凸优化和单调变分不等式的收缩算法

第十四讲: 定制 PPA 算法
意义的线性化交替方向法

Linearized alternating direction method
in sense of the customized PPA

南京大学数学系　　何炳生

hebma@nju.edu.cn

The context of this lecture is based on the manuscript [2]

1   Structured constrained convex optimization

We consider the following structured constrained convex optimization problem

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

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.

The task of solving the problem (1.1) is to find an $(x^*, y^*, \lambda^*) \in \Omega$, such that

$$\begin{cases} \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, \\ \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-B^T\lambda^*) \geq 0, \quad \forall\, (x, y, \lambda) \in \Omega, \quad (1.2) \\ (\lambda - \lambda^*)^T(Ax^* + By^* - b) \geq 0, \end{cases}$$

where

$$\Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m.$$

By denoting

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

and

$$\theta(u) = \theta_1(x) + \theta_2(y),$$

the first order optimal condition (1.2) can be written in a compact form such as

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \ \forall \, w \in \Omega. \quad (1.3)$$

Note that the mapping F is monotone. We use Ω^* to denote the solution set of the variational inequality (1.3). For convenience we use the notations

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

Applied the ADM-based customized PPA to the problem (1.1)

From given $v^k = (y^k, \lambda^k)$, the prediction step produces $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$.

The prediction step:

1. First, for given (y^k, λ^k), $\tilde{x}^k$ is the solution of the following problem

$$\tilde{x}^k = \text{Argmin} \left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k - b) \\ + \frac{\beta}{2}\|Ax + By^k - b\|^2 \end{array} \,\middle|\, x \in \mathcal{X} \right\} \quad (1.4a)$$

2. Set the multipliers by

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \quad (1.4b)$$

3. Finally, use the obtained $\tilde{x}^k$ and $\tilde{\lambda}^k$, find $\tilde{y}^k$ by

$$\tilde{y}^k = \text{Argmin} \left\{ \begin{array}{c} \theta_2(y) - (\tilde{\lambda}^k)^T(A\tilde{x}^k + By - b) \\ + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \end{array} \,\middle|\, y \in \mathcal{Y} \right\} \quad (1.4c)$$

As analyzed in the last chapter, we have

$$\tilde{w}^k \in \Omega, \; \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T \{F(\tilde{w}^k) + Q(\tilde{v}^k - v^k)\} \geq 0, \; \forall\, w \in \Omega, \;\; \text{(1.5)}$$

where

$$Q = \begin{pmatrix} 0 & 0 \\ \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \quad \text{and} \quad M = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \quad \text{(1.6)}$$

The new iterate v^{k+1} is given by

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

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

$$\|v^{k+1} - v^*\|_M^2 \leq \|v^k - v^*\|_M^2 - \|v^k - v^{k+1}\|_M^2.$$

2   Linearized ADM-based PPA Method

Note that the subproblems (1.4a) and (1.4c) in the last section are equivalent to the problems

$$\tilde{x}^k = \mathsf{Argmin}\big\{\theta_1(x) + \tfrac{\beta}{2}\|(Ax + By^k - b) - \tfrac{1}{\beta}\lambda^k\|^2 \big| x \in \mathcal{X}\big\} \qquad (2.1\text{a})$$

and

$$\tilde{y}^k = \mathsf{Argmin}\big\{\theta_2(y) + \tfrac{\beta}{2}\|(A\tilde{x}^k + By - b) - \tfrac{1}{\beta}\tilde{\lambda}^k\|^2 \big| y \in \mathcal{Y}\big\} \qquad (2.1\text{b})$$

respectively. In some structured optimization (1.1), the subproblem (2.1b) is easy because B is usually a scalar matrix. However, to obtain the solution of the subproblem (2.1a) is expensive in the case that A does not have a special form. In this lecture, we suppose that only the solution of the problem

$$\min\left\{\theta_1(x) + \frac{r}{2}\|x - a\|^2 \mid x \in \mathcal{X}\right\}$$

has a closed form, and consider to linearize the quadratic function of the subproblem (2.1a) ADM in sense of the customized PPA.

