凸优化和单调变分不等式的收缩算法

第十八讲: 多个分离算子凸优化
带回代的线性化交替方向法

Linearized Alternating direction method with back substitution for convex optimization containing more separable operators

南京大学数学系　　何炳生

hebma@nju.edu.cn

The context of this lecture is based on the paper [3]

1 Introduction

In this paper, we consider the general case of linearly constrained separable convex programming with $m \geq 3$:

$$\begin{array}{ll} \min & \sum_{i=1}^{m} \theta_i(x_i) \\ & \sum_{i=1}^{m} A_i x_i = b; \\ & x_i \in \mathcal{X}_i, \quad i = 1, \cdots, m; \end{array} \tag{1.1}$$

where $\theta_i : \Re^{n_i} \rightarrow \Re \ (i = 1, \ldots, m)$ are closed proper convex functions (not necessarily smooth); $\mathcal{X}_i \subset \Re^{n_i} \ (i = 1, \ldots, m)$ are closed convex sets; $A_i \in \Re^{l \times n_i} \ (i = 1, \ldots, m)$ are given matrices and $b \in \Re^l$ is a given vector. Throughout, we assume that the solution set of (1.1) is nonempty.

In fact, even for the special case of (1.1) with $m = 3$, the convergence of the extended ADM is still open. In the last lecture, we provided a novel approach towards the extension of ADM for the problem (1.1). More specifically, we show that if a new iterate is generated by correcting the output of the ADM with a Gaussian back substitution procedure, then the sequence of iterates is convergent to a solution of (1.1). The resulting method is called the ADM with Gaussian back substitution (ADM-GbS).

Alternatively, the ADM-GbS can be regarded as a prediction-correction type method whose predictor is generated by the ADM procedure and the correction is completed by a Gaussian back substitution procedure. The main task of each iteration in ADM-GbS is to solve the following sub-problem:

$$\min\{\theta_i(x_i) + \tfrac{\beta}{2}\|A_i x_i - b_i\|^2 \mid x_i \in \mathcal{X}_i\}, \quad i = 1, \ldots, m. \tag{1.2}$$

Thus, ADM-GbS is implementable only when the subproblems of (1.2) have their solutions in the closed form. Again, each iteration of the proposed method in this lecture consists of two steps–prediction and correction. In order to implement the prediction step, we only assume that the x_i-subproblem

$$\min\{\theta_i(x_i) + \tfrac{r_i}{2}\|x_i - a_i\|^2 \mid x_i \in \mathcal{X}_i\}, \quad i = 1, \ldots, m \tag{1.3}$$

has its solution in the closed form.

The first-order optimality condition of (1.1) and thus characterize (1.1) by a variational inequality (VI). As we will show, the VI characterization is convenient for the convergence analysis to be conducted.

By attaching a Lagrange multiplier vector $\lambda \in \Re^l$ to the linear constraint, the Lagrange function of (1.1) is:

$$L(x_1, x_2, \ldots, x_m, \lambda) = \sum_{i=1}^{m} \theta_i(x_i) - \lambda^T (\sum_{i=1}^{m} A_i x_i - b), \tag{1.4}$$

which is defined on

$$\mathcal{W} := \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m \times \Re^l.$$

Let $\left(x_1^*, x_2^*, \ldots, x_m^*, \lambda^*\right)$ be a saddle point of the Lagrange function (1.4). Then we have

$$\begin{aligned} L_{\lambda\in\Re^l}(x_1^*, x_2^*, \cdots, x_m^*, \lambda) &\leq L(x_1^*, x_2^*, \cdots, x_m^*, \lambda^*) \\ &\leq L_{x_i\in\mathcal{X}_i\ (i=1,\ldots,m)}(x_1, x_2, \ldots, x_m, \lambda^*). \end{aligned}$$

For $i \in \{1, 2, \cdots, m\}$, we denote by $\partial\theta_i(x_i)$ the subdifferential of the convex function $\theta_i(x_i)$ and by $f_i(x_i) \in \partial\theta_i(x_i)$ a given subgradient of $\theta_i(x_i)$.

It is evident that finding a saddle point of $L(x_1, x_2, \ldots, x_m, \lambda)$ is equivalent to finding

$w^* = (x_1^*, x_2^*, ..., x_m^*, \lambda^*) \in \mathcal{W}$, such that

$$\begin{cases} (x_1 - x_1^*)^T \{f_1(x_1^*) - A_1^T \lambda^*\} \geq 0, \\ \quad \vdots \\ (x_m - x_m^*)^T \{f_m(x_m^*) - A_m^T \lambda^*\} \geq 0, \\ (\lambda - \lambda^*)^T (\sum_{i=1}^m A_i x_i^* - b) \geq 0, \end{cases} \tag{1.5}$$

for all $w = (x_1, x_2, \cdots, x_m, \lambda) \in \mathcal{W}$. More compactly, (1.5) can be written into

$$(w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \mathcal{W}, \tag{1.6a}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_m \\ \lambda \end{pmatrix} \qquad \text{and} \qquad F(w) = \begin{pmatrix} f_1(x_1) - A_1^T \lambda \\ \vdots \\ f_m(x_m) - A_m^T \lambda \\ \sum_{i=1}^m A_i x_i - b \end{pmatrix}. \tag{1.6b}$$

Note that the operator $F(w)$ defined in (1.6b) is monotone due to the fact that θ_i's are all convex functions. In addition, the solution set of (1.6), denoted by $\mathcal{W}^*$, is also nonempty.

