求解线性约束凸优化问题分裂收缩算法
统一框架的五个两页短篇

数学之美, 不是纯数学的专利. 对最优化方法的研究, 我们一贯追求简单与统一. 简单, 他人才会拿去使用, 统一, 自己才有美的享受. 科学与工程计算中出现的凸优化问题, 很多是带线性约束的. 引入乘子以后, 问题就可以归结为求其拉格朗日(Lagrange)函数的鞍点. 鞍点, 犹如利益冲突双方的平衡点, 迭代求解, 好比双方谈判, 需要相向而行. 鞍点的等价数学表达形式是变分不等式(VI), 基于这种考虑, 过去的十多年里, 利用普通的大学数学和一般的优化原理, 我们在变分不等式框架下提出了一个分裂收缩算法统一框架, 它每步迭代包括预测和校正, 其收敛性证明特别简单. 所谓分裂, 就是每步迭代的预测通过求解一些形式简单的凸优化子问题去实现; 而收缩, 则是指通过校正保证迭代点在一定范数意义下离解集(鞍点的集合)越来越近. 我们据此做出的一些工作, 得到国际同行知名学者"非常简单但功能强大"(Very Simple yet Powerful) 和"优雅"(Elegant) 的赞誉. 利用这个框架验证一些算法的收敛性非常容易, 我们最近的工作表明, 用它来指导构造算法也流畅便捷. 对多块可分离的线性约束的凸优化问题, 只要了解高斯消去法 (Gauss Elimination) 求解线性方程组的基本程式, 就可以懂得如何设计出一簇预测-校正的分裂收缩方法. 十年前我们辛辛苦苦"凑"出来的每个方法, 都是现在能设计出的一簇方法中的一个特例.

我用英文写了五个卡片式短篇, 每篇两页, 介绍我们 2010 年以来在凸优化分裂算法方面的主要成果, 其中有些是最近几年我和一些学生的新的合作, 有的则是我七旬以后更成熟的观点. 我个人的经验是: 有一定研究基础的退休学者, 如果身体许可, 兴趣还有, 可以继续和学生合作, 晚一点"洗手"也是可以的.

以下是这五篇两页短文具体内容的简要说明.

- 短篇**A**介绍了我们的算法统一框架, 给出收敛性证明, 收敛性条件的等价表示以及由此得到的预测-校正的广义 PPA. 那个在收敛性证明中起关键作用的"积化和差"的恒等式, 是得益于文革前中学数理基础教育赋我的功底. 这一篇是综述, 后面四篇的进一步介绍, 阐述了统一框架的强大功能.

- 短篇**B**说明变分不等式意义下 H-模下的邻近点算法(PPA)是属于这个算法统一框架中的一类具体方法. 以一些典型问题为例介绍如何利用统一框架构造子问题容易求解的 PPA (Customized PPA 和 Balanced PPA). 这些法则, 对设计多块可分离线性约束凸优化问题的求解方法也是适用的.

- 短篇**C**讲述交替方向法(ADMM)也可以演绎成统一框架中的一个具体算法. 看似多此一举的演译, 却给收敛速率证明提供了很大的方便. 我和袁晓明合作的 ADMM 遍历意义下和点列意义下收敛速率的两篇文章, 靠的就是这里的预测-校正拆解方式. 这种演绎方式, 也被一批海外著名学者在他们的论文或专著的定理证明中采用.

- 短篇**D**报告求解三块问题的方法. 直接推广的交替方向法用来求解三个可分离块线性约束凸优化问题是不能保证收敛的. 统一框架可以用来指导构建求解多块问题的算法. 以三个可分离块的凸优化为例, 说明根据同一个 Gauss 形预测, 采用不同校正可以构造一簇预测-校正的分裂收缩算法.

- 短篇**E**以线性约束的大规模可分离凸优化问题为例, 在统一框架指导下设计了一类广义 PPA. 算法每步迭代的预测-校正基本步骤就像 Gauss 消去法求解线性方程组的"先消去-后回代"那样, 自然流畅, 不同的只是这里的"消去"是通过求解形式简单的凸优化子问题实现的.

这些短篇说明, 对线性约束的凸优化问题, 我们的统一框架不仅涵盖了邻近点算法 (PPA), 增广拉格朗日乘子法 (ALM) 和交替方向法 (ADMM) 等耳熟能详的基础算法, 还为这些方法无法胜任求解的 (三块和多于三块的) 可分离凸优化问题提供了系列求解方法. 如果要问我这一生哪些是最值得留下来的, 这里大概已经占了一半.

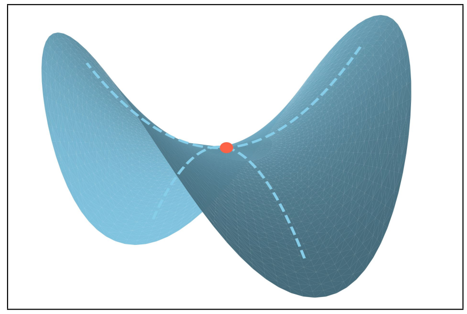
有点优化基础的研究生浏览这五个短篇, 大概花一天就可以, 要比较全面地弄懂理解, 一周时间应该也够. 七旬又六的我, 不时会考虑学术上自己究竟有什么值得留下的. 草草整理这些资料, 是怕以后想做却有心无力而留下遗憾. 谬误之处, 望读者不吝斧正.

何炳生　Bingsheng He　maths.nju.edu.cn/~hebma　2024 年10 月.

A. Splitting and contraction methods for convex optimization

Optimization, VI, Algorithmic framework and Generalized-PPA © Bingsheng He

1 Convex optimization and its related variational inequality



We consider the linearly constrained convex optimization

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

The Lagrangian function of the problem (1.1) is

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \tag{1.2}$$

which is defined on $\mathcal{U} \times \Re^m$. A pair of $(u^*,\lambda^*) \in \mathcal{U} \times \Re^m$ is called a saddle point of the Lagrangian function (1.2), if

$$L_{u\in\mathcal{U}}(u,\lambda^*) \ge L(u^*,\lambda^*) \ge L_{\lambda\in\Re^m}(u^*,\lambda). \tag{1.3}$$

The two inequalities in (1.3) can be written as $(u^*,\lambda^*) \in \mathcal{U} \times \Re^m$,

$$\begin{cases} L(u,\lambda^*) - L(u^*,\lambda^*) \ge 0, & \forall u \in \mathcal{U}, \\ L(u^*,\lambda^*) - L(u^*,\lambda) \ge 0, & \forall \lambda \in \Re^m. \end{cases} \iff \begin{cases} \theta(u) - \theta(u^*) + (u-u^*)^T(-\mathcal{A}^T\lambda^*) \ge 0, & \forall u \in \mathcal{U}, \\ (\lambda-\lambda^*)^T(\mathcal{A}u^* - b) \ge 0, & \forall \lambda \in \Re^m. \end{cases} \tag{1.4}$$

Combining the above two inequalities, the saddle point is described as the solution of the following VI :

$$\text{(VI)} \qquad w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w-w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{1.5}$$

where
$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \qquad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \qquad \text{and} \qquad \Omega = \mathcal{U} \times \Re^m. \tag{1.6}$$

Please notice that $(w-\tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0$. The problem (1.1) is translated to VI (1.5).