2.1 Linearized alternating direction method

The prediction step:

1. First, for given (x^k, y^k, λ^k), solving the x subproblem to get $\tilde{x}^k$ by

$$\tilde{x}^k = \text{Argmin} \left(\begin{cases} \theta_1(x) + \beta x^T A^T (Ax^k + By^k - b - \frac{1}{\beta}\lambda^k) \\ + \frac{r}{2}\|x - x^k\|^2 \quad | \ x \in \mathcal{X}\} \end{cases} \right).$$

$$(2.2a)$$

2. Set the new multipliers by

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \qquad (2.2b)$$

3. Finally, use the obtained $\tilde{x}^k$ and $\tilde{\lambda}^k$, solving the y subproblem to get $\tilde{y}^k$ by

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

$$(2.2c)$$

Request on the parameter r   For given $\beta > 0$, r should satisfy

$$rI - \beta A^T A \succeq 0. \tag{2.3}$$

The correction step: Update the new iterate w^{k+1} by

$$w^{k+1} = w^k - \gamma(w^k - \tilde{w}^k), \quad \gamma \in [1, 2). \tag{2.4}$$

To get the new iterate w^{k+1}, this method does not need to calculate the step size. However, it needs to estimate the max-eigenvalue of $A^T A$, *i. e.*, $\lambda_{\max}(A^T A)$.

2.2 Analysis in the PPA framework

Note that the solution of (2.2a), $\tilde{x}^k$ satisfies

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T$$
$$\{-A^T \lambda^k + \beta A^T(Ax^k + By^k - b) + r(\tilde{x}^k - x^k)\} \geq 0, \ \forall x \in \mathcal{X}. \tag{2.5}$$

Substituting $\tilde{\lambda}^k$ (see (2.2b)) in (2.5) (eliminating λ^k), we get

$$\tilde{x}^k \in \mathcal{X}, \;\; \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T$$
$$\{-A^T\tilde{\lambda}^k + (rI - \beta A^T A)(\tilde{x}^k - x^k)\} \geq 0, \; \forall\, x \in \mathcal{X}. \quad (2.6)$$

The solution of (2.2c), $\tilde{y}^k$ satisfies

$$\tilde{y}^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T$$
$$\{-B^T\tilde{\lambda}^k + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b)\} \geq 0, \; \forall\, y \in \mathcal{Y}. \quad (2.7)$$

Note that $\beta(A\tilde{x}^k + B\tilde{y}^k - b) = (\lambda^k - \tilde{\lambda}^k) + \beta B(\tilde{y}^k - y^k)$ (see (2.2b)).
Substituting it in (2.7), we obtain

$$\tilde{y}^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T$$
$$\{-B^T(2\tilde{\lambda}^k - \lambda^k) + \beta B^T B(\tilde{y}^k - y^k))\} \geq 0, \; \forall\, y \in \mathcal{Y}. \quad (2.8)$$

From (2.2b) we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \quad (2.9)$$

Combining the inequalities (2.6), (2.8) and (2.9), we obtain

$$
\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} \right.
$$

$$
+ \begin{pmatrix} (rI - \beta A^T A) & 0 & 0 \\ 0 & \beta B^T B & -B^T \\ 0 & -B & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} \tilde{x}^k - x^k \\ \tilde{y}^k - y^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \right\} \geq 0, \ \forall\, w \in \Omega. \text{(2.10)}
$$

The last variational inequality can be written in form of

$$
\tilde{w}^k \in \Omega, \ (w - \tilde{w}^k)^T \{ F(\tilde{w}^k) + G(\tilde{w}^k - w^k) \} \geq 0, \ \forall\, w \in \Omega, \quad \text{(2.11)}
$$

where

$$
G = \begin{pmatrix} (rI - \beta A^T A) & 0 & 0 \\ 0 & \beta B^T B & -B^T \\ 0 & -B & \frac{1}{\beta}I_m \end{pmatrix} \qquad \text{(2.12)}
$$

which is essential in the framework of the PPA contraction methods.

2.3   Convergence of the Linearized ADM-based PPA Method

Based on the analysis in the last subsection, we have the following lemma.

Lemma 2.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (2.2) *from the given* $w^k = (x^k, y^k, \lambda^k)$. *Then, we have*

$$(\tilde{w}^k - w^*)^T G(w^k - \tilde{w}^k) \geq 0, \ \forall w^* \in \Omega^*, \tag{2.13}$$

where G *is defined in* (2.12).

