

变分不等式框架下结构型 凸优化的分裂收缩算法

IV. 线性约束凸优化问题分裂收缩算法的统一框架

中学的数理基础 必要的社会实践
普通的大学数学 一般的优化原理

何炳生 南京大学数学系

Homepage: maths.nju.edu.cn/~hebma

天元数学东北中心 2023年10月17 – 27日

1 凸优化分裂收缩算法的统一框架

我们总是用变分不等式 (VI) 指导算法设计, 把线性约束的凸优化问题归结为下面的变分不等式:

$$w^* \in \Omega, \quad \theta(w) - \theta(w^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega. \quad (1.1)$$

Algorithms in a unified framework

A unified Algorithmic Framework for (1.1)

统一框架由预测-校正两部分组成

[Prediction Step.] 从给定的 v^k 出发, 求得预测点 $\tilde{w}^k \in \Omega$ 使其满足

$$\theta(w) - \theta(\tilde{w}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \quad (1.2a)$$

其中 Q 不一定对称, 但是 $Q^T + Q$ 正定.

[Correction Step.] 给一个合适的非奇异矩阵 M , 由下式确定新的迭代点

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \quad (1.2b)$$

Q 和 M 分别叫做预测矩阵和校正矩阵

Convergence Conditions

For the matrices Q and M , there is a positive definite matrix H such that

$$HM = Q. \quad (1.3a)$$

In addition,

$$G = Q^T + Q - M^T H M \succ 0. \quad (1.3b)$$

其实, 只要预测 (1.2a) 中的预测矩阵 Q 满足

$$Q^T + Q \succ 0,$$

我们总可以取

$$0 \prec G \prec Q^T + Q.$$

然后记

$$D = (Q^T + Q) - G,$$

则 $D \succ 0$. 令

$$M^T H M = D.$$

由矩阵方程组解得

$$\begin{cases} HM = Q, \\ M^T H M = D. \end{cases} \Leftrightarrow \begin{cases} HM = Q, \\ Q^T M = D. \end{cases} \Leftrightarrow \begin{cases} H = QD^{-1}Q^T, \\ M = Q^{-T}D. \end{cases}$$

就得到满足收敛条件的校正矩阵 M .

实际计算中, 我们只要校正矩阵 M .

H 和 G 只是用来验证收敛条件的.

换句话说, 只要

$$Q^T + Q \succ 0.$$

我们就可以选两个正定矩阵 $D \succ 0$ 和 $G \succ 0$, 使得

$$D + G = Q^T + Q.$$

将(1.2b)中的校正矩阵 M 取成

$$M = Q^{-T} D$$

条件(1.3)自然满足.

2 预测-校正方法的例子

We consider the min – max problem

$$\min_x \max_y \{ \Phi(x, y) = \theta_1(x) - y^T A x - \theta_2(y) \mid x \in \mathcal{X}, y \in \mathcal{Y} \}. \quad (2.4)$$

Let (x^*, y^*) be the solution of (2.4), then we have

根据鞍点的定义

$$(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}, \quad \Phi(x^*, y) \leq \Phi(x^*, y^*) \leq \Phi(x, y^*), \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

上面的两个不等式可以写成等价的

$$\begin{cases} x^* \in \mathcal{X}, & \Phi(x, y^*) - \Phi(x^*, y^*) \geq 0, \quad \forall x \in \mathcal{X}, \end{cases} \quad (2.5a)$$

$$\begin{cases} y^* \in \mathcal{Y}, & \Phi(x^*, y^*) - \Phi(x^*, y) \geq 0, \quad \forall y \in \mathcal{Y}. \end{cases} \quad (2.5b)$$

Using the notation of $\Phi(x, y)$, it can be written as

只要把 $\Phi(x, y)$ 的形式填进去

$$\begin{cases} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^T (-A^T y^*) \geq 0, \quad \forall x \in \mathcal{X}, (*) \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) + (y - y^*)^T (A x^*) \geq 0, \quad \forall y \in \mathcal{Y}. (\diamond) \end{cases}$$

Furthermore, it can be written as a variational inequality in the compact form:

$$u \in \Omega, \quad \theta(u) - \theta(u^*) + (u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega, \quad (2.6)$$

where

对上式中任意的 $u \in \Omega$ 分别取 $u = (x, y^*)$ 和 $u = (x^*, y)$, 就得到 (*) 和 (\diamond).

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad F(u) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}, \quad \Omega = \mathcal{X} \times \mathcal{Y}.$$

The output of Original PDHG algorithm [16] as predictor

For given (x^k, y^k) , PDHG [16] produces a pair of $(\tilde{x}^k, \tilde{y}^k)$. First,

$$\tilde{x}^k = \operatorname{argmin}\left\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\right\}, \quad (2.7a)$$

and then we obtain \tilde{y}^k via

$$\tilde{y}^k = \operatorname{argmax}\left\{\Phi(\tilde{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\right\}. \quad (2.7b)$$

Ignoring the constant term in the objective function, the subproblems (2.7) are reduced to

$$\left\{ \begin{array}{l} \tilde{x}^k = \operatorname{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \\ \tilde{y}^k = \operatorname{argmin}\{\theta_2(y) + y^T A \tilde{x}^k + \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. \end{array} \right. \quad (2.8a)$$

$$\left\{ \begin{array}{l} \tilde{x}^k = \operatorname{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \\ \tilde{y}^k = \operatorname{argmin}\{\theta_2(y) + y^T A \tilde{x}^k + \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. \end{array} \right. \quad (2.8b)$$

According to the basic lemma, the optimality condition of (2.8a) is $\tilde{x}^k \in \mathcal{X}$ and

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T y^k + r(\tilde{x}^k - x^k)\} \geq 0, \quad \forall x \in \mathcal{X}. \quad (2.9)$$

Similarly, from (2.8b) we get $\tilde{y}^k \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{A \tilde{x}^k + s(\tilde{y}^k - y^k)\} \geq 0, \quad \forall y \in \mathcal{Y}. \quad (2.10)$$

Combining (2.9) and (2.10), we have

$$\begin{aligned} \tilde{u}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T \tilde{y}^k \\ A \tilde{x}^k \end{pmatrix} \right. \\ \left. + \begin{pmatrix} r(\tilde{x}^k - x^k) + A^T (\tilde{y}^k - y^k) \\ s(\tilde{y}^k - y^k) \end{pmatrix} \right\} \geq 0, \quad \forall (x, y) \in \Omega. \end{aligned}$$

The compact form is $\tilde{u}^k \in \Omega$,

$$\theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T \{F(\tilde{u}^k) + Q(\tilde{u}^k - u^k)\} \geq 0, \quad \forall u \in \Omega, \quad (2.11a)$$

where

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}. \quad (2.11b)$$

对于这样的预测, 我们考虑比较简单的校正

$$u^{k+1} = u^k - M(u^k - \tilde{u}^k) \quad (2.12)$$

校正. 其中 M 为单位上三角矩阵或单位下三角矩阵. 收敛性条件 (1.3)

- $H \succ 0$ and $HM = Q$.
- $G = Q^T + Q - M^T H M \succ 0$.