For the multi-block separable convex optimization, we take the three-block problem

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) \mid Ax + By + Cz = b, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\} \tag{1.7}$$

as an example. Its corresponding variational inequality has the form (1.5), where $\Omega = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m$, and

$$u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z), \quad w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix}. \tag{1.8}$$

For the $F(w)$ in (1.8), we still have $(w-\tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0$, for any w and $\tilde{w}$ in the space containing Ω.

2 Algorithmic unified framework for monotone variational inequalities

We focus on how to solve the variational inequality (1.5), following is our algorithmic framework.

Algorithms in a unified framework **(Each iteration of the method consists of a prediction and a correction)**

[Prediction Step]. Start from a given v^k, find a predictor $\tilde{w}^k \in \Omega$, which satisfies

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

where the prediction matrix Q is not necessarily symmetric, but the kernel of $Q^T + Q$ is positive definite.

v is called the essential variable in the iteration which can be equal to w, or a part of the whole vector of w.

[Correction Step]. Find the correction matrix M which satisfied (2.2). The new iteration v^{k+1} is given by

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

Convergence Conditions **(It is easy to find the matrix M which satisfies the conditions, see the details in §4)**

[Convergence Conditions]. For the prediction matrix Q in (2.1a) and the correction matrix M in (2.1b), there is a positive definite matrix H, such that

$$HM = Q \quad \text{and} \quad G = Q^T + Q - M^T HM \succ 0. \tag{2.2}$$

3 Convergence proof of the methods in the algorithmic framework

Theorem A. For solving the VI (1.5), let $\{v^k\}$, $\{\tilde{w}^k\}$ be the sequences generated by (2.1). If the conditions (2.2) are satisfied, then we have

$$\tilde{w}^k \in \Omega, \ \ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge \tfrac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\tfrac{1}{2}\|v^k-\tilde{v}^k\|_G^2, \ \ \forall w\in\Omega. \tag{3.1}$$

and

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - \|v^k-\tilde{v}^k\|_G^2, \quad \forall v^*\in\mathcal{V}^*. \tag{3.2}$$

Proof. Treating the term $Q(v^k-\tilde{v}^k)$ in the RHS of (2.1a) by using $Q=HM$ (see (2.2)) and the correction formula (2.1b), we obtain $Q(v^k-\tilde{v}^k)=HM(v^k-\tilde{v}^k)=H(v^k-v^{k+1})$. Thus we get

$$\tilde{w}^k \in \Omega, \quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T H(v^k-v^{k+1}), \quad \forall w\in\Omega. \tag{3.3}$$

Applying the identity

$$(a-b)^T H(c-d) = \tfrac{1}{2}\big(\|a-d\|_H^2-\|a-c\|_H^2\big)+\tfrac{1}{2}\big(\|b-c\|_H^2-\|b-d\|_H^2\big) \tag{3.4}$$

to the RHS of (3.3) with $a=v$, $b=\tilde{v}^k$, $c=v^k$ and $d=v^{k+1}$, we obtain

$$(v-\tilde{v}^k)^T H(v^k-v^{k+1}) = \tfrac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\tfrac{1}{2}\big(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\big). \tag{3.5}$$

To the second part of the RHS of (3.5), by using $HM=Q$ and $2v^TQv=v^T(Q^T+Q)v$, it follows that

$$\begin{aligned}\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2 &\overset{(2.1b)}{=} \|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-M(v^k-\tilde{v}^k)\|_H^2 \\ &= (v^k-\tilde{v}^k)^T(Q^T+Q-M^THM)(v^k-\tilde{v}^k) \overset{(2.2)}{=} \|v^k-\tilde{v}^k\|_G^2.\end{aligned} \tag{3.6}$$

Substituting (3.6) in (3.5), and then in (3.3), we get the assertion (3.1) directly. Setting the $w\in\Omega$ in (3.1) by any fixed w^*, then using $(\tilde{w}^k-w^*)^TF(\tilde{w}^k)=(\tilde{w}^k-w^*)^TF(w^*)$ and $\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^TF(w^*)\ge 0$, we obtain (3.2) and the theorem is completely proved. □

This theorem is proved under weak conditions: $Q^T+Q\succeq 0,\ H\succeq 0,\ HM=Q,\ G=Q^T+Q-M^THM\succeq 0.$

Assertion (3.1) is useful for the convergence rate proof of ADMM, see SIAM Numer. Anal. 2012, 50:700-709.

4 The equivalent convergence conditions and the generalized PPA

Under the condition that the prediction matrix Q is nonsingular, it is easy to construct the correction matrix M which satisfies the convergence conditions (2.2). In fact, because $Q^T+Q\succ 0$, we can take

$$D\succ 0, \qquad G\succ 0 \qquad \text{and} \qquad D+G=Q^T+Q. \tag{4.1}$$

Afterwards, we let

$$HM=Q \qquad \text{and} \qquad M^THM=D. \tag{4.2}$$

Because $\left\{\begin{array}{l} HM=Q, \\ M^THM=D.\end{array}\right. \Leftrightarrow \left\{\begin{array}{l} Q^TM=D, \\ HM=Q.\end{array}\right. \Leftrightarrow \left\{\begin{array}{l} M=Q^{-T}D, \\ H=QD^{-1}Q^T,\end{array}\right.$ we get the matrices M, H and G,

which satisfy the conditions (2.2). There are infinite combinations of D and G which satisfy conditions (4.1).

Choosing matrix D that satisfies condition (4.1), we get $M=Q^{-T}D$, and $H=QD^{-1}Q^T$ is positive definite.
The correction $v^{k+1}=v^k-M(v^k-\tilde{v}^k)$ (can be achieved by solving $Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$.

The generalized PPA by choosing a special D. We can take a special pair of D and G in (4.1) by

$$D=G=\tfrac{1}{2}(Q^T+Q). \tag{4.3}$$

In this case, $M=\frac{1}{2}Q^{-T}(Q^T+Q)$. Because $D=G$, the contractive inequality (3.2) becomes

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2-\|v^k-\tilde{v}^k\|_D^2, \quad \forall v^*\in\mathcal{V}^*. \tag{4.4}$$

Moreover, since $D=M^THM$ (see (4.2)) and $M(v^k-\tilde{v}^k)=v^k-v^{k+1}$ (see (2.1b)), it follows that

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2-\|v^k-v^{k+1}\|_H^2, \quad \forall v^*\in\mathcal{V}^*. \tag{4.5}$$

The inequality (4.5) is just the main convergence result of the classical PPA, a favorable formula ! Because the each iteration consists of a prediction and a correction, we call the related method as a generalized PPA.

In practice, in the generalized PPA, we suggest to take $D=\alpha(Q^T+Q)$ and $\alpha\in[0.5,1)$.

