利用预测–校正统一框架构造凸优化的分裂收缩算法

[arXiv 2107.01897, arXiv 2108.08554, arXiv 2204.11522]

何 炳 生

南京大学数学系

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

统计与数据科学前沿理论及应用教育部重点实验室

华东师范大学　　2022 年 5 月 7 日

- 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 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, 校正采用平凡延拓。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。
- 我们最近的研究说明, 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。

1 Problems and the customized PPA

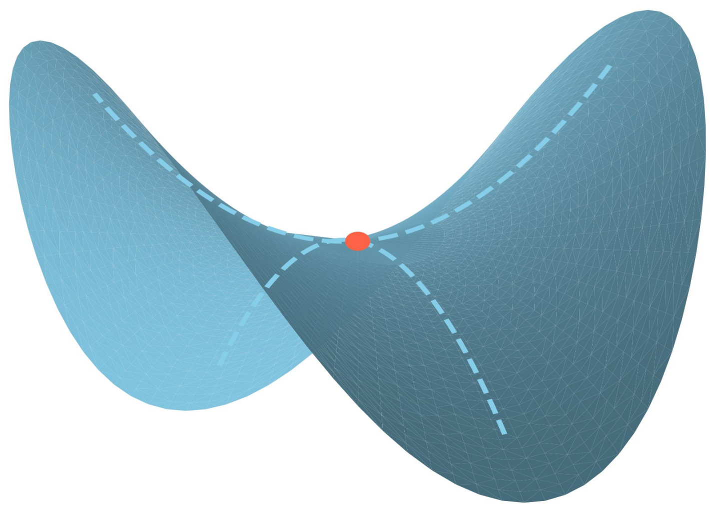
1.1 Problem and the related VI and PPA

We consider the linearly constrained convex optimization problem

$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{1.1}$$

The Lagrange function of (1.1) is

$$L(u, \lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \qquad (u, \lambda) \in \mathcal{U} \times \Re^m. \tag{1.2}$$



Saddle Point of the Lagrange Function

A pair of $(u^*,\lambda^*)\in\mathcal{U}\times\Re^m$ is called a saddle point if

$$L_{\lambda\in\Re^m}(u^*,\lambda)\le L(u^*,\lambda^*)\le L_{u\in\mathcal{U}}(u,\lambda^*).$$

The above inequalities can be written as

$$\left\{\begin{array}{ll} u^*\in\mathcal{U}, & L(u,\lambda^*)-L(u^*,\lambda^*)\ge 0,\ \ \forall\, u\in\mathcal{U}, \\ \lambda^*\in\Re^m, & L(u^*,\lambda^*)-L(u^*,\lambda)\ge 0,\ \ \forall\, \lambda\in\Re^m. \end{array}\right.$$

According to the definition of $L(u,\lambda)$ (see(1.2)), we get the following variational inequality:

$$\left\{\begin{array}{lr} u^*\in\mathcal{U}, & \theta(u)-\theta(u^*)+(u-u^*)^T(-\mathcal{A}^T\lambda^*)\ge 0, \quad \forall\, u\in\mathcal{U}, \\ \lambda^*\in\Re^m, & (\lambda-\lambda^*)^T(\mathcal{A}u^*-b)\ge 0, \quad \forall\, \lambda\in\Re^m. \end{array}\right.$$

Using a more compact form, the saddle-point can be characterized as the solution of the following VI:

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega. \tag{1.4a}$$

where

$$w=\left(\begin{array}{c} u \\ \lambda \end{array}\right),\quad F(w)=\left(\begin{array}{c} -\mathcal{A}^T\lambda \\ \mathcal{A}u-b \end{array}\right)\ \ \text{and}\ \ \Omega=\mathcal{U}\times\Re^m. \tag{1.4b}$$

Because F is a affine operator and

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

the matrix is skew-symmetric, and we have

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \equiv 0. \tag{1.5}$$

For solving the VI, we use the following customized PPA.

Lemma 1 *Let the vectors* $a, b \in \Re^n$, $H \in \Re^{n \times n}$ *be a positive definite matrix. If* $b^T H(a - b) \geq 0$, *then we have*

$$\|b\|_H^2 \leq \|a\|_H^2 - \|a - b\|_H^2.$$

The assertion follows from $\|a\|_H^2 = \|b + (a - b)\|_H^2 \;\geq\; \|b\|_H^2 + \|a - b\|_H^2.$

Lemma 2 *Let* $\mathcal{X} \subset \Re^n$ *be a closed convex set,* $\theta(x)$ *and* $f(x)$ *be convex functions and* $f(x)$ *is differentiable on an open set which contains* $\mathcal{X}$*.*

Assume that the solution set of the minimization problem $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ *is nonempty. Then,*

$$x^* \in \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{1.6a}$$

if and only if

$$x^* \in \mathcal{X}, \;\; \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \;\; \forall\, x \in \mathcal{X}. \tag{1.6b}$$

PPA for VI (1.4) For given v^k ($v = w$ or sub-vector of w) and $H \succ 0$, find

$$\begin{aligned} w^{k+1} \in \Omega, \;\; \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (v - v^{k+1})^T H(v^k - v^{k+1}), \; \forall\, w \in \Omega. \end{aligned} \tag{1.7}$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (1.4).

w^k is the solution of (1.4) if and only if $v^k = v^{k+1}$.

Setting $w = w^*$ in (1.7), we obtain

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

Note that (see the structure of $F(w)$ in (1.5))

$$(w^{k+1} - w^*)^T F(w^{k+1}) = (w^{k+1} - w^*)^T F(w^*),$$

and consequently (by using (1.4)) we obtain

$$(v^{k+1}-v^*)^T H(v^k-v^{k+1}) \geq \theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^T F(w^*) \geq 0,$$

and thus

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0. \tag{1.8}$$

Now, by denoting $a = v^k - v^*$ and $b = v^{k+1} - v^*$ and using Lemma 1, we get the nice convergence property of Proximal Point Algorithm:

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2. \tag{1.9}$$

Relaxed PPA for VI (1.4)

Set the output of (1.7) as $\tilde{w}^k$, we get

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \;\forall\, w \in \Omega. \tag{1.10}$$

Instead of (1.8), we have

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

Thus, we have

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_H^2. \tag{1.11}$$

By setting

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

we get

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \alpha(2 - \alpha)\|v^k - \tilde{v}^k\|_H^2.$$

Usually, by setting $\alpha \in [1.2, 1.6]$, The relaxed PPA is much fast than the classical PPA.

