凸优化和单调变分不等式的收缩算法

第十五讲: 统一框架下交替方向法的收敛速率

Convergence rate study of ADMM
under a unified framework

南京大学数学系　　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the paper [10, 12]

In this note, we discuss the ADMM and ADM-like methods for the separable convex optimization problem in a unified framework, and study the their global convergence and convergence rate.

1 Algorithms in a unified framework

The classical separable convex optimization problem has the following mathematical form:

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

where $\theta_1(x) : \Re^{n_1} \to \Re, \;\theta_2(y) : \Re^{n_2} \to \Re$ are convex functions (but not necessary smooth), $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$ and $b \in \Re^m$, $\mathcal{X} \subset \Re^{n_1}$, $\mathcal{Y} \subset \Re^{n_2}$ are given closed convex sets. Let $n = n_1 + n_2$.

The optimal condition of the linearly constrained convex optimization is resulted in a variational inequality: The Lagrangian function of the problem (1.1) is

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

which is defined on $\Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m$. Let $(x^*, y^*, \lambda^*) \in \mathcal{X} \times \mathcal{Y} \times \Re^m$ be an saddle point which satisfies

$$L_{\lambda \in \Re^m}(x^*, y^*, \lambda) \le L(x^*, y^*, \lambda^*) \le L_{x \in \mathcal{X}, y \in \mathcal{Y}}(x, y, \lambda^*).$$

These inequalities can be interpreted as

$$\left\{ \begin{array}{rll} x^* \in \mathcal{X}, & L(x, y^*, \lambda^*) - L(x^*, y^*, \lambda^*) \ge 0, & \forall x \in \mathcal{X} \\ y^* \in \mathcal{Y}, & L(x^*, y, \lambda^*) - L(x^*, y^*, \lambda^*) \ge 0, & \forall y \in \mathcal{Y} \\ \lambda^* \in \Re^m, & L(x^*, y^*, \lambda^*) - L(x^*, y^*, \lambda) \ge 0, & \forall \lambda \in \Re^m \end{array} \right. \tag{1.2}$$

The equivalent expression is the following variational inequality:

$$\left\{ \begin{array}{rrl} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \ge 0, & \forall x \in \mathcal{X} \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-B^T\lambda^*) \ge 0, & \forall y \in \mathcal{Y} \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T(Ax^* + By^* - b) \ge 0, & \forall \lambda \in \Re^m \end{array} \right. \tag{1.3}$$

This first order optimal condition (1.3) can be written in a compact form such as

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

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix} \tag{1.4b}$$

$$\theta(u) = \theta_1(x) + \theta_2(y) \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{1.4c}$$

We study the algorithms using the guidance of variational inequality. The analysis can be fund in [6] (Sections 4 and 5 therein). Notice that the augmented Lagrangian function of (1.1) is

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

The recursion of the classical alternating direction method of multipliers for the structured convex optimization (1.1) can be written as

$$\text{(ADMM)} \quad \left\{ \begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array} \right. \tag{1.5}$$

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 (1.5) is $v = (y, \lambda)$.

The linearized ADMM for (1.1) which taking $v = (y, \lambda)$ as the essential variable can be written as

$$\text{(L-ADMM)} \quad \left\{ \begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_D^2 \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array} \right. \tag{1.6}$$

where

$$D = sI - \beta B^T B, \quad s > \beta\rho(B^T B). \tag{1.7}$$

1.1 The algorithmic framework

We study the convergence and the convergence rate of the different ADMM methods. First, we interpret these methods in a prediction-correction framework.

A Prototype Algorithm for (1.4)

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \quad (1.8a)$$

where the matrix Q has the property: $Q^T + Q$ is (essentially) positive definite.

[Correction Step.] Determine a nonsingular matrix M and a scalar $\alpha > 0$, let

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \quad (1.8b)$$

Convergence Conditions

For the matrices Q and M, and the step size α determined in (1.8), the matrices

$$H = QM^{-1} \quad (1.9a)$$

and

$$G = Q^T + Q - \alpha M^T HM. \quad (1.9b)$$

are positive definite.