B. PPA belongs to the algorithmic unified framework

Proximal Point Algorithm, Algorithmic unified framework © Bingsheng He

1 Preliminary theorem for convex optimization

Theorem B. Let $\mathcal{X} \subset \Re^n$ be a closed convex set, $\theta(x)$ and $f(x)$ be convex functions and $f(x)$ be differentiable. Assume that the solution set of the minimization problem $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ is nonempty. Then,

$$x^* \in \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{1.1}$$

if and only if

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \mathcal{X}. \tag{1.2}$$

2 Convex optimization and its related variational inequality

1). min-max problem. Let (x^*, y^*) be the solution of the min-max problem

$$\min_x \max_y \{\Phi(x, y) = \theta_1(x) - y^T A x - \theta_2(y) \,|\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{2.1}$$

Using the notation of $\Phi(x, y)$, we have

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

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

$$\text{(VI)} \qquad u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega, \tag{2.2}$$

where
$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad \theta(u) = \theta_1(x) + \theta_2(y), \qquad F(u) = \begin{pmatrix} -A^T y \\ A x \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y}. \tag{2.3}$$

Please notice that for the $F(u)$ in (2.3), we have $(u - \tilde{u})^T(F(u) - F(\tilde{u})) \equiv 0$.

2). Linearly constrained convex optimization The Lagrangian function of the linearly constrained convex optimization problem $\min\{\theta(u) \mid \mathcal{A}u = b, \ u \in \mathcal{U}\}$ is

$$L(u, \lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \quad \text{which defined on } \mathcal{U} \times \Re^m.$$

The saddle point $(u^*, \lambda^*) \in \mathcal{U} \times \Re^m$ of the Lagrangian function can be characterized as

$$\begin{cases} u^* \in \mathcal{U}, \quad L(u, \lambda^*) - L(u^*, \lambda^*) \geq 0, \quad \forall\, u \in \mathcal{U}, & \text{(2.4a)} \\ \lambda^* \in \Re^m, \ L(u^*, \lambda^*) - L(u^*, \lambda) \geq 0, \quad \forall\, \lambda \in \Re^m. & \text{(2.4b)} \end{cases}$$

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

$$\text{(VI)} \qquad w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{2.5}$$

where
$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \qquad F(w) = \begin{pmatrix} -\mathcal{A}^T \lambda \\ \mathcal{A}u - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{U} \times \Re^m. \tag{2.6}$$

Please notice that $(w - \tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0$. We focus on solving the variational inequalities (2.2) and (2.5).

3 Algorithmic unified framework for monotone variational inequalities

Algorithms in a unified framework **(Each iteration of the method consists of a prediction and a correction)**

[Prediction Step]. Start from a given v^k, find a predictor $\tilde{w}^k \in \Omega$, which satisfies

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{3.1a}$$

where the prediction matrix Q is not necessarily symmetric, but the kernel of $Q^T + Q$ is positive definite.

v can be equal to w, or a part vector of the whole vector of w. In this note, v is the whole vector w.

[Correction Step]. Find the correction matrix M which satisfied (3.2). The new v^{k+1} is given by

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{3.1b}$$

Convergence Conditions

[Convergence Conditions]. For the prediction matrix Q in (3.1a) and the correction matrix M in (3.1b), there is a positive definite matrix H, such that

$$HM = Q \quad \text{and} \quad G = Q^T + Q - M^T H M \succ 0. \tag{3.2}$$

4 Proximal point algorithm for monotone variational inequalities

Definition (k-th iteration of the relaxed PPA) $\mathcal{H} \succ 0$. Start with a given w^k, find a $\tilde{w}^k$, such that

[Prediction] $\tilde{w}^k \in \Omega, \ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (w - \tilde{w}^k)^T \mathcal{H}(w^k - \tilde{w}^k), \ \ \forall\, w \in \Omega.$ (4.1a)

[Correction] $w^{k+1} = w^k - \alpha(w^k - \tilde{w}^k), \ \ \alpha \in (0,2). \qquad \big(\alpha \in [1.2, 1.8] \ \text{ is suggested}\big).$ (4.1b)

The relaxed PPA is a special algorithm of (3.1) with $Q = \mathcal{H}$ and $M = \alpha I$. By setting $H = \frac{1}{\alpha}\mathcal{H}$, we have

$$HM = \mathcal{H} = Q \qquad \text{and} \qquad G = Q^T + Q - M^T H M = (2-\alpha)\mathcal{H} \succ 0. \tag{4.2}$$

Taking $\alpha = 1$, we get the classical PPA. PPA belongs to (3.1) and satisfies the convergence conditions (3.2).

5 PPA (Customized PPA and Balanced PPA) for VI

The main task of the each iteration of the relaxed PPA is to find a predictor $\tilde{w}^k$ which satisfies (4.1a).

1). PPA for VI (2.2) (Bingsheng He and Xiaoming Yuan, SIAM Imaging Science 5, (2012) 119-149)

Start with given u^k, find a $\tilde{u}^k$, such that (4.1a) is satisfied, where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = u, \quad F(w) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}, \quad \mathcal{H} = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}, \quad rs > \|A^T A\|. \tag{5.1}$$

The concrete formula of (4.1a) with F and H given in (5.1) is

the uwave parts is $F(\tilde{w}^k)$

$$\begin{cases} \tilde{x}^k \in \mathcal{X}, \ \ \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{\underset{\sim}{-A^T \tilde{y}^k} + rI_n(\tilde{x}^k - x^k) + A^T(\tilde{y}^k - y^k)\} \ge 0, \quad \forall x \in \mathcal{X}, \\ \tilde{y}^k \in \mathcal{Y}, \ \ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{ \underset{\sim}{A\tilde{x}^k} + A(\tilde{x}^k - x^k) + sI_m(\tilde{y}^k - y^k)\} \ge 0, \quad \forall y \in \mathcal{Y}. \end{cases}$$

According to Theorem B, to complete the PPA iteration, we need only to solve the following sub-problems:

$$\begin{cases} \tilde{x}^k = \text{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{1}{2} r \|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, & (5.2a) \\ \tilde{y}^k = \text{argmin}\{\theta_2(y) + y^T A(2\tilde{x}^k - x^k) + \frac{1}{2} s \|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. & (5.2b) \end{cases}$$

2). Customized PPA and Balanced PPA for VI (2.5)

Start with given w^k, find a $\tilde{w}^k$, such that (4.1a) is satisfied, where

$$F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix}, \ \ \mathcal{H} = \mathcal{H}_1 = \begin{pmatrix} \beta\mathcal{A}^T\mathcal{A} + \delta I_n & \mathcal{A}^T \\ \mathcal{A} & \frac{1}{\beta} I_m \end{pmatrix} \ \text{ or } \ \mathcal{H} = \mathcal{H}_2 = \begin{pmatrix} rI_n & \mathcal{A}^T \\ \mathcal{A} & \frac{1}{r}\mathcal{A}\mathcal{A}^T + \delta I_m \end{pmatrix}. \tag{5.3}$$

Customized PPA for VI (2.5) (For customized PPA, see He, Yuan, Zhang 2013 COA, Gu, He, Yuan 2014 COA)