Proof. Setting $w = w^*$ in (2.10), we get

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

Since F is monotone and $\tilde{w}^k \in \Omega$, it follows that

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

The right hand side of the last inequality is non-negative because $\tilde{w}^k \in \Omega$ and $w^* \in \Omega^*$. the assertion follows directly. □

Lemma 2.2 *Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ be generated by (2.2) from the given $w^k = (x^k, y^k, \lambda^k)$. Then, we have*

$$(w^k - w^*)^T G(w^k - \tilde{w}^k) \geq \|w^k - \tilde{w}^k\|_G^2, \ \forall\, w^* \in \Omega^*, \qquad (2.14)$$

where G is defined in (2.12).

Proof. Assertion (2.14) follows from the last inequality directly. □

Since G is symmetric and positive semi-definite, we have

$$w^k - \tilde{w}^k = 0 \quad \text{or} \quad G(w^k - \tilde{w}^k) = 0,$$

whenever $\|w^k - \tilde{w}^k\|_G^2 = 0$. Therefore, it follows from (2.10) that $\tilde{w}^k$ is a solution of the variational inequality when $\|w^k - \tilde{w}^k\|_G^2 = 0$.

Theorem 2.1 *Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ be generated by (2.2) from the given*

$w^k = (x^k, y^k, \lambda^k)$ *and the new iterate* w^{k+1} *be given by* (2.4). *Then we have*

$$\|w^{k+1} - w^*\|_G^2 \le \|w^k - w^*\|_G^2 - \gamma(2-\gamma)\|w^k - \tilde{w}^k\|_G^2, \ \forall\, w^* \in \Omega^*. \quad (2.15)$$

Proof. By using (2.4) and (2.14), we obtain

$$
\begin{aligned}
&\|w^{k+1} - w^*\|_G^2 \\
&\overset{(2.4)}{=} \quad \|(w^k - w^*) - \gamma(w^k - \tilde{w}^k)\|_G^2 \\
&\overset{(2.14)}{\le} \quad \|w^k - w^*\|_G^2 - 2\gamma\|w^k - \tilde{w}^k\|_G^2 + \gamma^2\|w^k - \tilde{w}^k\|_G^2 \\
&= \quad \|w^k - w^*\|_G^2 - \gamma(2-\gamma)\|w^k - \tilde{w}^k\|_G^2.
\end{aligned}
$$

This is true for any $w^* \in \Omega^*$ and the theorem is proved. □

The inequality (2.15) is essential for the convergence of the Linearized alternating direction method. By using (2.4), the result of Theorem 2.1 can be written as

$$\|w^{k+1} - w^*\|_G^2 \le \|w^k - w^*\|_G^2 - \frac{2-\gamma}{\gamma}\|w^k - w^{k+1}\|_G^2, \ \forall\, w^* \in \Omega^*.$$

3 Convergence rate of L-ADM-based C-PPA Method

Lemma 3.1 *Let $\{w^k\}$ be the sequence generated by the customized PPA (2.2) with (2.4). Then, we have*

$$(\tilde{w}^k - \tilde{w}^{k+1})^T G\{(w^k - w^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})\} \geq 0. \qquad (3.1)$$

Proof. Set $w = \tilde{w}^{k+1}$ in (2.11), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T \{F(\tilde{w}^k) + G(\tilde{w}^k - w^k)\} \geq 0. \quad (3.2)$$

Note that (2.11) is also true for $k := k + 1$ and thus we have

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T \{F(\tilde{w}^{k+1}) + G(\tilde{w}^{k+1} - w^{k+1})\} \geq 0, \ \forall \, w \in \Omega.$$

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

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

$$(3.3)$$

Adding (3.2) and (3.3) and using the monotonicity of F, we get

$$(\tilde{w}^k - \tilde{w}^{k+1})^T G\{(w^k - w^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})\} \geq 0.$$

we obtain (3.1) immediately. $\square$

Lemma 3.2 *Let $\{w^k\}$ be the sequence generated by the customized PPA (2.2) with (2.4). Then, we have*

$$(w^k - \tilde{w}^k)^T G\{(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\}$$

$$\geq \frac{1}{\gamma}\|(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\|_G^2. \quad (3.4)$$

