典型凸优化问题的分裂收缩算法讲座

IV. 线性约束凸优化问题分裂收缩算法的统一框架

何 炳 生　　南京大学数学系

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

江苏省研究生视觉计算与可信人工智能暑期学校

2022年7月13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:

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

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

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

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

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

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

$$HM = Q. \tag{1.3a}$$

In addition,

$$G = Q^T + Q - M^T HM \succ 0. \tag{1.3b}$$

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

$$M^T HM = D.$$

由矩阵方程组解得

$$\left\{ \begin{array}{r} HM = Q, \\ M^T HM = D. \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{r} HM = Q, \\ Q^T M = D. \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{l} H = QD^{-1}Q^T, \\ M = Q^{-T}D. \end{array} \right.$$

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

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

取

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

条件 (1.3) 自然满足.

2 预测-校正方法的例子

We consider the $\min - \max$ problem

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

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

根据鞍点的定义

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

上面的两个不等式可以写成等价的

$$\left\{ \begin{array}{ll} x^* \in \mathcal{X}, \quad \Phi(x, y^*) - \Phi(x^*, y^*) \ge 0, \quad \forall\, x \in \mathcal{X}, & \text{(2.5a)} \\ y^* \in \mathcal{Y}, \quad \Phi(x^*, y^*) - \Phi(x^*, y) \ge 0, \quad \forall\, y \in \mathcal{Y}. & \text{(2.5b)} \end{array} \right.$$

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

只要把 $\Phi(x,y)$的形式填进去

$$\left\{ \begin{array}{llll} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T y^*) \ge 0, & \forall\, x \in \mathcal{X}, & (*) \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) \; + \; (y - y^*)^T(\; A x^*) \; \ge 0, & \forall\, y \in \mathcal{Y}. & (\diamond) \end{array} \right.$$

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

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

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 [17] as predictor

For given (x^k, y^k), PDHG [17] produces a pair of $(\tilde{x}^k, \tilde{y}^k)$. First,

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

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

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

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

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

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

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

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

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

Combining (2.9) and (2.10), we have

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

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

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

where

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

对于这样的预测, 我们考虑比较简单的校正

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

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

可以改写成等价的

(i) $H \succ 0$ and $H = QM^{-1}$.

(ii) $G = Q^T + Q - M^T HM \succ 0$.

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

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

必须对称, 推得

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

因此,

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

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

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

因此,

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

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

当 $(Q^T + Q)$ 正定时, 我们取

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

这样就能保证

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

• $H \succ 0$ and $HM = Q$. • $G = Q^T + Q - M^THM \succ 0$.	可以改写成	(i) $HM = Q$. (ii) $M^THM = D$.

$$\left\{\begin{array}{r} HM = Q, \\ M^THM = D. \end{array}\right. \Leftrightarrow \left\{\begin{array}{r} HM = Q, \\ Q^TM = D. \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} H = QD^{-1}Q^T, \\ M = Q^{-T}D. \end{array}\right. \tag{2.14}$$

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

$$Q^{-T} = \begin{pmatrix} rI & 0 \\ A & sI \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{r}I & 0 \\ -\frac{1}{rs}A & \frac{1}{s}I \end{pmatrix}$$

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

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 [15] 根据统一框架提出的方法. 计算效果有很大进步.

把 rs 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.