2 Linearized ADM with Gaussian back substitution

2.1 Linearized ADM Prediction

Step 1. ADM step (prediction step). Obtain $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_m^k, \tilde{\lambda}^k)$ in the forward (alternating) order by the following ADM procedure:

$$\left\{\begin{array}{l}
\tilde{x}_1^k = \arg\min\left\{\theta_1(x_1) + q_1^T A_1 x_1 + \frac{r_1}{2}\|x_1 - x_1^k\|^2 \;\middle|\; x_1 \in \mathcal{X}_1\right\};\\
\qquad \vdots\\
\tilde{x}_i^k = \arg\min\left\{\theta_i(x_i) + q_i^T A_i x_i + \frac{r_i}{2}\|x_i - x_i^k\|^2 \;\middle|\; x_i \in \mathcal{X}_i\right\};\\
\qquad \vdots\\
\tilde{x}_m^k = \arg\min\left\{\theta_m(x_m) + q_m^T A_m x_m + \frac{r_m}{2}\|x_m - x_m^k\|^2 \;\middle|\; x_m \in \mathcal{X}_m\right\};\\
\qquad \text{where} \quad q_i = \beta\left(\sum_{j=1}^{i-1} A_j \tilde{x}_j^k + \sum_{j=i}^{m} A_j x_j^k - b\right).\\
\tilde{\lambda}^k = \lambda^k - \beta\left(\sum_{j=1}^{m} A_j \tilde{x}_j^k - b\right).
\end{array}\right. \tag{2.1}$$

The prediction is implementable due to the assumption (1.3) of this lecture and

$$
\begin{aligned}
&\arg\min\Big\{\theta_i(x_i)+ q_i^T A_i x_i + \tfrac{r_i}{2}\|x_i - x_i^k\|^2 \,\big|x_i\in\mathcal{X}_i\Big\} \\
&= \;\; \arg\min\Big\{\theta_i(x_i) + \tfrac{r_i}{2}\|x_i - \big(x_i^k - \tfrac{1}{r_i}A_i^T q_i\big)\|^2 \,\big|x_i\in\mathcal{X}_i\Big\}.
\end{aligned}
$$

Assumption $r_i,\; i=1,\ldots,m$ is chosen that condition

$$
r_i\|x_i^k - \tilde{x}_i^k\|^2 \ge \beta\|A_i(x_i^k - \tilde{x}_i^k)\|^2 \tag{2.2}
$$

is satisfied in each iteration.

In the case that $A_i = I_{n_i}$, we take $r_i=\beta$, the condition (2.2) is satisfied. Note that in this case we have

$$
\begin{aligned}
&\arg\min_{x_i\in\mathcal{X}_i}\Big\{\theta_i(x_i)+\Big\{\beta(\sum_{j=1}^{i-1}A_j\tilde{x}_j^k + \sum_{j=i}^{m}A_j x_j^k - b)\Big\}^T A_i x_i + \frac{\beta}{2}\|x_i - x_i^k\|^2\Big\} \\
&= \;\; \arg\min_{x_i\in\mathcal{X}_i}\Big\{\theta_i(x_i)+\frac{\beta}{2}\Big\|\Big(\sum_{j=1}^{i-1}A_j\tilde{x}_j^k + A_i x_i + \sum_{j=i+1}^{m}A_j x_j^k - b\Big) - \frac{1}{\beta}\lambda^k\Big\|^2\Big\}.
\end{aligned}
$$

2.2 Correction by the Gaussian back substitution

To present the Gaussian back substitution procedure, we define the matrices:

$$M = \begin{pmatrix} r_1 I_{n_1} & 0 & \cdots & \cdots & 0 \\ \beta A_2^T A_1 & r_2 I_{n_2} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \beta A_m^T A_1 & \cdots & \beta A_m^T A_{m-1} & r_m I_{n_m} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_l \end{pmatrix}, \qquad (2.3)$$

and

$$H = \text{diag}\big(r_1 I_{n_1}, r_2 I_{n_2}, \ldots, r_m I_{n_m}, \frac{1}{\beta} I_l\big). \qquad (2.4)$$

