凸优化的一些典型问题及其求解方法

三. 统一框架下的分裂收缩算法设计

何 炳 生

南方科技大学数学系　　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

华东师范大学计算机科学与技术学院

2020 年 11 月 28 – 29 日

1 Mathematical Background

两大基本概念：变分不等式　和　邻近点 (PPA) 算法

Lemma 1 *Let $\mathcal{X} \subset \Re^n$ be a closed convex set, $\theta(x)$ and $f(x)$ be convex functions and $f(x)$ is differentiable. Assume that the solution set of the minimization problem $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ is nonempty. Then,*

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

if and only if

$$x^* \in \mathcal{X}, \;\; \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \ge 0, \;\; \forall\, x \in \mathcal{X}. \tag{1.1b}$$

1.1 Preliminaries of PPA for Variational Inequalities

The optimal condition of the linearly constrained optimization problem is characterized as a mixed monotone variational inequality:

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

PPA for monotone mixed VI in H-norm

For given w^k, find the proximal point w^{k+1} in H-norm which satisfies

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w^{k+1}) \\ &\geq (w - w^{k+1})^T H (w^k - w^{k+1}) \geq 0, \quad \forall\, w \in \Omega, \end{aligned} \tag{1.3}$$

where H is a symmetric positive definite matrix.

Convergence Property of Proximal Point Algorithm in H-norm

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \tag{1.4}$$

1.2 Splitting Methods in a Unified Framework

We study the algorithms using the guidance of variational inequality.

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{1.5}$$

Algorithms in a unified framework

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \;\; \forall w \in \Omega, \tag{1.6a}$$

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

[Correction Step.] The new iterate v^{k+1} by

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \tag{1.6b}$$

Convergence Conditions

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

$$HM = Q. \tag{1.7a}$$

Moreover, the matrix

$$G = Q^T + Q - \alpha M^T HM \tag{1.7b}$$

is positive semi-definite.

Convergence using the unified framework

Theorem 1 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (3.1) *and $\tilde{w}^k$ is obtained in the k-th iteration. If v^k, v^{k+1} and $\tilde{w}^k$ satisfy the conditions in the unified framework, then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha\|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{1.8}$$

定理 1 的主要结论

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

是跟 PPA 类似的收缩不等式, 所以说这类方法是 PPA Like 方法.

关于统一框架下算法及其收敛性证明可以参考下面的文章:

- 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.
- 何炳生, 我和乘子交替方向法 20 年,《运筹学学报》22 卷第1期, pp. 1-31, 2018.

PPA 类算法步步为营, 稳扎稳打; 缺点是思想保守, 影响速度与精度.

2 ADMM for problems with two separable blocks

This section concern the structured convex optimization problem

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{2.1}$$

The equivalent following variational inequality is :

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

where

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

and

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

The augmented Lagrange Function of (2.1) is

$$\mathcal{L}_\beta^{[2]}(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2, \tag{2.3}$$

where $\beta > 0$ is a penalty coefficient in the augmented Lagrange function.

The recursion of the alternating direction method of multipliers for the structured convex optimization (2.1) can be written as

$$\begin{cases} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{cases} \tag{2.4}$$

Thus, ADMM can be viewed as a relaxed Augmented Lagrangian Method. The main advantage of ADMM is that one can solve the x and y-subproblem separately. Note that the essential variable of ADMM (2.4) is $v = (y, \lambda)$. The variational inequality form of the k-th iteration is

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

2.1 Classical ADMM in the Unified Framework

This subsection shows that the ADMM scheme (2.4) is also a special case of the prototype algorithm, and the Convergence Condition is satisfied. Recall the model (2.1) can be explained as the VI .

In order to cast the ADMM scheme (2.4) into a special case of (1.6), let us first define the artificial vector $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ by

$$\tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1} \quad \text{and} \quad \tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b), \tag{2.5}$$

where (x^{k+1}, y^{k+1}) is generated by the ADMM (2.4).

According to the scheme (2.4), the defined artificial vector $\tilde{w}^k$ satisfies the following VI:

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

This can be written in form of (1.6a) as described in the following lemma.