Theorem 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*

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

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

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

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

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

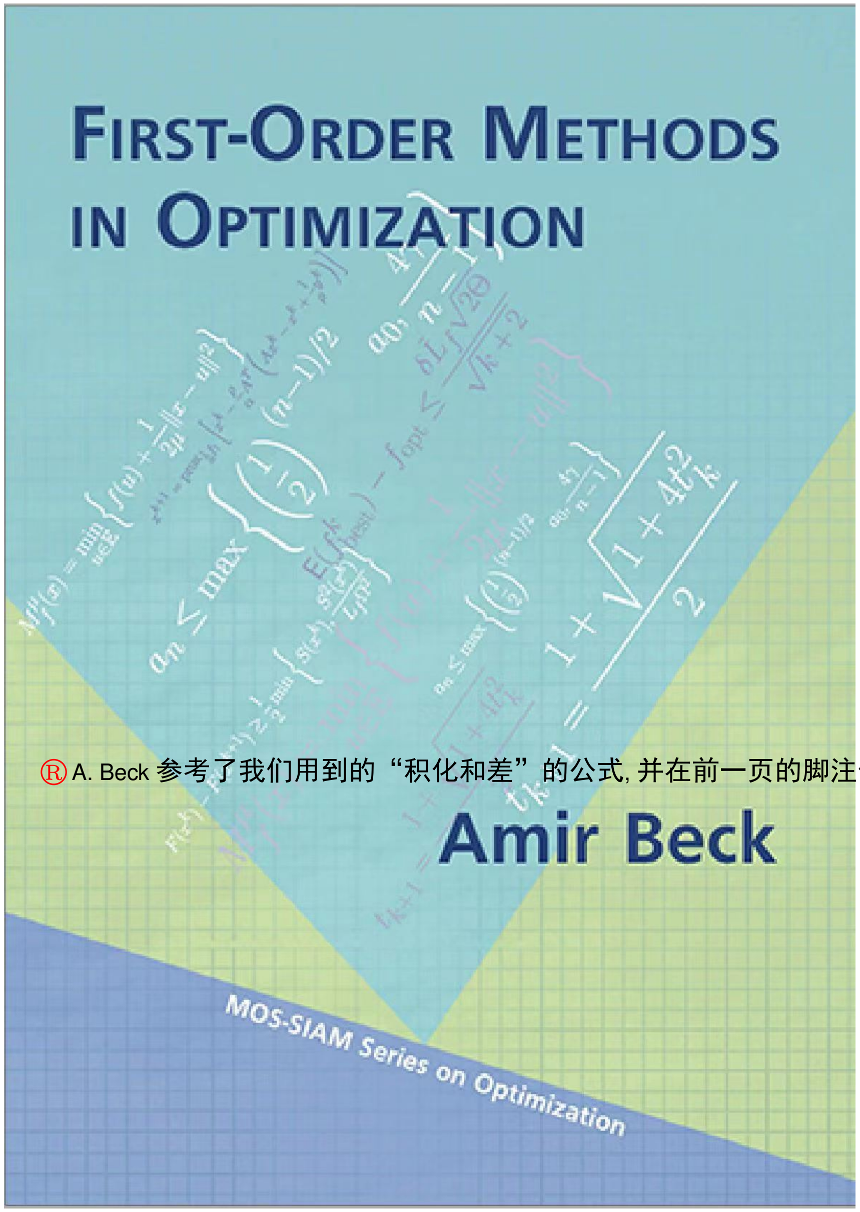
we thus obtain

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

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

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

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



Ⓡ 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

15.3. Convergence Analysis of AD-PMM 429

We will use the following notation:

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

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

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

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

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

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 $\mathbf{x} \in \mathrm{dom}(h_1)$, $\mathbf{z} \in \mathrm{dom}(h_2)$, and $\mathbf{y} \in \mathbb{R}^m$,

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

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

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

Using the above identity, we can conclude that

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

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

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

3.1 Convergence in a strictly contraction sense

Theorem 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*

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

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

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

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

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

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \alpha\|v^k - \tilde{v}^k\|_G^2. \tag{3.7}$$

The assertion (3.5) follows directly. $\square$

定理 1 中的结论 (3.1)

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

是为收敛收敛的证明而准备的.

否则, 我们可以通过在 (3.2)

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

中令 $w=w^*$, 得到

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

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

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

根据 (3.8) 就有

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

再把上式的右端化简一下,

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

将 (3.10) 代入 (3.9) 就得到引理的结论. □

3.2 Convergence rate

Convergence rate in an ergodic sense [11]

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

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

和

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

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

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

其中

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

就叫做变分不等式 (1.1) 的 ϵ 近似解. 它可以等价地表示成

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

人们感兴趣的是：对给定的 $\epsilon > 0$, 经过多少次迭代, 能够得到一个 $\tilde{w} \in \Omega$, 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 (1.3b) 中的

矩阵 G, 只要求它半正定.