1.2 Interpreting ADMM in the framework

The ADMM scheme (1.5) (and its linearized version (1.6)) is also a special case of the prototype algorithm (1.8). Recall the model (1.1) can be explained as the VI (1.4) with the specification given in (1.4b)-(1.4c).

In order to cast the ADMM scheme (1.5) (and (1.6)) into a special case of (1.8), let us first define the artificial vector $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ by

$$\left\{ \begin{array}{lcl} \tilde{x}^k & = & x^{k+1}, \\ \tilde{y}^k & = & y^{k+1}, \\ \tilde{\lambda}^k & = & \lambda^k - \beta(Ax^{k+1} + By^k - b), \end{array} \right. \tag{1.10}$$

where (x^{k+1}, y^{k+1}) is generated by the ADMM (1.5). Using these notations, we define the prediction by

$$\left\{ \begin{array}{l} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta(\tilde{x}^k, y, \lambda^k) \,|\, y \in \mathcal{Y}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \end{array} \right. \tag{1.11}$$

According to the scheme (1.5), the defined artificial vector $\tilde{w}^k$ satisfies the following VI:

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, & \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T(-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k)) \geq 0, & \forall\, y \in \mathcal{Y}, \\ \quad (A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases} \tag{1.12}$$

This can be written in form of (1.8a) as described in the following lemma.

Lemma 1.1 *For given v^k, let w^{k+1} be generated by* (1.5) *and $\tilde{w}^k$ be defined by* (1.10)*. Then, we have $\tilde{w}^k \in \Omega$ and*

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

where

$$Q = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix}. \tag{1.13}$$

Recall the essential variable of the ADMM scheme (1.5) is (y, λ). Moreover, using the definition of $\tilde{w}^k$, the λ^{k+1} updated by (1.5) can be represented as

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

Therefore, the output of ADMM scheme (1.5) can be written as the correction

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \qquad (1.14a)$$

which corresponds to the step (1.8b) in the framework with

$$M = \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \quad \text{and} \quad \alpha = 1. \qquad (1.14b)$$

Now we check that the Convergence Condition is satisfied by the ADMM scheme (1.5).

Indeed, for the matrix M in (1.14b), we have

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

Thus, by using (1.13) and (1.14b), we obtain $H = QM^{-1}$ and thus

$$H = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ \beta B & I \end{pmatrix} = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \qquad (1.15)$$

and consequently (since $\alpha = 1$)

$$\begin{aligned} G &= Q^T + Q - \alpha M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} \beta B^T B & -B^T \\ 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \\ &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \qquad (1.16) \end{aligned}$$

Therefore, H is symmetric and positive definite under the assumption that B is full column rank; and G is positive semi-definite. The Convergence Condition is satisfied; and thus the convergence of the ADMM scheme (1.5) is guaranteed.

1.3 Interpreting the Linearized ADMM in the framework

For the linearized version (1.6), using the same definition about $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ as in (1.10), the predictor is given by

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta(\tilde{x}^k, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_D^2 \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{cases} \tag{1.17}$$

Because $D = sI - \beta B^T B$, in comparing with (1.12), the defined artificial vector $\tilde{w}^k$ satisfies the following VI:

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, & \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T(-B^T\tilde{\lambda}^k + s(\tilde{y}^k - y^k)) \geq 0, & \forall\, y \in \mathcal{Y}, \\ (A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases} \tag{1.18}$$

This can be written in form of (1.8a) as described in the following lemma.

