) \le \Delta_d^{r-1} + \Delta_p^{r-1}$$

引理5：在基本假设下，对任意$\delta > 0$, 存在$\tau > 0$,使得$\left||Ex(y^r) - q|\right| < \delta$时, $\Delta_d^r \le \tau\left||Ex(y^r) - q|\right|^2$.

又存在$\zeta > 0$, 对任意r均有$\Delta_p^r \le \left||x^{r+1} - x^r|\right|^2$.

- 综上，$\{\Delta_d^r + \Delta_p^r\}$线性收敛。

- $\Delta_d^r + \Delta_p^r - \Delta_d^{r-1} - \Delta_p^{r-1} \le -c_1\left\|x^{r+1} - x^r\right\|^2 - c_2\left\|E\bar{x}^r - q\right\|^2,$

故$c_1\left\|x^{r+1} - x^r\right\|^2 \le \Delta_d^{r-1} + \Delta_p^{r-1} - \Delta_d^r - \Delta_p^r$

从而$\{x^{r+1} - x^r\}$线性收敛。

- $f(x^{r+1}) - d^* = [f(x^{r+1}) - d(y^r)] + [d(y^r) - d^*]$

$= \Delta_p^r - \Delta_d^r - \langle y^r, q - Ex^{r+1}\rangle - \frac{\rho}{2}\left\|Ex^{r+1} - q\right\|^2$

由上式$\left\{\left\|E\bar{x}^r - q\right\|\right\}$也线性收敛，结合引理3、4的结论

$\left\|x^r - \bar{x}^r\right\| \le \tau\sigma\left\|x^{r+1} - x^r\right\|$,知$\left\{\left\|Ex^r - q\right\|\right\}$也线性收敛。再由基本假设中$\{y^r\}$有界可得$\{f(x^r)\}$线性收敛至$p^* = d^*$.

- $\{y^r\}$的收敛也是类似的（可以证明等价于$\left\{\left\|Ex^r - q\right\|\right\}$的收敛）。

Proximal ADMM

在计算子问题的时候将目标函数光滑部分在x_k^r处线性化，然后加入proximal term $\frac{\beta}{2}||x_k - x_k^r||$，就得到如下公式：

$$\begin{aligned} x_k^{r+1} = \arg\min_{x_k} \Big\{ & h_k(x_k) + \langle y^r, q - E_k x_k \rangle + \langle A_k^T \nabla g_k(A_k x_k^r), x_k - x_k^r \rangle \\ & + \frac{\beta}{2}||x_k - x_k^r||^2 + \langle \rho E_k^T \left(\sum_{j<k} E_j x_j^{r+1} + \sum_{j \geq k} E_j x_j^r - q \right), x_k - x_k^r \rangle \Big\} \end{aligned}$$

在E_k不是列满秩的情况下，仍然可以证明Proximal ADMM的线性收敛性。

Jacobi 迭代

用Jacobi迭代来更新$\{x_k\}$

$$x_k^{r+1} = \arg\min_{x_k} \left(h_k(x_k) + g_k(A_k x_k) - \langle y^r, E_k x_k \rangle + \frac{\rho}{2} ||E_k x_k + \sum_{j \neq k} E_j x_j^r - q||^2 \right)$$

然而Jacobi迭代更新之后增广拉格朗日函数值不一定下降，所以作进一步改动

$$w_k^{r+1} = \arg\min_{x_k} \left(h_k(x_k) + g_k(A_k x_k) - \langle y^r, E_k x_k \rangle + \frac{\rho}{2} ||E_k x_k + \sum_{j \neq k} E_j x_j^r - q||^2 \right)$$

$$x_k^{r+1} = x_k^r + \frac{1}{K}(w_k^{r+1} - x_k^r)$$

这样我们就可以得到Jacobi迭代ADMM的线性收敛性。