Lemma 2 *For given v^k, let w^{k+1} be generated by* (2.4) *and $\tilde{w}^k$ be defined by* (2.5)*. Then, we have*

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

where

$$Q = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix}. \tag{2.6}$$

Recall the essential variable of the ADMM scheme (2.4) is (y, λ). Moreover, using the definition of $\tilde{w}^k$, the λ^{k+1} updated by (2.4) can be represented as

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

Therefore, the ADMM scheme (2.4) can be written as

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \begin{pmatrix} y^{k} - \tilde{y}^{k} \\ \lambda^{k} - \tilde{\lambda}^{k} \end{pmatrix}. \tag{2.7a}$$

which corresponds to the step (3.2b) with

$$M = \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \quad \text{and} \quad \alpha = 1. \tag{2.7b}$$

Now we check that the Convergence Condition is satisfied by the ADMM scheme (2.4). Indeed, for the matrix M in (2.7b), we have

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

Thus, by using (2.6) and (2.7b), we obtain

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

and consequently

$$\begin{aligned} G &= Q^T + Q - \alpha M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} \beta B^T B & -B^T \\ 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \\ &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix} \end{aligned} \tag{2.8}$$

Therefore, H is symmetric and positive definite under the assumption that B is full column rank; and G is positive semi-definite. The Convergence Condition is satisfied; and thus the convergence of the ADMM scheme (2.4) is guaranteed.

Note that Theorem of complicity is true for $G \succeq 0$. Thus the classical ADMM (2.4) has $O(1/t)$ convergence rate in the ergodic sense.

Since $\alpha = 1$, according to (1.8) and the form of G in (2.8), we have

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

Lemma 3 *For given v^k, let w^{k+1} be generated by* (2.4) *and $\tilde{w}^k$ be defined by* (2.5)*. Then, we have*

$$\tfrac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 \ge \|v^k - v^{k+1}\|_H^2. \tag{2.10}$$

Proof. According to (2.4) and (2.5), the optimal condition of the y-subproblem is

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

Because

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

it can be written as

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

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

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

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

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

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

$$\begin{array}{rcl} \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 & = & \frac{1}{\beta}\|(\lambda^k - \lambda^{k+1}) + \beta B(y^k - y^{k+1})\|^2 \\ & \geq & \frac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2 + \beta\|B(y^k - y^{k+1})\|^2 \\ & = & \|v^k - v^{k+1}\|_H^2. \end{array}$$

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

Substituting (2.10) in (2.9), we get the following nice property of the classical ADMM.

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.14}$$

which is the same as in the standard ADMM.

Notice that the sequence $\{\|v^k - v^{k+1}\|_H^2\}$ generated by the classical ADMM is monotone non-increasing [15]. In fact, in the last note, we have proved that

$$\|M(v^k - \tilde{v}^k)\|_H \le \|M(v^{k-1} - \tilde{v}^{k-1})\|_H, \quad \forall\, k \ge 1. \tag{2.15}$$

Because (see the correction formula (2.7))

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

it follows from (2.15) that

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

On the other hand, the inequality (2.14) tell us that

$$\sum_{k=0}^{\infty} \|v^k - v^{k+1}\|_H^2 \le \|v^0 - v^*\|_H^2.$$

Thus, we have

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

Therefore, ADMM (2.4) has $O(1/t)$ convergence rate in pointwise iteration-complexity.

2.2 ADMM in Sense of Customized PPA [2]

If we change the performance order of y and λ of the classical ADMM (2.4), it becomes

$$\left\{\begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\[2mm] \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\[2mm] y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^{k+1}) \,|\, y \in \mathcal{Y}\}. \end{array}\right. \tag{2.16}$$

In this way we can get a positive semidefinite matrix Q in (1.6a). We define

$$\tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1}, \quad \tilde{\lambda}^k = \lambda^{k+1}, \tag{2.17}$$

where $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is the output of (2.16) and thus it can be rewritten as

$$\left\{\begin{array}{l} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\[2mm] \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\[2mm] \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(\tilde{x}^k, y, \tilde{\lambda}^k) \,|\, y \in \mathcal{Y}\}. \end{array}\right. \tag{2.18}$$

Because $\tilde{\lambda}^k = \lambda^{k+1} = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, the optimal condition of the x-subproblem of (2.18) is

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.19}$$

Notice that

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

ignoring the constant term in the y optimization subproblem of (2.18), it turns to

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) - (\tilde{\lambda}^k)^T By + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \mid y \in \mathcal{Y}\},$$

and consequently, the optimal condition is $\tilde{y}^k \in \mathcal{Y}$,

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

For the term $[\,\cdot\,]$ in the last inequality, using $\beta(A\tilde{x}^k + By^k - b) = -(\tilde{\lambda}^k - \lambda^k)$, we have