可以改写成等价的

- (i) $H \succ 0$ and $H = QM^{-1}$.
- (ii) $G = Q^T + Q - Q^T M \succ 0$.

一. 校正矩阵 M 为单位下三角矩阵

其中的 K 是待定的.

$$M = \begin{pmatrix} I_n & 0 \\ K & I_m \end{pmatrix} \quad \text{则} \quad M^{-1} = \begin{pmatrix} I_n & 0 \\ -K & I_m \end{pmatrix}.$$

对条件 (i), 我们在统一框架下指导下求出这个 K 的具体形式. 由于 $H = QM^{-1}$ 正定, 首先必须是对称的. 由

$$H = QM^{-1} = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -K & I_m \end{pmatrix} = \begin{pmatrix} rI_n - A^T K & A^T \\ -sK & sI_m \end{pmatrix}$$

必须对称, 推得

$$-sK = A, \quad \Rightarrow \quad K = -\frac{1}{s}A.$$

因此,

$$M = \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix}, \quad H = \begin{pmatrix} rI_n + \frac{1}{s}A^T A & A^T \\ A & sI_m \end{pmatrix}.$$

对任意的 $r, s > 0$, 矩阵 H 是正定的.

对条件 (ii),

$$\begin{aligned}
 G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\
 &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ A & sI_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix} \\
 &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix} = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}.
 \end{aligned}$$

要矩阵 G 正定, 必须有 $rs > \|A^T A\|$.

采用 PDHG 预测, 单位下三角矩阵校正, 需要 $rs > \|A^T A\|$.

二. 校正矩阵 M 为单位上三角矩阵

同样, 其中的 K 是待定的.

$$M = \begin{pmatrix} I_n & K \\ 0 & I_m \end{pmatrix} \quad \text{则} \quad M^{-1} = \begin{pmatrix} I_n & -K \\ 0 & I_m \end{pmatrix}.$$

对条件 (i), 我们在统一框架下指导下求出这个 K 的具体形式. 由于 $H = QM^{-1}$ 正定,

首先必须是对称的. 由

$$H = QM^{-1} = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & -K \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} rI_n & -rK + A^T \\ 0 & sI_m \end{pmatrix}$$

必须对称, 推得

$$rK = A^T, \quad \Rightarrow \quad K = \frac{1}{r}A^T.$$

因此,

$$M = \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix}, \quad H = \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix}.$$

对任意的 $r, s > 0$, 矩阵 H 是正定的.

而对条件 (ii),

$$\begin{aligned}
 G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\
 &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ A & sI_m \end{pmatrix} \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} \\
 &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \\
 &= \begin{pmatrix} rI_n & 0 \\ 0 & sI_m - \frac{1}{r}AA^T \end{pmatrix}.
 \end{aligned}$$

要矩阵 G 正定, 必须有 $rs > \|A^T A\|$.

采用 PDHG 预测, 单位上三角矩阵校正, 需要 $rs > \|A^T A\|$.

虽然把不能保证收敛的 PDHG 方法改造成了收敛的方法, 但是, rs 的值没有降下来.

我们的目标, 是把预测 (2.8) 中的参数 rs 想办法降下来.

对于 (2.11) 中的 Q , 我们有

$$Q^T + Q = \begin{pmatrix} 2rI & A^T \\ A & 2sI \end{pmatrix}$$

只要 $rs > \frac{1}{4} \|A^T A\|$, 矩阵 $Q^T + Q$ 都是正定的.

当 $(Q^T + Q)$ 正定时, 我们取

$$D = \frac{1}{2}(Q^T + Q), \quad \text{并令} \quad M^T H M = D. \quad (2.13)$$

这样就能保证

$$G = Q^T + Q - M^T H M = \frac{1}{2}(Q^T + Q) \succ 0.$$

- $H \succ 0$ and $HM = Q$.

- $G = Q^T + Q - M^T H M \succ 0$.

可以改写成

- (i) $HM = Q$.

- (ii) $M^T H M = D$.

$$\begin{cases} HM = Q, \\ M^T H M = D. \end{cases} \Leftrightarrow \begin{cases} HM = Q, \\ Q^T M = D. \end{cases} \Leftrightarrow \begin{cases} H = Q D^{-1} Q^T, \\ M = Q^{-T} D. \end{cases} \quad (2.14)$$

换句话说, 当 $(Q^T + Q) \succ 0$, 取

$$D = \begin{pmatrix} rI & \frac{1}{2}A^T \\ \frac{1}{2}A & sI \end{pmatrix}, \quad M = Q^{-T} D$$

所有收敛性条件都满足. 而

$$\begin{aligned}
 Q^{-T} &= \begin{pmatrix} rI & 0 \\ A & sI \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{r}I & 0 \\ -\frac{1}{rs}A & \frac{1}{s}I \end{pmatrix} \\
 M &= Q^{-T}D = \begin{pmatrix} \frac{1}{r}I & 0 \\ -\frac{1}{rs}A & \frac{1}{s}I \end{pmatrix} \begin{pmatrix} rI & \frac{1}{2}A^T \\ \frac{1}{2}A & sI \end{pmatrix} \\
 &= \begin{pmatrix} I & \frac{1}{2r}A^T \\ -\frac{1}{2s}A & I - \frac{1}{2rs}AA^T \end{pmatrix} \tag{2.15}
 \end{aligned}$$

利用上面的校正矩阵 M

$$\begin{cases} x^{k+1} &= \tilde{x}^k - \frac{1}{2r}A^T(y^k - \tilde{y}^k) \\ y^{k+1} &= \tilde{y}^k + \frac{1}{2s}A[(x^k - \tilde{x}^k) + \frac{1}{r}A^T(y^k - \tilde{y}^k)]. \end{cases}$$

这是马峰他们 [14] 根据统一框架提出的方法. 计算效果有很大进步.

把 rs 的积降了 $\frac{3}{4}$, 有了很大进步.

3 Convergence proof in the unified framework

In this section, assuming the conditions (1.3) in the unified framework are satisfied, we prove some convergence properties.

定理 1 *Let $\{v^k\}$ be the sequence generated by a method for the problem (1.1) and \tilde{w}^k is obtained in the k -th iteration. If v^k, v^{k+1} and \tilde{w}^k satisfy the conditions in the unified framework, then we have*

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ & \geq \frac{1}{2} (\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2} \|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega. \end{aligned} \quad (3.1)$$

Proof. Using $Q = HM$ (see (1.3a)) and the relation (1.2b), the right hand side of (1.3a) can be written as $(v - \tilde{v}^k)^T H(v^k - v^{k+1})$ and hence

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \quad \forall w \in \Omega. \quad (3.2)$$

Applying the identity

$$Q(v^k - \tilde{v}^k) = HM(v^k - \tilde{v}^k) = H(v^k - v^{k+1}).$$

$$(a - b)^T H(c - d) = \frac{1}{2} \{\|a - d\|_H^2 - \|a - c\|_H^2\} + \frac{1}{2} \{\|c - b\|_H^2 - \|d - b\|_H^2\},$$

to the right hand side of (3.2) with

$$a = v, \quad b = \tilde{v}^k, \quad c = v^k, \quad \text{and} \quad d = v^{k+1},$$

we thus obtain

$$\begin{aligned} & 2(v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ &= (\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + (\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \end{aligned} \quad (3.3)$$

