第十二讲: 线性化的交替方向收缩算法

Linearized alternating direction method for separable convex programming

南京大学数学系  何炳生
hebma@nju.edu.cn
1 Structured constrained convex optimization

We consider the following structured constrained convex optimization problem

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

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

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

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

(1.2)

where

\[ \Omega = \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m. \]
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

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

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

\[ v = \begin{pmatrix} y \\ \lambda \end{pmatrix} \quad \text{and} \quad \mathcal{V}^* = \{(y^*, \lambda^*) \mid (x^*, y^*, \lambda^*) \in \Omega^*\}. \]
Alternating Direction Method is a simple but powerful algorithm that is well suited to distributed convex optimization [1]. This approach also has the benefit that one algorithm could be flexible enough to solve many problems.

**Applied ADM to the structured COP:** \((y^k, \lambda^k) \Rightarrow (y^{k+1}, \lambda^{k+1})\)

First, for given \((y^k, \lambda^k)\), \(x^{k+1}\) is the solution of the following problem

\[
x^{k+1} = \text{Argmin} \left\{ \theta_1(x) + \frac{\beta}{2} \| Ax + By^k - b - \frac{1}{\beta} \lambda^k \|^2 \middle| x \in \mathcal{X} \right\} \tag{1.4a}
\]

Use \(\lambda^k\) and the obtained \(x^{k+1}\), \(y^{k+1}\) is the solution of the following problem

\[
y^{k+1} = \text{Argmin} \left\{ \theta_2(y) + \frac{\beta}{2} \| Ax^{k+1} + By - b - \frac{1}{\beta} \lambda^k \|^2 \middle| y \in \mathcal{Y} \right\} \tag{1.4b}
\]

\[
\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{1.4c}
\]
In some structured convex optimization (1.1), $B$ is a scalar matrix. However, the solution of the subproblem (1.4a) does not have the closed form solution because of the general structure of the matrix $A$. In this case, we linearize the quadratic term of (1.4a)

$$\frac{\beta}{2} \| Ax + By^k - b - \frac{1}{\beta} \lambda^k \|^2$$

at $x^k$ and add a proximal term $\frac{r}{2} \| x - a \|^2$ to the objective function. In other words, instead of (1.4a), we solve the following $x$ subproblem:

$$\min \left\{ \theta_1(x) + \beta x^T A^T (Ax^k + By^k - b - \frac{1}{\beta} \lambda^k) + \frac{r}{2} \| x - x^k \|^2 \mid x \in \mathcal{X} \right\}.$$ 

Based on linearizing the quadratic term of (1.4a), in this lecture, we construct the linearized alternating direction method. We still assume that the solution of the problem

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

(1.5)

has a closed form.
2 Linearized Alternating Direction Method

In the Linearized ADM, \( x \) is not an intermediate variable. The \( k \)-th iteration of the Linearized ADM is from \( (x^k, y^k, \lambda^k) \) to \( (x^{k+1}, y^{k+1}, \lambda^{k+1}) \).

2.1 Linearized ADM

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

\[
x^{k+1} = \underset{x \in \mathcal{X}}{\text{Argmin}} \left( \begin{array}{c}
\left\{ \theta_1(x) + \beta x^T A^T (Ax^k + By^k - b - \frac{1}{\beta} \lambda^k) \right. \\
+ \frac{\nu}{2} \|x - x^k\|^2
\end{array} \right). \tag{2.1a}
\]

2. Then, use \( \lambda^{k+1} \) and the obtained \( x^{k+1}, y^{k+1} \) is the solution of the following problem

\[
y^{k+1} = \underset{y \in \mathcal{Y}}{\text{Argmin}} \left\{ \theta_2(y) + \frac{\beta}{2} \|Ax^{k+1} + By - b - \frac{1}{\beta} \lambda^{k}\|^2 \right\}. \tag{2.1b}
\]

3. Finally,

\[
\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{2.1c}
\]
Requirements on parameters $\beta$, $r$ For given $\beta > 0$, choose $r$ such that

$$rI_n - \beta A^T A \succeq 0.$$  \hfill (2.2)

Analysis of the optimal conditions of subproblems in (2.1)