Equivalent Characterization of the Solution Set of VI

Theorem 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*

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k. \tag{3.13}$$

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

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2\alpha(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{3.14}$$

Convergence rate in a pointwise iteration-complexity [13]

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

Theorem 4 *For the sequence generated by the prototype algorithm (1.2) where the Convergence Condition is satisfied, we have*

$$\|M(v^{k+1} - \tilde{v}^{k+1})\|_H \le \|M(v^k - \tilde{v}^k)\|_H, \quad \forall\, k > 0. \tag{3.15}$$

4 ADMM for problems with two separable blocks

This section concern the structured convex optimization problem namely,

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

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

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

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega. \tag{4.3a}$$

where

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

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

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

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

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

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

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.

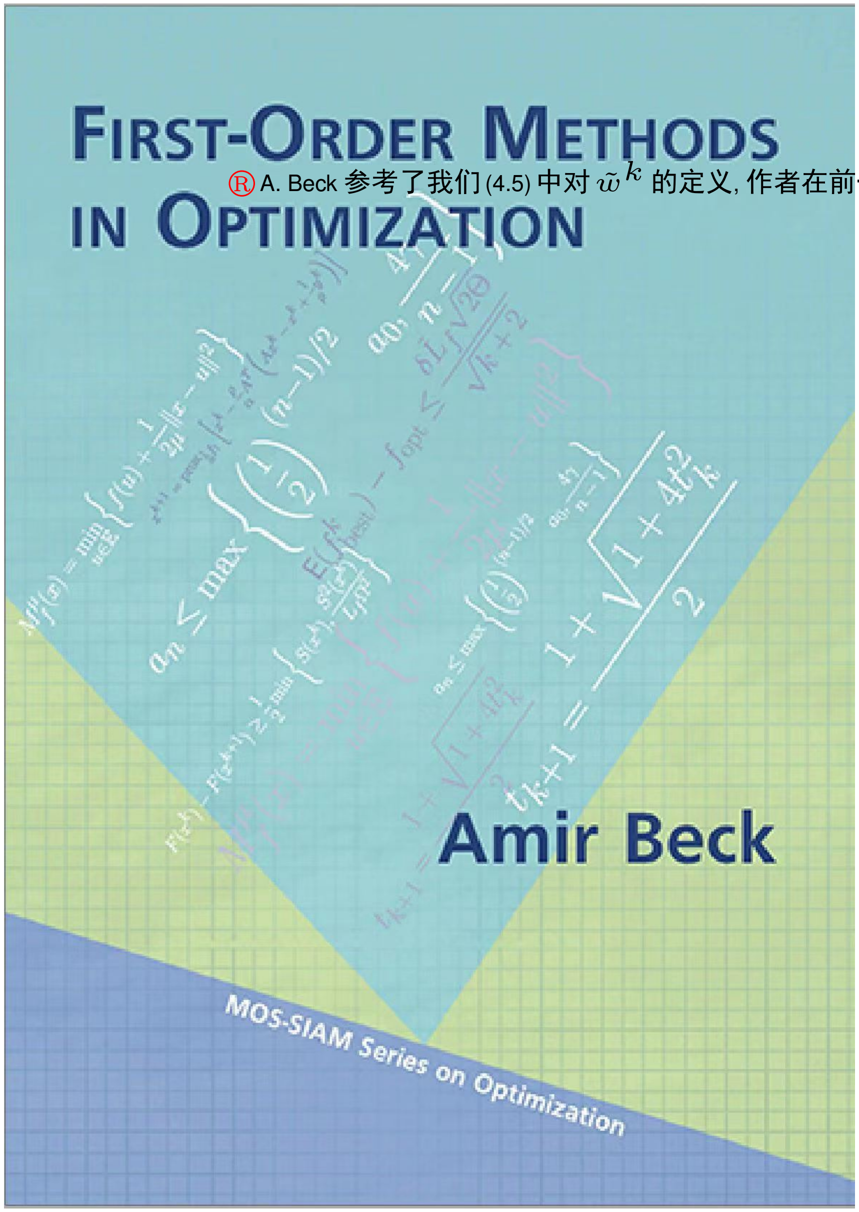
Prediction

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

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

$$\left\{\begin{array}{ll} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k\} \ge 0, & \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k)\} \ge 0, & \forall\, y \in \mathcal{Y}, \\ \qquad (A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \dfrac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. & \end{array}\right.$$

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



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

We will use the following notation:

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

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

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

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

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

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 $\mathbf{x} \in \mathrm{dom}(h_1)$, $\mathbf{z} \in \mathrm{dom}(h_2)$, 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 $\mathbf{P}$:

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

Using the above identity, we can conclude that

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

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

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

Lemma 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*

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

where

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