Lemma 1.2 *For given* v^k*, let* w^{k+1} *be generated by* (1.6) *and* $\tilde{w}^k$ *be defined by* (1.10)*. Then, we have* $\tilde{w}^k \in \Omega$ *and*

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

where

$$Q = \begin{pmatrix} sI_m & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix}. \tag{1.19}$$

In this way, the output of the linearized ADMM scheme (1.6) can be written as (1.14) with the same M and $\alpha = 1$. By using (1.19) and (1.14b), we obtain $H = QM^{-1}$ and thus

$$H = \begin{pmatrix} sI_m & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ \beta B & I \end{pmatrix} = \begin{pmatrix} sI_m & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{1.20}$$

and consequently (since $\alpha = 1$)

$$\begin{aligned} G &= Q^T + Q - M^T H M \;=\; Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2sI_m & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} sI_m & -B^T \\ 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \\ &= \begin{pmatrix} 2sI_m & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} sI_m + \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} \\ &= \begin{pmatrix} sI_m - \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \end{aligned} \tag{1.21}$$

Since $sI_m - \beta B^T B \succ 0$, both the matrices H (in (1.20)) and G (in (1.21)) are symmetric and positive definite. The Convergence Condition is satisfied; and thus the convergence of the ADMM scheme (1.5) is guaranteed.

2 Convergence proof in the unified framework

In this section, assuming the conditions (1.9) in the unified framework are satisfied, we prove some convergence properties.

Theorem 2.1 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (1.4) *and* $\tilde{w}^k$ *is obtained in the* k*-th iteration. If* v^k*,* v^{k+1} *and* $\tilde{w}^k$ *satisfy the conditions in the unified framework, then we have*

$$\begin{array}{rl} & \alpha\big(\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\big) \\ \geq & \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{\alpha}{2}\|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega. \end{array} \tag{2.1}$$

Proof. Using $Q = HM$ (see (1.9a)) and the relation (1.8b), the right hand side of (1.9a) can be written as $(v - \tilde{v}^k)^T \frac{1}{\alpha} H(v^k - v^{k+1})$ and hence

$$\alpha\{\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\} \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \ \forall w \in \Omega. \tag{2.2}$$

Applying the identity

$$(a-b)^T H(c-d) = \frac{1}{2}\{\|a-d\|_H^2 - \|a-c\|_H^2\} + \frac{1}{2}\{\|c-b\|_H^2 - \|d-b\|_H^2\},$$

to the right hand side of (2.2) with

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

we thus obtain

$$\begin{aligned}(v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ = \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \end{aligned} \tag{2.3}$$

For the last term of (2.3), we have

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&= \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\
&\overset{(1.9a)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - \alpha M(v^k - \tilde{v}^k)\|_H^2 \\
&= \quad 2\alpha(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - \alpha^2(v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\
&= \quad \alpha(v^k - \tilde{v}^k)^T(Q^T + Q - \alpha M^T HM)(v^k - \tilde{v}^k) \\
&\overset{(1.9b)}{=} \quad \alpha\|v^k - \tilde{v}^k\|_G^2.
\end{aligned} \tag{2.4}$$

Substituting (2.3), (2.4) in (2.2), the assertion of this theorem is proved. $\square$

2.1 Convergence in a strictly contraction sense

Theorem 2.2 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (1.4) *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*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha\|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.5}$$

Proof. Setting $w = w^*$ in (2.1), we get

$$\begin{aligned} &\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &\quad \ge \quad \alpha\|v^k - \tilde{v}^k\|_G^2 + 2\alpha\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \end{aligned} \tag{2.6}$$

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

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

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \alpha\|v^k - \tilde{v}^k\|_G^2. \tag{2.7}$$

The assertion (2.5) follows directly. $\Box$

2.2 Convergence rate in an ergodic sense

Equivalent Characterization of the Solution Set of VI

For the convergence rate analysis, we need another characterization of the solution set of VI (1.4). It can be described the following theorem and the proof can be found in [**?**] (Theorem 2.3.5) or [10] (Theorem 2.1). We include all the details for completeness.