Note that $x^{k+1}$, the solution of (2.1a), satisfies

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \left\{ -A^T \lambda^k \right. \left. + \beta A^T (Ax^k + By^k - b) + r(x^{k+1} - x^k) \right\} \geq 0, \ \forall \ x \in \mathcal{X}. \ (2.3a)$$

Similarly, the solution of (2.1b) $y^{k+1}$ satisfies

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \left\{ -B^T \lambda^k \right. \left. + \beta B^T (Ax^{k+1} + By^{k+1} - b) \right\} \geq 0, \ \forall \ y \in \mathcal{Y}. \ (2.3b)$$
Substituting $\lambda^{k+1}$ (see (2.1c)) in (2.3) (eliminating $\lambda^k$), we get $x^{k+1} \in X$,

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \left\{ -A^T \lambda^{k+1} + \beta A^T B(y^k - y^{k+1}) \right. + \left. (rI_n - \beta A^T A)(x^{k+1} - x^k) \right\} \geq 0, \; \forall x \in X,$$  \hspace{1cm} (2.4a)

and

$$y^{k+1} \in Y, \; \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \left\{ -B^T \lambda^{k+1} \right\} \geq 0, \; \forall y \in Y. \hspace{1cm} (2.4b)$$

For analysis convenience, we rewrite (2.4) as the following equivalent form:

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T \lambda^{k+1} \\ -B^T \lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) \right. + \left. \begin{pmatrix} rI_n - \beta A^T A \\ 0 \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\} \geq 0, \; \forall (x, y) \in X \times Y.$$
Combining the last inequality with (2.1c), we have
\[
\theta(u) - \theta(u^{k+1}) + \left( x - x^{k+1} \right)^T \begin{pmatrix} -A^T \lambda^{k+1} \\ -B^T \lambda^{k+1} \\ A x^{k+1} + B y^{k+1} - b \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B (y^k - y^{k+1}) \\
+ \begin{pmatrix} r I_{n_1} - \beta A^T A & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} \geq 0,
\]
for all \((x, y, \lambda) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m\). The above inequality can be rewritten as
\[
\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) + \beta \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B (y^k - y^{k+1}) \\
\geq \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} r I_{n_1} - \beta A^T A & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} x^k - x^{k+1} \\ y^k - y^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix}, \forall w \in \Omega_{(2.5)}
2.2 Convergence of Linearized ADM

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

Lemma 2.1 Let \( w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega \) be generated by (2.1) from the given \( w^k = (x^k, y^k, \lambda^k) \). Then, we have

\[
(w^{k+1} - w^*)^T G(w^k - w^{k+1}) \geq (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}), \quad \forall w^* \in \Omega^*, \quad (2.6)
\]

where

\[
\eta(y^k, y^{k+1}) = \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}), \quad (2.7)
\]

and

\[
G = \begin{pmatrix} rI_{n_1} - \beta A^T A & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \quad (2.8)
\]
Proof. Setting $w = w^*$ in (2.5), and using $G$ and $\eta(y^k, y^{k+1})$, we get

$$(w^{k+1} - w^*)^T G (w^k - w^{k+1}) \geq (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) + \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}).$$

Since $F$ is monotone, it follows that

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

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

The lemma is proved. \square

By using $\eta(y^k, y^{k+1})$ (see (2.7)), $Ax^* + By^* = b$ and (2.1c), we have

$$(w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) = (B(y^k - y^{k+1}))^T \beta \{(Ax^{k+1} + By^{k+1}) - (Ax^* + By^*)\} = (B(y^k - y^{k+1}))^T \beta (Ax^{k+1} + By^{k+1} - b) = (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}).$$

(2.9)
Substituting it in (2.6), we obtain

\((w^{k+1} - w^*)^T G(w^k - w^{k+1}) \geq (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}), \forall w^* \in \Omega^*. \) \hspace{1cm} (2.10)

**Lemma 2.2** Let \(w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega\) be generated by (2.1) from the given \(w^k = (x^k, y^k, \lambda^k)\). Then, we have

\((\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0. \) \hspace{1cm} (2.11)