1.2 Applications for separable problems

This section presents various applications of the proposed algorithms for the separable convex optimization problem

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{1.12}$$

Its VI-form is

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{1.13a}$$

where

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \tag{1.13b}$$

and

$$\theta(u) = \theta_1(x) + \theta_2(y), \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{1.13c}$$

The augmented Lagrangian Function of the problem (1.12) is

$$\mathcal{L}_\beta(x,y,\lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax+By-b) + \frac{\beta}{2}\|Ax+By-b\|^2. \tag{1.14}$$

Solving the problem (1.12) by using ADMM, the k-th iteration begins with given (y^k, λ^k), it offers the new iterate (y^{k+1}, λ^{k+1}) via

$$\text{(ADMM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x, y^k, \lambda^k) \;\big|\; x \in \mathcal{X}\big\}, & \text{(1.15a)} \\ y^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \;\big|\; y \in \mathcal{Y}\big\}, & \text{(1.15b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(1.15c)} \end{array} \right.$$

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad v = \begin{pmatrix} y \\ \lambda \end{pmatrix} \quad \text{and} \quad \mathcal{V}^* = \{(y^*, \lambda^*) \,|\, (x^*, y^*, \lambda^*) \in \Omega^*\}.$$

The main convergence result is

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

where

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

Ignoring some constant term in the objective function, ADMM (1.15) **is implemented by**

(ADMM)

$$\begin{cases} x^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T p^k \\ + \frac{\beta}{2}\|A(x - x^k)\|^2. \end{array} \middle| \, x \in \mathcal{X} \right\}, & \text{(1.16a)} \\ y^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T q^k \\ + \frac{\beta}{2}\|B(y - y^k)\|^2. \end{array} \middle| \, y \in \mathcal{Y} \right\}, & \text{(1.16b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(1.16c)} \end{cases}$$

where

$$\begin{aligned} p^k &= \lambda^k - \beta(Ax^k + By^k - b), \\ q^k &= \lambda^k - \beta(Ax^{k+1} + By^k - b). \end{aligned}$$

1.3 ADMM in PPA-sense

根据 PPA 算法的要求　设计的右端矩阵为对称正定.

In order to solve the separable convex optimization problem (1.12), we construct a method whose prediction-step is

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{1.17a}$$

where

$$H = \begin{pmatrix} \beta B^T B + \delta I_{n_2} & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad \text{(a small } \delta > 0\text{, say } \delta = 0.05\text{)}. \tag{1.17b}$$

Since H is positive definite, we can use the update form of Algorithm I to produce the new iterate $v^{k+1} = (y^{k+1}, \lambda^{k+1})$. (In the algorithm [1], we took $\delta = 0$).

The concrete form of (1.17) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k}\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + (\boldsymbol{\beta B^T B + \delta I_{n_2}})(\tilde{y}^k - y^k) \boldsymbol{-B^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b}) \quad \boldsymbol{-B}(\tilde{y}^k - y^k) \quad + \quad \boldsymbol{(1/\beta)}\,(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

The underline part is $F(\tilde{w}^k)$:

$$\boldsymbol{F(w)} = \begin{pmatrix} \boldsymbol{-A^T\lambda} \\ \boldsymbol{-B^T\lambda} \\ \boldsymbol{Ax + By - b} \end{pmatrix}$$

In fact, the prediction can be arranged by

$$\begin{cases} \tilde{x}^k \in \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\} & (1.18a) \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), & (1.18b) \\ \tilde{y}^k \in \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T[2\tilde{\lambda}^k - \lambda^k] \\ +\frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\}. & (1.18c) \end{cases}$$

这个预测与经典的交替方向法 (1.15) 相当, 采用松弛校正, 会加快速度.

According to Lemma 2, the solution of (1.18a), $\tilde{x}^k$ satisfies

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

By using (1.18b), $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, the above variational inequality can be written as

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

The equation (1.18b) can written as

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

The remainder part of the prediction (1.17), namely,

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{\underline{-B^T\tilde{\lambda}^k} + \boldsymbol{(\beta B^T B + \delta I_{n_2})}(\tilde{y}^k - y^k) \boldsymbol{-B^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0$$

can be achieved by

$$\tilde{y}^k = \text{Argmin}\Big\{\theta_2(y) - y^T B^T[2\tilde{\lambda}^k - \lambda^k] + \frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \;\Big|\; y \in \mathcal{Y}\Big\}.$$

1.4 Linearized ADMM-Like Method

当子问题 (1.18c) 求解有困难时, 用 $\frac{s}{2}\|y-y^k\|^2$ 代替 $\frac{1+\delta}{2}\beta\|B(y-y^k)\|^2$.

By using the linearized version of (1.18), the prediction step becomes

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T H(v^k-\tilde{v}^k), \ \forall w\in\Omega, \quad (1.19)$$

where

$$H=\begin{bmatrix} sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}, \quad \text{代替 (1.17) 中的} \quad \begin{bmatrix} (1+\delta)\beta B^TB & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}. \quad (1.20)$$

The concrete formula of (1.19) is

$$\begin{cases} \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + \ \boldsymbol{s}(\tilde{y}^k-y^k)\boldsymbol{-B^T}(\tilde{\lambda}^k-\lambda^k)\} \ge 0, \\ (\underline{A\tilde{x}^k+B\tilde{y}^k-b})\boldsymbol{-B}(\tilde{y}^k-y^k)+\boldsymbol{(1/\beta)}(\tilde{\lambda}^k-\lambda^k)=0. \end{cases} \quad (1.21)$$

The underline part is $F(\tilde{w}^k)$:

$$\boldsymbol{F(w)}=\begin{pmatrix} -\boldsymbol{A^T\lambda} \\ -\boldsymbol{B^T\lambda} \\ \boldsymbol{Ax+By-b} \end{pmatrix}$$

Then, we use the form

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

to update the new iterate v^{k+1}.

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (1.21), we need only use the following procedure:

$$\begin{cases} \tilde{x}^k \in \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\theta_2(y) - y^T B^T [\boldsymbol{2\tilde{\lambda}^k - \lambda^k}] + \dfrac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \end{cases}$$

用 $\frac{s}{2}\|y-y^k\|^2$ 代替 $\frac{1}{2}(\beta\|B(y-y^k)\|^2 + \delta\|y-y^k\|^2)$, 为保证收敛, 需要 $s > \beta\|B^TB\|$. 对给定的 $\beta > 0$, 太大的 s 会影响收敛速度. 只有当由二次项 $\frac{1}{2}\beta\|B(y-y^k)\|^2$ 引发求解困难, 才用线性化方法.

1.5 Balanced ADMM

We design the following PPA in sense of balanced ADMM whose essential variables are w:

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (w-\tilde{w}^k)^T H(w^k-\tilde{w}^k), \ \forall w \in \Omega, \quad (1.22a)$$

where

$$H = \begin{pmatrix} \beta A^T A & 0 & A^T \\ 0 & sI_{n_2} & B^T \\ A & B & (\frac{1}{\beta}+\delta)I_m + \frac{1}{s}BB^T \end{pmatrix}. \quad (1.22b)$$

The concrete formula of (1.22) is

$$\begin{cases} \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k} \\ \qquad +\boldsymbol{\beta A^T A}(\tilde{x}^k-x^k) \quad + \boldsymbol{A^T}(\tilde{\lambda}^k-\lambda^k)\} \ge 0, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T\{\underline{-B^T\tilde{\lambda}^k} \\ \qquad + \boldsymbol{s}(\tilde{y}^k-y^k)\boldsymbol{+B^T}(\tilde{\lambda}^k-\lambda^k)\} \ge 0, \\ (\underline{A\tilde{x}^k+B\tilde{y}^k-b})+\boldsymbol{A}(\tilde{x}^k-x^k)+\boldsymbol{B}(\tilde{y}^k-y^k) \\ \qquad +\boldsymbol{\{(\frac{1}{\beta}+\delta)I_m+\frac{1}{s}BB^T\}}(\tilde{\lambda}^k-\lambda^k)=0. \end{cases} \quad (1.23)$$

Then, we use the form

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

to update the new iterate v^{k+1}.

