典型凸优化问题的分裂收缩算法讲座

V. 根据统一框架设计预测-校正方法的一般原则

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14日

1 Prediction-Correction Framework

我们已经把线性约束的凸优化问题

$$\min\{\theta(u) \,|\, \mathcal{A}u = b \text{ (or} \geq b), \; u \in \mathcal{U}\} \tag{1.1}$$

化成相应的变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{1.2a}$$

其中

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \text{ and } \quad \Omega = \mathcal{U} \times \Lambda. \tag{1.2b}$$

$\Lambda = \Re^m (\mathcal{A}u = b)$ 或者 $\Lambda = \Re^m_+ (\mathcal{A}u \geq b)$. 因为仿射算子 F 的形式是

$$F(w) = \begin{pmatrix} 0 & -\mathcal{A}^T \\ \mathcal{A} & 0 \end{pmatrix}\begin{pmatrix} u \\ \lambda \end{pmatrix} - \begin{pmatrix} 0 \\ b \end{pmatrix},$$

其中的矩阵是反对称的, 因此我们有

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \equiv 0. \tag{1.3}$$

- 交替方向法 (ADMM) 是已经被广泛接受用来求解两个可分离块凸优化问题的有效算法。ADMM 直接推广用来求解三块和三块以上的可分离凸优化问题, 通常条件下收敛性无法得到保证。

- 过去的十多年，我们发表了一系列求解各类线性（等式和不等式）约束的（两块或多块）可分离凸优化问题的 ADMM 类分裂收缩算法, 并从中归纳出一个预测-校正的算法统一框架。利用这个框架，算法的收敛性证明只需要（通过简单的矩阵运算）验证两个条件。

- 在这个预测–校正统一框架指导下我们还构造了一些算法, 其基本套路是采用 ADMM 类分裂技术预测, 得到相应的预测矩阵: 对称正定矩阵 H, 或者非对称正定矩阵 Q。

- 当预测矩阵 H 对称正定时, 校正采用平凡延伸。在预测矩阵 Q 非对称时, 往往是靠"聪明"去凑能满足收敛性两个条件的校正矩阵 M。

- 实际上, 可以根据预测矩阵 Q 和收敛性的两个条件, 容易地"倒推"出一族校正矩阵 M, 因此也有了基于同一个预测的不同的校正方法。这让原本看起来颇有难度的设计校正矩阵 M 变得不再神秘, 使得构造不同的分裂收缩算法成为一个有分析依据指导的常规工作。这一讲的材料主要取自 [20].

Prediction-correction framework for the VI (1.2)

[Prediction Step.] With given (essential variable) v^k, find a vector $\tilde{w}^k \in \Omega$ such that

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \quad (1.4a)$$

where the matrix Q is not necessarily symmetric but $Q^T + Q$ is assumed to be positive definite. (We focus on the case that Q is asymmetric).

Q 预测矩阵

[Correction Step.] Find a nonsingular matrix M and update v by

M 校正矩阵

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \quad (1.4b)$$

Convergence conditions

For the matrices Q and M used in (1.4a) and (1.4b), there exists a matrix $H \succ 0$ such that

H 范数矩阵

$$HM = Q, \quad (1.5a)$$

and

G 效益矩阵

$$G := Q^T + Q - M^T HM \succ 0. \quad (1.5b)$$

1.1 Convergence

Theorem 1 *Let $\{v^k\}$ be the sequence generated by the prediction-correction framework (1.4) under the conditions (1.5). Then, it holds that*

H 范数矩阵 G 效益矩阵

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{1.6}$$

Proof. Using $Q = HM$ (see (1.5a)), the prediction step can be written as

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

Then, it follows from (1.4b) that

$Q = HM$

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

Setting $w = w^*$ in the above inequality, we get

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

Because $(\tilde{w}^k - w^*)^T F(\tilde{w}^k) = (\tilde{w}^k - w^*)^T F(w^*)$, it follows from the optimality of

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

$$(v^k - v^{k+1})^T H(\tilde{v}^k - v^*) \geq 0, \quad \forall v^* \in \mathcal{V}^*. \tag{1.7}$$