**Proof.** Since (2.4b) is true for the \(k\)-th iteration and the previous iteration, we have

\(\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0, \forall y \in \mathcal{Y}, \) \hspace{1cm} (2.12)

and

\(\theta_2(y) - \theta_2(y^k) + (y - y^k)^T \{-B^T \lambda^k\} \geq 0, \forall y \in \mathcal{Y}, \) \hspace{1cm} (2.13)
Setting \( y = y^k \) in (2.12) and \( y = y^{k+1} \) in (2.13), respectively, and then adding the two resulting inequalities, we get

\[
(\lambda^k - \lambda^{k+1})^T B (y^k - y^{k+1}) \geq 0.
\]

The assertion of this lemma is proved. \( \square \)

Under the assumption (2.2), the matrix \( G \) is positive semi-definite. In addition, if \( B \) is a full column rank matrix, \( G \) is positive definite. Even if in the positive semi-definite case, we also use \( \| w - \tilde{w} \|_G \) to denote

\[
\| w - \tilde{w} \|_G = \sqrt{(w - \tilde{w})^T G (w - \tilde{w})}.
\]

If \( B \) is a full column rank matrix, \( G \) is positive definite.

**Lemma 2.3** Let \( w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega \) be generated by (2.1) from the given \( w^k = (x^k, y^k, \lambda^k) \). Then, we have

\[
(w^{k+1} - w^*)^T G (w^k - w^{k+1}) \geq 0, \quad \forall \ w^* \in \Omega^* \tag{2.14}
\]
and consequently

\[ \|w^{k+1} - w^*\|^2_G \leq \|w^k - w^*\|^2_G - \|w^k - w^{k+1}\|^2_G, \quad \forall w^* \in \Omega^*. \tag{2.15} \]

**Proof.** Assertion (2.14) follows from (2.10) and (2.11) directly. By using (2.14), we have

\[
\|w^k - w^*\|^2_G = \|(w^{k+1} - w^*) + (w^k - w^{k+1})\|^2_G = \|w^{k+1} - w^*\|^2_G + 2(w^{k+1} - w^*)^T G (w^k - w^{k+1}) + \|w^k - w^{k+1}\|^2_G \geq \|w^{k+1} - w^*\|^2_G + \|w^k - w^{k+1}\|^2_G,
\]

and thus (2.15) is proved. \(\Box\)

The inequality (2.15) is essential for the convergence of the alternating direction method. Note that \(G\) is positive semi-definite.

\[ \|w^k - w^{k+1}\|^2_G = 0 \iff G(w^k - w^{k+1}) = 0. \]
The inequality (2.15) can be written as
\[
\|x^{k+1} - x^*\|_2^2 (rI - \beta A^T A) + \beta \|B(y^{k+1} - y^*)\|_2^2 + \frac{1}{\beta} \|\lambda^{k+1} - \lambda^*\|_2^2 \\
\leq \|x^k - x^*\|_2^2 (rI - \beta A^T A) + \beta \|B(y^k - y^*)\|_2^2 + \frac{1}{\beta} \|\lambda^k - \lambda^*\|_2^2 \\
- \left(\|x^k - x^{k+1}\|_2^2 (rI - \beta A^T A) + \beta \|B(y^k - y^{k+1})\|_2^2 + \frac{1}{\beta} \|\lambda^k - \lambda^{k+1}\|_2^2\right).
\]

It leads to that
\[
\lim_{k \to \infty} x^k = x^*, \quad \lim_{k \to \infty} By^k = By^* \quad \text{and} \quad \lim_{k \to \infty} \lambda^k = \lambda^*.
\]

The linearizing ADM is also known as the split inexact Uzawa method in image processing literature [9, 10].
3 Self-Adaptive ADM-based Contraction Method