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (1.23),

We need only use the following procedure:

$$\begin{cases} \tilde{x}^k \in \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|A(x - x^k)\|^2 \,|\, x \in \mathcal{X}\}, & \text{(1.24a)} \\ \tilde{y}^k \in \text{Argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}s\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}, & \text{(1.24b)} \\ \tilde{\lambda}^k = \lambda^k - H_0^{-1}[A(2\tilde{x}^k - x^k) + B(2\tilde{y}^k - y^k) - b]. & \text{(1.24c)} \end{cases}$$

where

$$H_0 = (\frac{1}{\beta} + \delta)I_m + \frac{1}{s}BB^T. \tag{1.24d}$$

In this way, we get a PPA for (1.13) in the sense of the balanced ADMM.

It needs a cholesky decomposition of H_0 (Levenberg-Marquardt Type)

If the subproblem (1.24a) is hard to solve, we can change it to (1.24) to

$$\left\{\begin{array}{ll} \tilde{x}^k \in \text{Argmin}\{\theta_1(x) - x^T A^T\lambda^k + \frac{1}{2}r\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & \text{(1.25a)} \\ \tilde{y}^k \in \text{Argmin}\{\theta_2(y) - y^T B^T\lambda^k + \frac{1}{2}s\|y - y^k\|^2 \mid y \in \mathcal{Y}\}, & \text{(1.25b)} \\ \tilde{\lambda}^k = \lambda^k - H_0^{-1}[A(2\tilde{x}^k - x^k) + B(2\tilde{y}^k - y^k) - b]. & \text{(1.25c)} \end{array}\right.$$

where

$$H_0 = \frac{1}{r}AA^T + \frac{1}{s}BB^T + \delta I_m. \qquad \text{(1.25d)}$$

We get a PPA for (1.13) with the following prediction:

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \qquad \text{(1.26a)}$$

where

$$H = \left(\begin{array}{ccc} rI_{n_1} & 0 & A^T \\ 0 & sI_{n_2} & B^T \\ A & B & \frac{1}{r}AA^T + \frac{1}{s}BB^T + \delta I_m \end{array}\right). \qquad \text{(1.26b)}$$

2 Prediction-Correction Framework

Prediction-correction framework for the VI (1.4)

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \quad (2.1a)$$

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).

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

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \quad (2.1b)$$

Convergence conditions

For the matrices Q and M used in (2.1a) and (2.1b), respectively, there exists a matrix $H \succ 0$ such that

$$HM = Q, \quad (2.2a)$$

and

$$G := Q^T + Q - M^T HM \succ 0. \quad (2.2b)$$

2.1 Convergence

Theorem 1 *Let $\{v^k\}$ be the sequence generated by the prediction-correction framework (2.1) under the conditions (2.2). Then, it holds that*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.3}$$

Proof. Using $Q = HM$ (see (2.2a)), the prediction step can be written as

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

Then, it follows from (2.1b) 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 $(\tilde{w}^k - w^*)^T F(\tilde{w}^k) = (\tilde{w}^k - w^*)^T F(w^*)$ (see (1.5)), it follows from the optimality of w^* that $\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \ge 0$ and thus

$$(v^k - v^{k+1})^T H(\tilde{v}^k - v^*) \ge 0, \quad \forall v^* \in \mathcal{V}^*. \tag{2.4}$$

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

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

we know from (2.4) that

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2. \tag{2.5}$$

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

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&\quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\
&\quad \overset{(2.1b)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
&\quad = \quad 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\
&\quad = \quad (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\
&\quad \overset{(2.2b)}{=} \quad \|v^k - \tilde{v}^k\|_G^2.
\end{aligned} \tag{2.6}$$

Substituting (2.6) in (2.5), the assertion (2.3) is proved. $\square$

3 Constructing M from the conditions

3.1 Construction from the condition $(2.2a)$

Note that the condition (2.2a), $HM = Q$, can be rewritten as

$$H = QM^{-1}. \tag{3.1}$$

Since the norm matrix H is required to be symmetric and positive definite, the condition (3.1) implies that H should be representable in form of

$$H = QD^{-1}Q^T, \tag{3.2}$$

in which the matrix D is a undetermined positive definite matrix. Indeed, by comparing (3.1) with (3.2), we know that $M^{-1} = D^{-1}Q^T$ and thus

$$M = Q^{-T}D. \tag{3.3}$$

Hence, although the matrix D in (3.2) is still unknown, choosing M as (3.3) can ensure the condition (2.2a).

Now, we investigate the restriction on D to ensure the condition (2.2b) with the matrix M given as (3.3). Notice that

$$M^T HM = \big(DQ^{-1}\big)\big(QD^{-1}Q^T\big)\big(Q^{-T}D\big) = D. \tag{3.4}$$

With (3.4), then the condition (2.2b) is reduced to

$$G := Q^T + Q - M^T HM = Q^T + Q - D \succ 0. \tag{3.5}$$

Hence, to ensure the condition (2.2b), the only restriction on the positive definite matrix D in (3.2) is

$$0 \prec D \prec Q^T + Q, \tag{3.6}$$

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:

$$\begin{cases} HM = Q, \\ M^T HM = D. \end{cases} \Leftrightarrow \begin{cases} HM = Q, \\ Q^T M = D. \end{cases} \Leftrightarrow \begin{cases} H = QD^{-1}Q^T, \\ M = Q^{-T}D. \end{cases} \tag{3.7}$$

Through this construction, both the conditions (2.1b) and (2.2b) are guaranteed to be satisfied. Note that once the matrix D is chosen according to (3.6), the matrices H, M and G are all uniquely determined. Then, with the specified matrix M in (3.3), the correction step (2.1b) and thus the prediction-correction framework (2.1) is also specified as a concrete contraction splitting algorithm for the VI(1.4)-(1.4b).