Setting $a = v^k$, $b = v^{k+1}, c = \tilde{v}^k$ and $d = v^*$ in the identity

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

we know from (1.7) that

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2. \tag{1.8}$$

For the right-hand side of the last inequality, we have

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&\overset{(1.4b)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
&= \quad 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\
&= \quad (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\
&\overset{(1.5b)}{=} \quad \|v^k - \tilde{v}^k\|_G^2.
\end{aligned} \tag{1.9}$$

Substituting (1.9) in (1.8), the assertion (1.6) is proved. $\square$

2 Constructing M from the conditions

2.1 Construction from the condition $(1.5a)$

Note that the condition (1.5a), $HM = Q$, can be rewritten as

$$H = QM^{-1}. \tag{2.1}$$

Since the norm matrix H is required to be symmetric and positive definite, the condition (2.1) implies that H should be representable in form of

$$H = QD^{-1}Q^T, \tag{2.2}$$

in which the matrix D is a undetermined positive definite matrix. Indeed, by comparing (2.1) with (2.2), we know that $M^{-1} = D^{-1}Q^T$ and thus

$$M = Q^{-T}D. \tag{2.3}$$

Hence, although the matrix D in (2.2) is still unknown, choosing M as (2.3) can ensure the condition (1.5a).

Now, we investigate the restriction on D to ensure the condition (1.5b) with the matrix M given as (2.3). Notice that

$$M^T HM = \big(DQ^{-1}\big)\big(QD^{-1}Q^T\big)\big(Q^{-T}D\big) = D. \tag{2.4}$$

With (2.4), then the condition (1.5b) is reduced to

$$G := Q^T + Q - M^T HM = Q^T + Q - D \succ 0. \tag{2.5}$$

Hence, to ensure the condition (1.5b), the only restriction on the positive definite matrix D in (2.2) is

$$0 \prec D \prec Q^T + Q, \tag{2.6}$$

In other words, whenever Q is given and it satisfies $Q^T + Q \succ 0$, then both H and M can be constructed via the following steps:

$$\left\{\begin{array}{r} HM = Q, \\ M^T HM = D. \end{array}\right. \quad \Leftrightarrow \quad \left\{\begin{array}{r} HM = Q, \\ Q^T M = D. \end{array}\right. \quad \Leftrightarrow \quad \left\{\begin{array}{r} H = QD^{-1}Q^T, \\ M = Q^{-T}D. \end{array}\right. \tag{2.7}$$

Through this construction, both the conditions (1.4b) and (1.5b) are guaranteed to be satisfied. Note that once the matrix D is chosen according to (2.6), the matrices H, M and G are all uniquely determined. Then, with the specified matrix M in (2.3), the correction step (1.4b) and thus the prediction-correction framework (1.4) is also specified as a concrete contraction splitting algorithm for the VI(1.2)-(1.2b).

2.2 Construction from the condition $(1.5b)$

Alternatively, we can from the condition (1.5b), $G = Q^T + Q - M^T HM \succ 0$, to construct the norm matrix H and the correction matrix M. Again, with a given Q satisfying $Q^T + Q \succ 0$, we can choose the profit matrix G such that

$$0 \prec G \prec Q^T + Q. \tag{2.8}$$

Denote

$$\Delta = Q^T + Q - G, \tag{2.9}$$

which is positive definite. According to (1.5b), we know that the matrices H and M should satisfy

$$M^T HM = \Delta.$$

Recall the condition (1.5a): $HM = Q$. Thus, with a chosen G satisfying (2.8), H and M can be constructed vis the following steps:

$$\left\{ \begin{array}{r} M^T HM = \Delta, \\ HM = Q. \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{r} Q^T M = \Delta, \\ HM = Q. \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{l} M = Q^{-T}\Delta, \\ H = Q\Delta^{-1}Q^T. \end{array} \right. \tag{2.10}$$

Then, with the constructed matrix M in (2.10), the correction step (1.4b) and thus the prediction-correction framework (1.4) can also be specified as a concrete splitting contraction algorithm for the VI(1.2)-(1.2b). Again, with a given G satisfying (2.8), the matrices H and M are both uniquely determined.

2.3 Choices of D and G

It is interesting to observe that the proposed two construction strategies can be related via the relationship

$$D \succ 0, \quad G \succ 0, \quad \text{and} \quad D + G = Q^T + Q. \tag{2.11}$$

Hence, once D is chosen for the construction strategy in Section 2.1, the corresponding G given by (2.11) can be used for the construction strategy in Section 2.2, and vice versa.