In the last section, we get $x^{k+1}$ by solving the following $x$-subproblem:

$$
\min \left\{ \theta_1(x) + \beta x^T A^T (Ax^k + By^k - b - \frac{1}{\beta} \lambda^k) + \frac{r}{2} \|x - x^k\|^2 \mid x \in X \right\}
$$

and it required that the parameter $r$ to satisfy

$$
 rI_n - \beta A^T A \succeq 0 \iff r > \beta \lambda_{\max}(A^T A).
$$

In some practical problem, a conservative estimation of $\lambda_{\max}(A^T A)$ will leads a slow convergence. In this section, based on the linearized ADM, we consider the self-adaptive contraction methods. Each iteration of the self-adaptive contraction methods consists of two steps—prediction step and correction step. From the given $w^k$, the prediction step produces a test vector $\bar{w}^k$ and the correction step offers the new iterate $w^{k+1}$. 
3.1 Prediction

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

\[
\tilde{x}^k = \text{Argmin} \left( \left\{ \theta_1(x) + \beta x^T A^T (Ax^k + By^k - b - \frac{1}{\beta} \lambda^k) + \frac{r}{2} \|x - x^k\|^2 \bigg| x \in X \right\} \right) \tag{3.1a}
\]

2. Then, use \(\lambda^k\) and the obtained \(\tilde{x}^k\), \(\tilde{y}^k\) is the solution of the problem

\[
\tilde{y}^k = \text{Argmin} \left\{ \theta_2(y) - (\lambda^k)^T (Ax^k + By - b) + \frac{\beta}{2} \|Ax^k + By - b\|^2 \bigg| y \in Y \right\} \tag{3.1b}
\]

3. Finally,

\[
\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b). \tag{3.1c}
\]

The subproblems in (3.1) are similar as in (2.1). Instead of \((x^{k+1}, y^{k+1}, \lambda^{k+1})\) in (2.1), we denote the output of (3.1) by \((\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)\).
**Requirements on parameters** $\beta, r$

For given $\beta > 0$, choose $r$ such that

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

(3.2)

If $rI - \beta A^T A \succ 0$, then (3.2) is satisfied. Thus, (2.2) is sufficient for (3.2).

**Analysis of the optimal conditions of subproblems in (3.1)**

Because we get $\tilde{w}^k$ in (3.1) via substituting $w^{k+1}$ in (2.1) by $\tilde{w}^k$. Therefore, similar as (2.5), we get

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

$$
+ \begin{pmatrix} rI_{n_1} - \beta A^T A & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} \tilde{x}^k - x^k \\ \tilde{y}^k - y^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \geq 0, \quad \forall w \in \Omega.
$$
The last variational inequality can be rewritten as

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T \{ F(\tilde{w}^k) + \eta(y^k, \tilde{y}^k) + H M (\tilde{w}^k - w^k) \} \geq 0, \quad \forall w \in \Omega, \quad (3.3)$$

where

$$\eta(y^k, \tilde{y}^k) = \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - \tilde{y}^k), \quad (3.4)$$

$$H = \begin{pmatrix} r I_{n_1} & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}, \quad (3.5)$$
and

\[
M = \begin{pmatrix}
I_{n_1} - \frac{\beta}{r} A^T A & 0 & 0 \\
0 & I_{n_2} & 0 \\
0 & 0 & I_m
\end{pmatrix}.
\]

(3.6)

Based on the above analysis, we have the following lemma.

**Lemma 3.1** Let \(\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega\) be generated by (3.1) from the given \(w^k = (x^k, y^k, \lambda^k)\). Then, we have

\[
(\tilde{w}^k - w^*)^T HM(\tilde{w}^k - w^k) \geq (\tilde{w}^k - w^*)^T \eta(y^k, \tilde{y}^k), \quad \forall w^* \in \Omega^*.
\]

(3.7)

**Proof.** Setting \(w = w^*\) in (3.3), we obtain

\[
(\tilde{w}^k - w^*)^T HM(\tilde{w}^k - w^k) \\
\geq (\tilde{w}^k - w^*)^T \eta(y^k, \tilde{y}^k) \\
+ \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k).
\]

(3.8)
Since $F$ is monotone and $\tilde{w}^k \in \Omega$, it follows that
\[
\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \\
\geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \geq 0.
\]
The last inequality is due to $\tilde{w}^k \in \Omega$ and $w^* \in \Omega^*$ is a solution of (1.3).
Substituting it in the right hand side of (3.8), the lemma is proved. \qed

