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

The context of this lecture is based on the publication [7]
1 Introduction

In the literature, the alternating direction method (ADM) proposed originally for the following linearly constrained separable convex programming whose objective function is separable into two individual convex functions without crossed variables:

\[ \min \theta_1(x_1) + \theta_2(x_2) \]

\[ A_1 x_1 + A_2 x_2 = b, \quad (1.1) \]

\[ x_1 \in \mathcal{X}_1 \quad \text{and} \quad x_2 \in \mathcal{X}_2, \]

where \( \theta_1 : \mathbb{R}^{n_1} \rightarrow \mathbb{R} \) and \( \theta_2 : \mathbb{R}^{n_2} \rightarrow \mathbb{R} \) are closed proper convex functions (not necessarily smooth); \( \mathcal{X}_1 \subset \mathbb{R}^{n_1} \) and \( \mathcal{X}_2 \subset \mathbb{R}^{n_2} \) are closed convex sets; \( A_1 \in \mathbb{R}^{L \times n_1} \) and \( A_2 \in \mathbb{R}^{L \times n_2} \) are given matrices; and \( b \in \mathbb{R}^L \) is a given vector. The iterative
scheme of ADM for solving (1.1) is as follows:

\[
\begin{align*}
    x_1^{k+1} &= \arg\min \left\{ \theta_1(x_1) + \frac{\beta}{2} \| (A_1 x_1 + A_2 x_2^k - b) - \frac{1}{\beta} \lambda^k \|^2 \mid x_1 \in X_1 \right\}, \\
    x_2^{k+1} &= \arg\min \left\{ \theta_2(x_2) + \frac{\beta}{2} \| (A_1 x_1^{k+1} + A_2 x_2 - b) - \frac{1}{\beta} \lambda^k \|^2 \mid x_2 \in X_2 \right\}, \quad (1.2) \\
    \lambda^{k+1} &= \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b),
\end{align*}
\]

where \( \lambda^k \in \mathbb{R}^L \) is the Lagrange multiplier associated with the linear constraint and \( \beta > 0 \) is the penalty parameter for the violation of the linear constraint.

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

\[
\begin{align*}
    \min & \quad \sum_{i=1}^{m} \theta_i(x_i) \\
    \sum_{i=1}^{m} A_i x_i &= b; \\
    x_i &\in X_i, \quad i = 1, \ldots, m; \quad (1.3)
\end{align*}
\]

where \( \theta_i : \mathbb{R}^{n_i} \to \mathbb{R} \ (i = 1, \ldots, m) \) are closed proper convex functions (not necessarily smooth); \( X_i \subset \mathbb{R}^{n_i} \ (i = 1, \ldots, m) \) are closed convex sets; \( A_i \in \mathbb{R}^{l \times n_i} \ (i = 1, \ldots, m) \) are given matrices and \( b \in \mathbb{R}^l \) is a given vector.
Because of the efficiency of ADM for (1.1), a natural idea for solving (1.3) is to extend the ADM (1.2) from the special case (1.1) to the general case (1.3).

In fact, even for the special case of (1.3) with $m = 3$, the convergence of the extended ADM is still open.

In this paper, we provide a novel approach towards the extension of ADM for the problem (1.3). 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 (1.3). In this sense, we prove the convergence of the extension of ADM for (1.3). The resulting method is called the ADM with Gaussian back substitution from now on.

Alternatively, the ADM with Gaussian back substitution 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. We prove the convergence of the ADM with Gaussian back substitution under the analytic framework of contractive type methods.