Technically, there are infinitely many such choices subject to (2.11). For example, we can choose

$$D = \alpha[Q^T + Q] \quad \text{and} \quad G = (1-\alpha)[Q^T + Q], \qquad \alpha \in (0,1).$$

We will elaborate on the choice $D = G = \frac{1}{2}[Q^T + Q]$ in Section 3.3.3.

2.4 Implementation of the correction step (1.4b)

Note that the correction step (1.4b) can be rewritten as

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

To implement the correction step (1.4b) with the constructed two choices for M, i.e., $M = Q^{-T}D$ in (2.3) and $M = Q^{-T}\Delta$ in (2.10), we need to solve one of the following systems of equations:

$$Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k), \tag{2.12}$$

and

$$Q^T(v^{k+1} - v^k) = \Delta(\tilde{v}^k - v^k). \tag{2.13}$$

Hence, although D and G (thus Δ) can be chosen arbitrarily with the only constraint (2.6) or (2.8), it is preferred to choose some model-tailored ones that can favor solving the systems of equations (2.12) or (2.13) more efficiently.

3 求解等式约束的三个可分离块的优化问题

We apply the strategies proposed in Sections 2.1 and 2.2 to a separable convex optimization problem, and showcase how to construct the norm matrix H and the correction matrix M when the matrix Q is given.

3.1 相应的变分不等式

We consider the three-block separable convex optimization model with linear constraints

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) \,|\, Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\}, \quad (3.1)$$

Clearly, it is a special case of the canonical convex programming problem (1.1), and the VI (1.2)-(1.2b) can be specified as the following:

$$w^*\in\Omega, \quad \theta(u)-\theta(u^*)+(w-w^*)^T F(w^*)\geq 0, \quad \forall\, w\in\Omega, \qquad (3.2)$$

where

$$w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix}, \tag{3.3a}$$

with

$$\theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z), \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m. \tag{3.3b}$$

Let the augmented Lagrangian function of the model (3.1) be

$$\begin{aligned} \mathcal{L}_\beta^{[3]}(x, y, z, \lambda) &= \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^T(Ax + By + Cz - b) \\ &\quad + \frac{\beta}{2}\|Ax + By + Cz - b\|^2 \end{aligned} \tag{3.4}$$

with $\lambda \in \Re^m$ the Lagrange multiplier and $\beta > 0$ the penalty parameter.

直接推广的交替方向法

$$\begin{cases} x^{k+1} \in \arg\min\{\mathcal{L}_{\beta}^{[3]}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} \in \arg\min\{\mathcal{L}_{\beta}^{[3]}(x^{k+1}, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\}, \\ z^{k+1} \in \arg\min\{\mathcal{L}_{\beta}^{[3]}(x^{k+1}, y^{k+1}, z, \lambda^k) \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \big(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b\big). \end{cases} \tag{3.5}$$

However, the splitting scheme (3.5) is coarse in sense of that its convergence is not guaranteed as shown in [2].

3.2 基于直接推广的 ADMM 的预测矩阵 Q

Our construction starts from the coarse splitting scheme (3.5) which can be rewritten as the prediction step (1.4a) and hence the corresponding prediction matrix Q can be discerned. For this purpose, we first consider the subproblems related to the primal variables in (3.5), and rewrite them as

$\tilde{u}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Namely, we have

$$\begin{cases} \tilde{x}^k \in \arg\min\big\{\mathcal{L}_\beta^{[3]}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \arg\min\big\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\big\}, \\ \tilde{z}^k \in \arg\min\big\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, \tilde{y}^k, z, \lambda^k) \mid z \in \mathcal{Z}\big\}. \end{cases} \tag{3.6}$$

Ignoring some constant terms, we can rewrite the formula above as

$$\begin{cases} \tilde{x}^k \in \arg\min\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \arg\min\big\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|A\tilde{x}^k + By + Cz^k - b\|^2 \mid y \in \mathcal{Y}\big\}, \\ \tilde{z}^k \in \arg\min\big\{\theta_3(z) - z^T C^T \lambda^k + \frac{\beta}{2}\|A\tilde{x}^k + B\tilde{y}^k + Cz - b\|^2 \mid z \in \mathcal{Z}\big\}. \end{cases}$$