In addition, because $Ax^* + By^* = b$ and $\beta(A\tilde{x}^k + B\tilde{y}^k - b) = \lambda^k - \tilde{\lambda}^k$,
we have
\[
(\tilde{w}^k - w^*)^T \eta(y^k, \tilde{y}^k) \\
= (B(y^k - \tilde{y}^k))^T \beta\{(A\tilde{x}^k + B\tilde{y}^k) - (Ax^* + By^*)\} \\
= (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k).
\]

**Lemma 3.2** Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ be generated by (3.1) from the given $w^k = (x^k, y^k, \lambda^k)$. Then, we have
\[
(w^k - w^*)^T H M (w^k - \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k), \quad \forall w^* \in \Omega^*,
\]
where
\[
\varphi(w^k, \tilde{w}^k) = (w^k - \tilde{w}^k)^T H M (w^k - \tilde{w}^k) + (\lambda^k - \tilde{\lambda}^k)^T B (y^k - \tilde{y}^k). \tag{3.11}
\]

**Proof.** From (3.7) and (3.9) we have
\[
(\tilde{w}^k - w^*)^T H M (w^k - \tilde{w}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T B (y^k - \tilde{y}^k).
\]
Assertion (3.10) follows from the last inequality and the definition of \( \varphi(w^k, \tilde{w}^k) \) directly. \(\square\)

3.2 The Primary Contraction Methods

The primary contraction methods use \( M(w^k - \tilde{w}^k) \) as search direction and the unit step length. In other words, the new iterate is given by
\[
w^{k+1} = w^k - M(w^k - \tilde{w}^k). \tag{3.12}
\]
According to (3.6), it can be written as

\[
\begin{pmatrix}
  x^{k+1} \\
  y^{k+1} \\
  \lambda^{k+1}
\end{pmatrix}
= 
\begin{pmatrix}
  \tilde{x}^k + \frac{\beta}{r} A^T A (x^k - \tilde{x}^k) \\
  \tilde{y}^k \\
  \tilde{\lambda}^k
\end{pmatrix}.
\tag{3.13}
\]

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 (3.2). In practical computation, we try to use the average of the eigenvalues of \(\beta A^T A\).

Using (3.10), we have

\[
\|w^k - w^*\|_H^2 - \|w^{k+1} - w^*\|_H^2 =
\|w^k - w^*\|_H^2 - \|(w^k - w^*) - M(w^k - \tilde{w}^k)\|_H^2
= 2(w^k - w^*)^T H M (w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2
\geq 2 \varphi(w^k, \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2.
\tag{3.14}
\]
Because \((\tilde{y}^k, \tilde{\lambda}^k) = (y^{k+1}, \lambda^{k+1})\), the inequality (2.11) is still holds and thus

\[(\lambda^k - \tilde{\lambda}^k)^T B (y^k - \tilde{y}^k) \geq 0.\]

Therefore, it follows from (3.14), (3.11) and the last inequality that

\[
\|w^k - w^*\|_H^2 - \|w^{k+1} - w^*\|_H^2 \\
\geq 2(w^k - \tilde{w}^k)^T HM (w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2. \tag{3.15}
\]

**Lemma 3.3** Under the condition (3.2), we have

\[
2(w^k - \tilde{w}^k)^T HM (w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2 \\
\geq (1 - \nu^2) r \|x^k - \tilde{x}^k\|^2 + \beta \|B(y^k - \tilde{y}^k)\|^2 + \frac{1}{\beta} \|\lambda^k - \tilde{\lambda}^k\|^2. \tag{3.16}
\]

**Proof.** First, we have

\[
2(w^k - \tilde{w}^k)^T HM (w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2 \\
= (w^k - \tilde{w}^k)^T (M^T H + HM - M^T HM)(w^k - \tilde{w}^k).
\]
By using the structure of the matrices $H$ and $M$ (see (3.5) and (3.6)), we obtain

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

$$= \begin{pmatrix} rI_n & 0 & 0 \\ 0 & \beta B^TB & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} - \begin{pmatrix} r\left(\frac{\beta}{r}AT\right)^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$  