The concrete formula of (4.1a) with $\mathcal{H} = \mathcal{H}_1$ given in (5.3) is

the uwave parts is $F(\tilde{w}^k)$

$$\begin{cases} \tilde{u}^k \in \mathcal{U}, \ \ \theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T \{\underset{\sim}{-\mathcal{A}^T \tilde{\lambda}^k} + (\beta\mathcal{A}^T\mathcal{A} + \delta I_n)(\tilde{u}^k - u^k) + \mathcal{A}^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \ \ \forall u \in \mathcal{U}, \\ \tilde{\lambda}^k \in \Re^m, \qquad (\lambda - \tilde{\lambda}^k)^T \{(\underset{\sim}{\mathcal{A}\tilde{u}^k - b}) + \mathcal{A}(\tilde{u}^k - u^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \ \ \forall \lambda \in \Re^m. \end{cases}$$

According to Theorem B, to complete PPA iteration, we need only to solve the following sub-problems:

$$\begin{cases} \tilde{u}^k = \text{argmin}\big\{\theta(u) - u^T \mathcal{A}^T \lambda^k + \frac{1}{2}(u - u^k)^T(\beta\mathcal{A}^T\mathcal{A} + \delta I_n)(u - u^k) \,|\, u \in \mathcal{U}\big\}, & (5.4a) \\ \tilde{\lambda}^k = \lambda^k - \beta\big(\mathcal{A}[2\tilde{u}^k - u^k] - b\big). & (5.4b) \end{cases}$$

Balanced PPA for VI (2.5) (For balanced PPA, see He and Yuan, arXiv:2108.08554; S.J. Xu, 2023, JAMC)

The concrete formula of (4.1a) with $\mathcal{H} = \mathcal{H}_2$ given in (5.3) is

the uwave parts is $F(\tilde{w}^k)$

$$\begin{cases} \tilde{u}^k \in \mathcal{U}, \ \ \theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T \{\underset{\sim}{-\mathcal{A}^T \tilde{\lambda}^k} + rI_n(\tilde{u}^k - u^k) + \mathcal{A}^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \ \ \forall u \in \mathcal{U}, \\ \tilde{\lambda}^k \in \Re^m, \qquad (\lambda - \tilde{\lambda}^k)^T \{(\underset{\sim}{\mathcal{A}\tilde{u}^k - b}) + \mathcal{A}(\tilde{u}^k - u^k) + (\frac{1}{r}\mathcal{A}\mathcal{A}^T + \delta I_m)(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \ \ \forall \lambda \in \Re^m. \end{cases}$$

According to Theorem B, to complete PPA iteration, we need only to solve the following sub-problems:

$$\begin{cases} \tilde{u}^k = \text{argmin}\big\{\theta(u) - u^T \mathcal{A}^T \lambda^k + \frac{r}{2}\|u - u^k\|^2 \,|\, u \in \mathcal{U}\big\}, & \boxed{\text{Primal problem is easier than(5.4a)}} \quad (5.5a) \\ \tilde{\lambda}^k = \lambda^k - (\frac{1}{r}\mathcal{A}\mathcal{A}^T + \delta I_m)^{-1}\big(\mathcal{A}[2\tilde{u}^k - u^k] - b\big). & \boxed{\text{Only once Cholesky-Decomposition}} \quad (5.5b) \end{cases}$$

Algorithm (5.5) is called the balanced algorithm, because its two subproblems share the difficulties.

C. Explaining ADMM as an algorithm within the framework

The convergence rate proof benefit from this explaination © Bingsheng He

1 Two block convex optimization and its related VI and ADMM

We consider the two block separable convex optimization problem

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\, x\in\mathcal{X}, y\in\mathcal{Y}\}, \tag{1.1}$$

where $\theta_1(x),\theta_2(y)$ are convex function, $A,\ B,\ b$ are the corresponding matrices and vector. $\mathcal{X}$ and $\mathcal{Y}$ are closed convex set. The saddle point of the Lagrangian function of the problem (1.1), say $w^*=(x^*,y^*,\lambda^*)$, can be characterized as the solution of the following variational inequality:

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{1.2a}$$

where

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad \theta(u)=\theta_1(x)+\theta_2(y),\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}. \tag{1.2b}$$

The k-th iteration of ADMM begins with a given $v^k=(y^k,\lambda^k)$, produces $(x^{k+1},y^{k+1},\lambda^{k+1})$ via

$$\begin{cases} x^{k+1}\in\arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\frac12\beta\|Ax+By^k-b\|^2 \;\big|\; x\in\mathcal{X}\big\}, & (1.3a)\\ y^{k+1}\in\arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac12\beta\|Ax^{k+1}+By-b\|^2 \;\big|\; y\in\mathcal{Y}\big\}, & (1.3b)\\ \lambda^{k+1}=\ \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). & (1.3c)\end{cases}$$

2 Split ADMM iteration into prediction and correction of the unified framework

We intentionally split ADMM (1.3) into a prediction-correction method.

The prediction of the k-th iteration begins with given $v^k=(y^k,\lambda^k)$, produces $(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)$ via

[Prediction]

$$\begin{cases} \tilde{x}^k\in\arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\frac12\beta\|Ax+By^k-b\|^2 \;\big|\; x\in\mathcal{X}\big\}, & (2.1a)\\ \tilde{y}^k\in\arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac12\beta\|A\tilde{x}^k+By-b\|^2 \;\big|\; y\in\mathcal{Y}\big\}, & (2.1b)\\ \tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b). & (2.1c)\end{cases}$$

According to the optimality of the convex optimization, the VI forms of three subproblems of (2.1) are

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

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

$$\begin{cases} \tilde{x}^k\in\mathcal{X},\ \underline{\theta_1(x)-\theta_1(\tilde{x}^k)}+(x-\tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\}\ge 0,\quad \forall x\in\mathcal{X}, & (2.2a)\\ \tilde{y}^k\in\mathcal{Y},\ \underline{\theta_2(y)-\theta_2(\tilde{y}^k)}+(y-\tilde{y}^k)^T\{\underline{-B^T\tilde{\lambda}^k}+\beta B^TB(\tilde{y}^k-y^k)\}\ge 0,\quad \forall y\in\mathcal{Y}, & (2.2b)\\ \tilde{\lambda}^k\in\Re^m,\ (\lambda-\tilde{\lambda}^k)^T\{\underline{(A\tilde{x}^k+B\tilde{y}^k-b)}-B(\tilde{y}^k-y^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)\}\ge 0,\ \forall\lambda\in\Re^m. & (2.2c)\end{cases}$$

Furthermore, by using the notations of VI (1.2), we get the compact **VI form of the prediction (2.1)** :

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\quad \forall w\in\Omega, \tag{2.3a}$$

where

$$Q=\begin{pmatrix} \beta B^TB & 0\\ -B & \frac{1}{\beta}I_m\end{pmatrix}. \tag{2.3b}$$

Let $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})$ be the output of the classical ADMM (1.3), then we have

$$x^{k+1}=\tilde{x}^k,\quad y^{k+1}=\tilde{y}^k\quad \text{and}\quad \lambda^{k+1}=\tilde{\lambda}+\beta B(y^k-\tilde{y}^k). \tag{2.4}$$

where $\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)$ are generated by the prediction (2.1). According to (2.4)