Note that for $\beta > 0$ and $r_i > 0$, the matrix M defined in (2.3) is a non-singular lower-triangular block matrix. In addition, according to (2.3) and (2.4), we have:

$$H^{-1}M^T = \begin{pmatrix} I_{n_2} & \frac{\beta}{r_1}A_1^T A_2 & \cdots & \frac{\beta}{r_1}A_1^T A_m & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & I_{n_{m-1}} & \frac{\beta}{r_{n_{m-1}}}A_{m-1}^T A_m & 0 \\ 0 & \cdots & 0 & I_{n_m} & 0 \\ 0 & \cdots & 0 & 0 & I_l \end{pmatrix}. \qquad (2.5)$$

which is a upper-triangular block matrix whose diagonal components are identity matrices. The Gaussian back substitution procedure to be proposed is based on the matrix $H^{-1}M^T$ defined in (2.5).

Step 2. Gaussian back substitution step (correction step). Correct the ADM output $\tilde{w}^k$ in the backward order by the following Gaussian back substitution procedure and generate the new iterate w^{k+1}:

$$H^{-1}M^T(w^{k+1} - w^k) = \alpha\big(\tilde{w}^k - w^k\big). \tag{2.6}$$

Recall that the matrix $H^{-1}M^T$ defined in (2.5) is a upper-triangular block matrix. The Gaussian back substitution step (2.6) is thus very easy to execute. In fact, as we mentioned, after the predictor is generated by the linearized ADM scheme (2.1) in the forward (alternating) order, the proposed Gaussian back substitution step corrects the predictor in the backward order. Since the Gaussian back substitution step is easy to perform, the computation of each iteration of the ADM with Gaussian back substitution is dominated by the ADM procedure (2.1).

To show the main idea with clearer notation, we restrict our theoretical discussion to the case with fixed $\beta > 0$. The main task of the Gaussian back substitution step (2.6) can be rewritten into

$$w^{k+1} = w^k - \alpha M^{-T}H(w^k - \tilde{w}^k). \tag{2.7}$$

As we will show, $-M^{-T}H(w^k - \tilde{w}^k)$ is a descent direction of the distance function

$\frac{1}{2}\|w - w^*\|_G^2$ with $G = MH^{-1}M^T$ at the point $w = w^k$ for any $w^* \in \mathcal{W}^*$. In this sense, the proposed linearized ADM with Gaussian back substitution can also be regarded as an ADM-based contraction method where the output of the linearized ADM scheme (2.1) contributes a descent direction of the distance function. Thus, the constant α in (2.6) plays the role of a step size along the descent direction $-(w^k - \tilde{w}^k)$. In fact, we can choose the step size dynamically based on some techniques in the literature (e.g. [4]), and the Gaussian back substitution procedure with the constant α can be modified accordingly into the following variant with a dynamical step size:

$$H^{-1}M^T(w^{k+1} - w^k) = \gamma\alpha_k^*\big(\tilde{w}^k - w^k\big), \tag{2.8}$$

where

$$\alpha_k^* = \frac{\|w^k - \tilde{w}^k\|_H^2 + \|w^k - \tilde{w}^k\|_Q^2}{2\|w^k - \tilde{w}^k\|_H^2}; \tag{2.9}$$

$$Q = \begin{pmatrix} \beta A_1^T A_1 & \beta A_1^T A_2 & \cdots & \beta A_1^T A_m & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \cdots & \beta A_2^T A_m & A_2^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta A_m^T A_1 & \beta A_m^T A_2 & \cdots & \beta A_m^T A_m & A_m^T \\ A_1 & A_2 & \cdots & A_m & \frac{1}{\beta} I_l \end{pmatrix}; \qquad (2.10)$$

and $\gamma \in (0, 2)$. Indeed, for any $\beta > 0$, the symmetric matrix Q is positive semi-definite. Then, for given w^k and the $\tilde{w}^k$ obtained by the ADM procedure (2.1), we have that

$$\|w^k - \tilde{w}^k\|_H^2 = \sum_{i=1}^{m} r_i \|x_i^k - \tilde{x}_i^k\|^2 + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2,$$

and

$$\|w^k - \tilde{w}^k\|_Q^2 = \beta \left\| \sum_{i=1}^{m} A_i(x_i^k - \tilde{x}_i^k) + \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \right\|^2,$$

where the norm $\|w\|_H^2$ ($\|w\|_Q^2$, respectively) is defined as $w^T H w$ ($w^T Q w$, respectively). Note that the step size α_k^* defined in (2.9) satisfies $\alpha_k^* \geq \frac{1}{2}$.

3 Convergence of the Linearized ADM-GbS

In this section, we prove the convergence of the proposed ADM with Gaussian back substitution for solving (1.1). Our proof follows the analytic framework of contractive type methods. Accordingly, we divide this section into three subsections.