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

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

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

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

which corresponds to the step (1.2b) with

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

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 正定

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

and consequently

验证 G 的半正定

$$
\begin{aligned}
G &= Q^T + Q - \alpha M^T H M = Q^T + Q - Q^T M \\
&= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} \beta B^T B & -B^T \\ 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \\
&= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \qquad (4.9)
\end{aligned}
$$

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 [11]

收缩性和点列意义下的收敛速率. 我们将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

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k) \qquad (4.10)$$

校正. 上式表示

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

在 (3.15) 中我们证明了

$$\|M(v^{k+1} - \tilde{v}^{k+1})\|_H \le \|M(v^k - \tilde{v}^k)\|_H, \quad \forall\, k > 0.$$

根据 (4.11) 就是

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

由 (4.12), 对任意的正整数 $t > 0$，

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

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 [13]

5 利用统一框架设计三个可分离块的 ADMM 类算法

三个可分离块的凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz=b,\ x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}\} \tag{5.1}$$

的求解方法. 这个问题的拉格朗日函数是

$$L(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b).$$

问题 (5.1) 同样可以归结为变分不等式问题

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{5.2a}$$

其中 $\theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z),\quad \Omega=\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\times\Re^m.$

$$w=\begin{pmatrix} x\\ y\\ z\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\\ z\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ -C^T\lambda\\ Ax+By+Cz-b\end{pmatrix}. \tag{5.2b}$$

相应的增广拉格朗日函数记为(与两个算子的符号有区别)

$$\mathcal{L}_\beta^{[3]}(x,y,z,\lambda) = \theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b) + \frac{\beta}{2}\|Ax+By+Cz-b\|^2. \tag{5.3}$$

直接推广的 ADMM 求解三块可分离问题不保证收敛

对三个可分离块的凸优化问题, 采用直接推广的乘子交替方向法, 第 k 步迭代是从给定的 $v^k=(y^k,z^k,\lambda^k)$ 出发, 通过

$$\left\{\begin{array}{lcl} x^{k+1} & \in & \arg\min\left\{\mathcal{L}_\beta^{[3]}(x,y^k,z^k,\lambda^k)\;\middle|\; x\in\mathcal{X}\right\}, \\ y^{k+1} & \in & \arg\min\left\{\mathcal{L}_\beta^{[3]}(x^{k+1},y,z^k,\lambda^k)\;\middle|\; y\in\mathcal{Y}\right\}, \\ z^{k+1} & \in & \arg\min\left\{\mathcal{L}_\beta^{[3]}(x^{k+1},y^{k+1},z,\lambda^k)\;\middle|\; z\in\mathcal{Z}\right\}, \\ \lambda^{k+1} & = & \lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b), \end{array}\right. \tag{5.4}$$

求得新的迭代点 $w^{k+1}=(x^{k+1},y^{k+1},z^{k+1},\lambda^{k+1})$. 当矩阵 A, B, C 中有两个是互相正交的时候, 用方法 (5.4) 求解问题 (5.1) 是收敛的, 因为这种三块的可分离问题, 实际上相当于两块可分离的问题, 对一般的三块可分离问题, 是不能保证收敛的 [5].