3.2 Construction from the condition $(2.2b)$

Alternatively, we can from the condition (2.2b), $G = Q^T + Q - M^T HM \succ 0$, 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

$$0 \prec G \prec Q^T + Q. \tag{3.8}$$

Denote

$$\Delta = Q^T + Q - G, \tag{3.9}$$

which is positive definite. According to (2.2b), we know that the matrices H and M should satisfy

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

Recall the condition (2.2a): $HM = Q$. Thus, with a chosen G satisfying (3.8), H and M can be constructed vis the following steps:

$$\begin{cases} M^T HM = \Delta, \\ \quad HM = Q. \end{cases} \Leftrightarrow \begin{cases} Q^T M = \Delta, \\ HM = Q. \end{cases} \Leftrightarrow \begin{cases} M = Q^{-T}\Delta, \\ H = Q\Delta^{-1}Q^T. \end{cases} \tag{3.10}$$

Then, with the constructed matrix M in (3.10), the correction step (2.1b) and thus the prediction-correction framework (2.1) can also be specified as a concrete splitting contraction algorithm for the VI(1.4)-(1.4b). Again, with a given G satisfying (3.8), the matrices H and M are both uniquely determined.

3.3 Choices of D and G

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

$$D \succ 0, \quad G \succ 0, \quad \text{and} \quad D + G = Q^T + Q. \tag{3.11}$$

Hence, once D is chosen for the construction strategy in Section 3.1, the corresponding G given by (3.11) can be used for the construction strategy in Section 3.2, and vice versa.

Technically, there are infinitely many such choices subject to (3.11). For example, we can choose

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

We will elaborate on the choice $D = G = \frac{1}{2}[Q^T + Q]$ in Section 4.3.3.

3.4 Implementation of the correction step (2.1b)

Note that the correction step (2.1b) can be rewritten as

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

To implement the correction step (2.1b) with the constructed two choices for M, i.e., $M = Q^{-T}D$ in (3.3) and $M = Q^{-T}\Delta$ in (3.10), we need to solve one of the following systems of equations:

$$Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k), \tag{3.12}$$

and

$$Q^T(v^{k+1} - v^k) = \Delta(\tilde{v}^k - v^k). \tag{3.13}$$

Hence, although D and G (thus Δ) can be chosen arbitrarily with the only constraint (3.6) or (3.8), it is preferred to choose some model-tailored ones that can favor solving the systems of equations (3.12) or (3.13) more efficiently.

4 Application to three-block separable convex optimization

In this section, we apply the strategies proposed in Sections 3.1 and 3.2 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.

4.1 Model

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

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)\,|\,Ax+By+Cz=b,\; x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\}, \quad (4.1)$$

Clearly, it is a special case of the canonical convex programming problem (1.1), and the VI (1.4)-(1.4b) can be specified as the following:

$$w^*\in\Omega, \quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\geq 0, \quad \forall\, w\in\Omega, \qquad (4.2)$$

where

$$w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix}, \tag{4.3a}$$

with

$$\theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z), \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m. \tag{4.3b}$$

Let the augmented Lagrangian function of the model (4.1) be

$$\begin{aligned} \mathcal{L}_\beta^{[3]}(x, y, z, \lambda) \;=\; & \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^T(Ax + By + Cz - b) \\ & + \frac{\beta}{2}\|Ax + By + Cz - b\|^2 \end{aligned} \tag{4.4}$$

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

Direct extension of ADMM

$$\begin{cases} x^{k+1} \in \arg\min\{\mathcal{L}_\beta^{[3]}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} \in \arg\min\{\mathcal{L}_\beta^{[3]}(x^{k+1}, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\}, \\ z^{k+1} \in \arg\min\{\mathcal{L}_\beta^{[3]}(x^{k+1}, y^{k+1}, z, \lambda^k) \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \big(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b\big). \end{cases} \tag{4.5}$$

However, the splitting scheme (4.5) is coarse in sense of that its convergence is not guaranteed as shown in [2].

4.2 Discerning the prediction matrix Q

Our construction starts from the coarse splitting scheme (4.5) which can be rewritten as the prediction step (2.1a) and hence the corresponding prediction matrix Q can be discerned. For this purpose, we first consider the subproblems related to the primal variables in (4.5), 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

$$\left\{ \begin{array}{l} \tilde{x}^k \in \arg\min\big\{\mathcal{L}_\beta^{[3]}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \arg\min\big\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\big\}, \\ \tilde{z}^k \in \arg\min\big\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, \tilde{y}^k, z, \lambda^k) \mid z \in \mathcal{Z}\big\}. \end{array} \right. \tag{4.6}$$

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

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

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

$$\left\{ \begin{array}{l} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \big\{ -A^T \lambda^k \\ \qquad\qquad\qquad +\beta A^T \big(A\tilde{x}^k + By^k + Cz^k - b \big) \big\} \geq 0, \;\; \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \big\{ -B^T \lambda^k \\ \qquad\qquad\qquad +\beta B^T \big(A\tilde{x}^k + B\tilde{y}^k + Cz^k - b \big) \big\} \geq 0, \;\; \forall\, y \in \mathcal{Y}, \\ \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T \big\{ -C^T \lambda^k \\ \qquad\qquad\qquad +\beta C^T \big(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b \big) \big\} \geq 0, \;\; \forall\, z \in \mathcal{Z}. \end{array} \right. \tag{4.7}$$

By defining

$$\tilde{\lambda}^k = \lambda^k - \beta \big(A\tilde{x}^k + By^k + Cz^k - b \big), \tag{4.8}$$

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 (4.3), we get $\tilde{w}^k \in \Omega$ and

$$\left\{\begin{array}{l} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \quad \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{\underline{-B^T\tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k)\} \ge 0, \quad \forall\, y \in \mathcal{Y}, \\ \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T\left\{\begin{array}{c} \underline{-C^T\tilde{\lambda}^k} + \beta C^T B(\tilde{y}^k - y^k) \\ \beta C^T C(\tilde{z}^k - z^k) \end{array}\right\} \ge 0, \quad \forall\, z \in \mathcal{Z}, \\ (\lambda - \tilde{\lambda}^k)^T\left\{\begin{array}{l} (\underline{A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b}) \\ -B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \end{array}\right\} \ge 0, \quad \forall\, \lambda \in \Lambda. \end{array}\right. \tag{4.9}$$

The sum of the underline parts of (4.9) is exactly $F(\tilde{w}^k)$, where $F(\cdot)$ is defined in (4.3). Thus, we have

$$\tilde{w}^k \in \Omega,\ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k),\ \forall w \in \Omega, \tag{4.10}$$

where the prediction matrix is

$$Q = \begin{pmatrix} \beta B^T B & 0 & 0 \\ \beta C^T B & \beta C^T C & 0 \\ -B & -C & \dfrac{1}{\beta} I_m \end{pmatrix}. \tag{4.11}$$

Moreover, for the prediction matrix Q in (4.11) which is determined by the splitting scheme (4.6) and the defined $\tilde{\lambda}^k$ by (4.8), we have

$$Q^T + Q = \begin{pmatrix} 2\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & 2\beta C^T C & -C^T \\ -B & -C & \frac{2}{\beta} I_m \end{pmatrix}, \tag{4.12}$$

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

4.3 Constructing the correction matrix M

With the prediction matrix Q given in (4.11), the prediction-correction framework (2.1) can be specified as a concrete algorithm for the model (4.1) once the correction step (2.1b) is specified. Now, we showcase how to specify the correction step (2.1b) by the construction strategies discussed in Sections 3.1, 3.2 and 3.3. 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.