3.1 Verification of the descent directions

In this subsection, we mainly show that $-(w^k - \tilde{w}^k)$ is a descent direction of the function $\frac{1}{2}\|w - w^*\|_G^2$ at the point $w = w^k$ whenever $\tilde{w}^k \neq w^k$, where $\tilde{w}^k$ is generated by the ADM scheme (2.1), $w^* \in \mathcal{W}^*$ and G is a positive definite matrix.

Lemma 3.1 *Let* $\tilde{w}^k = (\tilde{x}_1^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k)$ *be generated by the linearized ADM step* (2.1) *from the given vector* $w^k = (x_1^k, \ldots, x_m^k, \lambda^k)$*. Then, we have*

$$\tilde{w}^k \in \mathcal{W}, \quad (w - \tilde{w}^k)^T \{d_2(w^k, \tilde{w}^k) - d_1(w^k, \tilde{w}^k)\} \geq 0, \quad \forall\, w \in \mathcal{W}, \qquad (3.1)$$

where

$$d_1(w^k,\tilde{w}^k) = \begin{pmatrix} r_1 I_{n_1} & 0 & \cdots & \cdots & 0 \\ \beta A_2^T A_1 & r_2 I_{n_2} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \beta A_m^T A_1 & \cdots & \beta A_m^T A_{m-1} & r_m I_{n_m} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_l \end{pmatrix} \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}, \tag{3.2}$$

$$d_2(w^k,\tilde{w}^k) = F(\tilde{w}^k) + \beta \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \\ 0 \end{pmatrix} \Big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\Big). \tag{3.3}$$

Proof. Since $\tilde{x}_i^k$ is the solution of (2.1), for $i = 1, 2, \ldots, m$, according to the optimality condition, we have

$$\tilde{x}_i^k \in \mathcal{X}_i, \;\; (x_i - \tilde{x}_i^k)^T \big\{ f_i(\tilde{x}_i^k) - A_i^T [\lambda^k - \beta(\textstyle\sum_{j=1}^{i-1} A\tilde{x}_j^k + \sum_{j=i}^{m} A_j x_j^k - b)] \\ + r_i(\tilde{x}_i^k - x_i^k) \big\} \geq 0, \quad \forall\, x_i \in \mathcal{X}_i. \tag{3.4}$$

By using the fact

$$\tilde{\lambda}^k = \lambda^k - \beta(\sum_{j=1}^{m} A_j \tilde{x}_j^k - b),$$

the inequality (3.4) can be written as

$$\tilde{x}_i^k \in \mathcal{X}_i, \quad (x_i - \tilde{x}_i^k)^T \big\{ f_i(\tilde{x}_i^k) - A_i^T \tilde{\lambda}^k + \beta A_i^T \big(\sum_{j=i}^{m} A_j (x_j^k - \tilde{x}_j^k)\big) \\ + r_i(\tilde{x}_i^k - x_i^k) \big\} \geq 0, \;\; \forall\, x_i \in \mathcal{X}_i. \tag{3.5}$$

Summing the inequality (3.5) over $i = 1, \ldots, m$, we obtain $\tilde{x}^k \in \mathcal{X}$ and

$$\begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \left\{ \begin{pmatrix} f_1(\tilde{x}_1^k) - A_1^T \tilde{\lambda}^k \\ f_2(\tilde{x}_2^k) - A_2^T \tilde{\lambda}^k \\ \vdots \\ f_m(\tilde{x}_m^k) - A_m^T \tilde{\lambda}^k \end{pmatrix} + \beta \begin{pmatrix} A_1^T \left(\sum_{j=1}^m A_j (x_j^k - \tilde{x}_j^k)\right) \\ A_2^T \left(\sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k)\right) \\ \vdots \\ A_m^T \left(A_m (x_m^k - \tilde{x}_m^k)\right) \end{pmatrix} \right\}$$

$$\geq \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \begin{pmatrix} r_1 I_{n_1} & 0 & 0 & 0 \\ 0 & r_2 I_{n_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_m I_{n_m} \end{pmatrix} \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \end{pmatrix} \tag{3.6}$$

for all $x \in \mathcal{X}$. Adding the following term

$$\begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \beta \begin{pmatrix} 0 \\ A_2^T\left(\sum_{j=1}^{1} A_j(x_j^k - \tilde{x}_j^k)\right) \\ \vdots \\ A_m^T\left(\sum_{j=1}^{m-1} A_j(x_j^k - \tilde{x}_j^k)\right) \end{pmatrix}$$

to the both sides of (3.6), we get $\tilde{x}^k \in \mathcal{X}$ and for all $x \in \mathcal{X}$,

$$\begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \begin{pmatrix} f_1(\tilde{x}_1^k) - A_1^T\tilde{\lambda}^k + \beta A_1^T\left(\sum_{j=1}^{m} A_j(x_j^k - \tilde{x}_j^k)\right) \\ f_2(\tilde{x}_2^k) - A_2^T\tilde{\lambda}^k + \beta A_2^T\left(\sum_{j=1}^{m} A_j(x_j^k - \tilde{x}_j^k)\right) \\ \vdots \\ f_m(\tilde{x}_m^k) - A_m^T\tilde{\lambda}^k + \beta A_m^T\left(\sum_{j=1}^{m} A_j(x_j^k - \tilde{x}_j^k)\right) \end{pmatrix}$$