Throughout, we assume that the matrices $A_i^T A_i$ ($i = 2, \ldots, m$) are nonsingular and the solution set of (1.3) is nonempty.
2 The variational inequality characterization

In this section, we derive the first-order optimality condition of (1.3) and thus characterize (1.3) 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 $\lambda \in \mathbb{R}^l$ to the linear constraint, the Lagrange function of (1.3) is:

$$L(x_1, x_2, \ldots, x_m, \lambda) = \sum_{i=1}^{m} \theta_i(x_i) - \lambda^T \left( \sum_{i=1}^{m} A_i x_i - b \right), \quad (2.1)$$

which is defined on

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

Let $(x_1^*, x_2^*, \ldots, x_m^*, \lambda^*)$ be a saddle point of the Lagrange function (2.1). Then we have

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

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

\[
\begin{align*}
(x_1 - x_1^*)^T \{ f_1(x_1^*) - A_1^T \lambda^* \} & \geq 0, \\
(x_2 - x_2^*)^T \{ f_2(x_2^*) - A_2^T \lambda^* \} & \geq 0, \\
& \vdots \\
(x_m - x_m^*)^T \{ f_m(x_m^*) - A_m^T \lambda^* \} & \geq 0, \\
(\lambda - \lambda^*)^T (\sum_{i=1}^{m} A_i x_i^* - b) & \geq 0,
\end{align*}
\]

for all \( w = (x_1, x_2, \cdots, x_m, \lambda) \in \mathcal{W} \). More compactly, (2.2) can be written into the following VI:

\[
(w - w^*)^T F(w^*) \geq 0, \quad \forall \ w \in \mathcal{W},
\]
where

\[
\begin{pmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_m \\
    \lambda
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
    f_1(x_1) - A_1^T \lambda \\
    f_2(x_2) - A_2^T \lambda \\
    \vdots \\
    f_m(x_m) - A_m^T \lambda \\
    \sum_{i=1}^{m} A_i x_i - b
\end{pmatrix}.
\]

(2.3b)

Note that the operator \( F(w) \) defined in (2.3b) is monotone due to the fact that \( \theta_i \)'s are all convex functions. In addition, since we have assumed that the solution set of (1.3) is not empty, the solution set of (2.3), denoted by \( \mathcal{W}^* \), is also nonempty.

In addition to the notation of \( w = (x_1, x_2, \ldots, x_m, \lambda) \), for any integer number \( k \) we also use the following notation:

\[
v = (x_2, \ldots, x_m, \lambda).
\]

Moreover, we define

\[
\mathcal{V}^* = \{(x_2^*, \ldots, x_m^*, \lambda^*) \mid (x_1^*, x_2^*, \ldots, x_m^*, \lambda^*) \in \mathcal{W}^* \}.
\]
3 The ADM with Gaussian back substitution

In this section, we show the combination of the extended ADM scheme (1.2) with a Gaussian back substitution procedure, and derive the resulting ADM with Gaussian back substitution for solving (1.3). We also elucidate how to realize the Gaussian back substitution for some special cases of (1.3).

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

\[
M = \begin{pmatrix}
\beta A_2^T A_2 & 0 & \cdots & \cdots & 0 \\
\beta A_3^T A_2 & \beta A_3^T A_3 & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\beta A_m^T A_2 & \beta A_m^T A_3 & \cdots & \beta A_m^T A_m & 0 \\
0 & 0 & \cdots & 0 & \frac{1}{\beta} I_l
\end{pmatrix},
\]

and

\[
H = \text{diag}(\beta A_2^T A_2, \beta A_3^T A_3, \ldots, \beta A_m^T A_m, \frac{1}{\beta} I_l).
\]

Note that for any \( \beta > 0 \), under the assumption that all the matrices \( A_i^T A_i \)'s are
nonsingular, the matrix $M$ defined in (3.1) is a non-singular lower-triangular block matrix and $H$ defined in (3.2) is a symmetric positive definite matrix. In addition, according to (3.1) and (3.2), we easily have:

$$H^{-1}M^T = \begin{pmatrix}
I_{n_2} & (A_2^T A_2)^{-1} A_2^T A_3 & \cdots & (A_2^T A_2)^{-1} A_2^T A_m & 0 \\
0 & \ddots & \ddots & \vdots & \vdots \\
\vdots & \ddots & \ddots & (A_{m-1}^T A_{m-1})^{-1} A_{m-1}^T A_m & 0 \\
0 & \cdots & 0 & I_{n_m} & 0 \\
0 & \cdots & 0 & 0 & I_l
\end{pmatrix}$$

(3.3)

which is a upper-triangular block matrix whose diagonal components are identity matrices.
Algorithm: The ADM with Gaussian back substitution for (1.3):

Let $\beta > 0$ and $\alpha \in [0.5, 1)$, With the given iterate $v^k = (x^k_2, \ldots, x^k_m, \lambda^k)$.