Then, according to the optimality Lemma, we have $\tilde{u}^k \in \mathcal{U}$ and

$$\left\{\begin{array}{l} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\big\{-A^T\lambda^k \\ \qquad\qquad\qquad +\beta A^T\big(A\tilde{x}^k + By^k + Cz^k - b\big)\big\} \geq 0, \quad \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\big\{-B^T\lambda^k \\ \qquad\qquad\qquad +\beta B^T\big(A\tilde{x}^k + B\tilde{y}^k + Cz^k - b\big)\big\} \geq 0, \quad \forall\, y \in \mathcal{Y}, \\ \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T\big\{-C^T\lambda^k \\ \qquad\qquad\qquad +\beta C^T\big(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b\big)\big\} \geq 0, \quad \forall\, z \in \mathcal{Z}. \end{array}\right. \tag{3.7}$$

By defining

$$\tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big), \tag{3.8}$$

we have

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

Using the VI form (3.3), we get $\tilde{w}^k \in \Omega$ and

$$\left\{ \begin{array}{l} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{\underline{-A^T\tilde{\lambda}^k}\} \geq 0, \quad \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{\underline{-B^T\tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k)\} \geq 0, \quad \forall\, y \in \mathcal{Y}, \\ \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T \left\{ \begin{array}{c} \underline{-C^T\tilde{\lambda}^k} + \beta C^T B(\tilde{y}^k - y^k) \\ \beta C^T C(\tilde{z}^k - z^k) \end{array} \right\} \geq 0, \quad \forall\, z \in \mathcal{Z}, \\ (\lambda - \tilde{\lambda}^k)^T \left\{ \begin{array}{l} (\underline{A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b}) \\ -B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \end{array} \right\} \geq 0, \quad \forall\, \lambda \in \Lambda. \end{array} \right. \tag{3.9}$$

The sum of the underline parts of (3.9) is exactly $F(\tilde{w}^k)$, where $F(\cdot)$ is defined in (3.3). Thus, we have

$$\tilde{w}^k \in \Omega,\ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k),\ \forall w \in \Omega, \tag{3.10}$$

where the prediction matrix is

$$Q = \begin{pmatrix} \beta B^T B & 0 & 0 \\ \beta C^T B & \beta C^T C & 0 \\ -B & -C & \dfrac{1}{\beta} I_m \end{pmatrix}. \tag{3.11}$$

Moreover, for the prediction matrix Q in (3.11) which is determined by the splitting scheme (3.6) and the defined $\tilde{\lambda}^k$ by (3.8), we have

$$Q^T + Q = \begin{pmatrix} 2\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & 2\beta C^T C & -C^T \\ -B & -C & \frac{2}{\beta} I_m \end{pmatrix}, \tag{3.12}$$

which is positive definite whenever B and C are full column rank.

3.3 基于直接推广的ADMM 的预测矩阵 Q

With the prediction matrix Q given in (3.11), the prediction-correction framework (1.4) can be specified as a concrete algorithm for the model (3.1) once the correction step (1.4b) is specified. Now, we showcase how to specify the correction step (1.4b) by the construction strategies discussed in Sections 2.1, 2.2 and 2.3. Note that $v = (y, z, \lambda)$ below.

选择 $0 \prec D \prec Q^T + Q$, 可以提出自己想要的方法, 下面只是一些例子而已.

3.3.1 构造方法 I

Based on (3.11) and (3.12), and following the strategy in Section 2.1, we can choose

$$D = \begin{pmatrix} \nu\beta B^TB & 0 & 0 \\ 0 & \nu\beta C^TC & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \tag{3.13}$$

with $0 < \nu < 1$, which is positive definite whenever B and C are both full column rank. Recall the correction matrix M in (2.3). Then, a concrete splitting contraction algorithm for (3.1) can be generated as below.

Algorithm 1 for the model (3.1)

[Prediction Step.] Obtain $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$ via the direct extension of the ADMM (3.6) and define $\tilde{\lambda}^k$ by (3.8).

[Correction Step.] $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$.

For the correction step $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$, we know that

$$Q^T = \begin{pmatrix} \beta B^TB & \beta B^TC & -B^T \\ 0 & \beta C^TC & -C^T \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} = \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} B & C & -\frac{1}{\beta}I_m \\ 0 & C & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix},$$

and

$$D = \begin{pmatrix} \nu\beta B^TB & 0 & 0 \\ 0 & \nu\beta C^TC & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} = \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} \nu B & 0 & 0 \\ 0 & \nu C & 0 \\ 0 & 0 & I_m \end{pmatrix}.$$

That is, Q^T and D have a common matrix in their factorization forms above.