For the last term of (3.3), using $HM = Q$ and $2v^T Qv = v^T (Q^T + Q)v$, we have

$$\begin{aligned} & \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ &= \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ &\stackrel{(1.3a)}{=} \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ &= 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ &= (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\ &\stackrel{(1.3b)}{=} \|v^k - \tilde{v}^k\|_G^2. \end{aligned} \quad (3.4)$$

Substituting (3.3), (3.4) in (3.2), the assertion of this theorem is proved. \square

FIRST-ORDER METHODS IN OPTIMIZATION

Amir Beck

MOS-SIAM Series on Optimization

© A. Beck 参考了我们用到的“积化和差”的公式,并在前一页的脚注做了说明

We will use the following notation:

$$\begin{aligned} \tilde{\mathbf{x}}^k &= \mathbf{x}^{k+1}, \\ \tilde{\mathbf{z}}^k &= \mathbf{z}^{k+1}, \\ \tilde{\mathbf{y}}^k &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^k - \mathbf{c}). \end{aligned}$$

Using (15.15), (15.16), the subgradient inequality, and the above notation, we obtain that for any $\mathbf{x} \in \text{dom}(h_1)$ and $\mathbf{z} \in \text{dom}(h_2)$,

$$\begin{aligned} h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \left\langle \rho\mathbf{A}^T \left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right) + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \right\rangle &\geq 0, \\ h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \left\langle \rho\mathbf{B}^T \left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right) + \mathbf{Q}(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \right\rangle &\geq 0. \end{aligned}$$

Using the definition of $\tilde{\mathbf{y}}^k$, the above two inequalities can be rewritten as

$$\begin{aligned} h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \langle \mathbf{A}^T \tilde{\mathbf{y}}^k + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \rangle &\geq 0, \\ h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \langle \mathbf{B}^T \tilde{\mathbf{y}}^k + (\rho\mathbf{B}^T \mathbf{B} + \mathbf{Q})(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \rangle &\geq 0. \end{aligned}$$

Adding the above two inequalities and using the identity

$$\mathbf{y}^{k+1} - \mathbf{y}^k = \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c}),$$

we can conclude that for any $\mathbf{x} \in \text{dom}(h_1)$, $\mathbf{z} \in \text{dom}(h_2)$ and $\mathbf{y} \in \mathbb{R}^m$

$$H(\mathbf{x}, \mathbf{z}) - H(\tilde{\mathbf{x}}^k, \tilde{\mathbf{z}}^k) + \left\langle \begin{pmatrix} \mathbf{x} - \tilde{\mathbf{x}}^k \\ \mathbf{z} - \tilde{\mathbf{z}}^k \\ \mathbf{y} - \tilde{\mathbf{y}}^k \end{pmatrix}, \begin{pmatrix} \mathbf{A}^T \tilde{\mathbf{y}}^k \\ \mathbf{B}^T \tilde{\mathbf{y}}^k \\ -\mathbf{A}\tilde{\mathbf{x}}^k - \mathbf{B}\tilde{\mathbf{z}}^k + \mathbf{c} \end{pmatrix} - \begin{pmatrix} \mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) \\ \mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) \\ \frac{1}{\rho}(\mathbf{y}^k - \mathbf{y}^{k+1}) \end{pmatrix} \right\rangle \geq 0, \tag{15.17}$$

where $\mathbf{C} = \rho\mathbf{B}^T \mathbf{B} + \mathbf{Q}$. We will use the following identity that holds for any positive semidefinite matrix \mathbf{P} :

$$(\mathbf{a} - \mathbf{b})^T \mathbf{P}(\mathbf{c} - \mathbf{d}) = \frac{1}{2} (\|\mathbf{a} - \mathbf{d}\|_{\mathbf{P}}^2 - \|\mathbf{a} - \mathbf{c}\|_{\mathbf{P}}^2 + \|\mathbf{b} - \mathbf{c}\|_{\mathbf{P}}^2 - \|\mathbf{b} - \mathbf{d}\|_{\mathbf{P}}^2).$$

Using the above identity, we can conclude that

$$\begin{aligned} (\mathbf{x} - \tilde{\mathbf{x}}^k)^T \mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) &= \frac{1}{2} (\|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2 + \|\tilde{\mathbf{x}}^k - \mathbf{x}^k\|_{\mathbf{G}}^2) \\ &\geq \frac{1}{2} \|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \frac{1}{2} \|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2, \end{aligned} \tag{15.18}$$

as well as

$$(\mathbf{z} - \tilde{\mathbf{z}}^k)^T \mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) = \frac{1}{2} \|\mathbf{z} - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 - \frac{1}{2} \|\mathbf{z} - \mathbf{z}^k\|_{\mathbf{C}}^2 + \frac{1}{2} \|\mathbf{z}^k - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 \tag{15.19}$$

and

$$\begin{aligned} 2(\mathbf{y} - \tilde{\mathbf{y}}^k)^T (\mathbf{y}^k - \mathbf{y}^{k+1}) &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \|\tilde{\mathbf{y}}^k - \mathbf{y}^k\|^2 - \|\tilde{\mathbf{y}}^k - \mathbf{y}^{k+1}\|^2 \\ &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2 \|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2 \\ &\quad - \|\mathbf{y}^k + \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}) - \mathbf{y}^k - \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c})\|^2 \\ &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2 \|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2 - \rho^2 \|\mathbf{B}(\mathbf{z}^k - \tilde{\mathbf{z}}^k)\|^2. \end{aligned}$$

3.1 Convergence in a strictly contraction sense

定理 2 Let $\{v^k\}$ be the sequence generated by a method for the problem (1.1) and \tilde{w}^k is obtained in the k -th iteration. If v^k, v^{k+1} and \tilde{w}^k satisfy the conditions in the unified framework, then we have

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \quad (3.5)$$

Proof. Setting $w = w^*$ in (3.1), we get

$$\begin{aligned} & \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ & \geq \|v^k - \tilde{v}^k\|_G^2 + 2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \end{aligned} \quad (3.6)$$

By using the optimality of w^* and the monotonicity of $F(w)$, we have

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \geq 0$$

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \|v^k - \tilde{v}^k\|_G^2. \quad (3.7)$$

The assertion (3.5) follows directly. \square

定理 1 中的结论 (3.1)

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ & \geq \frac{1}{2} (\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2} \|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega. \end{aligned}$$

是为收敛速率的证明而准备的.