**Step 1. ADM step (prediction step).** For $i = 1, \ldots, m$, obtain $\tilde{x}^k_i$ in the forward (alternating) order by solving the following $x_i$-problem:

$$\min \left\{ \theta_i(x) + \frac{\beta}{2} \left\| (\sum_{j=1}^{i-1} A_j \tilde{x}^k_j + A_i x_i + \sum_{j=i+1}^{m} A_j x^k_j - b) - \frac{1}{\beta} \lambda^k \right\|_2^2 \mid x_i \in X_i \right\} \tag{3.4a}$$

and set

$$\tilde{\lambda}^k = \lambda^k - \beta (\sum_{j=1}^{m} A_j \tilde{x}^k_j - b). \tag{3.4b}$$

**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 $v^{k+1}$:

$$H^{-1} M^T (v^{k+1} - v^k) = \alpha (\tilde{w}^k - v^k). \tag{3.4c}$$

where the matrices $M$ and $H$ are defined by (3.1) and (3.2), respectively.

Note that in this method, the iteration is from $v^k$ to $v^{k+1}$, the variable $x_1$ is only an intermediate variable.
Recall that the matrix $H^{-1}M^T$ defined in (3.3) is a upper-triangular block matrix. The Gaussian back substitution step (3.4c) is thus very easy to execute. In fact, as we mentioned, after the predictor is generated by the ADM scheme (3.4a) 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 (3.4a).

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 (3.4c) can be rewritten into

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

(3.5)

As we will show, $-(v^k - \tilde{v}^k)$ is a descent direction of the distance function $\frac{1}{2}\|v - v^*\|_G^2$ with $G = MH^{-1}M^T$ at the point $v = v^k$ for any $v^* \in \mathcal{V}^*$. In this sense, the proposed ADM with Gaussian back substitution can also be regarded as an ADM-based contraction method where the output of the ADM scheme (3.4a) contributes a descent direction of the distance function. Thus, the constant $\alpha$ in (3.4c) plays the role of a step size along the
descent direction $-(v^k - \tilde{v}^k)$. In fact, we can choose the step size dynamically based on some techniques in the literature (e.g. [10]), and the Gaussian back substitution procedure with the constant $\alpha$ can be modified accordingly into the following variant with a dynamical step size:

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

(3.6)

where

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

(3.7)

$$Q = \begin{pmatrix}
\beta A_2^T A_2 & \beta A_2^T A_3 & \cdots & \beta A_2^T A_m & A_2^T \\
\beta A_3^T A_2 & \beta A_3^T A_3 & \cdots & \beta A_3^T A_m & A_3^T \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\beta A_m^T A_2 & \beta A_m^T A_3 & \cdots & \beta A_m^T A_m & A_m^T \\
A_2 & A_3 & \cdots & A_m & \frac{1}{\beta} I_t
\end{pmatrix};$$

(3.8)

and $\gamma \in (0, 2)$. Indeed, for any $\beta > 0$, the symmetric matrix $Q$ is positive semi-definite.
Then, for given $v^k$ and the $\tilde{v}^k$ obtained by the ADM procedure (3.4a), we have that

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

and

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

where the norm $\|v\|_H^2$ ($\|v\|_Q^2$, respectively) is defined as $v^T H v$ ($v^T Q v$, respectively).

In fact, it is easy to prove that the step size $\alpha_k^*$ defined in (3.7) satisfies $\frac{1}{2} \leq \alpha_k^* \leq \frac{m+1}{2}$.