[Correction] $\quad v^{k+1}=v^k-M(v^k-\tilde{v}^k),\quad$ where $\quad M=\begin{pmatrix} I & 0\\ -\beta B & I_m\end{pmatrix}.\quad$ (2.5)

By setting $H=\begin{pmatrix} \beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}$, we have $HM=Q$ and $\underline{G=Q^T+Q-M^THM}=\begin{pmatrix} 0 & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}$. (2.6)

ADMM (1.3) is split into the prediction-correction method (2.3)-(2.5). According to the framework, we have

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.7}$$

Theorem C1. Let the sequences $\{v^k\}$ and $\{\tilde{v}^k\}$ be generated by the prediction-correction method (2.3)-(2.5), then we have

$$\|v^k - \tilde{v}^k\|_G^2 = \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 \ge \|v^k - v^{k+1}\|_H^2, \tag{2.8}$$

and consequently,

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.9}$$

Proof. From (2.2b), we know that the optimal condition of the y-subproblem is

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k)\} \ge 0, \quad \forall\, y \in \mathcal{Y}. \tag{2.10}$$

According to (2.4), $\lambda^{k+1} = \tilde{\lambda}^k + \beta B(y^k - \tilde{y}^k)$ and $\tilde{y}^k = y^{k+1}$, the inequality (2.10) can be written as

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \ge 0, \quad \forall\, y \in \mathcal{Y}. \tag{2.11}$$

The above inequality is hold also for the last iteration, *i. e.*, we have

$$y^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^k) + (y - y^k)^T\{-B^T\lambda^k\} \ge 0, \quad \forall\, y \in \mathcal{Y}. \tag{2.12}$$

Setting $y = y^k$ and $y = y^{k+1}$ in in (2.11) and (2.12), respectively, and then adding them, we get

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \ge 0. \tag{2.13}$$

Using $\lambda^k - \tilde{\lambda}^k = (\lambda^k - \lambda^{k+1}) + \beta B(y^k - y^{k+1})$ (see (2.4)) and the inequality (2.13), we obtain

$$\begin{aligned}\frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 &= \frac{1}{\beta}\|(\lambda^k - \lambda^{k+1}) + \beta B(y^k - y^{k+1})\|^2 \\ &\ge \beta\|B(y^k - y^{k+1})\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2 \;=\; \|v^k - v^{k+1}\|_H^2. \end{aligned}\tag{2.14}$$

The assertion (2.8) follows from (2.14) directly. Consequently, the assertion (2.9) follows from (2.7) and (2.8). The theorem is proved. $\square$

Note that from (2.9) we have $\quad \sum_{k=1}^{\infty} \|v^k - v^{k+1}\|_H \le \|v^0 - v^*\|_H^2, \;\; \forall v^* \in \mathcal{V}^*. \quad (2.15)$

3 Important property of ADMM by using the algorithmic framework

Theorem C2. Let the sequences $\{v^k\}$ and $\{\tilde{v}^k\}$ be generated by the prediction (2.3) and correction (2.5), then we have

$$\|v^{k+1} - v^{k+2}\|_H \le \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{3.1}$$

Proof. According to (2.3), we have

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \;\; \forall w \in \Omega \tag{3.2}$$

and

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \ge (v - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}), \;\; \forall w \in \Omega. \tag{3.3}$$

Setting the w in (3.2) and (3.3) by $\tilde{w}^{k+1}$ and $\tilde{w}^k$, respectively, and then adding the resulting inequalities together and using $(\tilde{w}^k - \tilde{w}^{k+1})^T\big(F(\tilde{w}^k) - F(\tilde{w}^{k+1})\big) = 0$, we obtain

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \ge 0. \tag{3.4}$$

Adding $\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}$ to the both sides of (3.4), we get

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

Because $HM = Q$ and $M(v^k - \tilde{v}^k) = (v^k - v^{k+1})$ (see (2.5)), the above inequality becomes

$$(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} \ge \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2. \tag{3.5}$$

Using the identity $\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a - b) - \|a - b\|_H^2$ and (3.5), we get

$$\begin{aligned}&\|v^k - v^{k+1}\|_H^2 - \|v^{k+1} - v^{k+2}\|_H^2 \\ &\quad = 2(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} - \|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|_H^2 \\ &\quad \ge \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2 - \|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|_H^2. \end{aligned}\tag{3.6}$$

Since $(v^k - v^{k+1}) = M(v^k - \tilde{v}^k)$ and $Q^T + Q - M^T HM = G$, the RHS of (3.6) is $\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2$. The assertion (3.1) follows immediately. $\square$

Theorem C3. Let $\{v^k\}$ be the sequence generated by ADMM, then for any integer $t > 0$, we have

$$\|v^t - v^{t+1}\|_H^2 \le \frac{1}{t+1}\sum_{k=0}^{t} \|v^k - v^{k+1}\|_H \le \frac{1}{t+1}\sum_{k=0}^{\infty} \|v^k - v^{k+1}\|_H \le \frac{1}{t+1}\|v^0 - v^*\|_H^2, \; \forall v^* \in \mathcal{V}^*.$$

Proof. The assertion follows from (3.1) and (2.15) directly. $\square$ See H and Yuan, Numer. Math. 130 (2015) 567-577.

D. Constructing SC methods for three block convex optimization

Different corrections based on the same prediction © Bingsheng He

1 The variational inequality and the algorithmic framework

Finding the saddle point of the Lagrangian function is reduced to solving the variational inequality

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

For solving VI (1.1), we have proposed the following algorithmic unified framework:

[Prediction Step]. Start from a given v^k, find a predictor $\tilde{w}^k \in \Omega$, which satisfies

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

where the matrix Q is not necessarily symmetric, but the kernel of $Q^T + Q$ is positive definite.

[Correction Step]. The new iteration (corrector) v^{k+1} is given by

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

where $M = Q^{-T}D$, D is chosen which is satisfied $D \succ 0,\ G \succ 0,\ D + G = Q^T + Q$.

Because $M = Q^{-T}D$, the correction (1.2b) can be achieved by $\quad Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$.

2 Prediction

We take the three block separable convex optimization as an example for explanation. The problem and the VI are described in **A**. We take the direct extension of ADMM for producing its prediction.