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

Finally, the optimal condition of the y-subproblem can be written as $\tilde{y}^k \in \mathcal{Y}$ and

$$\theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T\left[-B^T\tilde{\lambda}^k+\beta B^T B(\tilde{y}^k-y^k)-B^T(\tilde{\lambda}^k-\lambda^k)\right] \geq 0,\ \forall y \in \mathcal{Y}. \tag{2.20}$$

From the λ update form in (2.18) we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \tag{2.21}$$

Combining (2.19), (2.20) and (2.21), and using the notations of the related VI, we get following lemma.

Lemma 4 *For given v^k, let $\tilde{w}^k$ be generated by* (2.18)*. Then, we have*

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

where

$$Q = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{2.22}$$

Because Q is symmetric and positive semidefinite, according to (3.4), we can take

$$M = I \qquad \alpha \in (0,2) \qquad \text{and thus} \qquad H = Q.$$

In this way, we get the new iterate by

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k).$$

The generated sequence $\{v^k\}$ has the convergence property

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

Ensure the matrix H to be positive definite If we add an additional proximal term $\frac{\delta\beta}{2}\|B(y-y^k)\|^2$ to the y-subproblem of (2.18) with any small $\delta > 0$, it becomes

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta^{(2)}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta^{(2)}(\tilde{x}^k, y, \tilde{\lambda}^k) + \frac{\delta\beta}{2}\|B(y-y^k)\|^2 \,|\, y \in \mathcal{Y}\}. \end{cases} \tag{2.23}$$

In the ADMM based customized PPA (2.18), the y-subproblem can be written as

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \frac{\beta}{2}\|By - p^k\|^2 \,|\, y \in \mathcal{Y}\}, \tag{2.24}$$

where

$$p^k = b + \frac{1}{\beta}\tilde{\lambda}^k - A\tilde{x}^k.$$

If we add an additional term $\frac{\delta\beta}{2}\|B(y-y^k)\|^2$ (with any small $\delta>0$) to the objective function of the y-subproblem, we will get $\tilde{y}^k$ via

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|By - p^k\|^2 + \tfrac{\delta\beta}{2}\|B(y-y^k)\|^2 \,|\, y \in \mathcal{Y}\}.$$

By a manipulation, the solution point of the above subproblem is obtained via

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{(1+\delta)\beta}{2}\|By - q^k\|^2 \,|\, y \in \mathcal{Y}\}, \tag{2.25}$$

where

$$q^k = \tfrac{1}{1+\delta}(p^k + \delta B y^k).$$

In this way, the matrix Q in (2.22) will turn to

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

Take $H=Q$, for any $\delta>0$, H is positive definite when B is a full rank matrix. In other words, instead of (2.24), using (2.25) to get $\tilde{y}^k$, it will ensure the positivity of H theoretically. However, in practical computation, it works still well by using $\delta=0$.

ADMM in sense of customized PPA

1. Produce a predictor $\tilde{w}^k$ via (2.23) with given $v^k = (y^k, \lambda^k)$,
2. Update the new iterate by $v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k)$, $\alpha = 1.5 \in (0, 2)$.

Theorem 2 *The sequence $\{v^k\}$generated by the ADMM in Sense of PPA satisfies*

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

where

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

Since the correction formula is $v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k)$, the contraction inequality can be written as

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

Notice that the sequence $\{\|v^k - v^{k+1}\|_H^2\}$ generated by the ADMM in sense of PPA is also monotone non-increasing. Again, because (see (**??**))

$$\|M(v^k - \tilde{v}^k)\|_H \le \|M(v^{k-1} - \tilde{v}^{k-1})\|_H, \quad \forall\, k \ge 1. \tag{2.26}$$

it follows from (2.26) and the correction formula that

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

Thus, we have

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

Therefore, ADMM (in Sense of Customized PPA) has $O(1/t)$ convergence rate in pointwise iteration-complexity.

2.3 Symmetric ADMM [9]

In the problem (2.1), x and y are a pair of fair variables. It is nature to consider a symmetric method: Update the Lagrangian Multiplier after solving each x and y-subproblem. .