4 Convergence of the proposed method

In this section, we prove the convergence of the proposed ADM with Gaussian back substitution for solving (1.3). Our proof follows the analytic framework of contractive type methods, and it consists of the following three phases:

1.) Prove that $-M^{-T} H (v^k - \tilde{v}^k)$ is a descent direction of the function $\frac{1}{2}\|v - v^*\|_G^2$ at
the point \( v = v^k \) whenever \( \tilde{v}^k \neq v^k \), where \( \tilde{v}^k \) is generated by the ADM scheme (3.4a), \( v^* \in \mathcal{V}^* \) and \( G \) is a positive definite matrix.

2.) Prove that the sequence generated by the proposed ADM with Gaussian back substitution is contractive with respect to \( \mathcal{V}^* \).

3.) Derive the convergence based on the Fejér monotonicity of the sequence generated by the proposed ADM with Gaussian back substitution.

Accordingly, we divide this section into three subsections to address the tasks listed above.

### 4.1 Verification of the descent directions

We mainly show that \(- (v^k - \tilde{v}^k)\) is a descent direction of the function \( \frac{1}{2} \| v - v^* \|_G^2 \) at the point \( v = v^k \) whenever \( \tilde{v}^k \neq v^k \), where \( \tilde{v}^k \) is generated by the ADM scheme (3.4a), \( v^* \in \mathcal{V}^* \) and \( G \) is a positive definite matrix. For this purpose, we first prove two lemmas.

**Lemma 4.1** Let \( \tilde{\omega}^k = (\tilde{x}^k_1, \tilde{x}^k_2, \ldots, \tilde{x}^k_m, \tilde{\lambda}^k) \) be generated by the ADM step (3.4a) from the given vector \( v^k = (x^k_2, \ldots, x^k_m, \lambda^k) \). Then, we have

\[
\tilde{\omega}^k \in \mathcal{W}, \quad (w - \tilde{\omega}^k)^T \left\{ d_2(v^k, \tilde{\omega}^k) - d_1(v^k, \tilde{v}^k) \right\} \geq 0, \quad \forall \ w \in \mathcal{W}, \quad (4.1)
\]
where

\[
d_1(v^k, \tilde{v}^k) = \begin{pmatrix}
0 & 0 & \cdots & \cdots & 0 \\
\beta A_2^T A_2 & 0 & \cdots & \cdots & 0 \\
\beta A_3^T A_2 & \beta A_3^T A_3 & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\beta A_m^T A_2 & \beta A_m^T A_3 & \cdots & \beta A_m^T A_m & 0 \\
0 & 0 & \cdots & 0 & \frac{1}{\beta} I
\end{pmatrix}
\begin{pmatrix}
x_2^k - \tilde{x}_2^k \\
x_3^k - \tilde{x}_3^k \\
\vdots \\
x_m^k - \tilde{x}_m^k \\
\lambda^k - \tilde{\lambda}^k
\end{pmatrix},
\]

and

\[
d_2(v^k, \tilde{w}^k) = F(\tilde{w}^k) + \beta \begin{pmatrix}
A_1^T \\
A_2^T \\
\vdots \\
A_m^T \\
0
\end{pmatrix} \left( \sum_{j=2}^{m} A_j (x_j^k - \tilde{x}_j^k) \right).
\]
Proof. Since $\tilde{x}_i^k$ is the solution of (3.4a), for $i = 1, 2, \ldots, m$, according to the optimality condition, we have

$$\tilde{x}_i^k \in \mathcal{X}_i, \quad (x_i - \tilde{x}_i^k)^T \left\{ f_i(\tilde{x}_i^k) - A_i^T \lambda^k + \beta A_i^T (\sum_{j=1}^{i} A\tilde{x}_j^k + \sum_{j=i+1}^{m} A_j x_j^k - b) \right\} \geq 0, \quad \forall x_i \in \mathcal{X}_i. \quad (4.4)$$

By using the fact (see (3.4b)) $\lambda^k = \tilde{\lambda}^k + \beta (\sum_{j=1}^{m} A_j \tilde{x}_j^k - b)$, thus, we have