[Prediction]

$$\left\{\begin{array}{ll} \tilde{x}^k \in \arg\min\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\big\}, & (2.1a)\\ \tilde{y}^k \in \arg\min\big\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + By + Cz^k - b\|^2 \mid y \in \mathcal{Y}\big\}, & (2.1b)\\ \tilde{z}^k \in \arg\min\big\{\theta_3(z) - z^T C^T \lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + B\tilde{y}^k + Cz - b\|^2 \mid z \in \mathcal{Z}\big\}, & (2.1c)\\ \tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big). & (2.1d) \end{array}\right.$$

According to the basic theorem of optimization, it follows from (2.1) that

$$\left\{\begin{array}{ll} \tilde{x}^k \in \mathcal{X},\ \underline{\theta_1(x) - \theta_1(\tilde{x}^k)} + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\} \ge 0,\ \forall x \in \mathcal{X}, & (2.2a)\\ \tilde{y}^k \in \mathcal{Y},\ \underline{\theta_2(y) - \theta_2(\tilde{y}^k)} + (y - \tilde{y}^k)^T\{\underline{-B^T\tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k)\} \ge 0,\ \forall y \in \mathcal{Y}, & (2.2b)\\ \tilde{z}^k \in \mathcal{Z},\ \underline{\theta_3(z) - \theta_3(\tilde{z}^k)} + (z - \tilde{z}^k)^T\{\underline{-C^T\tilde{\lambda}^k} + \beta C^T B(\tilde{y}^k - y^k) + \beta C^T C(\tilde{z}^k - z^k)\} \ge 0,\ \forall z \in \mathcal{Z}, & (2.2c)\\ \underline{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)} - B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. & (2.2d) \end{array}\right.$$

The inequalities (2.2) can be written together as the prediction (1.2a) where

$$w = \begin{pmatrix} x\\ y\\ z\\ \lambda \end{pmatrix},\quad v = \begin{pmatrix} y\\ z\\ \lambda \end{pmatrix} \quad\text{and}\quad Q = \begin{pmatrix} \beta B^T B & 0 & 0\\ \beta C^T B & \beta C^T C & 0\\ -B & -C & \frac{1}{\beta}I \end{pmatrix}. \tag{2.3}$$

3 Correction by using the kernel matrix

Note that the matrix

$$Q^T + Q = \begin{pmatrix} B^T & 0 & 0\\ 0 & C^T & 0\\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} 2\beta I & \beta I & -I\\ \beta I & 2\beta I & -I\\ -I & -I & \frac{2}{\beta}I \end{pmatrix} \begin{pmatrix} B & 0 & 0\\ 0 & C & 0\\ 0 & 0 & I \end{pmatrix}$$

is positive definite (whenever B and C are both full column rank) and

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

The center part of the matrices $Q^T + Q$ and Q^T are called the kernel matrices, and denoted respectively by

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} 2\beta I & \beta I & -I\\ \beta I & 2\beta I & -I\\ -I & -I & \frac{2}{\beta}I \end{pmatrix} \text{ and } \mathcal{Q}^T = \begin{pmatrix} \beta I & \beta I & -I\\ 0 & \beta I & -I\\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}. \text{ Note that } \mathcal{Q}^{-T} = \begin{pmatrix} \frac{1}{\beta}I & -\frac{1}{\beta}I & 0\\ 0 & \frac{1}{\beta}I & I\\ 0 & 0 & \beta I \end{pmatrix}. \tag{3.1}$$

$\mathcal{Q}^{-T}$ has the simple form because the kernel matrix $\mathcal{Q}^T$ is a upper triangular matrix ! Since the k-th iteration begins with a given (By^k, Cz^k, λ^k), for starting the next iteration, the correction of the k-th iteration needs only to offer $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$. This is very simple ! See some examples !

3.1 Constructing method I $\quad D$ has a simple form

$$D=\begin{pmatrix} B^T & 0 & 0\\ 0 & C^T & 0\\ 0 & 0 & I\end{pmatrix}\begin{pmatrix} \nu\beta I & 0 & 0\\ 0 & \nu\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix} B & 0 & 0\\ 0 & C & 0\\ 0 & 0 & I\end{pmatrix}, \quad \nu\in(0,1). \tag{3.2}$$

Since the kernel matrix of Q^T+Q can be decomposed as

$$\begin{pmatrix} 2\beta I & \beta I & -I\\ \beta I & 2\beta I & -I\\ -I & -I & \frac{2}{\beta}I\end{pmatrix}=\begin{pmatrix} \nu\beta I & 0 & 0\\ 0 & \nu\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}+\begin{pmatrix} (2-\nu)\beta I & \beta I & -I\\ \beta I & (2-\nu)\beta I & -I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix},$$

the kernel matrices of D and G in the right hand side of the above equation are positive definite. The solution of the system of equation $Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$ can be obtained by solving

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

By using the inverse of the kernel matrix of Q^T in (3.1), the correction form can be simplified to

$$\begin{pmatrix} By^{k+1}\\ Cz^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix} By^k\\ Cz^k\\ \lambda^k\end{pmatrix}-\begin{pmatrix} \nu I & -\nu I & 0\\ 0 & \nu I & \frac{1}{\beta}I\\ 0 & 0 & I\end{pmatrix}\begin{pmatrix} B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}. \tag{3.3}$$

3.2 Constructing method II $\quad G$ has a simple form

$$D=\begin{pmatrix} B^T & 0 & 0\\ 0 & C^T & 0\\ 0 & 0 & I\end{pmatrix}\begin{pmatrix} (2-\nu)\beta I & \beta I & -I\\ \beta I & (2-\nu)\beta I & -I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix} B & 0 & 0\\ 0 & C & 0\\ 0 & 0 & I\end{pmatrix}, \quad \nu\in(0,1). \tag{3.4}$$

Since the kernel matrix of Q^T+Q can be decomposed as

$$\begin{pmatrix} 2\beta I & \beta I & -I\\ \beta I & 2\beta I & -I\\ -I & -I & \frac{2}{\beta}I\end{pmatrix}=\begin{pmatrix} (2-\nu)\beta I & \beta I & -I\\ \beta I & (2-\nu)\beta I & -I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix}+\begin{pmatrix} \nu\beta I & 0 & 0\\ 0 & \nu\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix},$$

the kernel matrices of D and G in the right hand side of the above equation are positive definite. The solution of the system of equation $Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$ can be obtained by solving

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

By using the inverse of the kernel matrix of Q^T in (3.1), the correction form can be simplified to

$$\begin{pmatrix} By^{k+1}\\ Cz^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix} By^k\\ Cz^k\\ \lambda^k\end{pmatrix}-\begin{pmatrix} (1-\nu)I & -(1-\nu)I & 0\\ 0 & (1-\nu)I & 0\\ -\beta I & -\beta I & I\end{pmatrix}\begin{pmatrix} B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}. \tag{3.5}$$

3.3 Constructing method III $\quad D$ is proportional to (Q^T+Q)

$$D=\alpha\big[Q^T+Q\big]=\alpha\begin{pmatrix} B^T & 0 & 0\\ 0 & C^T & 0\\ 0 & 0 & I\end{pmatrix}\begin{pmatrix} 2\beta I & \beta I & -I\\ \beta I & 2\beta I & -I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix} B & 0 & 0\\ 0 & C & 0\\ 0 & 0 & I\end{pmatrix}, \quad \alpha\in(0,1). \tag{3.6}$$

The solution of the system of equations $Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$ can be obtained by solving

$$\begin{pmatrix} \beta I & \beta I & -I\\ 0 & \beta I & -I\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix} By^{k+1}-By^k\\ Cz^{k+1}-Cz^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}=\alpha\begin{pmatrix} 2\beta I & \beta I & -I\\ \beta I & 2\beta I & -I\\ -I & -I & \frac{2}{\beta}I\end{pmatrix}\begin{pmatrix} B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}, \quad \alpha\in(0,1).$$