Theorem 2.3 *The solution set of VI*(Ω, F, θ) *is convex and it can be characterized as*

$$\Omega^* = \bigcap_{w\in\Omega} \big\{\tilde{w} \in \Omega : \big(\theta(u) - \theta(\tilde{u})\big) + (w-\tilde{w})^T F(w) \geq 0\big\}. \qquad (2.8)$$

Proof. Indeed, if $\tilde{w} \in \Omega^*$, we have

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

By using the monotonicity of F on Ω, this implies that

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

Thus, $\tilde{w}$ belongs to the right-hand set in (2.8). Conversely, suppose $\tilde{w}$ belongs to the latter set of (2.8). Let $w \in \Omega$ be arbitrary. The vector

$$\bar{w} = \alpha\tilde{w} + (1-\alpha)w$$

belongs to Ω for all $\alpha \in (0,1)$. Thus we have

$$\theta(\bar{u}) - \theta(\tilde{u}) + (\bar{w} - \tilde{w})^T F(\bar{w}) \geq 0. \tag{2.9}$$

Because $\theta(\cdot)$ is convex, we have

$$\theta(\bar{u}) \leq \alpha\theta(\tilde{u}) + (1-\alpha)\theta(u) \quad \Rightarrow \quad (1-\alpha)(\theta(u) - \theta(\tilde{u})) \geq \theta(u) - \theta(\tilde{u}).$$

Substituting it in (2.9) and using $\bar{w} - \tilde{w} = (1-\alpha)(w - \tilde{w})$, we get

$$(\theta(u) - \theta(\tilde{u})) + (w - \tilde{w})^T F(\alpha\tilde{w} + (1-\alpha)w) \geq 0$$

for all $\alpha \in (0,1)$. Letting $\alpha \to 1$, it yields

$$(\theta(u) - \theta(\tilde{u})) + (w - \tilde{w})^T F(\tilde{w}) \geq 0.$$

Thus $\tilde{w} \in \Omega^*$. Now, we turn to prove the convexity of Ω^*. For each fixed but arbitrary $w \in \Omega$, the set

$$\{\tilde{w} \in \Omega : \; \theta(\tilde{u}) + \tilde{w}^T F(w) \leq \theta(u) + w^T F(w)\}$$

and its equivalent expression

$$\{\tilde{w} \in \Omega : \ \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \geq 0\}$$

is convex. Since the intersection of any number of convex sets is convex, it follows that the solution set of VI(Ω, F, θ) is convex. $\Box$

In Theorem 2.3, we have proved the equivalence of

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

and

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

We use the late one to define the approximate solution of VI (1.4). Namely, for given $\epsilon > 0$, $\tilde{w} \in \Omega$ is called an ϵ-approximate solution of VI(Ω, F, θ), if it satisfies

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

where

$$\mathcal{D}_{(\tilde{w})} = \{w \in \Omega \,|\, \|w - \tilde{w}\| \leq 1\}.$$

We need to show that for given $\epsilon > 0$, after t iterations, it can offer a $\tilde{w} \in \mathcal{W}$, such that

$$\tilde{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \left\{ \theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \right\} \leq \epsilon. \tag{2.10}$$

Theorem 2.1 is also the base for the convergence rate proof. Using the monotonicity of F, we have

$$(w - \tilde{w}^k)^T F(w) \geq (w - \tilde{w}^k)^T F(\tilde{w}^k).$$

Substituting it in (2.1), we obtain

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2\alpha} \|v - v^k\|_H^2 \geq \frac{1}{2\alpha} \|v - v^{k+1}\|_H^2, \ \forall w \in \Omega. \tag{2.11}$$

Note that the above assertion is hold for $G \succeq 0$.

Theorem 2.4 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (1.4) *and* $\tilde{w}^k$ *is obtained in the* k*-th iteration. Assume that* v^k*,* v^{k+1} *and* $\tilde{w}^k$ *satisfy the conditions in the unified framework and let* $\tilde{w}_t$ *be defined by*

$$\tilde{w}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{w}^k. \tag{2.12}$$

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

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2\alpha(t+1)} \|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{2.13}$$

Proof. First, it holds that $\tilde{w}^k \in \Omega$ for all $k \ge 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (2.12) implies that $\tilde{w}_t \in \Omega$. Summing the inequality (2.11) over $k = 0, 1, \ldots, t$, we obtain

$$(t+1)\theta(u) - \sum_{k=0}^{t} \theta(\tilde{u}^k) + \Big((t+1)w - \sum_{k=0}^{t} \tilde{w}^k\Big)^T F(w) + \frac{1}{2\alpha} \|v - v^0\|_H^2 \ge 0, \quad \forall w \in \Omega.$$