We take $\mu \in (0,1)$ (usually $\mu = 0.9$), the method is described as

$$\left\{\begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}_{\beta}^{[2]}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\[2ex] \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), \\[2ex] y^{k+1} = \text{Argmin}\{\mathcal{L}_{\beta}^{[2]}(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\[2ex] \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b). \end{array}\right. \tag{2.27}$$

This method is called **Alternating direction method of multipliers with symmetric multipliers updating**, or **Symmetric Alternating Direction Method of Multipliers**.

The convergence of the proposed method is established via the unified framework.

For establishing the main result, we introduce an artificial vector $\tilde{w}^k$ by

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}, \tag{2.28}$$

where (x^{k+1}, y^{k+1}) is generated by the ADMM (2.27).

According to (2.28), the optimal condition of the x-subproblem of (2.27) is

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.29}$$

Notice that the objective function of the y-subproblem in (2.27) is

$$\begin{aligned} &\mathcal{L}_\beta^{[2]}(\tilde{x}^k, y, \lambda^{k+\frac{1}{2}}) \\ &\quad = \quad \theta_1(\tilde{x}^k) + \theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(A\tilde{x}^k + By - b) + \tfrac{\beta}{2}\|A\tilde{x}^k + By - b\|^2. \end{aligned}$$

Ignoring the constant term in the y-subproblem, it turns to

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \tfrac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \mid y \in \mathcal{Y}\}.$$

Consequently, according to Lemma 1, we have

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

Using

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

and $\beta(A\tilde{x}^k + By^k - b) = (\tilde{\lambda}^k - \lambda^k)$, we get

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

Finally, the optimal condition of the y-subproblem can be written as $\tilde{y}^k \in \mathcal{Y}$, and

$$\theta_2(y) - \theta_2(\tilde{y}^k) \;+\; (y - \tilde{y}^k)^T \{ -B^T \tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) \\ -\mu B^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \;\; \forall y \in \mathcal{Y}. \quad (2.30)$$

According to the definition of $\tilde{w}^k$ in (2.28), we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \quad (2.31)$$

Combining (2.29), (2.30) and (2.31), and using the notations of the related VI, we get following lemma.

Lemma 5 *For given v^k, let w^{k+1} be generated by* (2.27) *and $\tilde{w}^k$ be defined by* (2.28)*. Then, we have*

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

where

$$Q = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \quad (2.32)$$

Using this notation and by a manipulation, the update form of λ in (2.27) can be represented as

$$\begin{aligned} \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \mu\big[-\beta B(y^k - \tilde{y}^k) + \beta(Ax^{k+1} + By^k - b)\big] \\ &= \lambda^k - \big[-\mu\beta B(y^k - \tilde{y}^k) + 2\mu(\lambda^k - \tilde{\lambda}^k)\big]. \end{aligned} \tag{2.33}$$

Thus, together with $y^{k+1} = \tilde{y}^k$, we have the following useful relationship

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

This can be rewritten into a compact form:

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

with

$$M = \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix}. \tag{2.34}$$

These relationships greatly simplify our analysis and presentation.

In order to use the unified framework, we only need to verify the positiveness of H and G. For the matrix M given by (2.34), we have

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

For $H = QM^{-1}$, it follows that

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

Thus

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

Notice that

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

Therefore, H is positive definite for any $\mu \in (0,1)$ when B is a full column rank matrix.

It remains to check the positiveness of $G = Q^T + Q - M^T HM$. Note that

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

Using (2.32) and the above equation, we have

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

Thus

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

Because the matrix

$$\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

is positive definite, for any $\mu \in (0,1)$, G is essentially positive definite (positive definite when B is a full column rank matrix). The convergence conditions (1.7) are satisfied.

Take $\mu = 0.9$, it will accelerate the convergence much. For the numerical experiments of this method, it is refereed to consult [9].

The symmetric ADMM is a special version of the unified framework (3.4) - (1.7) whose $\alpha = 1$,

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

Both the matrices H and G are positive definite for $\mu \in (0,1)$. According to Theorem 1, we have

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

3 Two special prediction-correction methods

We study the optimization algorithms using the guidance of variational inequality.

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{3.1}$$

3.1 Algorithms I $\quad Q = H$, H is positive definite

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

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

where the matrix H is symmetric and positive definite.

[Correction Step.] The new iterate v^{k+1} by

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0, 2) \tag{3.2b}$$

H is a symmetric positive definite matrix. 预测往往对参数有要求

The sequence $\{v^k\}$ generated by the prediction-correction method (3.2) satisfies

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha(2-\alpha)\|v^k - \tilde{v}^k\|_H^2. \quad \forall v^* \in \mathcal{V}^*.$$

The above inequality is the Key for convergence analysis !