$$\geq \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \left\{ \begin{pmatrix} r_1(x_1^k - \tilde{x}_1^k) \\ r_2(x_2^k - \tilde{x}_2^k) \\ \vdots \\ r_m(x_m^k - \tilde{x}_m^k) \end{pmatrix} + \begin{pmatrix} 0 \\ \beta A_2^T\left(\sum_{j=1}^{1} A_j(x_j^k - \tilde{x}_j^k)\right) \\ \vdots \\ \beta A_m^T\left(\sum_{j=1}^{m-1} A_j(x_j^k - \tilde{x}_j^k)\right) \end{pmatrix} \right\} \tag{3.7}$$

Because that $\sum_{j=1}^m A_j\tilde{x}_j^k - b = \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k)$, we have

$$(\lambda - \tilde{\lambda}^k)^T(\sum_{j=1}^m A_j\tilde{x}_j^k - b) = (\lambda - \tilde{\lambda}^k)^T\frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k).$$

Adding (3.8) and the last equality together, we get $\tilde{w}^k \in \mathcal{W}$, and for all $w \in \mathcal{W}$

$$\begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} f_1(\tilde{x}_1^k) - A_1^T\tilde{\lambda}^k + \beta A_1^T\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ f_2(\tilde{x}_2^k) - A_2^T\tilde{\lambda}^k + \beta A_2^T\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ f_m(\tilde{x}_m^k) - A_m^T\tilde{\lambda}^k + \beta A_m^T\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ \sum_{j=1}^m A_j\tilde{x}_j^k - b \end{pmatrix}$$

$$\geq \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} r_1(x_1^k - \tilde{x}_1^k) \\ r_2(x_2^k - \tilde{x}_2^k) \\ \vdots \\ r_m(x_m^k - \tilde{x}_m^k) \\ \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix} + \begin{pmatrix} 0 \\ \beta A_2^T\big(\sum_{j=1}^1 A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ \beta A_m^T\big(\sum_{j=1}^{m-1} A_j(x_j^k - \tilde{x}_j^k)\big) \\ 0 \end{pmatrix} \right\} \tag{3.8}$$

Use the notations of $d_1(w^k, \tilde{w}^k)$ and $d_2(w^k, \tilde{w}^k)$, the assertion is proved. $\square$

Lemma 3.2 *Let* $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k)$ *be generated by the ADM step (2.1) from the given vector* $w^k = (x_2^k, \ldots, x_m^k, \lambda^k)$*. Then, we have*

$$(\tilde{w}^k - w^*)^T d_1(w^k, \tilde{w}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\Big), \quad \forall\, w^* \in \mathcal{W}^*, \tag{3.9}$$

where $d_1(w^k, \tilde{w}^k)$ *is defined in* (3.2)*.*

Proof. Since $w^* \in \mathcal{W}$, it follows from (3.1) that

$$(\tilde{w}^k - w^*)^T d_1(w^k, \tilde{w}^k) \geq (\tilde{w}^k - w^*)^T d_2(w^k, \tilde{w}^k). \tag{3.10}$$

We consider the right-hand side of (3.10). By using (3.3), we get

$$\begin{aligned} &(\tilde{w}^k - w^*)^T d_2(w^k, \tilde{w}^k) \\ &= \Big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\Big)^T \beta \Big(\sum_{j=1}^m A_j(\tilde{x}_j^k - x_j^*)\Big) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \end{aligned} \tag{3.11}$$

Then, we look at the right-hand side of (3.11). Since $\tilde{w}^k \in \mathcal{W}$, by using the monotonicity of F, we have

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

Because that

$$\sum_{j=1}^{m} A_j x_j^* = b \qquad \text{and} \qquad \beta(\sum_{j=1}^{m} A_j \tilde{x}_j^k - b) = \lambda^k - \tilde{\lambda}^k,$$

it follows from (3.11) that

$$(\tilde{w}^k - w^*)^T d_2(w^k, \tilde{w}^k) \quad \geq \quad (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=2}^{m} A_j (x_j^k - \tilde{x}_j^k)\Big). \tag{3.12}$$

Substituting (3.12) into (3.10), the assertion (3.9) follows immediately. $\square$

Since (see (2.3) and (3.2))

$$d_1(w^k, \tilde{w}^k) = M(w^k - \tilde{w}^k), \tag{3.13}$$

from (3.9) follows that

$$(\tilde{w}^k - w^*)^T M(w^k - \tilde{w}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=1}^{m} A_j (x_j^k - \tilde{x}_j^k)\Big), \quad \forall\, w^* \in \mathcal{W}^*. \tag{3.14}$$

Now, based on the last two lemmas, we are at the stage to prove the main theorem.