Use the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1} \sum_{k=0}^{t} \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2\alpha(t+1)} \|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{2.14}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{u}^k,$$

we have that

$$\theta(\tilde{u}_t) \le \frac{1}{t+1} \sum_{k=0}^{t} \theta(\tilde{u}^k).$$

Substituting it in (2.14), the assertion of this theorem follows directly. $\Box$

Recall (2.10). The conclusion (2.13) thus indicates obviously that the method is able to generate an approximate solution (i.e., $\tilde{w}_t$) with the accuracy $O(1/t)$ after t iterations. That is, in the case $G \succeq 0$, the convergence rate $O(1/t)$ of the method is established.

- **For the unified framework and the convergence proof, the reader can consult: 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.**

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

2.3 Convergence rate in pointwise iteration-complexity

In this subsection, we show that if the matrix G defined in (1.9b) is positive definite, a worst-case $O(1/t)$ convergence rate in a nonergodic sense can also be established for the prototype algorithm (1.8). Note in general a nonergodic convergence rate is stronger than the ergodic convergence rate.

We first need to prove the following lemma.

Lemma 2.1 *For the sequence generated by the prototype algorithm (1.8) where the Convergence Condition is satisfied, we have*

$$\begin{aligned} & (v^k - \tilde{v}^k)^T M^T H M \{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \\ \geq \quad & \frac{1}{2\alpha} \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \end{aligned} \tag{2.15}$$

Proof. First, set $w = \tilde{w}^{k+1}$ in (1.8a), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{2.16}$$

Note that (1.8a) is also true for $k := k+1$ and thus we have

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

Set $w = \tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}). \tag{2.17}$$

Combining (2.16) and (2.17) and using the monotonicity of F, we get

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq 0. \tag{2.18}$$

Adding the term

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

to the both sides of (2.18), and using $v^T Q v = \frac{1}{2} v^T (Q^T + Q) v$, we obtain

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

Substituting $(v^k - v^{k+1}) = \alpha M(v^k - \tilde{v}^k)$ in the left-hand side of the last inequality and using $Q = HM$, we obtain (2.15) and the lemma is proved. $\square$

Now, we are ready to prove (2.19), the key inequality in this section.

Theorem 2.5 *For the sequence generated by the prototype algorithm (1.8) where the Convergence Condition is satisfied, we have*

$$\|M(v^{k+1} - \tilde{v}^{k+1})\|_H \leq \|M(v^k - \tilde{v}^k)\|_H, \quad \forall\, k > 0. \tag{2.19}$$

Proof. Setting $a = M(v^k - \tilde{v}^k)$ and $b = M(v^{k+1} - \tilde{v}^{k+1})$ in the identity

$$\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a-b) - \|a-b\|_H^2,$$

we obtain

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

Inserting (2.15) into the first term of the right-hand side of the last equality, we obtain

$$
\begin{aligned}
&\|M(v^k-\tilde{v}^k)\|_H^2 - \|M(v^{k+1}-\tilde{v}^{k+1})\|_H^2 \\
&\quad \geq \quad \frac{1}{\alpha}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)} - \|M[(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})]\|_H^2 \\
&\quad = \quad \frac{1}{\alpha}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|_G^2 \geq 0,
\end{aligned}
$$

where the last inequality is because of the positive definiteness of the matrix $(Q^T+Q)-\alpha M^T HM \succeq 0$. The assertion (2.19) follows immediately. □

Note that it follows from $G \succ 0$ and Theorem 2.2 there is a constant $c_0 > 0$ such that

$$
\|v^{k+1}-v^*\|_H^2 \leq \|v^k-v^*\|_H^2 - c_0\|M(v^k-\tilde{v}^k)\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.20}
$$

Now, with (2.20) and (2.19), we can establish the worst-case $O(1/t)$ convergence rate in a nonergodic sense for the prototype algorithm (1.8).