Set $\alpha = 1$ in (3.2b), the prediction (3.2a) becomes: $w^{k+1} \in \Omega$ such that

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

The generated sequence $\{v^k\}$ satisfies

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

上式是跟 (1.4) 类似的不等式, 是关于核心变量 v 的 PPA 方法.

3.2 Algorithms II $\quad Q$ is the sum of two matrices

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

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

where

$$Q = D + K, \tag{3.3b}$$

D is a block diagonal positive definite matrix

K is skew-symmetric (反对称) $\quad Q^T + Q = 2D$

[Correction Step.] For the positive matrix D, the new iterate v^{k+1} is given by

$$v^{k+1} = v^k - \gamma \alpha_k^* M(v^k - \tilde{v}^k), \tag{3.4a}$$

where $\quad M = D^{-1}Q, \quad \gamma \in (0,2)$, and the optimal step size is given by

$$\alpha_k^* = \frac{\|v^k - \tilde{v}^k\|_D^2}{\|M(v^k - \tilde{v}^k)\|_D^2}. \tag{3.4b}$$

Since $M^T DM = M^T Q$, we have

$$\|M(v^k - \tilde{v}^k)\|_D^2 = \big[M(v^k - \tilde{v}^k)\big]^T \big[Q(v^k - \tilde{v}^k)\big]$$

and thus

$$\alpha_k^* = \frac{\|v^k - \tilde{v}^k\|_D^2}{\big[M(v^k - \tilde{v}^k)\big]^T \big[Q(v^k - \tilde{v}^k)\big]}.$$

步长计算很容易实现

The sequence $\{v^k\}$ generated by the prediction-correction Algorithm II satisfies

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_D^2 - \gamma(2-\gamma)\alpha_k^* \|v^k - \tilde{v}^k\|_H^2. \quad \forall v^* \in \mathcal{V}^*.$$

上式是跟 (1.4) 类似的不等式, 预测-校正方法都具有 PPA Like 收敛性质.

所以, 这个报告中所说的方法, 都是邻近点类 (PPA Like) 算法.

Convergence of the prediction-correction method II

Lemma 6 *For given* v^k*, let the predictor* $\tilde{w}^k$ *be generated by* (3.3a)*, then we have*

$$(v^k - v^*)^T Q(v^k - \tilde{v}^k) \ge \|v^k - \tilde{v}^k\|_D^2, \tag{3.5}$$

where Q *is given in the right hand side of* (3.3a) *and* D *is given in* (3.3b)*.*

Proof. Set $w = w^*$ in (3.3a), we get

$$(\tilde{v}^k - v^*)^T Q(v^k - \tilde{v}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \tag{3.6}$$

Because

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

and

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

the right hand side of (3.6) is non-negative. Thus, we have

$$\{(v^k - v^*) - (v^k - \tilde{v}^k)\}^T Q(v^k - \tilde{v}^k) \ge 0$$

and

$$(v^k - v^*)^T Q(v^k - \tilde{v}^k) \geq (v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{3.7}$$

For the right hand side of the above inequality, by using $Q = D + K$ and the skew-symmetry of K, we obtain

$$\begin{aligned}(v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) &= (v^k - \tilde{v}^k)^T (D+K)(v^k - \tilde{v}^k) \\ &= \|v^k - \tilde{v}^k\|_D^2.\end{aligned}$$

The lemma is proved. $\square$

Theorem 3 *For given* v^k*, let the predictor* $\tilde{w}^k$ *be generated by* (3.3a)*. If the new iterate* v^{k+1} *is given by*

$$v^{k+1}(\alpha) = v^k - \alpha M(v^k - \tilde{v}^k), \quad \gamma \in (0, 2), \tag{3.8}$$

then we have

$$\|v^{k+1} - v^*\|_D^2 \leq \|v^k - v^*\|_D^2 - q_k^{II}(\alpha), \quad \forall v^* \in \mathcal{V}^*, \tag{3.9}$$

where

$$q_k^{II}(\alpha) = 2\alpha\|w^k - \tilde{w}^k\|_D^2 - \alpha^2\|M(w^k - \tilde{w}^k)\|_D^2. \tag{3.10}$$