Theorem 3.1 **(Main Theorem)** *Let* $\tilde{w}^k = (\tilde{x}_1^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k)$ *be generated by the ADM step* (2.1) *from the given vector* $w^k = (x_1^k, \ldots, x_m^k, \lambda^k)$*. Then, we have*

$$\begin{aligned} & (w^k - w^*)^T M (w^k - \tilde{w}^k) \\ \geq \quad & \frac{1}{2}\|w^k - \tilde{w}^k\|_H^2 + \frac{1}{2}\|w^k - \tilde{w}^k\|_Q^2, \ \forall\, w^* \in \mathcal{W}^*, \end{aligned} \tag{3.15}$$

where M*,* H*, and* Q *are defined in* (2.3)*,* (2.4) *and* (2.10)*, respectively.*

Proof First, it follows from (3.14) that

$$\begin{aligned} & (w^k - w^*)^T M (w^k - \tilde{w}^k) \\ & \geq \ (w^k - \tilde{w}^k)^T M (w^k - \tilde{w}^k) + (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=1}^m A_j (x_j^k - \tilde{x}_j^k)\Big), \end{aligned} \tag{3.16}$$

for all $w^* \in \mathcal{W}^*$.

Now, we treat the terms of the right hand side of (3.16). Using the matrix M (see (2.3)), we have

$$(w^k - \tilde{w}^k)^T M (w^k - \tilde{w}^k)$$

$$= \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} r_1 I_{n_1} & 0 & \cdots & \cdots & 0 \\ \beta A_2^T A_1 & r_2 I_{n_2} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \beta A_m^T A_1 & \cdots & \beta A_m^T A_{m-1} & r_m I_{n_m} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_l \end{pmatrix} \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \tag{3.17}$$

For the second term of the right-hand side of (3.16), by a manipulations, we obtain

$$
\begin{aligned}
&(\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=1}^{m} A_j (x_j^k - \tilde{x}_j^k)\Big) \\
= \quad &\begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}^T
\begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ A_1 & A_2 & \dots & A_m & 0 \end{pmatrix}
\begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \quad (3.18)
\end{aligned}
$$

Adding (3.17) and (3.18) together, it follows that

$$(w^k - \tilde{w}^k)^T M (w^k - \tilde{w}^k) + (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=1}^{m} A_j (x_j^k - \tilde{x}_j^k)\Big)$$

$$= \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \begin{pmatrix} r_1 I_{n_1} & 0 & \cdots & \cdots & 0 \\ \beta A_2^T A_1 & r_2 I_{n_2} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \beta A_m^T A_1 & \cdots & \beta A_m^T A_{m-1} & r_m I_{n_m} & 0 \\ A_1 & A_2 & \cdots & A_m & \frac{1}{\beta} I_l \end{pmatrix} \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \begin{pmatrix} 2r_1 I_{n_1} & \beta A_1^T A2 & \cdots & \beta A_1^T A_m & A_1^T \\ \beta A_2^T A_1 & 2r_2 I_{n_2} & \ddots & \vdots & A_2^T \\ \vdots & \ddots & \ddots & \beta A_{m-1}^T A_m & \vdots \\ \beta A_m^T A_1 & \cdots & \beta A_m^T A_{m-1} & 2r_m I_{n_m} & A_m^T \\ A_1 & A_2 & \cdots & A_m & \frac{2}{\beta} I_l \end{pmatrix} \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

Use the notation of the matrices H, Q and the condition (2.2) to the right-hand side of the last equality, we obtain

$$
\begin{aligned}
&(w^k - \tilde{w}^k)^T M(w^k - \tilde{w}^k) + (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=1}^{m} A_j(x_j^k - \tilde{x}_j^k)\Big) \\
= \quad & \frac{1}{2}\|w^k - \tilde{w}^k\|_H^2 + \frac{1}{2}\|w^k - \tilde{w}^k\|_Q^2.
\end{aligned}
$$

Substituting the last equality in (3.16), the theorem is proved. $\square$

It follows from (3.15) that

$$
\big\langle MH^{-1}M^T(w^k - w^*), M^{-T}H(\tilde{w}^k - w^k)\big\rangle \le -\frac{1}{2}\|w^k - \tilde{w}^k\|_{(H+Q)}^2.
$$

In other words, by setting

$$
G = MH^{-1}M^T, \tag{3.19}
$$

$MH^{-1}M^T(w^k - w^*)$ is the gradient of the distance function $\frac{1}{2}\|w - w^*\|_G^2$, and $M^{-T}H(\tilde{w}^k - w^k)$ is a descent direction of $\frac{1}{2}\|w - w^*\|_G^2$ at the current point w^k whenever $\tilde{w}^k \neq w^k$.