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

Substituting it in (4.4), we obtain

$$\tilde{x}_i^k \in \mathcal{X}_i, \quad (x_i - \tilde{x}_i^k)^T \left\{ f_i(\tilde{x}_i^k) - A_i^T \tilde{\lambda}^k + \beta A_i^T (\sum_{j=i+1}^{m} A_j (x_j^k - \tilde{x}_j^k)) \right\} \geq 0, \quad \forall x_i \in \mathcal{X}_i. \quad (4.5)$$
It follows from (4.5) that $\tilde{x}^k \in \mathcal{X}$ and

$$
\begin{pmatrix}
    x_1 - \tilde{x}_1^k \\
    x_2 - \tilde{x}_2^k \\
    \vdots \\
    x_m - \tilde{x}_m^k
\end{pmatrix}^T \begin{pmatrix}
    f_1(\tilde{x}_1^k) - A_1^T \tilde{x}_1^k \\
    f_2(\tilde{x}_2^k) - A_2^T \tilde{x}_2^k \\
    \vdots \\
    f_m(\tilde{x}_m^k) - A_m^T \tilde{x}_m^k
\end{pmatrix} + \beta \begin{pmatrix}
    A_1^T (\sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k)) \\
    A_2^T (\sum_{j=3}^m A_j (x_j^k - \tilde{x}_j^k)) \\
    \vdots \\
    A_m^T (\sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k)) + 0
\end{pmatrix} \geq 0,
$$

(4.6)

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

$$
\begin{pmatrix}
    x_1 - \tilde{x}_1^k \\
    x_2 - \tilde{x}_2^k \\
    \vdots \\
    x_m - \tilde{x}_m^k
\end{pmatrix}^T \beta \begin{pmatrix}
    0 \\
    A_2^T (\sum_{j=2}^2 A_j (x_j^k - \tilde{x}_j^k)) \\
    \vdots \\
    A_m^T (\sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k))
\end{pmatrix}
$$

to the both sides of (4.6), we get $\tilde{x}^k \in \mathcal{X}$ and
\[
\begin{pmatrix}
  x_1 - \tilde{x}_1^k \\
  x_2 - \tilde{x}_2^k \\
  \vdots \\
  x_m - \tilde{x}_m^k
\end{pmatrix}^T
\begin{pmatrix}
  f_1(\tilde{x}_1^k) - A_1^T \tilde{\lambda}^k + \beta A_1^T \left( \sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right) \\
  f_2(\tilde{x}_2^k) - A_2^T \tilde{\lambda}^k + \beta A_2^T \left( \sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right) \\
  \vdots \\
  f_m(\tilde{x}_m^k) - A_m^T \tilde{\lambda}^k + \beta A_m^T \left( \sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right)
\end{pmatrix} \\
\geq \begin{pmatrix}
  x_1 - \tilde{x}_1^k \\
  x_2 - \tilde{x}_2^k \\
  \vdots \\
  x_m - \tilde{x}_m^k
\end{pmatrix}^T
\begin{pmatrix}
  0 \\
  \beta A_2^T \left( \sum_{j=2}^2 A_j (x_j^k - \tilde{x}_j^k) \right) \\
  \vdots \\
  \beta A_m^T \left( \sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right)
\end{pmatrix}
, \quad \forall \ x \in \mathcal{X}. \quad (4.7)
\]

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

\[
(\lambda - \tilde{\lambda}^k)^T \left( \sum_{j=1}^m A_j \tilde{x}_j^k - b \right) = (\lambda - \tilde{\lambda}^k)^T \frac{1}{\beta} (\lambda^k - \tilde{\lambda}^k).
\]
Adding (4.7) and the last equality together, we get $\tilde{w}^k \in \mathcal{W}$ and