4.3.1 Construction I

Based on (4.11) and (4.12), and following the strategy in Section 3.1, we can choose

$$D = \begin{pmatrix} \nu\beta B^T B & 0 & 0 \\ 0 & \nu\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \tag{4.13}$$

with $0 < \nu < 1$, which is positive definite whenever B and C are both full column rank. Recall the correction matrix M in (3.3). Then, a concrete splitting contraction algorithm for (4.1) can be generated as below.

Algorithm 1 for the model (4.1)

[Prediction Step.] Obtain $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$ via the direct extension of the ADMM (4.6) and define $\tilde{\lambda}^k$ by (4.8).

[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

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

and

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

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

Hence, to implement the correction step (2.1b), 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.

4.3.2 Construction 2

Based on (4.11) and (4.12), and following the strategy in Section 3.2, we can choose

$$G = \begin{pmatrix} (1-\nu)\beta B^T B & 0 & 0 \\ 0 & (1-\nu)\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}, \tag{4.15}$$

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 (4.15) is precisely the matrix D defined in (4.13). 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 (3.10). Then, another contraction splitting algorithm for (4.1) can be generated as below.

Algorithm 2 for the model (4.1)

[Prediction Step.] Obtain $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$ via the direct extension of the ADMM (4.6) define $\tilde{\lambda}^k$ by (4.8).

[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

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

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

Hence, to implement the correction step (2.1b), 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

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

$$= \begin{pmatrix} (1+\nu)B & C & -\frac{1}{\beta}I_m \\ B & (1+\nu)C & -\frac{1}{\beta}I_m \\ -\beta B & -\beta C & I_m \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{z}^k - z^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}. \tag{4.17}$$

The above system of equations equivalent to

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

$$= \begin{pmatrix} (1+\nu)I_m & I_m & -\frac{1}{\beta}I_m \\ I_m & (1+\nu)I_m & -\frac{1}{\beta}I_m \\ -\beta I_m & -\beta I_m & I_m \end{pmatrix} \begin{pmatrix} B\tilde{y}^k - By^k \\ C\tilde{z}^k - Cz^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}. \tag{4.18}$$

Similar as (4.3.1), with the choice of G in (4.15), implementing the resulting correction step (2.1b) essentially only requires solving the equation (4.18) in terms of (By^k, Cz^k, λ^k). By a manipulation, the correction form can be simplified to

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

4.3.3 Construction 3

Recall the relationship between the matrices D and G in (3.11), and $Q^T + Q$ given in (4.12). Essentially, the proposed construction strategies in Sections 3.1

and 3.2 take the same matrix

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

as D and G, respectively, and then the other one is determined by (3.11). As mentioned in Section 3.3, any other choice of D and G subject to the relationship (3.11) is also eligible. Let us consider the following specific one:

$$D = G = \frac{1}{2}\left[Q^T + Q\right] = \begin{pmatrix} \beta B^T B & \frac{1}{2}\beta B^T C & -\frac{1}{2} B^T \\ \frac{1}{2}\beta C^T B & \beta C^T C & -\frac{1}{2} C^T \\ -\frac{1}{2} B & -\frac{1}{2} C & \frac{1}{\beta} I_m \end{pmatrix}, \quad (4.19)$$

which are both positive definite whenever B and C are full column rank. Recall the correction matrix M in (3.10). Then, one more contraction splitting algorithm for (4.1) can be generated as below.

Algorithm 3 for the model (4.1)

[Prediction Step.] Obtain $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$ via the direct extension of the ADMM (4.6) and define $\tilde{\lambda}^k$ by (4.8).

[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

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

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

$$
\begin{aligned}
&\begin{pmatrix} B & C & -\frac{1}{\beta}I_m \\ 0 & C & -\frac{1}{\beta}I_m \\ 0 & 0 & I_m \end{pmatrix}
\begin{pmatrix} y^{k+1}-y^k \\ z^{k+1}-z^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} \\
= &\begin{pmatrix} B & \frac{1}{2}C & -\frac{1}{2\beta}I_m \\ \frac{1}{2}B & C & -\frac{1}{2\beta}I_m \\ -\frac{1}{2}\beta B & -\frac{1}{2}\beta C & I_m \end{pmatrix}
\begin{pmatrix} \tilde{y}^k-y^k \\ \tilde{z}^k-z^k \\ \tilde{\lambda}^k-\lambda^k \end{pmatrix}.
\end{aligned} \tag{4.21}
$$

The equivalent form

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

5 p-block separable convex optimization problems

In the following we consider the multiple-block convex optimization:

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i)\;\Big|\;\sum_{i=1}^{p}A_ix_i = b\;(\text{or} \ge b),\quad x_i\in\mathcal{X}_i\Big\}. \tag{5.1}$$

The Lagrangian function is

$$L(x_1,\ldots,x_p,\lambda)=\sum_{i=1}^{p}\theta_i(x_i)-\lambda^T\Big(\sum_{i=1}^{p}A_ix_i-b\Big),$$

which is defined on $\Omega=\prod_{i=1}^{p}\mathcal{X}_i\times\Lambda$, where

$$\Lambda=\begin{cases}\Re^m, & \text{if } \sum_{i=1}^{p}A_ix_i=b,\\[2ex] \Re^m_+, & \text{if } \sum_{i=1}^{p}A_ix_i\ge b.\end{cases}$$

Let $(x_1^*, \ldots, x_p^*, \lambda^*) \in \Omega$ be a saddle point of the Lagrangian function, then

$$L_{\lambda \in \Lambda}(x_1^*, \ldots, x_p^*, \lambda) \le L(x_1^*, \ldots, x_p^*, \lambda^*) \le L_{x_i \in \mathcal{X}_i}(x_1, \ldots, x_p, \lambda^*).$$

The optimality condition of (5.1) can be written as the following VI:

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

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T \lambda \\ \vdots \\ -A_p^T \lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}, \tag{5.2b}$$

and

$$\theta(x) = \sum_{i=1}^p \theta_i(x_i), \qquad \Omega = \prod_{i=1}^p \mathcal{X}_i \times \Lambda.$$

Again, we denote by Ω^* the solution set of the VI (5.2).

Primal-dual prediction for p-block problems

A Primal-Dual prediction for (5.1) . A natural consideration of Prediction

With given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, find $\tilde{w}^k \in \Omega$ via

$$\left\{\begin{array}{l} \tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^TA_1^T\lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\ \tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^TA_2^T\lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\ \qquad\vdots \\ \tilde{x}_i^k \in \arg\min_{x_i\in\mathcal{X}_i}\{\theta_i(x_i) - x_i^TA_i^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1}A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\ \qquad\vdots \\ \tilde{x}_p^k \in \arg\min_{x_p\in\mathcal{X}_p}\{\theta_p(x_p) - x_p^TA_p^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}; \\ \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^{p}A_j\tilde{x}_j^k - b\big)\big]. \end{array}\right. \tag{5.3}$$

预测先原始再对偶. 对可分离的原始变量子问题逐一按序求解.