3.2 The contractive property

In this subsection, we mainly prove that the sequence generated by the proposed ADM with Gaussian back substitution is contractive with respect to the set $\mathcal{W}^*$. Note that we follow the definition of contractive type methods. With this contractive property, the convergence of the proposed linearized ADM with Gaussian back substitution can be easily derived with subroutine analysis.

Theorem 3.2 *Let* $\tilde{w}^k = (\tilde{x}_1^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k)$ *be generated by the ADM step* (2.1) *from the given vector* $w^k = (x_1^k, \ldots, x_m^k, \lambda^k)$. *Let the matrix* G *be given by* (3.19). *For the new iterate* w^{k+1} *produced by the Gaussian back substitution* (2.7), *there exists a constant* $c_0 > 0$ *such that*

$$\|w^{k+1} - w^*\|_G^2 \le \|w^k - w^*\|_G^2 - c_0\big(\|w^k - \tilde{w}^k\|_H^2 + \|w^k - \tilde{w}^k\|_Q^2\big), \ \forall\, w^* \in \mathcal{W}^*, \tag{3.20}$$

where H *and* Q *are defined in* (2.4) *and* (2.10), *respectively.*

Proof. For $G = MH^{-1}M^T$ and any $\alpha \geq 0$, we obtain

$$
\begin{aligned}
&\|w^k - w^*\|_G^2 - \|w^{k+1} - w^*\|_G^2 \\
&\quad = \quad \|w^k - w^*\|_G^2 - \|(w^k - w^*) - \alpha M^{-T} H(w^k - \tilde{w}^k)\|_G^2 \\
&\quad = \quad 2\alpha(w^k - w^*)^T M(w^k - \tilde{w}^k) - \alpha^2\|w^k - \tilde{w}^k\|_H^2. \qquad (3.21)
\end{aligned}
$$

Substituting the result of Theorem 3.1 into the right-hand side of the last equation, we get

$$
\begin{aligned}
&\|w^k - w^*\|_G^2 - \|w^{k+1} - w^*\|_G^2 \\
&\quad \geq \quad \alpha\big(\|w^k - \tilde{w}^k\|_H^2 + \|w^k - \tilde{w}^k\|_Q^2\big) - \alpha^2\|w^k - \tilde{w}^k\|_H^2 \\
&\quad = \quad \alpha(1-\alpha)\|w^k - \tilde{w}^k\|_H^2 + \alpha\|w^k - \tilde{w}^k\|_Q^2,
\end{aligned}
$$

and thus

$$
\begin{aligned}
&\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 \\
&\qquad -\alpha\big((1-\alpha)\|w^k - \tilde{w}^k\|_H^2 + \|w^k - \tilde{w}^k\|_Q^2\big), \ \forall\, w^* \in \mathcal{W}^*. \qquad (3.22)
\end{aligned}
$$

Set $c_0 = \alpha(1-\alpha)$. Recall that $\alpha \in [0.5, 1)$. The assertion is proved. $\Box$

Corollary 3.1 *The assertion of Theorem 3.2 also holds if the Gaussian back substitution is (2.8).*

Proof. Analogous to the proof of Theorem 3.2, we have that

$$\begin{aligned} &\|w^k - w^*\|_G^2 - \|w^{k+1} - w^*\|_G^2 \\ &\geq\; 2\gamma\alpha_k^*(w^k - w^*)^T M(w^k - \tilde{w}^k) - (\gamma\alpha_k^*)^2\|w^k - \tilde{w}^k\|_H^2, \end{aligned} \tag{3.23}$$

where α_k^* is given by (2.9). According to (2.9), we have that

$$\alpha_k^*\big(\|w^k - \tilde{w}^k\|_H^2\big) = \frac{1}{2}\big(\|w^k - \tilde{w}^k\|_H^2 + \|w^k - \tilde{w}^k\|_Q^2\big).$$

Then, it follows from the above equality and (3.15) that

$$\begin{aligned} \|w^k - w^*\|_G^2 - \|w^{k+1} - w^*\|_G^2 & \\ \geq\quad & \gamma\alpha_k^*\big(\|w^k - \tilde{w}^k\|_H^2 + \|w^k - \tilde{w}^k\|_Q^2\big) \\ & -\frac{1}{2}\gamma^2\alpha_k^*\big(\|w^k - \tilde{w}^k\|_H^2 + \|w^k - \tilde{w}^k\|_Q^2\big) \\ =\quad & \frac{1}{2}\gamma(2-\gamma)\alpha_k^*\big(\|w^k - \tilde{w}^k\|_H^2 + \|w^k - \tilde{w}^k\|_Q^2\big). \end{aligned}$$

Because $\alpha_k^* \geq \frac{1}{2}$, it follows from the last inequality that

$$\begin{aligned} \|w^{k+1} - w^*\|_G^2 &\leq \|w^k - w^*\|_G^2 \\ &\quad -\frac{1}{4}\gamma(2-\gamma)\big(\|w^k - \tilde{w}^k\|_H^2 + \|w^k - \tilde{w}^k\|_Q^2\big), \ \forall\, w^* \in \mathcal{W}^*. \end{aligned} \tag{3.24}$$

Since $\gamma \in (0, 2)$, the assertion of this corollary follows from (3.24) directly. $\square$

3.3 Convergence

The proposed lemmas and theorems are adequate to establish the global convergence of the proposed ADM with Gaussian back substitution, and the analytic framework is quite typical in the context of contractive type methods.