$$
\begin{pmatrix}
    x_1 - \tilde{x}_1^k \\
    x_2 - \tilde{x}_2^k \\
    \vdots \\
    x_m - \tilde{x}_m^k \\
    \lambda - \tilde{\lambda}^k 
\end{pmatrix}
\begin{pmatrix}
    f_1(\tilde{x}_1^k) - A_1^T \tilde{\lambda}^k + \beta A_1^T \left( \sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right) \\
    f_2(\tilde{x}_2^k) - A_2^T \tilde{\lambda}^k + \beta A_2^T \left( \sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right) \\
    \vdots \\
    f_m(\tilde{x}_m^k) - A_m^T \tilde{\lambda}^k + \beta A_m^T \left( \sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right) \\
    \sum_{i=1}^m A_i \tilde{x}_i^k - b 
\end{pmatrix}
\begin{pmatrix}
    0 \\
    \beta A_2^T \left( \sum_{j=2}^2 A_j (x_j^k - \tilde{x}_j^k) \right) \\
    \vdots \\
    \beta A_m^T \left( \sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \right) \\
    \frac{1}{\beta} (\lambda^k - \tilde{\lambda}^k) 
\end{pmatrix}, \; \forall w \in \mathcal{W}.
$$

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

\[\Box\]

Note that the $d_1(v^k, \tilde{v}^k)$ depends only on $v^k$ and $\tilde{v}^k$, while $d_2(v^k, \tilde{w}^k)$ is determined by both $v^k$ and $\tilde{w}^k$. 
Lemma 4.2 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 (3.4a) from the given vector \( v^k = (x_2^k, \ldots, x_m^k, \lambda^k) \). Then, we have

\[
(\tilde{w}^k - w^*)^T d_1(v^k, \tilde{v}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{j=2}^{m} A_j (x_j^k - \tilde{x}_j^k) \right), \quad \forall v^* \in \mathcal{V}^*,
\]

where \( d_1(v^k, \tilde{v}^k) \) is defined in (4.2).

**Proof.** Since \( w^* \in \mathcal{W} \), it follows from (4.1) that

\[
(\tilde{w}^k - w^*)^T d_1(v^k, \tilde{v}^k) \geq (\tilde{w}^k - w^*)^T d_2(v^k, \tilde{w}^k).
\]

We consider the right-hand side of (4.9). By using (4.3), we get

\[
(\tilde{w}^k - w^*)^T d_2(v^k, \tilde{w}^k) = \left( \sum_{j=2}^{m} A_j (x_j^k - \tilde{x}_j^k) \right)^T \beta \left( \sum_{j=1}^{m} A_j (\tilde{x}_j^k - x_j^*) \right) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k).
\]

Then, we look at the right-hand side of (4.10). 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
\[ \sum_{j=1}^{m} A_j x_j^* = b \quad \text{and} \quad \beta \left( \sum_{j=1}^{m} A_j \tilde{x}_j^k - b \right) = \lambda^k - \tilde{\lambda}^k, \]
it follows from (4.10) that
\[ (\tilde{w}^k - w^*)^T d_2(v^k, \tilde{w}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{j=2}^{m} A_j(x_j^k - \tilde{x}_j^k) \right). \quad (4.11) \]

Substituting (4.11) into (4.9), the assertion (4.8) follows immediately. \( \square \)

Since (see (3.1) and (4.2))
\[ d_1(v^k, \tilde{v}^k) = \begin{pmatrix} 0 \\ M(v^k - \tilde{v}^k) \end{pmatrix}, \quad (4.12) \]
we have
\[ (\tilde{v}^k - v^*)^T M(v^k - \tilde{v}^k) = (\tilde{w}^k - w^*)^T d_1(v^k, \tilde{v}^k) \]
and consequently from (4.8) follows that

\[
(v^k - v^*)^T M (v^k - \tilde{v}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{j=2}^{m} A_j (x^k_j - \tilde{x}^k_j) \right), \quad \forall v^* \in \mathcal{V}^*. \quad (4.13)
\]

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

**Theorem 4.1 (Main Theorem)** 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 (3.4a) from the given vector \( v^k = (x_2^k, \ldots, x_m^k, \lambda^k) \). Then, we have

\[
(v^k - v^*)^T M (v^k - \tilde{v}^k) \geq \frac{1}{2} \| v^k - \tilde{v}^k \|_H^2 + \frac{1}{2} \| v^k - \tilde{v}^k \|_Q^2, \quad \forall v^* \in \mathcal{V}^*, \quad (4.14)
\]

where \( M, H, \) and \( Q \) are defined in (3.1), (3.2) and (3.8), respectively.