Proof. Adding the term $\|(w^k - w^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})\|_G^2$ to the both sides of (3.1), we obtain

$$(w^k - w^{k+1})^T G\{(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\}$$

$$\geq \|(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\|_G^2. \quad (3.5)$$

Substituting the term $\left(w^k - w^{k+1}\right)$ in the left hand side of (3.5) by $\gamma(w^k - \tilde{w}^k)$ (see (2.4)), we obtain (3.4) and the lemma is proved. $\quad\square$

Lemma 3.3 *Let $\{w^k\}$ be the sequence generated by the customized PPA (2.2) with (2.4). Then, we have*

$$\|w^{k+1} - \tilde{w}^{k+1}\|_G^2 \leq \|w^k - \tilde{w}^k\|_G^2. \tag{3.6}$$

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

$$\|a\|_G^2 - \|b\|_G^2 = 2a^T G(a - b) - \|a - b\|_G^2,$$

we obtain

$$\|w^k - \tilde{w}^k\|_G^2 - \|w^{k+1} - \tilde{w}^{k+1}\|_G^2$$
$$= \ 2(w^k - \tilde{w}^k)^T G\{(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\}$$
$$-\|(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\|_G^2.$$

By using (3.4) to the first term of the right hand side of the last equality, we obtain

$$\|w^k - \tilde{w}^k\|_G^2 - \|w^{k+1} - \tilde{w}^{k+1}\|_G^2 \geq \frac{2 - \gamma}{\gamma} \|(w^k - \tilde{w}^k) - (w^{k+1} - \tilde{w}^{k+1})\|_G^2.$$

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

Having the assertion (2.15) and Lemma 3.3, we are ready to present the $O(1/t)$ convergence rate of the customized PPA in the residue sense.

Theorem 3.1 *Let $\{w^k\}$ be the sequence generated by the customized PPA (2.2) with (2.4). Then, we have*

$$\|w^k - \tilde{w}^k\|_G^2 \leq \frac{1}{(k+1)\gamma(2-\gamma)} \|w^0 - w^*\|_G^2, \quad \forall w^* \in \Omega^*. \qquad (3.7)$$

Proof. First, it follows from (2.15) that

$$\gamma(2-\gamma) \sum_{t=0}^{\infty} \|w^t - \tilde{w}^t\|_G^2 \leq \|w^0 - w^*\|_G^2, \quad \forall w^* \in \Omega^*. \qquad (3.8)$$

According to Lemma 3.3, the sequence $\{\|w^t - \tilde{w}^t\|_H^2\}$ is non-increasing. Therefore, we have

$$(k+1)\|w^k - \tilde{w}^k\|_G^2 \le \sum_{i=0}^{k} \|w^i - \tilde{w}^i\|_G^2. \tag{3.9}$$

The assertion of this theorem follows from (3.8) and (3.9) directly. $\quad\square$

The solution set of the variational inequality $\text{VI}(\Omega, F, \theta)$ is convex and closed. Theorem 3.1 indicates that ADMM has $O(1/k)$ iteration convergence rate. Let

$$d = \inf\{\|w^0 - w^*\|_G \mid w^* \in \Omega^*\}.$$

For any given $\epsilon > 0$, in order to enforce the error $\|w^k - \tilde{w}^k\|_G^2 \le \epsilon$, according to (3.7), it needs at most $k = \lfloor d^2/\gamma(2-\gamma)\epsilon \rfloor$ iterations.

4   ADM-based Contraction Method

In ADM-based contraction methods, we use the $\tilde{w}^k$ generated by (2.2) to construct a search direction.

4.1   Contraction Method

The prediction step of the contraction method in this subsection is the same as (2.2). Therefore, we have (2.10) and rewrite it as

$$
\tilde{w}^k \in \Omega, \;\; \theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T \tilde{\lambda}^k \\ -B^T \tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} \right.
$$

$$
+ \begin{pmatrix} rI_n & 0 & 0 \\ 0 & \beta B^T B & -B^T \\ 0 & -B & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} (I_n - \frac{\beta}{r}A^T A)(\tilde{x}^k - x^k) \\ \tilde{y}^k - y^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \left. \right\} \geq 0, \; \forall w \in \Omega.
$$