Analysis for the P-D Prediction First, for the primal part of the predictor,

$$\tilde{x}_i^k \in \arg\min\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2} \| \sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 | x_i \in \mathcal{X}_i\}.$$

According to the optimality lemma, the optimal condition is $\tilde{x}_i^k \in \mathcal{X}_i$ and

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{-A_i^T\lambda^k + \beta A_i^T\big(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)\big)\} \geq 0,$$

for all $x_i \in \mathcal{X}_i$. It can be written as $\tilde{x}_i^k \in \mathcal{X}_i$ and

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{\underline{-A_i^T\tilde{\lambda}^k} + \beta A_i^T\big(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)\big) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \tag{5.4a}$$

for all $x_i \in \mathcal{X}_i$. The dual part of the predictor, $\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^{p} A_j\tilde{x}_j^k - b\big)\big]$,

$$\tilde{\lambda}^k = \arg\min\{\big\|\lambda - \big[\lambda^k - \beta\big(\textstyle\sum_{j=1}^{p} A_j\tilde{x}_j^k - b\big)\big]\big\|^2 \mid \lambda \in \Lambda\}.$$

The optimal condition is

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T\{\big(\underline{\textstyle\sum_{j=1}^{p} A_j\tilde{x}_j^k - b}\big) + \tfrac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \quad \forall \lambda \in \Lambda. \tag{5.4b}$$

Summating (5.4a) and (5.4b), for the predictor $\tilde{w}^k$ generated by (5.3), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \geq (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \tag{5.5a}$$

where

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{5.5b}$$

6 Translation of the correction variables

The optimization problem (5.1) has been translated to VI (5.2), namely,

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

For the easy analysis, we need to denote the following notations:

$$P = \begin{pmatrix} \sqrt{\beta}A_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta}A_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta}A_p & 0 \\ 0 & \cdots & \cdots & 0 & (1/\sqrt{\beta})I_m \end{pmatrix}, \quad \xi = Pw = \begin{pmatrix} \sqrt{\beta}A_1x_1 \\ \sqrt{\beta}A_2x_2 \\ \vdots \\ \sqrt{\beta}A_px_p \\ (1/\sqrt{\beta})\lambda \end{pmatrix}. \tag{6.1}$$

Accordingly, we define

$$\Xi = \big\{\xi \mid \xi = Pw, \ w \in \Omega\big\},$$

and

$$\Xi^* = \big\{\xi^* \mid \xi^* = Pw^*, \ w^* \in \Omega^*\big\}.$$

Using the notation P in (6.1), for the matrix Q in (5.5b), we have

$$Q = P^T \mathcal{Q} P, \qquad \text{where} \qquad \mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{6.2}$$

Thus, for the right hand side of (5.5a), we have

$$\begin{aligned} (w - \tilde{w}^k)^T Q (w^k - \tilde{w}^k) &= (w - \tilde{w}^k)^T P^T \mathcal{Q} P (w^k - \tilde{w}^k) \\ &= (\xi - \tilde{\xi}^k)^T \mathcal{Q} (\xi^k - \tilde{\xi}^k). \end{aligned}$$

Then, it follows from (5.5) that we have the following VI for the P-D prediction:

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) &+ (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ &\geq (\xi - \tilde{\xi}^k)^T \mathcal{Q} (\xi^k - \tilde{\xi}^k), \quad \forall\, w \in \Omega. \end{aligned} \tag{6.3}$$

where $\mathcal{Q}$ is given in (6.2).

Prediction-Correction Framework for VI (5.2).

1. (Prediction Step) With given w^k and $\xi^k = Pw^k$, find $\tilde{w}^k \in \Omega$ such that

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k), \quad \forall\, w \in \Omega, \tag{6.4a}$$

   with $\mathcal{Q} \in \Re^{(p+1)m\times(p+1)m}$, and the matrix $\mathcal{Q}^T + \mathcal{Q}$ is positive definite.

2. (Correction Step) With the predictor $\tilde{w}^k$ by (6.4a) and $\tilde{\xi}^k = P\tilde{w}^k$, the new iterate ξ^{k+1} is updated by

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k), \tag{6.4b}$$

   where $\mathcal{M} \in \Re^{(p+1)m\times(p+1)m}$ is a non-singular matrix.

Theorem 2 *For the matrices* $\mathcal{Q}$ *and* $\mathcal{M}$ *in the algorithm* (6.4)*, if there is a positive definite matrix* $\mathcal{H} \in \Re^{(p+1)m\times(p+1)m}$ *such that*

$$\mathcal{H}\mathcal{M} = \mathcal{Q} \tag{6.5a}$$

and

$$\mathcal{G} := \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0, \tag{6.5b}$$

then we have

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall\, \xi^* \in \Xi^*. \tag{6.6}$$

Proof. Setting w in (6.4a) as any fixed $w^* \in \Omega^*$, and using

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) \equiv (\tilde{w}^k - w^*)^T F(w^*),$$

we get

$$(\tilde{\xi}^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*), \;\; \forall\, w^* \in \Omega^*.$$

The right-hand side of the last inequality is non-negative. Thus, we have

$$(\xi^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \geq (\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k), \quad \forall\, \xi^* \in \Xi^*. \qquad (6.7)$$

Then, by simple manipulations, we obtain

$$\begin{aligned}
&\|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \\
&\stackrel{(6.4b)}{=} \quad \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|(\xi^k - \xi^*) - \mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&\stackrel{(6.5a)}{=} \quad 2(\xi^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) - \|\mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&\stackrel{(6.7)}{\geq} \quad 2(\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) - \|\mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&= \quad (\xi^k - \tilde{\xi}^k)^T [(\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T \mathcal{H} \mathcal{M}](\xi^k - \tilde{\xi}^k) \\
&\stackrel{(6.5b)}{=} \quad \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2.
\end{aligned}$$

The assertion of this theorem is proved. $\square$

We call (6.5) the convergence conditions for the algorithm framework (6.4).

The inequality (6.6) is the key for the convergence proofs, for details, see [14]

7 A special correction

For given $\mathcal{Q}$ which satisfies $\mathcal{Q}^T + \mathcal{Q} \succ 0$, we chose $\mathcal{D}$ and $\mathcal{G}$, such that

$$\mathcal{D} \succ 0, \qquad \mathcal{G} \succ 0, \qquad \mathcal{D} + \mathcal{G} = \mathcal{Q}^T + \mathcal{Q}.$$

Then, the correction matrix $\mathcal{M}$ in (6.4b) is given by

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

选择了想要的 $0 \prec \mathcal{D}$, 构造 $\mathcal{M}$ 不再神秘！ 下面先介绍以前在 [14] 中 “凑” 出来的 $\mathcal{M}$

First, we give some correction examples which satisfy conditions (6.5) in Theorem 2.