否则, 我们可以通过在 (3.2)

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \quad \forall w \in \Omega.$$

中令 $w = w^*$, 得到

$$(v^k - v^{k+1})^T H(\tilde{v}^k - v^*) \geq 0. \quad (3.8)$$

将恒等式

$$(a - b)^T H(c - d) = \frac{1}{2} \{ \|a - d\|_H^2 - \|b - d\|_H^2 \} - \frac{1}{2} \{ \|a - c\|_H^2 - \|b - c\|_H^2 \}$$

用于 (3.8) 的左端, 令 $a = v^k$, $b = v^{k+1}$, $c = \tilde{v}^k$ 和 $d = v^*$, 我们得到

$$\begin{aligned} & (v^k - v^{k+1})^T H(\tilde{v}^k - v^*) \\ & = \frac{1}{2} \{ \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \} - \frac{1}{2} \{ \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \}. \end{aligned}$$

根据 (3.8) 就有

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2. \quad (3.9)$$

再把上式的右端化简一下,

$$\begin{aligned} & \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ &= \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ &\stackrel{(1.2b)}{=} \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ &= 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ &= (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\ &\stackrel{(1.3b)}{=} \|v^k - \tilde{v}^k\|_G^2. \end{aligned} \quad (3.10)$$

将 (3.10) 代入 (3.9) 就得到引理的结论. \square

3.2 Convergence rate (两篇主要理论文章)

Convergence rate in an ergodic sense [10]

为了证明算法遍历意义下的迭代复杂性, 我们需要对变分不等式 (1.1) 的解集做新的刻

画. 由于 (1.1) 中的仿射算子 F 恰有

$$(w - w^*)^T F(w^*) = (w - w^*)^T F(w),$$

变分不等式问题

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega,$$

和

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w) \geq 0, \quad \forall w \in \Omega$$

是等价的. 我们用后者定义变分不等式 (1.1) 的近似解. 对给定的 $\epsilon > 0$, 如果 \tilde{w} 满足

$$\tilde{w} \in \Omega, \quad \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq -\epsilon, \quad \forall w \in \mathcal{D}(\tilde{w}), \quad (3.11a)$$

其中

$$\mathcal{D}(\tilde{w}) = \{w \in \Omega \mid \|w - \tilde{w}\| \leq 1\}, \quad (3.11b)$$

就叫做变分不等式 (1.1) 的 ϵ 近似解. 它可以等价地表示成

$$\tilde{w} \in \Omega, \quad \sup_{w \in \mathcal{D}(\tilde{w})} \{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\} \leq \epsilon. \quad (3.12)$$

人们感兴趣的是: 对给定的 $\epsilon > 0$, 经过多少次迭代得到一个 $\tilde{w} \in \Omega$, 使得 (3.12) 成立.

这就是我们要讨论的遍历意义下的收敛速率. 讨论遍历意义下的收敛性, 对 (1.3) 中的矩阵 H 和 G , 只要求它半正定.

Theorem 1 is also the base for the convergence rate proof. Using the monotonicity of F , we have

$$(w - \tilde{w}^k)^T F(w) = (w - \tilde{w}^k)^T F(\tilde{w}^k).$$

Substituting it in (3.1), we obtain

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2} \|v - v^k\|_H^2 \geq \frac{1}{2} \|v - v^{k+1}\|_H^2, \quad \forall w \in \Omega. \quad (3.13)$$

Note that the above assertion is hold for $G \succeq 0$.

定理 3 *Let $\{v^k\}$ be the sequence generated by a method for the problem (1.1) and \tilde{w}^k is obtained in the k -th iteration. Assume that v^k , v^{k+1} and \tilde{w}^k satisfy the conditions in the unified framework and let \tilde{w}_t be defined by*

$$\tilde{w}_t = \frac{1}{t+1} \sum_{k=0}^t \tilde{w}^k. \quad (3.14)$$

Then, for any integer number $t > 0$, $\tilde{w}_t \in \Omega$ and

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)} \|v - v^0\|_H^2, \quad \forall w \in \Omega. \quad (3.15)$$

Proof. First, it holds that $\tilde{w}^k \in \Omega$ for all $k \geq 0$. Together with the convexity of Ω , (3.14) implies that $\tilde{w}_t \in \Omega$. Rewriting the inequality (3.13) in its equivalent form

$$\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^T F(w) + \frac{1}{2} \|v - v^{k+1}\|_H^2 \leq \frac{1}{2} \|v - v^k\|_H^2, \quad \forall w \in \Omega.$$

Summing the last inequality over $k = 0, 1, \dots, t$, we obtain

$$\sum_{k=0}^t \theta(\tilde{u}^k) - (t+1)\theta(u) + \left(\sum_{k=0}^t \tilde{w}^k - (t+1)w \right)^T F(w) \leq \frac{1}{2} \|v - v^0\|_H^2, \quad \forall w \in \Omega.$$

Use the notation of \tilde{w}_t , it can be written as

$$\frac{1}{t+1} \sum_{k=0}^t \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2\alpha(t+1)} \|v - v^0\|_H^2, \quad \forall w \in \Omega. \quad (3.16)$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1} \sum_{k=0}^t \tilde{u}^k,$$

we have

$$\theta(\tilde{u}_t) \leq \frac{1}{t+1} \sum_{k=0}^t \theta(\tilde{u}^k).$$

Substituting it in (3.16), the assertion of this theorem follows directly. \square

Recall (3.12). The conclusion (3.15) thus indicates obviously that the method is able to generate an approximate solution (i.e., \tilde{w}_t) with the accuracy $O(1/t)$ after t iterations. That is, in the case $G \succeq 0$, the convergence rate $O(1/t)$ of the method is established.

我们2012年发表在SIAM Numerical Analysis的论文[10]就是用这种方式证明了交替方向法在遍历意义下 $O(1/t)$ 的收敛速率. 这被认为我们在交替方向法方面的一个比较重要的结果, 这里只是说明, 该结果对符合统一框架收敛条件的方法都是成立的.

Convergence rate in a pointwise iteration-complexity [12]

我们2015年发表在Numerische Mathematik的论文[12]证明了交替方向法在点列意义下一个比较重要的结果.

$$\|v^{k+1} - v^{k+2}\|_H \leq \|v^k - v^{k+1}\|_H.$$

这个性质已经被一些学者用来研发加速ADMM. 下面证明这个结果对符合统一框架收敛条件的方法也都成立. 下面的证明也只需要矩阵 H 和 G 半正定.

定理 4 For solving the variational inequality (1.1), let $\{w^k\}$, $\{\tilde{w}^k\}$ be the sequence generated by (1.2). If the conditions (1.3) are satisfied, then we have

$$\|v^{k+1} - v^{k+2}\|_H^2 \leq \|v^k - v^{k+1}\|_H^2 - \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2. \quad (3.17)$$

Proof Note that we have

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega$$

and

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (v - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}), \quad \forall w \in \Omega.$$

Set the vector w in the above two inequalities by \tilde{w}^{k+1} and \tilde{w}^k , respectively, we get

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k)$$

and

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}).$$

Adding the above two inequalities, it follows that

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq 0.$$

Adding $\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}^T Q \{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}$ to the both sides of the last inequality, we get

$$(v^k - v^{k+1})^T Q \{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \frac{1}{2} \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T + Q)}^2,$$

and thus

$$(v^k - v^{k+1})^T H \{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} \geq \frac{1}{2} \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T + Q)}^2. \quad (3.18)$$

Finally, by using $\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a - b) - \|a - b\|_H^2$ and (3.18), we get

$$\begin{aligned} & \|v^k - v^{k+1}\|_H^2 - \|v^{k+1} - v^{k+2}\|_H^2 \\ &= 2(v^k - v^{k+1})^T H \{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} \\ &\quad - \|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|_H^2 \\ &\geq \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T + Q)}^2 - \|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|_H^2 \\ &= \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T + Q - M^T H M)}^2 \\ &= \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2. \end{aligned}$$