Again, the above variational inequality can be written in form of

$$\tilde{w}^k \in \Omega, \ \theta(u) - \theta(\tilde{u}^k)$$

$$+(w - \tilde{w}^k)^T \{F(\tilde{w}^k) + Q(\tilde{w}^k - w^k)\} \geq 0, \ \forall \, w \in \Omega, \quad (4.1)$$

where

$$Q = HM, \tag{4.2}$$

$$H = \begin{pmatrix} rI_n & 0 & 0 \\ 0 & \beta B^T B & -B^T \\ 0 & -B & \frac{1}{\beta} I_m \end{pmatrix} \tag{4.3}$$

and

$$M = \begin{pmatrix} I_n - \frac{\beta}{r} A^T A & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & 0 & I_m \end{pmatrix}. \tag{4.4}$$

Request on the parameter r   For given $\beta > 0$, r should satisfy

$$\beta \|A^T A(x^k - \tilde{x}^k)\| \le \nu r \|x^k - \tilde{x}^k\|, \quad \nu \in (0, 1). \qquad (4.5)$$

If the condition (2.3) is satisfied, *i.e.*, $rI_n - \beta A^T A \succ 0$, then the condition (4.5) is hold. In conversely it is not true. A conservative estimate for $\|A^T A\|$ will leads slow convergence. In the iteration process, we can check if the condition (4.5) is satisfied. This section considers the contraction in H-norm, where H (defined in (4.3)) is symmetric and positive semi-definite.

4.2   Convergence of ADM-based contraction method

Based on the analysis in the last subsection, we have the following lemma.

Lemma 4.1 *Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ be generated by* (2.2) *from the given* $w^k = (x^k, y^k, \lambda^k)$. *Then, we have*

$$(\tilde{w}^k - w^*)^T Q(w^k - \tilde{w}^k) \ge 0, \quad \forall w^* \in \Omega^*, \qquad (4.6)$$

where matrix Q is defined in (4.2).

Proof. Setting $w = w^*$ in (4.1), we get

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

Since F is monotone, it follows that

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

The last inequality is due to $w^{k+1} \in \Omega$ and $w^* \in \Omega^*$ (see (1.3)). The lemma is proved. $\square$

Lemma 4.2 *Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ be generated by (2.2) from the given $w^k = (x^k, y^k, \lambda^k)$. Then, we have*

$$(w^k - w^*)^T HM(w^k - \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k), \ \forall w^* \in \Omega^*, \quad (4.7)$$

where H is defined in (4.3) *and*

$$\varphi(w^k, \tilde{w}^k) = (w^k - \tilde{w}^k)^T HM(w^k - \tilde{w}^k) \qquad (4.8)$$

Proof. It follows from (4.2) and (4.6) that

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

Assertion (4.7) and the definition of $\varphi(w^k, \tilde{w}^k)$ directly. $\square$

Even though H is positive semi-definite, we still use $\|w - \tilde{w}\|_H$ to denote that

$$\|w - \tilde{w}\|_H = \sqrt{(w - \tilde{w})^T H(w - \tilde{w})}.$$

4.3 The primary contraction methods

In the primary method, we take the unit step length and use

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k) \qquad (4.9)$$

to update the new iterate w^{k+1}. According to (4.4), it can be written as

$$
\begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \tilde{x}^k + \frac{\beta}{r} A^T A(x^k - \tilde{x}^k) \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix}.
\tag{4.10}
$$

In the primary contraction method, only the x-part of the corrector is different from the predictor. In the method of Section 2, we need $r \geq \beta \|A^T A\|$. By using the method in this section, we need only a r to satisfy the condition (4.5). In practical computation, we try to use the average of the eigenvalues of $\beta A^T A$.

By using (4.7), we obtain

$$
\begin{aligned}
\|w^k &- w^*\|_H^2 - \|w^{k+1} - w^*\|_H^2 \\
&= 2(w^k - w^*)^T HM(w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2 \\
&\geq 2\varphi(w^k, \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2.
\end{aligned}
\tag{4.11}
$$

Note that (see (4.8))

$$2\varphi(w^k, \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2$$
$$= (w^k - \tilde{w}^k)^T (M^T H + HM - M^T HM)(w^k - \tilde{w}^k).$$