Proof. First, we define the profit function by

$$\vartheta_k^{II}(\alpha) = \|v^k - v^*\|_D^2 - \|v^{k+1}(\alpha) - v^*\|_D^2. \tag{3.11}$$

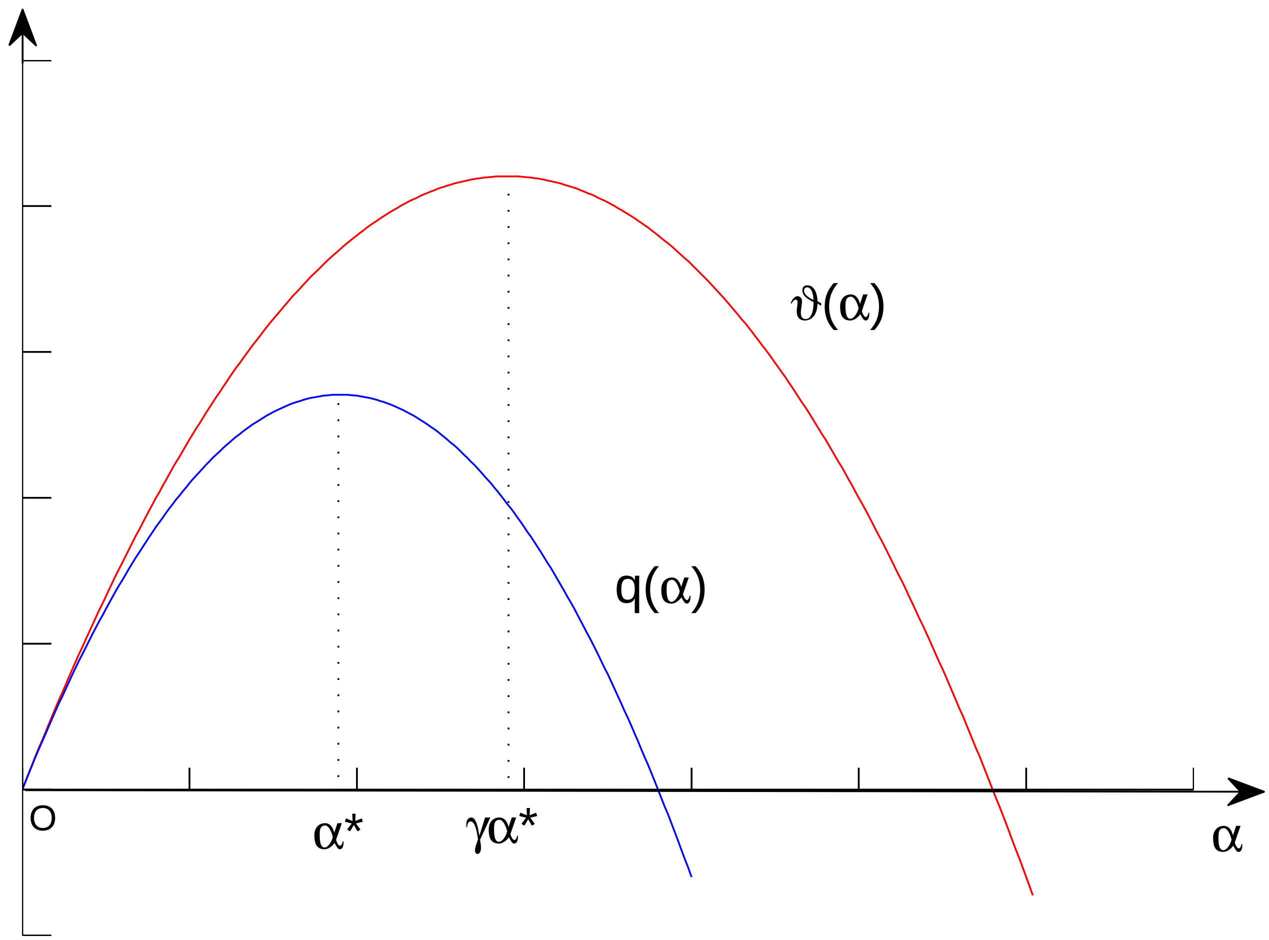
Thus, it follows from (3.8) that

$$\begin{aligned}\vartheta_k^{II}(\alpha) &= \|v^k - v^*\|_D^2 - \|(v^k - v^*) - \alpha M(v^k - \tilde{v}^k)\|_D^2 \\ &= 2\alpha(v^k - v^*)^T DM(v^k - \tilde{v}^k) - \alpha^2\|M(v^k - \tilde{v}^k)\|_D^2.\end{aligned}$$

By using $DM = Q$ and (3.5), we get

$$\vartheta_k^{II}(\alpha) \;\geq\; 2\alpha\|v^k - \tilde{v}^k\|_D^2 - \alpha^2\|M(v^k - \tilde{v}^k)\|_D^2 \;=\; q_k^{II}(\alpha). \qquad \Box$$

$q_k^{II}(\alpha)$ reaches its maximum at α_k^* which is given by $(3.4b)$.