Theorem 2.6 *Let $\{v^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prototype algorithm (1.8) under the Convergence Condition. For any integer $t > 0$, we have*

$$
\|M(v^t-\tilde{v}^t)\|_H^2 \leq \frac{1}{(t+1)c_0}\|v^0-v^*\|_H^2. \tag{2.21}
$$

Proof. First, it follows from (2.20) that

$$\sum_{k=0}^{\infty} c_0 \|M(v^k - \tilde{v}^k)\|_H^2 \le \|v^0 - v^*\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{2.22}$$

According to Theorem 2.5, the sequence $\{\|M(v^k - \tilde{v}^k)\|_H^2\}$ is monotonically non-increasing. Therefore, we have

$$(t+1)\|M(v^t - \tilde{v}^t)\|_H^2 \le \sum_{k=0}^{t} \|M(v^k - \tilde{v}^k)\|_H^2. \tag{2.23}$$

The assertion (2.21) follows from (2.22) and (2.23) immediately. □

Let $d := \inf\{\|v^0 - v^*\|_H \mid v^* \in \mathcal{V}^*\}$. Then, for any given $\epsilon > 0$, Theorem 2.6 shows that the it needs at most $\lfloor d^2/c_0\epsilon \rfloor$ iterations to ensure that $\|M(v^k - \tilde{v}^k)\|_H^2 \le \epsilon$. Recall that v^k is a solution of VI(Ω, F, θ) if $\|M(v^k - \tilde{v}^k)\|_H^2 = 0$ (see (1.8a) and due to $Q = HM$). A worst-case $O(1/t)$ convergence rate in pointwise iteration-complexity is thus established for the prototype algorithm (1.8).

Notice that, for a differentiable unconstrained convex optimization $\min f(x)$, it holds that

$$f(x) - f(x^*) = \nabla f(x^*)^T (x - x^*) + O(\|x - x^*\|^2) = O(\|x - x^*\|^2).$$

3 Classical ADMM in the Unified Framework

This section interprets the classical ADMM (1.5) and the linearized lADMM (1.6).

ADMM (1.5) Note that Theorem 2.4 is true for $G \succeq 0$. Thus the classical ADMM (1.5) has $O(1/t)$ convergence rate in the ergodic sense.

Since $\alpha = 1$, according to (2.5) and the form of G in (1.16), we have

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

Lemma 3.1 *For given v^k, let w^{k+1} be generated by* (1.5) *and $\tilde{w}^k$ be defined by* (1.10)*. Then, we have*

$$\tfrac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 \ge \|v^k - v^{k+1}\|_H^2. \tag{3.2}$$

Proof. According to (1.5) and (1.10), 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

$$y^{k+1} \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \ge 0, \; \forall\, y \in \mathcal{Y}. \quad (3.3)$$

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

$$y^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(y^k) + (y - y^k)^T\{-B^T\lambda^k\} \ge 0, \; \forall\, y \in \mathcal{Y}. \quad (3.4)$$

Setting $y = y^k$ in (3.3) and $y = y^{k+1}$ in (3.4), and then adding them, we get

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \ge 0. \quad (3.5)$$

Using $\lambda^k - \tilde{\lambda}^k = (\lambda^k - \lambda^{k+1}) + \beta B(y^k - y^{k+1})$ and the inequality (3.5), we obtain

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

The assertion of this lemma is proved. □

Substituting (3.2) in (3.1), we get the following nice property of the classical ADMM as in Lecture 11.

Theorem 3.1 *For given* v^k*, let* w^{k+1} *be generated by* (1.5)*. Then, we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{3.6}$$

Notice that the sequence $\{\|v^k - v^{k+1}\|_H^2\}$ generated by the classical ADMM is monotone non-increasing. In fact, in Theorem 2.5, we have proved that (see (2.19))

$$\|M(v^k - \tilde{v}^k)\|_H \le \|M(v^{k-1} - \tilde{v}^{k-1})\|_H, \quad \forall\, k \ge 1. \tag{3.7}$$

Because (see the correction formula (1.14)), $v^k - v^{k+1} = M(v^k - \tilde{v}^k)$, it follows from (3.7) that

$$\|v^k - v^{k+1}\|_H^2 \le \|v^{k-1} - v^k\|_H^2. \tag{3.8}$$

On the other hand, the inequality (3.6) tell us that

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

Thus, we have

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

Therefore, ADMM (1.5) has $O(1/t)$ convergence rate in pointwise iteration-complexity [12].