By using the structure of the matrices H and M, we obtain

$$M^T H + HM - M^T HM = H - (I - M^T)H(I - M)$$

$$= \begin{pmatrix} rI_n & 0 & 0 \\ 0 & \beta B^T B & -B^T \\ 0 & -B & \frac{1}{\beta} I_m \end{pmatrix} - \begin{pmatrix} r(\frac{\beta}{r} A^T A)^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Therefore,

$$2\varphi(w^k, \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2$$
$$= \|w^k - \tilde{w}^k\|_H^2 - r\left(\frac{\beta^2}{r^2}\right)\|A^T A(x^k - \tilde{x}^k)\|^2.$$

Under the condition (4.5), we have

$$\left(\tfrac{\beta^2}{r^2}\right)\|A^T A(x^k - \tilde{x}^k)\|^2 \leq \nu^2 \|x^k - \tilde{x}^k\|^2.$$

Consequently, we have

$$2\varphi(w^k, \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2 \geq (1 - \nu^2)\|w^k - \tilde{w}^k\|_H^2. \qquad (4.12)$$

Theorem 4.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (2.2) *from the given* $w^k = (x^k, y^k, \lambda^k)$ *and the new iterate* w^{k+1} *is given by* (4.9). *The sequence* $\{w^k\}$ *generated by the elementary contraction method satisfies*

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - (1 - \nu^2)\|w^k - \tilde{w}^k\|_H^2. \qquad (4.13)$$

Theorem 4.1 is essential for the convergence of the primary contraction method. It will lead $\lim_{k\to\infty} \|w^k - \tilde{w}^k\|_H^2 = 0$ and

$$\lim_{k\to\infty} x^k = x^*, \qquad \lim_{k\to\infty} By^k = By^* \qquad \text{and} \qquad \lim_{k\to\infty} \lambda^k = \lambda^*.$$

4.4   The general contraction methods

The general contraction method   For given w^k, we use

$$w(\alpha) = w^k - \alpha M(w^k - \tilde{w}^k) \qquad (4.14)$$

to update the α-dependent new iterate. For any $w^* \in \Omega^*$, we define

$$\vartheta(\alpha) := \|w^k - w^*\|_H^2 - \|w(\alpha) - w^*\|_H^2 \qquad (4.15)$$

and

$$q(\alpha) = 2\alpha\varphi(w^k, \tilde{w}^k) - \alpha^2 \|M(w^k - \tilde{w}^k)\|_H^2. \qquad (4.16)$$

Theorem 4.2 *Let $w(\alpha)$ be defined by (4.14). For any $w^* \in \Omega^*$ and $\alpha \geq 0$, we have*

$$\vartheta(\alpha) \geq q(\alpha). \qquad (4.17)$$

Proof. It follows from (4.14) and (4.15) that

$$
\begin{aligned}
\vartheta(\alpha) &= \|w^k - w^*\|_H^2 - \|(w^k - w^*) - \alpha M(w^k - \tilde{w}^k)\|_H^2 \\
&= 2\alpha(w^k - w^*)^T HM(w^k - \tilde{w}^k) - \alpha^2\|M(w^k - \tilde{w}^k)\|_H^2.
\end{aligned}
$$

By using (4.7) and the definition of $q(\alpha)$, the theorem is proved. $\quad\square$

Note that $q(\alpha)$ in (4.16) is a quadratic function of α and it reaches its maximum at

$$
\alpha^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|M(w^k - \tilde{w}^k)\|_H^2}. \tag{4.18}
$$

From (4.12) we know that under the condition (2.3), it holds that $\alpha_k^* \geq \frac{1}{2}$. In practical computation, we use

$$
w^{k+1} = w^k - \gamma\alpha_k^* M(w^k - \tilde{w}^k), \tag{4.19}
$$

to update the new iterate with $\gamma \in [1, 2)$. According to (4.15) and (4.17), we have

$$
\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - q(\gamma\alpha_k^*). \tag{4.20}
$$

Using (4.16) and (4.18), we obtain

$$
\begin{aligned}
q(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*\varphi(w^k,\tilde{w}^k) - (\gamma\alpha_k^*)^2\|M(w^k - \tilde{w}^k)\|_H^2 \\
&= \gamma(2-\gamma)\alpha_k^*\varphi(w^k,\tilde{w}^k).
\end{aligned}
\tag{4.21}
$$

Note that $\alpha_k^* > 1/2$ and (see (4.12))

$$
\varphi(w^k,\tilde{w}^k) \geq \frac{1}{2}\big(\|M(w^k - \tilde{w}^k)\|_H^2 + (1-\nu)\|w^k - \tilde{w}^k\|_H^2\big). \tag{4.22}
$$

Combining (4.20), (4.21) and (4.22), we get the following theorem for the general contraction method.