Theorem 3.3 *Let $\{w^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the proposed ADM with Gaussian back substitution. Then we have*

1. *The sequence $\{w^k\}$ is bounded.*
2. $\lim_{k\to\infty} \|w^k - \tilde{w}^k\| = 0$,

3. *Any cluster point of* $\{\tilde{w}^k\}$ *is a solution point of* (1.6).

4. *The sequence* $\{\tilde{w}^k\}$ *converges to some* $w^\infty \in \mathcal{W}^*$.

Proof. The first assertion follows from (3.20) directly. In addition, from (3.20) we get

$$\sum_{k=0}^{\infty} c_0 \|w^k - \tilde{w}^k\|_H^2 \le \|w^0 - w^*\|_G^2$$

and thus we get $\lim_{k\to\infty} \|w^k - \tilde{w}^k\|_H^2 = 0$, and consequently

$$\lim_{k\to\infty} \|x_i^k - \tilde{x}_i^k\| = 0, \quad i = 2, \ldots, m, \tag{3.25}$$

and

$$\lim_{k\to\infty} \|\lambda^k - \tilde{\lambda}^k\| = 0. \tag{3.26}$$

The second assertion is proved.

Substituting (3.25) into (3.5), for $i = 1, 2, \ldots, m$, we have

$$\tilde{x}_i^k \in \mathcal{X}_i, \quad \lim_{k\to\infty} (x_i - \tilde{x}_i^k)^T \{ f_i(\tilde{x}_i^k) - A_i^T \tilde{\lambda}^k \} \ge 0, \;\; \forall\, x_i \in \mathcal{X}_i. \tag{3.27}$$

It follows from (2.1) and (3.26) that

$$\lim_{k\to\infty}(\sum_{j=1}^{m} A_j\tilde{x}_j^k - b) = 0. \tag{3.28}$$

Combining (3.27) and (3.28) we get

$$\tilde{w}^k \in \mathcal{W}, \quad \lim_{k\to\infty}(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge 0, \quad \forall\, w \in \mathcal{W}, \tag{3.29}$$

and thus any cluster point of $\{\tilde{w}^k\}$ is a solution point of (1.6). The third assertion is proved.

It follows from the first assertion and $\lim_{k\to\infty}\|w^k - \tilde{w}^k\|_H^2 = 0$ that $\{\tilde{w}^k\}$ is also bounded. Let w^∞ be a cluster point of $\{\tilde{w}^k\}$ and the subsequence $\{\tilde{v}^{k_j}\}$ converges to w^∞. It follows from (3.29) that

$$\tilde{w}^{k_j} \in \mathcal{W}, \quad \lim_{k\to\infty}(w-\tilde{w}^{k_j})^T F(\tilde{w}^{k_j}) \ge 0, \quad \forall\, w \in \mathcal{W} \tag{3.30}$$

and consequently

$$\left\{\begin{array}{l} (x_i - x_i^{\infty})^T\{f_i(x_i^{\infty}) - A_i^T\lambda^{\infty}\} \ge 0, \quad \forall x_i \in \mathcal{X}_i,\ i = 1,\ldots,m, \\ \\ \sum_{j=1}^m A_j x_j^{\infty} - b = 0. \end{array}\right.$$

This means that $w^{\infty} \in \mathcal{W}^*$ is a solution point of (1.6) .

Since $\{w^k\}$ is Fej$\acute{e}$r monotone and $\lim_{k\to\infty}\|w^k - \tilde{w}^k\| = 0$, the sequence $\{\tilde{w}^k\}$ cannot have other cluster point and $\{\tilde{w}^k\}$ converges to $w^{\infty} \in \mathcal{W}^*$. $\square$

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] 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

[3] B.S. He and X. M. Yuan: Linearized Alternating Direction Method with Gaussian Back Substitution for Separable Convex Programming. http://www.optimization-online.org/DB_HTML/2011/10/3192.html. Numerical Algebra, Control and Optimization (Special Issue in Honor of Prof. Xuchu He's 90th Birthday), 3(2), 247-260, 2013.

[4] C. H. Ye and X. M. Yuan, A Descent Method for Structured Monotone Variational Inequalities, Optimization Methods and Software, **22**, 329-338, 2007.