ADMM (1.6) According to Theorem 2.2, for the v^{k+1} generated by (1.6) and $\tilde{v}^k$ defined by (1.10), we have

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*, \tag{3.9}$$

where

$$H = \begin{pmatrix} sI_m & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} sI_m - \beta B^T B & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix}.$$

Since

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

it follows from (3.9) that

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|M^{-1}(v^k - v^{k+1})\|_G^2, \quad \forall v^* \in \mathcal{V}^*, \qquad (3.10)$$

Using $M = \begin{bmatrix} I & 0 \\ -\beta B & I \end{bmatrix}$ (see (1.14b)) and defining $S = M^{-T} G M^{-1}$, we have

$$\begin{aligned} S &= \begin{bmatrix} I & \beta B \\ 0 & I \end{bmatrix} \begin{bmatrix} sI_m - \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{bmatrix} \begin{bmatrix} I & 0 \\ \beta B & I \end{bmatrix} \\ &= \begin{bmatrix} sI_m & B^T \\ B & \frac{1}{\beta} I \end{bmatrix}. \end{aligned}$$

Since $sI_m \succ \beta\rho(B^T B)$, the matrix S is positive definite and it follows from (3.10) that

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

4 ADMM in Sense of Customized PPA [3]

If we change the performance order of y and λ of the classical ADMM (1.5), it becomes

$$\begin{cases} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+1}) \,|\, y \in \mathcal{Y}\}. \end{cases} \tag{4.1}$$

In this way we can get a positive semidefinite matrix Q in (1.8a). We define

$$\tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1}, \quad \tilde{\lambda}^k = \lambda^{k+1}, \tag{4.2}$$

where $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is the output of (4.1) and thus it can be rewritten as

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta(\tilde{x}^k, y, \tilde{\lambda}^k) \,|\, y \in \mathcal{Y}\}. \end{cases} \tag{4.3}$$

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 (4.3) is

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, \quad \forall\, x \in \mathcal{X}. \qquad (4.4)$$

Notice that

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

ignoring the constant term in the y optimization subproblem of (4.3), it turns to

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

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

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

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

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \left[-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k) \right] \geq 0, \ \forall y \in \mathcal{Y}. \tag{4.5}$$

From the λ update form in (4.3) we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \tag{4.6}$$

Combining (4.4), (4.5) and (4.6), and using the notations of (1.4), we get following lemma.

Lemma 4.1 *For given v^k, let $\tilde{w}^k$ be generated by* (4.3)*. 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

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

Because Q is symmetric and positive semidefinite, according to (1.8), 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 (4.3) with any small $\delta > 0$, it becomes

$$\left\{ \begin{array}{l} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta^{(2)}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta^{(2)}(\tilde{x}^k, y, \tilde{\lambda}^k) + \frac{\delta\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}. \end{array} \right. \tag{4.8}$$

In the ADMM based customized PPA (4.3), the y-subproblem can be written as

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \frac{\beta}{2}\|By - p^k\|^2 \,|\, y \in \mathcal{Y}\}, \tag{4.9}$$

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

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

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

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{(1+\delta)\beta}{2}\|By - q^k\|^2 \mid y \in \mathcal{Y}\}, \tag{4.10}$$

where

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

In this way, the matrix Q in (4.7) 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 (4.9), using (4.10) to get $\tilde{y}^k$, it will ensure the positivity of H theoretically. However, in practical computation, it works still well by using $\delta = 0$.

5 Symmetric ADMM [7]

In the problem (1.1), 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

$$\begin{cases} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b). \end{cases} \tag{5.1}$$

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

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}, \tag{5.2}$$

where (x^{k+1}, y^{k+1}) is generated by the ADMM (5.1).