Hence, to implement the correction step (1.4b), i.e.,

$$Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k),$$

essentially we only need to consider the even easier equation

$$\begin{pmatrix} B & C & -\frac{1}{\beta}I_m \\ 0 & C & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = \begin{pmatrix} \nu B & 0 & 0 \\ 0 & \nu C & 0 \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{z}^k - z^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}.$$

The above system of equations equivalent to

$$\begin{pmatrix} I_m & I_m & -\frac{1}{\beta}I_m \\ 0 & I_m & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} By^{k+1} - By^k \\ Cz^{k+1} - Cz^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = \begin{pmatrix} \nu I_m & 0 & 0 \\ 0 & \nu I_m & 0 \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} B\tilde{y}^k - By^k \\ C\tilde{z}^k - Cz^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}.$$

We can get $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$ by a back-substitution.

3.3.2 构造方法 II

Based on (3.11) and (3.12), and following the strategy in Section 2.2, we can choose

$$G = \begin{pmatrix} (1-\nu)\beta B^T B & 0 & 0 \\ 0 & (1-\nu)\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}, \qquad (3.15)$$

with $\nu \in (0,1)$, which can be guaranteed to be positive definite whenever B and C are full column rank. Note that the matrix G in (3.15) is precisely the matrix D defined in (3.13). Furthermore, we have

$$\Delta = Q^T + Q - G = \begin{pmatrix} (1+\nu)\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & (1+\nu)\beta C^T C & -C^T \\ -B & -C & \frac{1}{\beta} I_m \end{pmatrix}.$$

Recall the correction matrix M in (2.10). Then, another contraction splitting algorithm for (3.1) can be generated as below.

Algorithm 2 for the model (3.1)

[Prediction Step.] Obtain $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$ via the direct extension of the ADMM (3.6) define $\tilde{\lambda}^k$ by (3.8).

[Correction Step.] $Q^T(v^{k+1} - v^k) = \Delta(\tilde{v}^k - v^k)$.

For the correction step $Q^T(v^{k+1} - v^k) = \Delta(\tilde{v}^k - v^k)$, we know that

$$
\begin{aligned}
Q^T &= \begin{pmatrix} \beta B^T B & \beta B^T C & -B^T \\ 0 & \beta C^T C & -C^T \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \\
&= \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} B & C & -\frac{1}{\beta} I_m \\ 0 & C & -\frac{1}{\beta} I_m \\ 0 & 0 & I_m \end{pmatrix},
\end{aligned}
$$

and

$$\Delta = \begin{pmatrix} (1+\nu)\beta B^TB & \beta B^TC & -B^T \\ \beta C^TB & (1+\nu)\beta C^TC & -C^T \\ -B & -C & \frac{1}{\beta}I_m \end{pmatrix}$$

$$= \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} (1+\nu)B & C & -\frac{1}{\beta}I_m \\ B & (1+\nu)C & -\frac{1}{\beta}I_m \\ -\beta B & -\beta C & I_m \end{pmatrix}.$$

That is, Q^T and Δ have a common matrix in their factorization forms above.

Hence, to implement the correction step (1.4b), i.e.,

$$Q^T(v^{k+1} - v^k) = \Delta(\tilde{v}^k - v^k),$$

essentially we only need to consider the even easier equation

$$\begin{pmatrix} B & C & -\frac{1}{\beta}I_m \\ 0 & C & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix}$$

$$= \begin{pmatrix} (1+\nu)B & C & -\frac{1}{\beta}I_m \\ B & (1+\nu)C & -\frac{1}{\beta}I_m \\ -\beta B & -\beta C & I_m \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{z}^k - z^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}. \quad (3.17)$$

The above system of equations equivalent to

$$\begin{pmatrix} I_m & I_m & -\frac{1}{\beta}I_m \\ 0 & I_m & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} By^{k+1} - By^k \\ Cz^{k+1} - Cz^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix}$$

$$= \begin{pmatrix} (1+\nu)I_m & I_m & -\frac{1}{\beta}I_m \\ I_m & (1+\nu)I_m & -\frac{1}{\beta}I_m \\ -\beta I_m & -\beta I_m & I_m \end{pmatrix} \begin{pmatrix} B\tilde{y}^k - By^k \\ C\tilde{z}^k - Cz^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}. \quad (3.18)$$