This is the equivalent form of (3.17) and the proof is complete. \square

4 ADMM for problems with two separable blocks

This section concern the structured convex optimization problem namely,

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\}. \quad (4.1)$$

The Lagrangian function and the augmented Lagrange Function of (4.1) are

$$L^{[2]}(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T (Ax + By - b).$$

and

$$\mathcal{L}_\beta^{[2]}(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T (Ax + By - b) + \frac{\beta}{2} \|Ax + By - b\|^2, \quad (4.2)$$

respectively. Recall the model (4.1) can be explained as the VI

$$w^* \in \Omega, \quad \theta(w) - \theta(w^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega. \quad (4.3a)$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad (4.3b)$$

$$F(w) = \begin{pmatrix} -A^T \lambda \\ -B^T \lambda \\ Ax + By - b \end{pmatrix}, \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \mathfrak{R}^m. \quad (4.3c)$$

Using the augmented Lagrange function, the recursion of the alternating direction method of multipliers for the structured convex optimization (4.1) can be written as

$$\begin{cases} x^{k+1} \in \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} \in \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{cases} \quad (4.4)$$

Note that the essential variable of ADMM (4.4) is $v = (y, \lambda)$.

统一框架下的 ADMM.

ADMM scheme (4.4) is also a special case which belongs to the unified algorithmic framework (1.2) and the Convergence Condition is satisfied.

In order to cast the ADMM scheme (4.4) into a special case of (1.2), let us first define the artificial vector $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ by

$$\tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1} \quad \text{and} \quad \tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b), \quad (4.5)$$

where (x^{k+1}, y^{k+1}) is generated by the ADMM (4.4).

我们注意到 A. Beck 在他的专著 First-Order Methods in convex optimization [1], 也采用了这种转换.

Prediction

$$\begin{cases} \tilde{x}^k \in \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2} \|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, \\ \tilde{y}^k \in \text{Argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2} \|A\tilde{x}^k + By - b\|^2 \mid y \in \mathcal{Y}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b). \end{cases} \quad (4.6)$$

According to the scheme (4.4), the defined artificial vector \tilde{w}^k satisfies the following VI:
 $\tilde{w}^k \in \Omega,$

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T \tilde{\lambda}^k\} \geq 0, & \forall x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{-B^T \tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k)\} \geq 0, & \forall y \in \mathcal{Y}, \\ (A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

This can be written in form of (1.2a) as described in the following lemma.

FIRST-ORDER METHODS IN OPTIMIZATION

© A. Beck 参考了我们 (4.5) 中对 $\tilde{\omega}^k$ 的定义, 作者在前一页的脚注做了说明

Amir Beck

MOS-SIAM Series on Optimization

We will use the following notation:

$$\begin{aligned}\tilde{\mathbf{x}}^k &= \mathbf{x}^{k+1}, \\ \tilde{\mathbf{z}}^k &= \mathbf{z}^{k+1}, \\ \tilde{\mathbf{y}}^k &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^k - \mathbf{c}).\end{aligned}$$

Using (15.15), (15.16), the subgradient inequality, and the above notation, we obtain that for any $\mathbf{x} \in \text{dom}(h_1)$ and $\mathbf{z} \in \text{dom}(h_2)$,

$$\begin{aligned}h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \left\langle \rho \mathbf{A}^T \left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho} \mathbf{y}^k \right) + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \right\rangle &\geq 0, \\ h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \left\langle \rho \mathbf{B}^T \left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c} + \frac{1}{\rho} \mathbf{y}^k \right) + \mathbf{Q}(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \right\rangle &\geq 0.\end{aligned}$$

Using the definition of $\tilde{\mathbf{y}}^k$, the above two inequalities can be rewritten as

$$\begin{aligned}h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \langle \mathbf{A}^T \tilde{\mathbf{y}}^k + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \rangle &\geq 0, \\ h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \langle \mathbf{B}^T \tilde{\mathbf{y}}^k + (\rho \mathbf{B}^T \mathbf{B} + \mathbf{Q})(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \rangle &\geq 0.\end{aligned}$$

Adding the above two inequalities and using the identity

$$\mathbf{y}^{k+1} - \mathbf{y}^k = \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c}),$$

we can conclude that for any $\mathbf{x} \in \text{dom}(h_1)$, $\mathbf{z} \in \text{dom}(h_2)$ and $\mathbf{y} \in \mathbb{R}^m$

$$H(\mathbf{x}, \mathbf{z}) - H(\tilde{\mathbf{x}}^k, \tilde{\mathbf{z}}^k) + \left\langle \begin{pmatrix} \mathbf{x} - \tilde{\mathbf{x}}^k \\ \mathbf{z} - \tilde{\mathbf{z}}^k \\ \mathbf{y} - \tilde{\mathbf{y}}^k \end{pmatrix}, \begin{pmatrix} \mathbf{A}^T \tilde{\mathbf{y}}^k \\ \mathbf{B}^T \tilde{\mathbf{y}}^k \\ -\mathbf{A}\tilde{\mathbf{x}}^k - \mathbf{B}\tilde{\mathbf{z}}^k + \mathbf{c} \end{pmatrix} - \begin{pmatrix} \mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) \\ \mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) \\ \frac{1}{\rho}(\mathbf{y}^k - \mathbf{y}^{k+1}) \end{pmatrix} \right\rangle \geq 0, \quad (15.17)$$

where $\mathbf{C} = \rho \mathbf{B}^T \mathbf{B} + \mathbf{Q}$. We will use the following identity that holds for any positive semidefinite matrix \mathbf{P} :

$$(\mathbf{a} - \mathbf{b})^T \mathbf{P}(\mathbf{c} - \mathbf{d}) = \frac{1}{2} (\|\mathbf{a} - \mathbf{d}\|_{\mathbf{P}}^2 - \|\mathbf{a} - \mathbf{c}\|_{\mathbf{P}}^2 + \|\mathbf{b} - \mathbf{c}\|_{\mathbf{P}}^2 - \|\mathbf{b} - \mathbf{d}\|_{\mathbf{P}}^2).$$

Using the above identity, we can conclude that

$$\begin{aligned}(\mathbf{x} - \tilde{\mathbf{x}}^k)^T \mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) &= \frac{1}{2} (\|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2 + \|\tilde{\mathbf{x}}^k - \mathbf{x}^k\|_{\mathbf{G}}^2) \\ &\geq \frac{1}{2} \|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \frac{1}{2} \|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2,\end{aligned} \quad (15.18)$$

as well as

$$(\mathbf{z} - \tilde{\mathbf{z}}^k)^T \mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) = \frac{1}{2} \|\mathbf{z} - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 - \frac{1}{2} \|\mathbf{z} - \mathbf{z}^k\|_{\mathbf{C}}^2 + \frac{1}{2} \|\mathbf{z}^k - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 \quad (15.19)$$

and

$$\begin{aligned}2(\mathbf{y} - \tilde{\mathbf{y}}^k)^T (\mathbf{y}^k - \mathbf{y}^{k+1}) &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \|\tilde{\mathbf{y}}^k - \mathbf{y}^k\|^2 - \|\tilde{\mathbf{y}}^k - \mathbf{y}^{k+1}\|^2 \\ &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2 \|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2 \\ &\quad - \|\mathbf{y}^k + \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}) - \mathbf{y}^k - \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c})\|^2 \\ &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2 \|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2 - \rho^2 \|\mathbf{B}(\mathbf{z}^k - \tilde{\mathbf{z}}^k)\|^2.\end{aligned}$$