Theorem 4.3 *The sequence $\{w^k = (x^k, y^k, \lambda^k)\}$ generated by the general contraction method* (4.19) *satisfies*

$$
\begin{aligned}
\|w^{k+1} - w^*\|_H^2 &\leq \|w^k - w^*\|_H^2 - \frac{\gamma(2-\gamma)(1-\nu)}{4}\|w^k - \tilde{w}^k\|_H^2 \\
&\quad - \frac{\gamma(2-\gamma)}{4}\|M(w^k - \tilde{w}^k)\|_H^2.
\end{aligned}
\tag{4.23}
$$

The inequality (4.23) in Theorem 4.3 is essential for the convergence of the

general contraction method.

On the other hand, by using (4.18) and (4.19), we have

$$
\begin{aligned}
q(\gamma\alpha_k^*) &= \gamma(2-\gamma)\alpha_k^*\varphi(w^k,\tilde{w}^k) \\
&= \gamma(2-\gamma)(\alpha_k^*)^2\|M(w^k-\tilde{w}^k)\|_H^2 \\
&= \frac{2-\gamma}{\gamma}\|w^k-w^{k+1}\|_H^2.
\end{aligned}
\tag{4.24}
$$

According to (4.15), (4.17) and the above inequality, we have

Theorem 4.4 *The sequence $\{w^k=(x^k,y^k,\lambda^k)\}$ generated by the general contraction method* (4.19) *satisfies*

$$
\|w^{k+1}-w^*\|_H^2 \le \|w^k-w^*\|_H^2 - \frac{2-\gamma}{\gamma}\|w^k-w^{k+1}\|_H^2.
\tag{4.25}
$$

Especially, by taking $\gamma=1$ in (4.19), then we have

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

Remark If we dynamically take $\gamma_k = 1/\alpha_k^*$, because $\alpha^* > 1/2$, we have

$\gamma_k \in (0, 2)$ and $\gamma_k \alpha_k^* \equiv 1$. In this way, we get the primary contraction method.

According to (4.25), since $\gamma = 1/\alpha_k^*$, we have

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - (2\alpha_k^* - 1)\|w^k - w^{k+1}\|_H^2. \quad (4.26)$$

Based on the above inequality, by using (4.18) and (4.9), we derive

$$
\begin{aligned}
&\|w^k - w^*\|_H^2 - \|w^{k+1} - w^*\|_H^2 \\
\geq\quad & (2\alpha_k^* - 1)\|w^k - w^{k+1}\|_H^2 \\
=\quad & \frac{2\varphi(w^k, \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2}{\|M(w^k - \tilde{w}^k)\|_H^2}\|w^k - w^{k+1}\|_H^2 \\
=\quad & 2\varphi(w^k, \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2,
\end{aligned}
$$

the same result as (4.11). Finally, from the last inequality, we can obtain the assertion in Theorem 4.1 for the primary contraction method.

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, Nov. 2010.

[2] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A relaxed customized proximal point algorithm for separable convex programming. http://www.optimization-online.org/DB_HTML/2011/08/3141.html

[3] D. Gabay, "Applications of the Method of Multipliers to Variational Inequalities," in *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, M. Fortin and R. Glowinski, eds., North Holland, Amsterdam, The Netherlands, 299-331, 1983.

[4] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems viafinite element approximations, Computers and Mathematics with Applications, 2:17-40, 1976.

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

[6] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, pp. 69-76, 1997.

[7] B.S. He, "Inexact implicit methods for monotone general variational inequalities", Math. Program. Vol. 86, Series A, pp. 199-217, 1999

[8] B.S. He, L-Z Liao, D.R. Han and H. Yang, *A new inexact alternating directions method for monotone variational inequalities*, Math. Progr., **92**, pp. 103-118, 2002

[9] B. S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, Pacific J. Optimization, 4, 195-212, 2008.

[10] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim., 14 (1976), pp. 877-898.