在对直接推广的 ADMM (5.4) 证明不了收敛性的时候, 我们就着手对三块可分离的问题

提出一些修正算法. 修正方法的原则是尽量对ADMM少做改动, 保持它原来的好品性. 特别是对问题不加(诸如目标函数强凸等)任何额外条件, 对经典ADMM中需要调比选取的大于零的β, 仍然让它可以自由选取.

带高斯回代的ADMM方法

带高斯回代的ADMM方法[8]是2012年发表的. 同样, 在综述文章[18]中用统一框架处理就更简单. 直接推广的乘子交替方向法(5.4) 对三个算子的问题不能保证收敛, 是因为它们处理有关核心变量的 y 和 z-子问题不公平. 采取补救的办法是将(5.4)提供的 $(y^{k+1}, z^{k+1}, \lambda^{k+1})$ 当成预测点, 再进行校正. 具体地说, 先用直接推广的ADMM

$$\left\{\begin{array}{l} x^{k+1} \in \operatorname{argmin}\big\{\theta_1(x) - x^TA^T\lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \big| x \in \mathcal{X}\big\}, \\ y^{k+1} \in \operatorname{argmin}\big\{\theta_2(y) - y^TB^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 \big| y \in \mathcal{Y}\big\}, \\ z^{k+1} \in \operatorname{argmin}\big\{\theta_3(z) - z^TC^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By^{k+1} + Cz - b\|^2 \big| z \in \mathcal{Z}\big\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{array}\right.$$

计算出 $(x^{k+1}, y^{k+1}, z^{k+1}, \lambda^{k+1})$, 然后只需要再对 (y, z) 校正. 校正公式为

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \end{pmatrix} := \begin{pmatrix} By^{k} \\ Cz^{k} \end{pmatrix} - \nu \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} By^{k} - By^{k+1} \\ Cz^{k} - Cz^{k+1} \end{pmatrix}. \tag{5.5}$$

其中 $\nu \in (0, 1)$, 右端的 (y^{k+1}, z^{k+1}) 是由 (5.4) 提供的. 想法是不公平, 需要找补, 调整.

由于为下一步迭代只需要准备 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$, 我们只要停机的最后一步用根据 (5.5) 左边利用数值代数中的方法求得. 校正中从下到上的的过程我们把它叫做高斯回代.

部分平行并加正则项的 ADMM 方法

如果对 y, z 子问题平行, 又不想做后处理, 就给它们俩预先都加个正则项

$$\left\{ \begin{array}{l} x^{k+1} = \arg\min \left\{ \mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X} \right\}, \qquad (\tau > 0 \text{ 为参数}) \\ y^{k+1} = \arg\min \left\{ \mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\beta \| B(y - y^k) \|^2 \middle| y \in \mathcal{Y} \right\}, \\ z^{k+1} = \arg\min \left\{ \mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\beta \| C(z - z^k) \|^2 \middle| z \in \mathcal{Z} \right\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{array} \right. \tag{5.6}$$

上述做法相当于

$$\left\{\begin{array}{l} x^{k+1} \in \arg\min\left\{\theta_1(x) - x^TA^T\lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \,\middle|\, x\in\mathcal{X}\right\},\\ y^{k+1} \in \arg\min\left\{\begin{array}{c}\theta_2(y) - y^TB^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2\\ +\frac{\tau}{2}\beta\|B(y-y^k)\|^2\end{array}\middle| y\in\mathcal{Y}\right\},\\ z^{k+1} \in \arg\min\left\{\begin{array}{c}\theta_3(z) - z^TC^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By^k + Cz - b\|^2\\ +\frac{\tau}{2}\beta\|C(z-z^k)\|^2\end{array}\middle| z\in\mathcal{Z}\right\},\\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b),\end{array}\right.$$