引理 1 For given v^k , let w^{k+1} be generated by (4.4) and \tilde{w}^k be defined by (4.5). Then, we have

$$\tilde{w}^k \in \Omega, \theta(w) - \theta(\tilde{w}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \forall w \in \Omega,$$

where

$$Q = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix}. \quad (4.7)$$

Recall the essential variable of the ADMM scheme (4.4) is (y, λ) . Moreover, using the definition of \tilde{w}^k , the λ^{k+1} updated by (4.4) can be represented as

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b) \\ &= \lambda^k - [-\beta B(y^k - \tilde{y}^k) + \beta(A\tilde{x}^k + By^k - b)] \\ &= \lambda^k - [-\beta B(y^k - \tilde{y}^k) + (\lambda^k - \tilde{\lambda}^k)]. \end{aligned}$$

Therefore, the ADMM scheme (4.4) can be written as

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \quad (4.8a)$$

which corresponds to the step (1.2b) with

$$M = \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \quad \text{and} \quad \alpha = 1. \quad (4.8b)$$

验证收敛性条件.

Now we check that the Convergence Condition is satisfied by the ADMM scheme (4.4). Indeed, for the matrix M in (4.8b), we have

$$M^{-1} = \begin{pmatrix} I & 0 \\ \beta B & I \end{pmatrix}.$$

Thus, by using (4.7) and (4.8b), we obtain

验证 H 半正定

$$H = QM^{-1} = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ \beta B & I \end{pmatrix} = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}, \quad (4.9)$$

and consequently

验证 G 的半正定

$$\begin{aligned}
 G &= Q^T + Q - \alpha M^T H M = Q^T + Q - Q^T M \\
 &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} \beta B^T B & -B^T \\ 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \\
 &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \quad (4.10)
 \end{aligned}$$

Therefore, H is symmetric and positive definite under the assumption that B is full column rank; and G is positive semi-definite. The Convergence Condition is satisfied; and thus the convergence of the ADMM scheme (4.4) is guaranteed.

见论文 [10]

收缩性和点列意义下的收敛速率.

我们将经典的 ADMM 按统一框架故意解释成预测-校正方法. 经过 (4.6) 预测以后, 再由

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k) \quad (4.11)$$

校正. 这符合统一框架的模式 (1.2). 在 (4.9) 和 (4.10) 中我们分别验证了矩阵 H 和 G 是

半正定的. 因此, 根据定理 4 就有

$$\|v^{k+1} - v^{k+2}\|_H \leq \|v^k - v^{k+1}\|_H, \quad \forall k > 0. \quad (4.12)$$

由 (4.12), 对任意的正整数 $t > 0$,

$$\begin{aligned} \|v^t - v^{t+1}\|_H^2 &\leq \frac{1}{t+1} \sum_{k=0}^t \|v^k - v^{k+1}\|_H^2 \\ &\leq \frac{1}{t+1} \sum_{k=0}^{\infty} \|v^k - v^{k+1}\|_H^2 \\ &\stackrel{(4.12)}{\leq} \frac{1}{t+1} \|v^0 - v^*\|_H^2. \end{aligned}$$

人们往往用 $\|v^t - v^{t+1}\|_H^2$ 的大小做停机准则的参考.

见论文 [12]

5 利用统一框架验证 ADMM 类算法的收敛性

5.1 交换顺序的交替方向法

将经典的 ADMM (4.4) 中求解 y -子问题和校正 λ 的顺序交换, 通过

$$\begin{cases} x^{k+1} \in \operatorname{Argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} \in \operatorname{Argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^{k+1}) \mid y \in \mathcal{Y}\}, \end{cases} \quad (5.1)$$

得到的 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ 作为预测点, 然后

$$\begin{cases} y^{k+1} := y^k - \gamma(y^k - y^{k+1}), \\ \lambda^{k+1} := \lambda^k - \gamma(\lambda^k - \lambda^{k+1}). \end{cases} \quad \text{(松弛延拓)} \quad (5.2)$$

这里 $\gamma \in (0, 2)$. 赋值号 “:=” 表示 (5.2) 右端的 (y^{k+1}, λ^{k+1}) 是由算法的前半部分 (5.1) 产生的. (5.2) 左端才是下一步迭代开始所需要的 (y^{k+1}, λ^{k+1}) . 对多数问题, 这样往往能加快收敛.

注意到, (5.1) 中核心变量还是 $v = (y, \lambda)$. 先把 (5.1) 产生的 w^{k+1} 本身定义成预测

点 \tilde{w}^k , 即

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}, \quad (5.3)$$

利用 $\mathcal{L}_\beta^{[2]}(x, y, \lambda)$ 的表达式, 交换顺序的交替方向法迭代公式 (5.1) 可以表示成

$$\begin{cases} \tilde{x}^k \in \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2} \|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, & (5.4a) \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), & (5.4b) \\ \tilde{y}^k \in \operatorname{argmin}\{\theta_2(y) - y^T B^T \tilde{\lambda}^k + \frac{\beta}{2} \|A\tilde{x}^k + By - b\|^2 \mid y \in \mathcal{Y}\}. & (5.4c) \end{cases}$$

我们用统一框架来证明交换顺序的交替方向法 (5.1)-(5.2) 的收敛性. 先给出由 (5.4) 求得的 \tilde{w}^k 满足的形如 (1.2a) 的预测公式.

首先, 根据第一讲的定理 1, (5.4a) 的最优性条件是

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T \lambda^k + \beta A^T (A\tilde{x}^k + By^k - b)\} \geq 0, \quad \forall x \in \mathcal{X}.$$

利用 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, 上式就是

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T \tilde{\lambda}^k\} \geq 0, \quad \forall x \in \mathcal{X}. \quad (5.5a)$$

类似地, 根据第一讲的定理 1, (5.4c) 的最优性条件是

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{-B^T \tilde{\lambda}^k + \beta B^T (A\tilde{x}^k + B\tilde{y}^k - b)\} \geq 0, \quad \forall y \in \mathcal{Y}.$$

由于 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b)$, 我们有

$$\begin{aligned} & -B^T \tilde{\lambda}^k + \beta B^T (A\tilde{x}^k + B\tilde{y}^k - b) \\ &= -B^T \tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) + \beta B^T (A\tilde{x}^k + B\tilde{y}^k - b) \\ &= -B^T \tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - B^T (\tilde{\lambda}^k - \lambda^k). \end{aligned}$$

因此, y -子问题 (5.4c) 的最优性条件是

$$\begin{aligned} \tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{ \underline{-B^T \tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k) \\ - B^T (\tilde{\lambda}^k - \lambda^k) \} \geq 0, \quad \forall y \in \mathcal{Y}. \quad (5.5b) \end{aligned}$$

对于 (5.4b) 中给出的 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b)$, 可以表示成

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0,$$

也就是

$$\tilde{\lambda}^k \in \mathfrak{R}^m, \quad (\lambda - \tilde{\lambda}^k)^T \left\{ \underline{(A\tilde{x}^k + B\tilde{y}^k - b)} \right. \\ \left. - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \right\} \geq 0, \quad \forall \lambda \in \mathfrak{R}^m. \quad (5.5c)$$

将(5.5a), (5.5b)和(5.5c)组合在一起, 注意到下划线部分是(4.3)中的 $F(\tilde{w}^k)$, 我们得到下面的引理.