Therefore,

$$2(w^k - \tilde{w}^k)^T H M (w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|^2_H$$

$$= \|w^k - \tilde{w}^k\|^2_H - r\left(\frac{\beta^2}{r^2}\right)\|AT(Ax^k - \tilde{x}^k)\|^2.$$  

(3.17)

Under the condition (3.2), we have

$$\left(\frac{\beta^2}{r^2}\right)\|AT(Ax^k - \tilde{x}^k)\|^2 \leq \nu^2\|x^k - \tilde{x}^k\|^2.$$  

Substituting it in (3.17), the assertion of this lemma is proved.  

\[\square\]

**Theorem 3.1** Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ be generated by (3.1) from the given
\( w^k = (x^k, y^k, \lambda^k) \) and the new iterate \( w^{k+1} \) is given by (3.12). The sequence 
\( \{ w^k = (x^k, y^k, \lambda^k) \} \) generated by the elementary contraction method satisfies

\[
\| w^{k+1} - w^* \|_H^2 \leq \| w^k - w^* \|_H^2 - (1 - \nu^2) \| w^k - \tilde{w}^k \|_H^2. \tag{3.18}
\]

**Proof.** From (3.15) and (3.16) we obtain

\[
\| w^k - w^* \|_H^2 - \| w^{k+1} - w^* \|_H^2 \\
\geq (1 - \nu^2)r \| x^k - \tilde{x}^k \|^2 + \beta \| B(y^k - \tilde{y}^k) \|^2 + \frac{1}{\beta} \| \lambda^k - \tilde{\lambda}^k \|^2 \\
\geq (1 - \nu^2) \| w^k - \tilde{w}^k \|_H^2.
\]

The assertion of this theorem is proved. \( \square \)

Theorem 3.1 is essential for the convergence of the primary contraction method.
## 3.3 The general contraction method

For given \( w^k \), we use

\[
w(\alpha) = w^k - \alpha M(w^k - \tilde{w}^k)
\]

(3.19)

to update the \( \alpha \)-dependent new iterate. For any \( w^* \in \Omega^* \), we define

\[
\vartheta(\alpha) := \| w^k - w^* \|_H^2 - \| w(\alpha) - w^* \|_H^2
\]

(3.20)

and

\[
q(\alpha) = 2\alpha \varphi(w^k, \tilde{w}^k) - \alpha^2 \| M(w^k - \tilde{w}^k) \|_H^2,
\]

(3.21)

where \( \varphi(w^k, \tilde{w}^k) \) is defined in (3.11).

**Theorem 3.2** Let \( w(\alpha) \) be defined by (3.19). For any \( w^* \in \Omega^* \) and \( \alpha \geq 0 \), we have

\[
\vartheta(\alpha) \geq q(\alpha),
\]

(3.22)

where \( \vartheta(\alpha) \) and \( q(\alpha) \) are defined in (3.20) and (3.21), respectively.
Proof. It follows from (3.19) and (3.20) that

$$\vartheta(\alpha) = \|w^k - w^*\|_H^2 - \|(w^k - w^*) - \alpha M(w^k - \tilde{w}^k)\|_H^2$$

$$= 2\alpha (w^k - w^*)^T H M (w^k - \tilde{w}^k) - \alpha^2 \|M(w^k - \tilde{w}^k)\|_H^2.$$ 

By using (3.10) and the definition of $q(\alpha)$, the theorem is proved. \(\square\)

Note that $q(\alpha)$ in (3.21) is a quadratic function of $\alpha$ and it reaches its maximum at

$$\alpha^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|M(w^k - \tilde{w}^k)\|_H^2}. \quad (3.23)$$

In practical computation, we use

$$w^{k+1} = w^k - \gamma \alpha^*_k M(w^k - \tilde{w}^k), \quad (3.24)$$

to update the new iterate, where $\gamma \in [1, 2)$ is a relaxation factor. By using (3.20) and (3.22), we have

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - q(\gamma \alpha^*_k). \quad (3.25)$$
Note that

\[ q(\gamma \alpha_k^*) = 2\gamma \alpha_k^* \varphi(w^k, \tilde{w}^k) - (\gamma \alpha_k^*)^2 \| M(w^k - \tilde{w}^k) \|_H^2. \]  