According to (5.2), the optimal condition of the x-subproblem of (5.1) is

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{5.3}$$

Notice that the objective function of the y-subproblem in (5.1) is

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

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 the optimality equivalent lemma , we have

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

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

$$\theta_2(y) - \theta_2(\tilde{y}^k) \;+\; (y - \tilde{y}^k)^T \big\{ -B^T \tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) \\ -\mu B^T(\tilde{\lambda}^k - \lambda^k) \big\} \ge 0, \;\; \forall y \in \mathcal{Y}. \tag{5.4}$$

According to the definition of $\tilde{w}^k$ in (5.2), we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \tag{5.5}$$

Combining (5.3), (5.4) and (5.5), and using the notations of (1.4), we get following lemma.

Lemma 5.1 *For given v^k, let w^{k+1} be generated by* (5.1) *and $\tilde{w}^k$ be defined by* (5.2)*. 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

$$Q = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{5.6}$$

Using this notation and by a manipulation, the update form of λ in (5.1) can be represented as

$$\begin{aligned} \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \mu\big[-\beta B(y^k - \tilde{y}^k) + \beta(Ax^{k+1} + By^k - b)\big] \\ &= \lambda^k - \big[-\mu\beta B(y^k - \tilde{y}^k) + 2\mu(\lambda^k - \tilde{\lambda}^k)\big]. \end{aligned} \tag{5.7}$$

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

$$M = \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix}. \tag{5.8}$$

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 (5.8), 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 (5.6) 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 (1.9) are satisfied.

Take $\mu = 0.9$, it will accelerate the convergence much. For the numerical experiments of this method, it is refereed to consult [7].

The symmetric ADMM is a special version of the unified framework (1.8) - (1.9) 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 2.2, we have

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

Conclusion Remarks

Theorem 2.1 is the base of this note. Using $\alpha M(v^k - \tilde{v}^k) = (v^k - v^{k+1})$ and $Q = HM$, we get

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

The elementary identity

$$(a-b)^T H(c-d) = \frac{1}{2}\{\|a-d\|_H^2 - \|a-c\|_H^2\} + \frac{1}{2}\{\|c-b\|_H^2 - \|d-b\|_H^2\},$$

plays a key role in the proof. we thus obtain

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

This basic technique has also been adopted by A. Beck [1] in his monographs. He has made a special footnote (see Page 428 in [1]). The technique is simple yet powerful !

References

[1] A. Beck, First-Order Methods in Optimization, MOS-SIAM Seriers on Optimization, 2017.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, manuscript, 2010.

[3] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A proximal point algorithms revisit on the alternating direction method of multipliers, Science China Mathematics, 56 (2013), 2179-2186.

[4] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appli., 2 (1976), pp. 17-40.

[5] R. Glowinski and P. Le Tallec, Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics, SIAM Studies in Applied Mathematics, Philadelphia, PA, 1989.

[6] 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.

[7] 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.

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

[9] B. S. He, M. H. Xu and X. M. Yuan, Solving large-scale least squares covariance matrix problems by alternating direction methods, SIAM J. Matrix Analy. Appli., 32, (2011), pp. 136-152.

[10] B.S. He and X.M. Yuan, On the $O(1/n)$ Convergence Rate of the Douglas-Rachford Alternating Direction Method， SIAM J. Numer. Anal. 50, 700-709, 2012.

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

[12] B.S. He and X.M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers, Numer. Math., (2015) 130:567–577.

[13] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appli., 4(1969), pp. 303-320.

[14] S. Kontogiorgis and R. R. Meyer, A variable-penalty alternating directions method for convex optimization, Math. Program., 83 (1998), pp. 29-53.

[15] A. Nemirovski, *Prox-method with rate of convergence $O(1/t)$ for variational inequality with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM J. Optim. 15 (2005), pp. 229-251.

[16] Y. E. Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence O($1/k^2$)*, Doklady AN SSSR, 269 (1983), pp. 543-547.

[17] P. Tseng, *On accelerated proximal gradient methods for convex-concave optimization*, manuscript, University of Washington, USA, 2008.