By using the inverse of the kernel matrix of Q^T in (3.1), the correction form can be simplified to

$$\begin{pmatrix} By^{k+1}\\ Cz^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix} By^k\\ Cz^k\\ \lambda^k\end{pmatrix}-\alpha\begin{pmatrix} I & -I & 0\\ 0 & I & \frac{1}{\beta}I\\ -\beta I & -\beta I & 2I\end{pmatrix}\begin{pmatrix} B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}, \quad \alpha\in(0,1). \tag{3.7}$$

E. Generalized PPA for multi-block convex optimization

Application of the generalized Proximal Point Algorithm © Bingsheng He

1 The variational inequality and the algorithmic framework

Consider the multi-block convex optimization problem

$$\min\{\sum_{i=1}^p \theta_i(x_i) \mid \sum_{i=1}^p A_i x_i = b,\ x_i \in \mathcal{X}_i\}. \tag{1.1}$$

The Lagrangian function of the problem (1.1) is $L(x_1,\ldots,x_p,\lambda) = \sum_{i=1}^p \theta_i(x_i) - \lambda^T(\sum_{i=1}^p A_i x_i - b)$, which is defined on $\Omega = \prod_{i=1}^p \mathcal{X}_i \times \Re^m$. Note tha the saddle point of the Lagrangian function, say $(x_1^*,\ldots,x_p^*,\lambda^*) \in \Omega$, can be described the solution of the following variational inequality:

$$\text{(VI)} \qquad w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{1.2}$$

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

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_p^T\lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}. \tag{1.3}$$

We still use Ω^* to represent the solution set of the variational inequality (1.2). For solving VI (1.2), we have proposed the following algorithmic unified framework which consists a prediction and a correction.

[Prediction]. Start from a given w^k, find a predictor $\tilde w^k \in \Omega$, which satisfies

$$\tilde w^k \in \Omega, \quad \theta(x) - \theta(\tilde x^k) + (w - \tilde w^k)^T F(\tilde w^k) \ge (w - \tilde w^k)^T Q(w^k - \tilde w^k), \quad \forall w \in \Omega, \tag{1.4a}$$

where the matrix Q is not necessarily symmetric, but the kernel of $Q^T + Q$ is positive definite.

[Correction]. The new iterate (corrector) w^{k+1} is given by

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

where $M = Q^{-T}D$, D is chosen which is satisfied $D \succ 0,\ G \succ 0,\ D + G = Q^T + Q$.

Because $M = Q^{-T}D$, the correction can be achieved by $\quad Q^T(w^{k+1} - w^k) = D(\tilde w^k - w^k). \quad (1.5)$

2 Prediction

[24] B.S.He, S.J.Xu, X.M.Yuan, Handbook of Numerical Analysis, 24 (2023) 511-557.

Start from a given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, obtain $\tilde w^k = (\tilde x_1^k, \tilde x_2^k, \cdots, \tilde x_p^k, \tilde\lambda^k)$ via:

$$\left\{\begin{array}{l} \tilde x_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T\lambda^k + \frac{1}{2}\beta\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\ \tilde x_2^k \in \arg\min\{\theta_2(x_2) - x_2^T A_2^T\lambda^k + \frac{1}{2}\beta\|A_1(\tilde x_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\ \qquad\vdots \\ \tilde x_p^k \in \arg\min\{\theta_p(x_p) - x_p^T A_p^T\lambda^k + \frac{1}{2}\beta\|\sum_{j=1}^{p-1} A_j(\tilde x_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2 \mid x_p \in \mathcal{X}_p\}; \\ \tilde\lambda^k = \lambda^k - \beta\big(\sum_{j=1}^p A_j\tilde x_j^k - b\big). \end{array}\right. \tag{2.1}$$

The optimal condition of the x_i subproblem is

$$\tilde x_i^k \in \mathcal{X}_i, \ \ \theta_i(x_i) - \theta_i(\tilde x_i^k) + (x_i - \tilde x_i^k)^T\{-A_i^T\tilde\lambda^k + \beta\sum_{j=1}^i A_j(\tilde x_j^k - x_j^k) + A_i^T(\tilde\lambda^k - \lambda^k)\} \ge 0, \ \ \forall x_i \in \mathcal{X}_i.$$

The formula which yields $\tilde\lambda^k$ can be rewritten as

$$\tilde\lambda^k \in \Re^m, \quad (\lambda - \tilde\lambda^k)^T\{\big(\sum_{j=1}^p A_j\tilde x_j^k - b\big) + \tfrac{1}{\beta}(\tilde\lambda^k - \lambda^k)\} \ge 0, \quad \forall\lambda \in \Re^m.$$

By using the notations $\theta(x)$ and $F(w)$, we obtain the following variational inequality for the prediction.

Lemma 2.1 *Let $\tilde w^k \in \Omega$ be generated by* (2.1) *with given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$. Then we have*

$$\tilde w^k \in \Omega, \ \ \theta(x) - \theta(\tilde x^k) + (w - \tilde w^k)^T F(\tilde w^k) \ge (w - \tilde w^k)^T Q(w^k - \tilde w^k), \ \ \forall\, w \in \Omega, \tag{2.2a}$$

where

$$Q = \begin{pmatrix} \beta A_1^TA_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^TA_1 & \beta A_2^TA_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^TA_1 & \beta A_p^TA_2 & \cdots & \beta A_p^TA_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta}I_m \end{pmatrix}. \tag{2.2b}$$

3 Correction (1.5) by using the kernel matrix also see arXiv:2107.01897v2[math.OC].

$$Q^T+Q=\begin{pmatrix} A_1^T & 0 & \cdots & 0 & 0\\ 0 & A_2^T & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p^T & 0\\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}\begin{pmatrix} 2\beta I & \beta I & \cdots & \beta I & I\\ \beta I & 2\beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & I\\ \beta I & \cdots & \beta I & 2\beta I & I\\ I & \cdots & I & I & \frac{2}{\beta}I \end{pmatrix}\begin{pmatrix} A_1 & 0 & \cdots & 0 & 0\\ 0 & A_2 & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p & 0\\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

is positive definite (whenever each $A_i, i=1,\ldots,p$, is full column rank matrix) and

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

The center part of the matrices Q^T+Q and Q^T are called their kernel matrix and denoted

$$\mathcal{Q}^T+\mathcal{Q}=\begin{pmatrix} 2\beta I & \beta I & \cdots & \beta I & I\\ \beta I & 2\beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & I\\ \beta I & \cdots & \beta I & 2\beta I & I\\ I & \cdots & I & I & \frac{2}{\beta}I \end{pmatrix} \quad\text{and}\quad \mathcal{Q}^T=\begin{pmatrix} \beta I & \beta I & \cdots & \beta I & 0\\ 0 & \beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & 0\\ 0 & \cdots & 0 & \beta I & 0\\ I & \cdots & I & I & \frac{1}{\beta}I \end{pmatrix}, \tag{3.1}$$

respectively. Note that $\mathcal{Q}^{-T}$ has the simple form because $\mathcal{Q}^T$ is a nonsingular upper triangular matrix !

Since the k-th iteration begins with a given $(A_1x_1^k, A_2x_2^k,\ldots,A_px_p^k,\lambda^k)$, for starting the next iteration, the correction this iteration needs only to offer $(A_1x_1^{k+1}, A_2x_2^{k+1},\ldots,A_px_p^{k+1},\lambda^{k+1})$. There are infinite combinations of D and G that meet the conditions, we only take the following example for illustration.