(3.26)

Using (3.23) and (3.24), we obtain

\[ q(\gamma \alpha_k^*) = \gamma(2 - \gamma)(\alpha_k^*)^2 \| M(w^k - \tilde{w}^k) \|_H^2 = \frac{2 - \gamma}{\gamma} \| w^k - w^{k+1} \|_H^2, \]

and consequently it follows from (3.25) that

\[ \| w^{k+1} - w^* \|_H^2 \leq \| w^k - w^* \|_H^2 - \frac{2 - \gamma}{\gamma} \| w^k - w^{k+1} \|_H^2, \quad \forall w^* \in \Omega^*. \]  

(3.27)

On the other hand, it follows from (3.23) and (3.26) that

\[ q(\gamma \alpha_k^*) = \gamma(2 - \gamma)\alpha_k^* \varphi(w^k, \tilde{w}^k). \]  

(3.28)
By using (3.11) and (3.16), we obtain

\[ 2\varphi(w^k, \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|^2_H \]
\[ = 2(w^k - \tilde{w}^k)^T H M(w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|^2_H \]
\[ + 2(\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k) \]
\[ \geq (1 - \nu^2)r\|x^k - \tilde{x}^k\|^2 + \beta\|B(y^k - \tilde{y}^k)\|^2 + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 \]
\[ + 2(\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k) \]
\[ = (1 - \nu^2)r\|x^k - \tilde{x}^k\|^2 + \beta\|B(y^k - \tilde{y}^k) + \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k)\|^2. \]

Thus, we have \( 2\varphi(w^k, \tilde{w}^k) > \|M(w^k - \tilde{w}^k)\|^2_H \) and consequently

\[ \alpha^*_k > \frac{1}{2}. \]
In addition, because
\[
\varphi(w^k, \tilde{w}^k) = (w^k - \tilde{w}^k)^T H M (w^k - \tilde{w}^k) + (\lambda^k - \tilde{\lambda}^k)^T B (y^k - \tilde{y}^k)
\]
\[
= \begin{pmatrix}
  x^k - \tilde{x}^k \\
y^k - \tilde{y}^k \\
\lambda^k - \tilde{\lambda}^k
\end{pmatrix}^T \begin{pmatrix}
  r I_{n_1} - \beta A^T A & 0 & 0 \\
  0 & \beta B^T B & \frac{1}{2} B^T \\
  0 & \frac{1}{2} B & \frac{1}{\beta} I_m
\end{pmatrix} \begin{pmatrix}
  x^k - \tilde{x}^k \\
y^k - \tilde{y}^k \\
\lambda^k - \tilde{\lambda}^k
\end{pmatrix}
\]
\[
\geq \|x^k - \tilde{x}^k\| \cdot (r\|x^k - \tilde{x}^k\| - \beta\|A^T A (x^k - \tilde{x}^k)\|)
\]
\[
+ \frac{1}{2} \begin{pmatrix}
y^k - \tilde{y}^k \\
\lambda^k - \tilde{\lambda}^k
\end{pmatrix}^T \begin{pmatrix}
\beta B^T B & 0 \\
0 & \frac{1}{\beta} I_m
\end{pmatrix} \begin{pmatrix}
y^k - \tilde{y}^k \\
\lambda^k - \tilde{\lambda}^k
\end{pmatrix}.
\]

Under the condition (3.2), it follows from the last inequality that
\[
\varphi(w^k, \tilde{w}^k) \geq \min\{1 - \nu, \frac{1}{2}\} \|w^k - \tilde{w}^k\|_H^2
\]
\[
\geq \frac{1 - \nu}{2} \|w^k - \tilde{w}^k\|_H^2.
\] (3.29)
By using (3.25), (3.28), (3.29) and \( \alpha^*_k \geq \frac{1}{2} \), we obtain the following theorem for the general contraction method.

**Theorem 3.3** The sequence \( \{w^k = (x^k, y^k, \lambda^k)\} \) generated by the general contraction method satisfies

\[
\|w^{k+1} - w^*\|_H^2 \\
\leq \|w^k - w^*\|_H^2 - \frac{\gamma(2-\gamma)(1-\nu)}{4}\|w^k - \tilde{w}^k\|_H^2, \quad \forall w^* \in \Omega^*. \tag{3.30}
\]

The inequality (3.30) in Theorem 3.3 is essential for the convergence of the general contraction method.