Similar as (3.3.1), with the choice of G in (3.15), implementing the resulting correction step (1.4b) essentially only requires solving the equation (3.18) in terms of (By^k, Cz^k, λ^k). By a manipulation, the correction form can be simplified to

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} By^{k} \\ Cz^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 \\ 0 & \nu I_m & 0 \\ -\beta I_m & -\beta I_m & I_m \end{pmatrix} \begin{pmatrix} B(y^k - \tilde{y}^k) \\ C(z^k - \tilde{z}^k) \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

3.3.3 构造方法 III

Recall the relationship between the matrices D and G in (2.11), and $Q^T + Q$ given in (3.12). Essentially, the proposed construction strategies in Sections 2.1

and 2.2 take the same matrix

$$\begin{pmatrix} \nu\beta B^TB & 0 & 0 \\ 0 & \nu\beta C^TC & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix}$$

as D and G, respectively, and then the other one is determined by (2.11). As mentioned in Section 2.3, any other choice of D and G subject to the relationship (2.11) is also eligible. Let us consider the following specific one:

$$D = G = \frac{1}{2}\left[Q^T + Q\right] = \begin{pmatrix} \beta B^TB & \frac{1}{2}\beta B^TC & -\frac{1}{2}B^T \\ \frac{1}{2}\beta C^TB & \beta C^TC & -\frac{1}{2}C^T \\ -\frac{1}{2}B & -\frac{1}{2}C & \frac{1}{\beta}I_m \end{pmatrix}, \quad (3.19)$$

which are both positive definite whenever B and C are full column rank. Recall the correction matrix M in (2.10). Then, one more contraction splitting algorithm for (3.1) can be generated as below.

Algorithm 3 for the model (3.1)

[Prediction Step.] Obtain $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$ via the direct extension of the ADMM (3.6) and define $\tilde{\lambda}^k$ by (3.8).

[Correction Step.] $Q^T(v^{k+1} - v^k) = \frac{1}{2}[Q^T + Q](\tilde{v}^k - v^k)$.

For the correction step $Q^T(v^{k+1} - v^k) = \frac{1}{2}[Q^T + Q](\tilde{v}^k - v^k)$, we know that

$$Q^T = \begin{pmatrix} \beta B^T B & \beta B^T C & -B^T \\ 0 & \beta C^T C & -C^T \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} = \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} B & C & -\frac{1}{\beta} I_m \\ 0 & C & -\frac{1}{\beta} I_m \\ 0 & 0 & I_m \end{pmatrix},$$

and

$$\frac{1}{2}[Q^T + Q] = \begin{pmatrix} \beta B^T B & \frac{1}{2}\beta B^T C & -\frac{1}{2}B^T \\ \frac{1}{2}\beta C^T B & \beta C^T C & -\frac{1}{2}C^T \\ -\frac{1}{2}B & -\frac{1}{2}C & \frac{1}{\beta}I_m \end{pmatrix}$$

$$= \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} B & \frac{1}{2}C & -\frac{1}{2\beta}I_m \\ \frac{1}{2}B & C & -\frac{1}{2\beta}I_m \\ -\frac{1}{2}\beta B & -\frac{1}{2}\beta C & I_m \end{pmatrix}.$$

That is, Q^T and $\frac{1}{2}[Q^T + Q]$ have a common matrix in their factorization forms above.

Hence, to implement $Q^T(v^{k+1}-v^k)=\frac{1}{2}[Q^T+Q](\tilde{v}^k-v^k)$, essentially we only need to consider the even easier equation

$$\begin{pmatrix} B & C & -\frac{1}{\beta}I_m \\ 0 & C & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix}\begin{pmatrix} y^{k+1}-y^k \\ z^{k+1}-z^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} = \begin{pmatrix} B & \frac{1}{2}C & -\frac{1}{2\beta}I_m \\ \frac{1}{2}B & C & -\frac{1}{2\beta}I_m \\ -\frac{1}{2}\beta B & -\frac{1}{2}\beta C & I_m \end{pmatrix}\begin{pmatrix} \tilde{y}^k-y^k \\ \tilde{z}^k-z^k \\ \tilde{\lambda}^k-\lambda^k \end{pmatrix}. \tag{3.21}$$