D is proportional to *G* $\quad D=\alpha\big(Q^T+Q\big)$ and $G=(1-\alpha)\big(Q^T+Q\big)$, $\alpha\in(0,1)$

$$D=\alpha\begin{pmatrix} A_1^T & 0 & \cdots & 0 & 0\\ 0 & A_2^T & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p^T & 0\\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}\begin{pmatrix} 2\beta I & \beta I & \cdots & \beta I & I\\ \beta I & 2\beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & I\\ \beta I & \cdots & \beta I & 2\beta I & I\\ I & \cdots & I & I & \frac{2}{\beta}I \end{pmatrix}\begin{pmatrix} A_1 & 0 & \cdots & 0 & 0\\ 0 & A_2 & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p & 0\\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{3.2}$$

The solution of the system of equations $Q^T(v^{k+1}-v^k)=D(\tilde v^k-v^k)$ can be obtained by solving

$$\begin{pmatrix} \beta I & \beta I & \cdots & \beta I & 0\\ 0 & \beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & 0\\ 0 & \cdots & 0 & \beta I & 0\\ I & \cdots & I & I & \frac{1}{\beta}I \end{pmatrix}\begin{pmatrix} A_1x_1^{k+1}-A_1x_1^k\\ A_2x_2^{k+1}-A_2x_2^k\\ \vdots\\ A_px_p^{k+1}-A_px_p^k\\ \lambda^{k+1}-\lambda^k \end{pmatrix}=\alpha\begin{pmatrix} 2\beta I & \beta I & \cdots & \beta I & I\\ \beta I & 2\beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & I\\ \beta I & \cdots & \beta I & 2\beta I & I\\ I & \cdots & I & I & \frac{2}{\beta}I \end{pmatrix}\begin{pmatrix} A_1\tilde x_1^k-A_1x_1^k\\ A_2\tilde x_2^k-A_2x_2^k\\ \vdots\\ A_p\tilde x_p^k-A_px_p^k\\ \tilde\lambda^k-\lambda^k \end{pmatrix}.$$

Because $\begin{pmatrix} \beta I & \beta I & \cdots & \beta I & 0\\ 0 & \beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & 0\\ 0 & \cdots & 0 & \beta I & 0\\ I & \cdots & I & I & \frac{1}{\beta}I \end{pmatrix}^{-1}=\begin{pmatrix} \frac{1}{\beta}I & -\frac{1}{\beta}I & 0 & \cdots & 0\\ 0 & \frac{1}{\beta}I & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & -\frac{1}{\beta}I & 0\\ 0 & \cdots & 0 & \frac{1}{\beta}I & 0\\ -I & 0 & \cdots & 0 & \beta I \end{pmatrix}$, finally, we get

$$\begin{pmatrix} A_1x_1^{k+1}\\ A_2x_2^{k+1}\\ \vdots\\ A_px_p^{k+1}\\ \lambda^{k+1} \end{pmatrix}=\begin{pmatrix} A_1x_1^{k}\\ A_2x_2^{k}\\ \vdots\\ A_px_p^{k}\\ \lambda^{k} \end{pmatrix}-\alpha\begin{pmatrix} I & -I & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & I & -I & 0\\ I & \cdots & I & 2I & \frac{1}{\beta}I\\ -\beta I & 0 & \cdots & 0 & I \end{pmatrix}\begin{pmatrix} A_1x_1^k-A_1\tilde x_1^k\\ A_2x_2^k-A_2\tilde x_2^k\\ \vdots\\ A_px_p^k-A_p\tilde x_p^k\\ \lambda^k-\tilde\lambda^k \end{pmatrix}, \quad \alpha\in(0,1). \tag{3.3}$$

References

[1] B.S. He, My 20 years research on alternating direction method of multipliers, Operations Research Transactions 22: 1–31 (2018). 中文：何炳生, 我和乘子交替方向法20 年. 运筹学学报22(1): 1–31 (2018). DOI: 10.15960/j.cnkj.isnn.1007-6093.2018.01.001

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

[3] B.S. He and H. Yang, Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, Operations Research Letters 23, (1998), 151—161.

[4] B.S. He, H. Yang, and S.L. Wang, Alternating directions method with self- adaptive penalty parameters for monotone variational inequalities JOTA 106: 349–368 (2000).

[5] B.S. He, L.Z. Liao, D.R. Han, and H. Yang, A new inexact alternating directions method for monotone variational inequalities, Math. Progr, 92: 103–118 (2002).

[6] B.S. He, X.M. Yuan, On the O(1/n) convergence rate of the Douglas-Rachford Alternating Direction Method, SIAM J. Numerical Analysis, 2012, 50: 700-709.

[7] 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.

[8] Glowinski R. *Numerical methods for nonlinear variational problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[9] B.S. He, F. Ma and X.M. Yuan, Optimally linearizing the alternating direction method of multipliers for convex programming, COA, 75 (2020), 361-388.

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

[11] B.S. He, X.M. Yuan and W.X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, COA, 56(2013), 559-572.

[12] 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, COA, 59(2014), 135-161

[13] B. S. He, PPA-like contraction methods for convex optimization: a framework using variational inequality approach, J. Oper. Res. Soc. China 3(2015), 391-420.

[14] B.S. He and X.M. Yuan, Balanced Augmented Lagrangian Method for Convex Optimization. manusript, 2021. arXiv:2108.08554

[15] S. J. Xu, A dual-primal balanced augmented Lagrangian method for lineraly constrained convex programming, J. Appl. Math. and Computing 69 (2023), 1015-1035

[16] 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.

[17] 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.

[18] 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 necessary convergent, Mathematical Programming, 155 (2016) 57-79.

[19] B.S. He, M. Tao and X.M. Yuan, Alternating Direction Method with Gaussian Back Substitution for Separable Convex Programming, SIAM J. Optim. 22(2012), 313-340.

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

[21] 何炳生, 修正乘子交替方向法求解三个可分离算子的凸优化. 运筹学学报, 2015, 19(3): 57–70.

[22] B.S. He and X.M. Yuan, On the optimal proximal parameter of an ADMM-like splitting method for separable convex programming. Mathematical methods in image processing and inverse problems, 139-163, Springer Proc. Math. Stat., 360. Springer, Singapore, 2021.

[23] 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.

[24] B.S. He, S.J. Xu and X.M. Yuan, Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints, Handbook of Numerical Analysis, 24 (2023) 511-557. See arXiv: 2107.01897v2 [math.OC].

[25] 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], to be appear in Communication on Optimization Theory.

[26] 何炳生, 利用统一框架设计凸优化的分裂收缩算法, 高等学校计算数学学报, 2022, 44: 1-35.

[27] 何炳生, 凸优化分裂收缩算法统一框架的新进展–从好不容易凑出一个方法到并不费劲构造一族算法, 高等学校计算数学学报, 2024, 46: 1-24.

[28] 何炳生, 变分不等式框架下凸优化的分裂收缩算法. 科学出版社 2025 出版.