In order to simplify the notations to be used, we define the following $p \times p$ block matrices:

$$\mathcal{L} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}, \quad \mathcal{I} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{7.1}$$

We also define the $1 \times p$ block matrix

$$\mathcal{E} = \begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix}. \tag{7.2}$$

Using the notations (7.1)-(7.2), the matrix $\mathcal{Q}$ in (6.2) has the form

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} \quad \text{and} \quad \mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix}. \tag{7.3}$$

In order to construct a convergent algorithm, we need only to give the matrices $\mathcal{M}$ and $\mathcal{H}$ and to verify the convergence conditions (6.5)

By setting

$$\mathcal{M} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{7.4}$$

For the above matrices $\mathcal{Q}$ and $\mathcal{M}$, the remaining tasks is to find a positive definite matrix $\mathcal{H}$, such that the convergence conditions (6.5)are satisfied.

(7.4) 中的 $\mathcal{M}$ 是我们在 [14] 中 “凑” 出来的.

How to improvise a correction matrix $\mathcal{M}$? Because $\mathcal{H}\mathcal{M} = \mathcal{Q}$,

$$\mathcal{H} = \mathcal{Q}\mathcal{M}^{-1}.$$

There is a block lower triangular matrix $\mathcal{M}$ which satisfies the convergence condition ?

The inverse of a triangular matrix is also a triangular matrix.

$$\mathcal{M}^{-1} = \begin{pmatrix} X & 0 \\ Y & I_m \end{pmatrix}.$$

$\mathcal{Q}\mathcal{M}^{-1}$ should be a symmetric matrix

$$\mathcal{H} = \mathcal{Q}\mathcal{M}^{-1} = \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} \begin{pmatrix} X & 0 \\ Y & I_m \end{pmatrix} = \begin{pmatrix} \mathcal{L}X + \mathcal{E}^T Y & \mathcal{E}^T \\ Y & I_m \end{pmatrix}.$$

Thus, $Y = \mathcal{E}$ and $X = S^{-1}\mathcal{L}^T$, S is a undetermined positive definite matrix.

$$\mathcal{M}^{-1} = \begin{pmatrix} S^{-1}\mathcal{L}^T & 0 \\ \mathcal{E} & I_m \end{pmatrix} \quad \text{and thus} \quad \mathcal{M} = \begin{pmatrix} \mathcal{L}^{-T}S & 0 \\ -\mathcal{E}\mathcal{L}^{-T}S & I_m \end{pmatrix}.$$

继续“凑”下去, 发现 $S = \nu I$ 就可以了, 我们因此也凑出了 $\mathcal{H}$.

Lemma 3 *For the matrices $\mathcal{Q}$ and $\mathcal{M}$ given by* (7.3) *and* (7.4)*, respectively, the matrix*

$$\mathcal{H} = \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix} \quad \textit{with} \quad \nu \in (0,1) \tag{7.5}$$

is positive definite, and it satisfies $\mathcal{H}\mathcal{M} = \mathcal{Q}$.

Proof. It is easy to check the positive definiteness of $\mathcal{H}$. In addition, for the block matrix $\mathcal{Q}$ in (6.2), we have

$$\begin{aligned} \mathcal{H}\mathcal{M} &= \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix} \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} = \mathcal{Q}. \end{aligned}$$

The assertions of this lemma are proved. $\square$

这样凑出来的 $\mathcal{M}$ 和 $\mathcal{H}$, 能否满足 $Q^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0$？ 还需要检查一下.

Lemma 4 *Let* $\mathcal{Q}$, $\mathcal{M}$ *and* $\mathcal{H}$ *be defined in* (6.2), (7.4) *and* (7.5), *respectively. Then the matrix*

$$\mathcal{G} := (\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T \mathcal{H} \mathcal{M} \tag{7.6}$$

is positive definite.

Proof. By elementary matrix multiplications, we know that

$$\mathcal{M}^T \mathcal{H} \mathcal{M} = \mathcal{Q}^T \mathcal{M} = \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E} & I_m \end{pmatrix} \begin{pmatrix} \nu \mathcal{L}^{-T} & 0 \\ -\nu \mathcal{E} \mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} \nu \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} = \mathcal{D}.$$

Then, it follows from $\mathcal{L}^T + \mathcal{L} = \mathcal{I} + \mathcal{E}^T \mathcal{E}$ (see (7.1)-(7.2)) that

$$\begin{aligned} \mathcal{G} &= (\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T \mathcal{H} \mathcal{M} \\ &= \begin{pmatrix} \mathcal{L}^T + \mathcal{L} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix} - \begin{pmatrix} \nu \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T \mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix}. \end{aligned}$$

Thus, the matrix $\mathcal{G}$ is positive definite for any $\nu \in (0, 1)$. $\square$

Finally, correction step can be written

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k). \tag{7.7}$$

Lemma 3 and Lemma 4 have verified the convergence conditions (6.5) and thus the key convergence inequality (6.6) holds. The algorithm (5.3) & (7.7) is convergent.

Recall the respective definitions $\mathcal{L}$ and $\mathcal{E}$ in (7.1) and (7.2). We have

$$\mathcal{L}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix}$$

and

$$\mathcal{E}\mathcal{L}^{-T} = \begin{pmatrix} I_m & 0 & \cdots & 0 \end{pmatrix}.$$

Thus

$$\mathcal{M} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu I_m & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{7.8}$$

By a manipulation, we have

$$\begin{pmatrix} A_1 x_1^{k+1} \\ A_2 x_2^{k+1} \\ \vdots \\ A_p x_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1 x_1^{k} \\ A_2 x_2^{k} \\ \vdots \\ A_p x_p^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & 0 \\ 0 & \nu I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -\nu I_m \\ 0 & \cdots & 0 & \nu I_m \end{pmatrix} \begin{pmatrix} A_1 x_1^k - A_1 \tilde{x}_1^k \\ A_2 x_2^k - A_2 \tilde{x}_2^k \\ \vdots \\ A_p x_p^k - A_p \tilde{x}_p^k \end{pmatrix}, \tag{7.9}$$

and

$$\lambda^{k+1} = \tilde{\lambda}^k + \nu\beta(A_1 x_1^k - A_1 \tilde{x}_1^k). \tag{7.10}$$

8 More Choices based on the predictions

只要 $\mathcal{Q}^{-T}$ 结构简单, 构造校正矩阵 $\mathcal{M}$ 的方法并不神秘！是非常容易的.

The matrix $\mathcal{Q}$ in (6.2) has the form

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} \quad \text{and thus} \quad \mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix}.$$

To further analyze the correction steps associated with the correction matrix $\mathcal{M}$, let us take a closer look at the matrix $\mathcal{Q}^{-T}$.