The equivalent form

$$\begin{pmatrix} I_m & I_m & -\frac{1}{\beta}I_m \\ 0 & I_m & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix}\begin{pmatrix} By^{k+1}-By^k \\ Cz^{k+1}-Cz^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} = \begin{pmatrix} I_m & \frac{1}{2}I_m & -\frac{1}{2\beta}I_m \\ \frac{1}{2}I_m & I_m & -\frac{1}{2\beta}I_m \\ -\frac{1}{2}\beta I_m & -\frac{1}{2}\beta I_m & I_m \end{pmatrix}\begin{pmatrix} B\tilde{y}^k-By^k \\ C\tilde{z}^k-Cz^k \\ \tilde{\lambda}^k-\lambda^k \end{pmatrix}.$$

4 ADMM with wider application & easy extensions

Let us consider the general separable convex optimization model

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

ADMM for (4.1) From (y^k,λ^k) to (y^{k+1},λ^{k+1})

$$\begin{cases} x^{k+1} \in \arg\min\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k-b\|^2 \,|\, x\in\mathcal{X}\}, \\ y^{k+1} \in \arg\min\{\theta_2(y)-y^TB^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2 \,|\, y\in\mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). \end{cases} \tag{4.2}$$

$$\begin{aligned} x^{k+1} &\in \arg\min\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k-b\|^2 \,|\, x\in\mathcal{X}\} \\ &\in \arg\min\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|(Ax^k+By^k-b)+A(x-x^k)\|^2 \,|\, x\in\mathcal{X}\} \\ &\in \arg\min\{\theta_1(x)-x^TA^T[\lambda^k-\beta(Ax^k+By^k-b)]+\frac{\beta}{2}\|A(x-x^k)\|^2 \,|\, x\in\mathcal{X}\}. \end{aligned}$$

Ignoring some constant terms in the objective functions of the corresponding subproblems, we can rewrite the ADMM (4.2) as

$$\begin{cases} x^{k+1} \in \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2}\|A(x-x^k)\|^2 \mid x \in \mathcal{X}\}, \\ y^{k+1} \in \operatorname{argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \\ \frac{\beta}{2}\|A(x^{k+1}-x^k) + B(y-y^k)\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\}, \\ \lambda^{k+1} = \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} - b\big) \end{cases} \tag{4.3}$$

where

$$\lambda^{k+\frac{1}{2}} := \lambda^k - \beta(Ax^k + By^k - b).$$

The λ update form can be also denoted by

$$\lambda^{k+1} = P_{\Re^m}\big[\lambda^k - \beta\big(Ax^{k+1} + By^{k+1} - b\big)\big].$$

为了说明我们后面提出的方法和 ADMM 的关系,
我们把经典的 ADMM 改写成等价的 (4.3).

4.1 ADMM with wider applications

Let us consider the general two-block separable convex optimization model

$$\min\left\{\theta_1(x)+\theta_2(y) \mid Ax+By=b \ (\text{or} \geq b), x\in\mathcal{X}, y\in\mathcal{Y}\right\}. \tag{4.4}$$

The linear constraints can be a system of linear equations or linear inequalities.

We define

$$\Lambda=\begin{cases}\Re^m, & \text{if } Ax+By=b,\\ \Re^m_+, & \text{if } Ax+By\geq b.\end{cases}$$

The projection on Λ is denoted by $P_\Lambda[\cdot]$.

For such special Λ, the projection on Λ is clear !

The only difference: $P_{\Re^m}(\lambda)=\lambda, \quad P_{\Re^m_+}(\lambda)=\max\{\lambda,0\}.$

4.1.1 Primal-dual extension of ADMM with wider application

A Primal-Dual Extension of the ADMM for (4.4).