取 $\gamma \in [1, 2)$ 的示意图

Since we take $\alpha = \gamma\alpha_k^*$, it follows from (3.10) that

$$q_k^{II}(\alpha) = 2\gamma\alpha_k^* \|v^k - \tilde{v}^k\|_D^2 - \gamma^2(\alpha_k^*)^2 \|M(v^k - \tilde{v}^k)\|_D^2. \tag{3.12}$$

By using (3.4b), we get

$$\begin{aligned} & (\alpha_k^*)^2 \|M(v^k - \tilde{v}^k)\|_D^2 \\ = \quad & \alpha_k^* \frac{\|v^k - \tilde{v}^k\|_D^2}{\|M(v^k - \tilde{v}^k)\|_D^2} \|M(v^k - \tilde{v}^k)\|_D^2 \\ = \quad & \alpha_k^* \|v^k - \tilde{v}^k\|_D^2. \end{aligned}$$

Substituting it in (3.12) we get $\quad q_k^{II}(\alpha) \geq \gamma(2-\gamma)\alpha_k^* \|v^k - \tilde{v}^k\|_D^2.$

$$\boxed{\|v^{k+1} - v^*\|_D^2 \leq \|v^k - v^*\|_D^2 - \gamma(2-\gamma)\alpha_k^* \|v^k - \tilde{v}^k\|_D^2. \quad \forall v^* \in \mathcal{V}^*.}$$

4 Applications for separable problems

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

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

Its VI-form is

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall \, w \in \Omega. \tag{4.2}$$

where

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

and

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

The augmented Lagrangian Function of the problem (4.1) is

$$\mathcal{L}_\beta(x,y,\lambda) = \theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)+\frac{\beta}{2}\|Ax+By-b\|^2. \quad (4.4)$$

Solving the problem (4.1) 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

$$\text{(ADMM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x,y^k,\lambda^k) \mid x\in\mathcal{X}\}, & \text{(4.5a)} \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1},y,\lambda^k) \mid y\in\mathcal{Y}\}, & \text{(4.5b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1}+By^{k+1}-b). & \text{(4.5c)} \end{array} \right.$$

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

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - \|v^k-v^{k+1}\|_H^2, \quad H = \begin{pmatrix} \beta B^TB & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix}$$

根据算法 I 的要求　设计预测公式.

4.1 ADMM in PPA-sense

In order to solve the separable convex optimization problem (4.1), we construct a method whose prediction-step is

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{4.6a}$$

where

$$H = \begin{pmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad \text{(a small } \delta > 0\text{, say } \delta = 0.05\text{).} \tag{4.6b}$$

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 [2], we took $\delta = 0$).

The concrete form of (4.6) is

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

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

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x,y^k,\lambda^k)\,|\,x\in\mathcal{X}\}, & \text{(4.7a)} \\ \tilde{\lambda}^k = \lambda^k-\beta(A\tilde{x}^k+By^k-b), & \text{(4.7b)} \\ \tilde{y}^k = \text{Argmin}\left\{ \begin{array}{c} \theta_2(y)-y^TB^T\boldsymbol{[2\tilde{\lambda}^k-\lambda^k]} \\ +\frac{1+\delta}{2}\beta\|B(y-y^k)\|^2 \end{array} \,\middle|\, y\in\mathcal{Y}\right\}. & \text{(4.7c)} \end{cases}$$

这个预测与经典的交替方向法 (4.5) 相当, 采用(3.2b) 校正, 会加快速度.