引理 2 求解变分不等式(4.3), 对给定的 v^k , 由(5.4)提供的 \tilde{w}^k 满足

$$\tilde{w}^k \in \Omega, \quad \theta(w) - \theta(\tilde{w}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \quad \forall w \in \Omega,$$

其中

$$H = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \quad (5.6)$$

这里的矩阵 H 是半正定的, 取 $\alpha = 1$ 的平凡校正, 就是方法(5.1). 它的收敛效果与经典的ADMM(4.4)别无二致. 但是, 如果做平凡的松弛延伸, 让

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha = 1.5 \in (0, 2),$$

就相当于方法(5.1)-(5.1), 收敛速度往往会有30%的提高.

从理论上讲, (5.6)中的矩阵 H 即使在 B 列满秩的时候也是半正定的, 但这并不影响

计算和收敛性态. 当然, 我们也可以通过在子问题 (5.4c) 的目标函数中增添一项 $\frac{\delta}{2} \|y - y^k\|^2$, 就能使相应的 H 矩阵变成

$$H = \begin{pmatrix} \beta B^T B + \delta I_{n_2} & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}.$$

对任意的 $\beta, \delta > 0$, 上面的 H 矩阵是正定的.

5.2 对称的交替方向法

人们习惯于用经典的乘子交替方向法 (4.4) 求解问题 (4.1). 从问题 (4.1) 本身看, 原始变量 x 和 y 是平等的, 在算法设计上平等对待 x 和 y 子问题, 也是最自然不过的考虑. 因此我们采用对称的交替方向法 [6], 它的 k 步迭代也是从给定的 (y^k, λ^k) 开始, 通过

$$(S\text{-ADMM}) \quad \begin{cases} x^{k+1} \in \operatorname{argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & (5.7a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), & (5.7b) \\ y^{k+1} \in \operatorname{argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \mid y \in \mathcal{Y}\}, & (5.7c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b). & (5.7d) \end{cases}$$

得到新的迭代点 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$, 其中 $\mu \in (0, 1)$ (通常取 $\mu = 0.9$).

当 $\mu = 1$ 时, 方法 (5.7) 是可以举出不收敛的反例的. 我们用统一框架证明对称型乘子交替方向法 (5.7) 的收敛性, 也是把方法 (5.7) 拆解成预测-校正两部分. 对由 (5.7) 产生的 x^{k+1} 和 y^{k+1} , 我们按如下方式定义预测点 \tilde{w}^k :

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}. \quad (5.8)$$

利用 $\mathcal{L}_\beta^{[2]}(x, y, \lambda)$ 的表达式, 对称型的乘子交替方向法迭代公式可以表示成等价的

$$\left\{ \begin{array}{l} \tilde{x}^k \in \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2} \|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, \quad (5.9a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(A\tilde{x}^k + By^k - b), \quad (5.9b) \\ \tilde{y}^k \in \operatorname{argmin}\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2} \|A\tilde{x}^k + By - b\|^2 \mid y \in \mathcal{Y}\}, \quad (5.9c) \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \quad (5.9d) \end{array} \right.$$

下面我们先找出 (5.9) 给出的 \tilde{w}^k 在统一框架中形如 (1.2a) 的预测公式.

根据第一讲的定理 1, (5.9a) 的最优性条件是

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T \lambda^k + \beta A^T (A\tilde{x}^k + By^k - b)\} \geq 0, \quad \forall x \in \mathcal{X}.$$

利用 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, 上式就是

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T \tilde{\lambda}^k\} \geq 0, \quad \forall x \in \mathcal{X}. \quad (5.10a)$$

类似地, 根据第一讲的定理 1, (5.9c) 的最优性条件是

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{-B^T \lambda^{k+\frac{1}{2}} + \beta B^T (A\tilde{x}^k + B\tilde{y}^k - b)\} \geq 0, \quad \forall y \in \mathcal{Y}.$$

利用 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, 我们有

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \mu(\lambda^k - \tilde{\lambda}^k) = \tilde{\lambda}^k + (\mu - 1)(\tilde{\lambda}^k - \lambda^k),$$

和

$$\beta(A\tilde{x}^k + By^k - b) = -(\tilde{\lambda}^k - \lambda^k).$$

因此,

$$\begin{aligned} & -B^T \lambda^{k+\frac{1}{2}} + \beta B^T (A\tilde{y}^k + B\tilde{y}^k - b) \\ &= -B^T [\tilde{\lambda}^k + (\mu - 1)(\tilde{\lambda}^k - \lambda^k)] + \beta B^T B(\tilde{y}^k - y^k) + \beta B^T (A\tilde{x}^k + By^k - b) \\ &= -B^T \tilde{\lambda}^k + (1 - \mu)B^T (\tilde{\lambda}^k - \lambda^k) + \beta B^T B(\tilde{y}^k - y^k) - B^T (\tilde{\lambda}^k - \lambda^k) \\ &= -B^T \tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - \mu B^T (\tilde{\lambda}^k - \lambda^k). \end{aligned}$$

子问题 (5.9c) 的最优性条件是

$$\begin{aligned} \tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{ \underbrace{-B^T \tilde{\lambda}^k}_{\text{}} + \beta B^T B(\tilde{y}^k - y^k) \\ - \mu B^T (\tilde{\lambda}^k - \lambda^k) \} \geq 0, \quad \forall y \in \mathcal{Y}. \end{aligned} \quad (5.10b)$$

对于 (5.9d) 中定义的 $\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b)$, 由于

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0,$$

可以表示成

$$\begin{aligned} \tilde{\lambda}^k \in \mathfrak{R}^m, \quad (\lambda - \tilde{\lambda}^k)^T \{ \underbrace{(A\tilde{x}^k + B\tilde{y}^k - b)}_{\text{}} \\ - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) \} \geq 0, \quad \forall \lambda \in \mathfrak{R}^m. \end{aligned} \quad (5.10c)$$

将 (5.10a), (5.10b) 和 (5.10c) 组合在一起并利用 (4.3) 中的记号, 我们有下面的引理.