若令 $\lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b)$, 这个方法就是

$$\left\{\begin{array}{l} x^{k+1} = \mathrm{argmin}\{\theta_1(x) - x^TA^T\lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 | x\in\mathcal{X}\},\\ \quad \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b)\\ y^{k+1} = \mathrm{argmin}\{\theta_2(y) - y^TB^T\lambda^{k+\frac{1}{2}} + \frac{\mu\beta}{2}\|B(y-y^k)\|^2 \,|\, y\in\mathcal{Y}\},\\ z^{k+1} = \mathrm{argmin}\{\theta_3(z) - z^TC^T\lambda^{k+\frac{1}{2}} + \frac{\mu\beta}{2}\|C(z-z^k)\|^2 \,|\, z\in\mathcal{Z}\},\\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b),\end{array}\right. \tag{5.7}$$

其中 $\mu = \tau + 1$. Osher 课题组在论文 [6] 中根据我们 [9] 中的 $\mu > 2$ 取了 $\mu = 2.01$.

μ 大收敛慢. [9] 中给出 $\mu > 2$. [14] 中证明 $\mu > 1.5$ 即可, 但对 $\mu < 1.5$ 有不收敛的例子.

This method is accepted by Osher's research group

- 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.

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 21, NO. 7, JULY 2012 3239

A Convex Model for Nonnegative Matrix Factorization and Dimensionality Reduction on Physical Space

Ernie Esser, Michael Möller, Stanley Osher, Guillermo Sapiro, *Senior Member, IEEE*, and Jack Xin

$$\min_{T\geq 0, V_j\in D_j, e\in E} \zeta \sum_i \max_j (T_{i,j}) + \langle R_w \sigma C_w, T\rangle$$

$$\text{such that} \quad YT - X_s = V - X_s \text{diag}(e). \quad (15)$$

Since the convex functional for the extended model (15) is slightly more complicated, it is convenient to use a variant of ADMM that allows the functional to be split into more than two parts. The method proposed by He *et al.* in [34] is appropriate for this application. Again, introduce a new variable Z

Using the ADMM-like method in [34], a saddle point of the augmented Lagrangian can be found by iteratively solving the subproblems with parameters $\delta > 0$ and $\mu > 2$, shown in the

tion refinement step. Due to the different algorithm used to solve the extended model, there is an additional numerical parameter μ, which for this application must be greater than two according to [34]. We set μ equal to 2.01. There are also model parame-

[33] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," 2009 [Online]. Available: http://arxiv.org/PS cache/arxiv/pdf/0912/0912.3599v1.pdf

[34] B. He, M. Tao, and X. Yuan, "A splitting method for separate convex programming with linking linear constraints," Tech. Rep., 2011 [Online]. Available: http://www.optimization-online.org/DB_FILE/2010/06/ 2665.pdf

最新进展：最优正则化因子的选择– OO6235 的结论

Recent Advance in : Bingsheng He, Xiaoming Yuan: On the Optimal Proximal Parameter of an ADMM-like Splitting Method for Separable Convex Programming http://www.optimization-online.org/DB_HTML/2017/ 10/6235.html [14].

Our new assertion: In (5.6)

- if $\tau > 0.5$, the method is still convergent;
- if $\tau < 0.5$, there is divergent example.

Equivalently in (5.7) :

- if $\mu > 1.5$, the method is still convergent;
- if $\mu < 1.5$, there is divergent example.

For convex optimization problem (5.1) with three separable objective functions, the parameters in the equivalent methods (5.6) and (5.7) :

- **0.5** is the threshold factor of the parameter τ in (5.6) !
- **1.5** is the threshold factor of the parameter μ in (5.7) !

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] A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vison, 40, 120-145, 2011.

[5] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent, Mathematical Programming, Series A, 155 (2016) 57-79.

[6] 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.

[7] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] P. Tseng, On accelerated proximal gradient methods for convex-concave optimization, manuscript, 2008.

[17] 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.

[18] 何炳生. 我和乘子交替方向法 20 年. 运筹学学报, 2018, 22: 1–31.