Both the inequalities (3.18) and (3.30) can be written as

\[
\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - c_0\|w^k - \tilde{w}^k\|_H^2, \quad \forall w^* \in \Omega^*,
\]
where $c_0 > 0$ is a constant. Therefore, we have

$$r\|x^{k+1} - x^*\|^2 + \beta\|B(y^{k+1} - y^*)\|^2 + \frac{1}{\beta} \|\lambda^{k+1} - \lambda^*\|^2$$

$$\leq r\|x^k - x^*\|^2 + \beta\|B(y^k - y^*)\|^2 + \frac{1}{\beta} \|\lambda^k - \lambda^*\|^2$$

$$-c_0 \left( r\|x^k - \tilde{x}^k\|^2 + \beta\|B(y^k - \tilde{y}^k)\|^2 + \frac{1}{\beta} \|\lambda^k - \tilde{\lambda}^k\|^2 \right).$$

It leads to that

$$\lim_{k \to \infty} \left( r\|x^k - \tilde{x}^k\|^2 + \beta\|B(y^k - \tilde{y}^k)\|^2 + \frac{1}{\beta} \|\lambda^k - \tilde{\lambda}^k\|^2 \right),$$

and

$$\lim_{k \to \infty} x^k = x^*, \quad \lim_{k \to \infty} By^k = By^* \quad \text{and} \quad \lim_{k \to \infty} \lambda^k = \lambda^*.$$
4 Applications in $l_1$-norm problems

An important $l_1$-norm problem in the area machine learning is the $l_1$ regularized linear regression, also called the lasso [8]. This involves solving

$$\min \tau \|x\|_1 + \frac{1}{2} \|Ax - b\|_2^2,$$

(4.1)

where $\tau > 0$ is a scalar regularization parameter that is usually chosen by cross-validation. In typical applications, there are many more features than training examples, and the goal is to find a parsimonious model for the data. The problem (1.1) can be reformulated to problem

$$\min \tau \|x\|_1 + \frac{1}{2} \|y\|_2^2$$

$$Ax - y = b$$

(4.2)

which is a form of (1.1). Applied the alternating direction method (1.4) to the problem (4.2), the $x$-subproblem is

$$x^{k+1} = \text{Argmin} \{ \tau \|x\|_1 + \frac{\beta}{2} \|(Ax - y^k) - \frac{1}{\beta} \lambda^k\|_2^2 \},$$
and the solution does not have a closed form. Applied the linearized alternating
direction method (2.1) to the problem (4.2), the $x$-subproblem (2.1a) is
\[
\tilde{x}^k = \text{Argmin}\{\tau \|x\|_1 + \frac{r}{2}\|x - [x^k + \frac{1}{r}\lambda^k - \frac{\beta}{r}A^T(Ax^k - y^k)]\|^2\}. \quad (4.3)
\]
This problem is of the form of (1.5) and its solution has the following closed form:
\[
\tilde{x}^k = a - P_{B^{\tau/r}_\infty}[a], \quad \text{where} \quad a = x^k + \frac{1}{r}\lambda^k - \frac{\beta}{r}A^T(Ax^k - y^k)
\]
and
\[
B^{\tau/r}_\infty = \{\xi \in \mathbb{R}^n \mid -(\tau/r)e \leq \xi \leq (\tau/r)e\}.
\]
By using the linearized alternating direction method in Section 2, for given $\beta > 0$, it needs $r > \beta\lambda_{\text{max}}(A^TA)$. By using the self-adaptive ADM-based contraction method in Section 3, it needs $r$ to satisfy
\[
\beta\|A^TA(x^k - \tilde{x}^k)\| \leq \nu r\|x^k - \tilde{x}^k\|, \quad \nu \in (0, 1).
\]
Because $A$ is a generic matrix, the above condition is satisfied even if $r$ is much less than $\beta\lambda_{\text{max}}(A^TA)$. 

References