引理 3 求解变分不等式 (4.3). 对给定的 v^k , 设 \tilde{w}^k 是由 (5.9) 提供的, 则有

$$\tilde{w}^k \in \Omega, \quad \theta(w) - \theta(\tilde{w}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \quad (5.11a)$$

其中

$$Q = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \quad (5.11b)$$

接着我们要导出形如 (1.2b) 的校正关系式. 利用 (5.8), 由 (5.7d) 给出的 λ^{k+1} 可以表示成

$$\begin{aligned}
 \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \mu[-\beta B(y^k - \tilde{y}^k) + \beta(A\tilde{x}^k + By^k - b)] \\
 &= [\lambda^k - \mu(\lambda^k - \tilde{\lambda}^k)] - \mu[-\beta B(y^k - \tilde{y}^k) + \beta(Ax^{k+1} + By^k - b)] \\
 &= \lambda^k - [-\mu\beta B(y^k - \tilde{y}^k) + 2\mu(\lambda^k - \tilde{\lambda}^k)]. \tag{5.12}
 \end{aligned}$$

跟 $y^{k+1} = \tilde{y}^k$ 结合在一起, 就有

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

因此, 我们有下面的引理.

引理 4 求解变分不等式 (4.3). 对给定的 v^k , 设 w^{k+1} 由 (5.7) 提供. 那么对由 (5.8) 定义的 \tilde{w}^k , 我们有

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{5.13a}$$

其中

$$M = \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix}. \tag{5.13b}$$

我们已经把对称的 ADMM (5.7) 拆解成 (5.11) 的预测和 (5.13) 的校正. 剩下的事情就是根

据统一框架中的收敛性条件 (1.3) 验证算法的收敛性. 对 (5.13b) 中的矩阵 M , 算出

$$M^{-1} = \begin{pmatrix} I & 0 \\ \frac{1}{2}\beta B & \frac{1}{2\mu}I_m \end{pmatrix}.$$

由 $H = QM^{-1}$ 得到

$$H = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} I & 0 \\ \frac{1}{2}\beta B & \frac{1}{2\mu}I_m \end{pmatrix} = \begin{pmatrix} (1 - \frac{1}{2}\mu)\beta B^T B & -\frac{1}{2}B^T \\ -\frac{1}{2}B & \frac{1}{2\mu\beta}I_m \end{pmatrix}.$$

因此

$$H = \frac{1}{2} \begin{pmatrix} \sqrt{\beta}B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}}I \end{pmatrix} \begin{pmatrix} (2 - \mu)I & -I \\ -I & \frac{1}{\mu}I \end{pmatrix} \begin{pmatrix} \sqrt{\beta}B & 0 \\ 0 & \sqrt{\frac{1}{\beta}}I \end{pmatrix} \quad (5.14)$$

注意到

$$\begin{pmatrix} (2 - \mu) & -1 \\ -1 & \frac{1}{\mu} \end{pmatrix} = \begin{cases} \succ 0, & \mu \in (0, 1); \\ \succeq 0, & \mu = 1. \end{cases}$$

所以, 对所有的 $\mu \in (0, 1)$, 当 B 列满秩时矩阵 H 是对称正定的.

再看矩阵 $G = Q^T + Q - M^T H M$. 因为 $M^T H M = M^T Q$, 由

$$M^T Q = \begin{pmatrix} I & -\mu\beta B^T \\ 0 & 2\mu I_m \end{pmatrix} \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} = \begin{pmatrix} (1+\mu)\beta B^T B & -2\mu B^T \\ -2\mu B & 2\mu \frac{1}{\beta} I_m \end{pmatrix}.$$

得到

$$\begin{aligned} G &= (Q^T + Q) - M^T H M \\ &= \begin{pmatrix} 2\beta B^T B & -(1+\mu)B^T \\ -(1+\mu)B & 2\frac{1}{\beta} I_m \end{pmatrix} - \begin{pmatrix} (1+\mu)\beta B^T B & -2\mu B^T \\ -2\mu B & 2\mu \frac{1}{\beta} I_m \end{pmatrix} \\ &= (1-\mu) \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix}. \end{aligned} \tag{5.15}$$

同样, 对所有的 $\mu \in (0, 1)$, 当 B 列满秩时矩阵 G 正定. 所以, 根据统一框架 (1.2a)-(1.2b), 方法是收敛的, 我们有下面的定理.

定理 5 求解变分不等式 (4.3). 对给定的 v^k , 设 w^{k+1} 由 (5.7) 提供. 我们有

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*,$$

其中 \tilde{w}^k 由 (5.8) 定义,

$$H = \begin{pmatrix} (1 - \frac{1}{2}\mu)\beta B^T B & -\frac{1}{2}B^T \\ -\frac{1}{2}B & \frac{1}{2\mu\beta}I_m \end{pmatrix}$$

和

$$G = (1 - \mu) \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta}I_m \end{pmatrix}.$$

由于 $\mu \in (0, 1)$, 矩阵 H 和 G 在 B 列满秩时都是正定的.

在矩阵 B 不一定列满秩的时候, 矩阵 H 和 G 半正定. 方法都具备定理 3 和定理 4 中的相关收敛速率性质.

References

- [1] A. Beck, First-Order Methods in convex optimization, MOS-SIAM Series on Optimization, SIAM-Society for Industrial and Applied Mathematics, 2017.
- [2] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, *Foundations and Trends in Machine Learning* Vol. 3, No. 1 (2010) 1 – 122.
- [3] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A proximal point algorithms revisit on the alternating direction method of multipliers, *Science China Mathematics*, 56 (2013), 2179-2186.
- [4] E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, *IEEE Trans. Imag. Process.*, 21(7), 3239-3252, 2012.
- [5] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.
- [6] B. S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman-Rachford splitting method for convex programming, *SIAM Journal on Optimization* **24**(2014), 1011-1040.
- [7] B. S. He, M. Tao and X.M. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, *SIAM Journal on Optimization* **22**(2012), 313-340.
- [8] B.S. He, M. Tao and X.M. Yuan, A splitting method for separable convex programming, *IMA Journal of Numerical Analysis*, **31**, 394-426, 2015.
- [9] B.S. He, H. Yang, and S.L. Wang, Alternating directions method with self-adaptive penalty

- parameters for monotone variational inequalities, *JOTA* **23**(2000), 349–368.
- [10] B. S. He and X. M. Yuan, On the $O(1/t)$ convergence rate of the alternating direction method, *SIAM J. Numerical Analysis* **50**(2012), 700-709.
- [11] B.S. He and X.M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, *SIAM J. Imag. Science* **5**(2012), 119-149.
- [12] B.S. He and X.M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers, *Numerische Mathematik*, 130 (2015) 567-577.
- [13] B.S. He and X.M. Yuan, On the optimal proximal parameter of an ADMM-like splitting method for separable convex programming. *Mathematical methods in image processing and inverse problems*, 139 – 163, Springer Proc. Math. Stat., 360, Springer, Singapore, 2021, and *Optimization-Online* 6235, 2017.
- [14] Feng Ma, Yiming Bi and Bin Ga, A prediction – correction-based primal – dual hybrid gradient method for linearly constrained convex minimization, *Numerical Algorithms* (2019) 82:641 – 662.
- [15] P. Tseng, On accelerated proximal gradient methods for convex-concave optimization, manuscript, 2008.
- [16] M. Zhu and T. F. Chan, An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration, CAM Report 08-34, UCLA, Los Angeles, CA, 2008.
- [17] 何炳生. 我和乘子交替方向法 20 年. *运筹学学报*, 2018, 22: 1–31.