当子问题 (4.7c) 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 (4.7), the prediction step becomes

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T H(v^k-\tilde{v}^k), \ \forall w\in\Omega, \quad (4.8)$$

where

$$H=\begin{bmatrix} sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}, \text{ 代替 (4.6) 中的 } \begin{bmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}. \quad (4.9)$$

The concrete formula of (4.8) is

$$\left\{\begin{array}{l} \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T \\ \qquad\qquad \{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T \\ \qquad\qquad \{\underline{-B^T\tilde{\lambda}^k} + \boldsymbol{s}(\tilde{y}^k-y^k)\boldsymbol{-B^T}(\tilde{\lambda}^k-\lambda^k)\} \ge 0, \\ (\underline{A\tilde{x}^k+B\tilde{y}^k-b})\boldsymbol{-B}(\tilde{y}^k-y^k)+\boldsymbol{(1/\beta)}(\tilde{\lambda}^k-\lambda^k)=0. \end{array}\right. \quad (4.10)$$

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 (4.10),

we need only use the following procedure:

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

用 $\frac{s}{2}\|y-y^k\|^2$ 代替 $\frac{1+\delta}{2}\beta\|B(y-y^k)\|^2$, 为保证收敛, 需要 $s > \beta\|B^TB\|$.

对给定的 $\beta > 0$，要求 $s > \beta\|B^TB\|$，太大的 s 会影响收敛速度

4.3 Method without $s > \beta\|B^TB\|$

当矩阵 B^TB 的条件不好, 又必须线性化, 就采取以下的方法

For solving the same problem, we give the following prediction:

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{4.11a}$$

where

$$Q = \begin{pmatrix} sI & B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} = D + K. \tag{4.11b}$$

Because

$$D = \begin{pmatrix} sI & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \quad \text{and} \quad K = \begin{pmatrix} 0 & B^T \\ -B & 0 \end{pmatrix},$$

根据这样的预测, 可以用算法 II 的校正公式 (3.4) 产生新的迭代点.

How to implement the prediction? The concrete formula of (4.11) is

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

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

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

This can be implemented by

$$\left\{\begin{array}{l}\tilde{x}^k=\text{Argmin}\{\mathcal{L}_\beta(x,y^k,\lambda^k)\,|\,x\in\mathcal{X}\}, \\ \tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b), \\ \tilde{y}^k=\text{Argmin}\{\theta_2(y)-y^TB^T\boldsymbol{\lambda^k}+\frac{s}{2}\|y-y^k\|^2\,|\,y\in\mathcal{Y}\}.\end{array}\right.$$

The y-subproblem is easy. 对给定的 $\beta>0$, 可以取任意的 $s>0$.

对可分离目标函数的优化问题, 我们在 §4 中提出三种预测-校正方法

- 如果子问题中求解过程中, 二次项不带来任何困难的时候, 建议采用 §4.1 中的方法.
- 如果子问题中求解中, 必须对一个子问题中的二次项线性化, 并且矩阵条件好的时候, 建议采用 §4.2 中的方法.
- 如果必须线性化, 矩阵条件又不好的时候, 建议分别采用 §**??**.3 和 §4.3 中的方法.

希望这些框架能为针对实际问题设计算法提供帮助.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Foundations and Trends in Machine Learning Vol. 3, No. 1 (2010) 1 – 122.

[2] 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.

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

[4] 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.

[5] E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

[6] D. Gabay, Applications of the method of multipliers to variational inequalities, *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, edited by M. Fortin and R. Glowinski, North Holland, Amsterdam, The Netherlands, 1983, pp. 299–331.

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

[8] D. Hallac, Ch. Wong, S. Diamond, A. Sharang, R. Sosič, S. Boyd and J. Leskovec, SnapVX: A Network-Based Convex Optimization Solver, Journal of Machine Learning Research 18 (2017) 1-5.

[9] B. S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman-Rachford splitting method for

convex programming, *SIAM Journal on Optimization* **24**(2014), 1011-1040.

[10] 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.

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

[12] B.S. He, H. Yang, and S.L. Wang, Alternating directions method with self-adaptive penalty parameters for monotone variational inequalities, *JOTA* **23**(2000), 349–368.

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

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

[15] 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.

[16] M. R. Hestenes, Multiplier and gradient methods, *JOTA* **4**, 303-320, 1969.

[17] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, *Rev. Francaise d'Inform. Recherche Oper.*, **4**, 154-159, 1970.

[18] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.* **15** (2004), 229–251.

[19] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in Optimization, R. Fletcher, ed., Academic Press, New York, NY, pp. 283-298, 1969.

Thank you very much for your attention ！

Thank you very much for reading ！