From (Ax^k, By^k, λ^k) to $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$:

1. (Prediction Step) With given (Ax^k, By^k, λ^k), find $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via

$$\left\{\begin{array}{l} \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|A(x-x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|A(\tilde{x}^k - x^k) + B(y-y^k)\|^2 \mid y \in \mathcal{Y}\big\}, \\ \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(A\tilde{x}^k + B\tilde{y}^k - b\big)\big]. \end{array}\right. \tag{4.5a}$$

2. (Correction Step) Generate the new iterate $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} Ax^{k+1} \\ By^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} Ax^k \\ By^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 \\ 0 & \nu I_m & 0 \\ -\nu\beta I_m & 0 & I_m \end{pmatrix} \begin{pmatrix} Ax^k - A\tilde{x}^k \\ By^k - B\tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{4.5b}$$

这是一类预测–校正方法. 需要额外的校正, 但校正花费很小！

预测先做 Primal 部分, 再做 Dual 部分, 顺序也可以倒过来.

4.1.2 Dual-Primal extension of ADMM with wider application

A Dual-Primal Extension of the ADMM for (4.4).

From (Ax^k, By^k, λ^k) to $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$:

1. (Prediction Step) With given (Ax^k, By^k, λ^k), find $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via

$$\left\{ \begin{array}{l} \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(Ax^k + By^k - b\big)\big], \\ \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^T A^T \tilde{\lambda}^k + \frac{1}{2}\beta\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^T B^T \tilde{\lambda}^k + \frac{1}{2}\beta\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \mid y \in \mathcal{Y}\big\}. \end{array} \right. \tag{4.6a}$$

2. (Correction Step) Generate the new iterate $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} Ax^{k+1} \\ By^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} Ax^k \\ By^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 \\ 0 & \nu I_m & 0 \\ -\beta I_m & -\beta I_m & I_m \end{pmatrix} \begin{pmatrix} Ax^k - A\tilde{x}^k \\ By^k - B\tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{4.6b}$$

预测采用不同顺序, 校正公式也略有不同. 校正同样是花费很小的.

无论是 primal-dual, 还是 dual-primal 方法, 都可以向多块问题直接推广.

References

[1] 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.

[2] 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* 2016.

[3] R. Glowinski and A. Marrocco, A, Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problémes non linéaires. RAIRO Anal. Numer. R2 (1975), 41-76.

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

[5] G. Y. Gu, B. S. He and X. M. Yuan, Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach, Comput. Optim. Appl., 59 (2014), 135-161.

[6] B. S. He, From the projection and contraction methods for variatioonal inequality to the splittingcontraction methods fro convex optimization (in Chinese), Numerical Mathematics, A Journal of Chinese Universities, 38 (2016), 74-96.

[7] B. S. He, My 20 years research on alternating directions method of multipliers (in

Chinese), Operations Research Trasactions, 22 (2018), 1-31.

[8] B. S. He, A uniform framework of contraction methods for convex optimization and monotone variational inequality (in Chinese). Sci Sin Math, 48 (2018), 255-272, doi: 10.1360/N012017-00034

[9] B. S. He, Using a unified framework to design the splitting and contraction methods for convex optimization (in Chinese), Numerical Mathematics, A Journal of Chinese Universities, 44 (2022), 1-35.

[10] B. S. He, H. Liu, Z. R. Wang and X. M. Yuan, A strictly Peaceman-Rachford splitting method for convex programming, SIAM J. Optim. 24 (2014), 1011-1040.

[11] 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.

[12] B. S. He, M. Tao and X. M. Yuan, A splitting method for separable convex programming, IMA Journal of Numerical Analysis, 31 (2015), 394-426.

[13] B. S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, SIAM Journal on Imaging Science, 5 (2012) 119-149.

[14] B. S. He, S. J. Xu and X. M. Yuan, Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints, arXiv:2107.01897v2[math.OC].

[15] B. S. He and X. M. Yuan, On the $O(1/n)$ convergence rate of the alternating direction method, SIAM J. Numerical Analysis, 50 (2012), 700-709.

[16] B. S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, SIAM Journal on Imaging Science, 5 (2012) 119-149.

[17] B. S. He and X. M. Yuan, A class of ADMM-based algorithms for three-block separable convex programming, Comput. Optim. Appl. 70 (2018), 791-826.

[18] 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.

[19] B. S. He and X. M. Yuan, Balanced Augmented Lagrangian Method for Convex Programming, arXiv 2108.08554 [math.OC]

[20] B. S. He and X. M. Yuan, On construction of splitting contraction algorithms in a prediction-correction framework for separable convex optimization, arXiv 2204.11522 [math.OC]