According to the primal-dual prediction (5.3), the matrix $\mathcal{Q}$ in (6.2), we have

$$\mathcal{Q}^{-T} = \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E} & I_m \end{pmatrix}^{-1} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{8.1}$$

and

$$\begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

The calculation $\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D}$ is essentially very easy for different $\mathcal{D}$!

Since

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix},$$

it can be decomposed as

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix}.$$

The both matrices in the right hand side are positive definite. If we chose

$$\mathcal{D} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} \quad \text{and thus} \quad \mathcal{G} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix},$$

it is just the correction in Section §7.

Conversely, we can also choose

$$\mathcal{D} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix} \quad \text{and thus} \quad \mathcal{G} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}$$

and thus get the another correction method.

There are many positive definite decompositions of $\mathcal{Q}^T + \mathcal{Q}$, for example,

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} (1-\nu)\mathcal{I} & 0 \\ 0 & (1-\nu)I_m \end{pmatrix} + \begin{pmatrix} \nu\mathcal{I} + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & (1+\nu)I_m \end{pmatrix}.$$

and

$$\mathcal{Q}^T + \mathcal{Q} = \mathcal{D} + \mathcal{G} = \alpha(\mathcal{Q}^T + \mathcal{Q}) + (1-\alpha)(\mathcal{Q}^T + \mathcal{Q}), \quad \alpha \in (0,1).$$

9 Conclusions

- 我的学术报告中常用的一个题目是“构造凸优化的分裂收缩算法– 用好 VI 和 PPA 两大法宝”, 是指构造变分不等式意义下的 PPA 算法, 文章首先发表在 [16]. 后来又做了一些人为地将预测矩阵设计成对称正定矩阵的方法 [1, 5], 包括我们 2021 年才提出的均困平衡的增广拉格朗日乘子法 [19]. 有时我们也称这样的方法为按需定制的 PPA - (Customized PPA).
- 对预测矩阵 Q 为非对称的预测-校正方法, 利用统一框架的套路证明收敛性, 最初出现在我和袁晓明 (Xiaoming Yuan) 2012 年 SIAM 数值分析的文章 [15] 中, 后面我们发表的一些论文 [2, 10, 11, 12, 18], 都用这个套路证明收敛性. 把它归结为统一框架, 是在南京大学讨论班上, 那是在我 2013 年即将退休之前, 以后便常常出现在我的“讲习班”讲义和报告的 PPT 中.
- 第一次在正式出版物里提到这个统一框架, 是在 2016 年《高校计算数学学报》的我的中文文章 [6] 中. 2018 年我在《运筹学学报》的综述文章“我和乘子交替方向法 20 年”[7] 中指出, 我们发表的方法都可以用这个框架非常简单地证明收敛性. 英文出版物中首次出现统一框架的是我和袁晓明 2018 年在 COAP 的文章 [17].
- 从 2018 年开始, 我在自己的报告和论文 [9] 中, 经常讲用统一框架去构造算法主要还是按收敛条件去“凑”. 如何根据确定的预测矩阵 $\boldsymbol{Q}$ 凑出满足收敛条件的校正矩阵 $\boldsymbol{M}$. 似乎给人一种难以效仿的神秘感觉.

- 2022 年初我在南师大做报告时有人问过. 最近我又在中科大和南航做线上报告, 教学相长, 得到一些新的看法, 觉得有必要将回答整理成下面的材料与听众共享.
- 我们从预测矩阵满足 $Q^T + Q \succ 0$ 出发. 根据条件 (2.2a), $HM = Q$, 我们有

$$H = QM^{-1}.$$

因为 H 是正定矩阵, 必须对称. 从上式又看到, H 有个左因子 Q, 那它必须有个右因子 Q^T, 中间夹一个"待定的"正定矩阵. 我们设这个正定矩阵为 D^{-1}, 则有

$$H = QD^{-1}Q^T.$$

比较上面两式, 我们得到 $M^{-1} = D^{-1}Q^T$, 因此

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

这样, 条件 (2.2a) 满足. 这个我们大概在 10 年前就知道. 当时往往考虑选择的 D 应该是个块对角矩阵.

- 至此, 我们还不知道矩阵 D 具体形式是什么. 计算一下收敛性条件中的 M^THM,

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

上式已经出现在我 2018 的暑期讲习班的讲义中, 没有向前再迈一步.

- 利用上式和 (2.2b), 这个待定的正定矩阵 D 只需要满足

$$\mathbf{0} \prec \boldsymbol{D} \prec \boldsymbol{Q}^{\boldsymbol{T}} + \boldsymbol{Q} \qquad （因此, 0 \prec G = Q^T + Q - D）$$

  就可以了. 明确这一条, 得益于为 2022 年以来在 南师大, 南航 和 中科大 讲课, 迫使我深入思考, 把方法讲明白.

- 在选了满足上述条件的矩阵 D 以后, 根据确定的 Q 和 D, 找未知矩阵 H 和 M 使得

$$HM = Q \qquad 和 \qquad M^T HM = D,$$

  我们的目的就达到了.

- 这样的 M 和 H: 可以通过求解下面的矩阵方程组得到.

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

- 选择不同的满足条件的矩阵 D (这非常容易), 就有不同的校正方法. 譬如说,

$$D = \alpha[Q^T + Q], \ \alpha \in (0, 1).$$

- 报告的第 5节开始, 对一般线性约束凸优化问题, 采用 primal-dual 预测, 子问题的求解方式是ADMM 类型的逐个向前. 我们需要的 $\boldsymbol{Q}^{-\boldsymbol{T}}$ 形式非常简单. 是的, 它需要额外的校正. 可喜的是, 校正花费很少, 又特别容易实现!

- 我们特别推崇“预测-校正”, 尤其是那种代价很小的校正. 生机勃勃的果树, 修剪就是校正. 社会治理也是一种校正, 当然也考虑成本！ 交替按序预测, 降低了问题难度; 全局整体校正, 把握了收敛方向.
- 预测-校正方法既可以用来求解等式约束的问题, 又可以用来求解不等式约束的问题. 适用从一块到任意多块的可分离问题, 算法结构和收敛性证明完全统一.
- 适用范围广的算法会不会影响效率？ 对经典 ADMM 擅长的两块可分离的等式约束凸优化问题, 我们用 5- 节本文提到的带校正的交替方向法去求解, 与网上他人提供的 ADMM 代码比较, 发现这种担心是多余的.
- **Question A.** In the prediction step, how to arrange a “good” prediction matrix whose matrix $\mathcal{Q}$ satisfies
$$\mathcal{Q}^T + \mathcal{Q} \succeq \mathcal{I}.$$
- **Question B** For the given prediction matrix $\mathcal{Q}$, what are the criteria for choosing matrix $\mathcal{D}$ which satisfies
$$0 \prec \mathcal{D} \prec \mathcal{Q}^T + \mathcal{Q}.$$

希望各位以质疑的态度审视我的观点, 对的就相信, 不对的请批评指正.

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]