**Proof** First, for all \( v^* \in \mathcal{V}^* \), it follows from (4.13) that

\[
(v^k - v^*)^T M (v^k - \tilde{v}^k) \geq (v^k - \tilde{v}^k)^T M (v^k - \tilde{v}^k) + (\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{j=2}^{m} A_j (x^k_j - \tilde{x}^k_j) \right). \quad (4.15)
\]

Now, we treat the first term of (4.15). Using the matrix \( M \) (see (3.1)), we have
\[(v^k - \tilde{v}^k)^T M (v^k - \tilde{v}^k) = \left( \begin{array}{c}
  x_2^k - \tilde{x}_2^k \\
  \vdots \\
  x_m^k - \tilde{x}_m^k \\
  \lambda^k - \tilde{\lambda}^k 
\end{array} \right)^T \left( \begin{array}{cccc}
  \beta A_2^T A_2 & 0 & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  \beta A_m^T A_2 & \cdots & \beta A_m^T A_m & 0 \\
  0 & \cdots & 0 & \frac{1}{\beta} I_l 
\end{array} \right) \left( \begin{array}{c}
  x_2^k - \tilde{x}_2^k \\
  \vdots \\
  x_m^k - \tilde{x}_m^k \\
  \lambda^k - \tilde{\lambda}^k 
\end{array} \right) \right)^T \left( \begin{array}{c}
  x_2^k - \tilde{x}_2^k \\
  \vdots \\
  x_m^k - \tilde{x}_m^k \\
  \lambda^k - \tilde{\lambda}^k 
\end{array} \right) \right)
\]

Let us deal with the second term of the right-hand side of (4.15). By manipulations, we have

\[(\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{j=2}^{m} A_j (x_j^k - \tilde{x}_j^k) \right) = \left( \begin{array}{c}
  x_2^k - \tilde{x}_2^k \\
  \vdots \\
  x_m^k - \tilde{x}_m^k \\
  \lambda^k - \tilde{\lambda}^k 
\end{array} \right)^T \left( \begin{array}{cccc}
  0 & \cdots & 0 & 0 \\
  \vdots & \ddots & \vdots & \vdots \\
  0 & \cdots & 0 & 0 \\
  A_2 & \cdots & A_m & 0 
\end{array} \right) \left( \begin{array}{c}
  x_2^k - \tilde{x}_2^k \\
  \vdots \\
  x_m^k - \tilde{x}_m^k \\
  \lambda^k - \tilde{\lambda}^k 
\end{array} \right) \right)^T \left( \begin{array}{c}
  x_2^k - \tilde{x}_2^k \\
  \vdots \\
  x_m^k - \tilde{x}_m^k \\
  \lambda^k - \tilde{\lambda}^k 
\end{array} \right) \right)
\]
Adding (4.16) and (4.17) together, it follows that

\[(v^k - \tilde{v}^k)^T M (v^k - \tilde{v}^k) + (\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{j=2}^{m} A_j(x_j^k - \tilde{x}_j^k) \right)\]

\[
\begin{pmatrix}
x_2^k - \tilde{x}_2^k \\
\vdots \\
x_m^k - \tilde{x}_m^k \\
\lambda^k - \tilde{\lambda}^k
\end{pmatrix}
\begin{pmatrix}
\beta A_2^T A_2 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\beta A_m^T A_2 & \cdots & \beta A_m^T A_m & 0 \\
A_2 & \cdots & A_m & \frac{1}{\beta} I_l
\end{pmatrix}
\begin{pmatrix}
x_2^k - \tilde{x}_2^k \\
\vdots \\
x_m^k - \tilde{x}_m^k \\
\lambda^k - \tilde{\lambda}^k
\end{pmatrix}
\]

\[
= \frac{1}{2}
\begin{pmatrix}
x_2^k - \tilde{x}_2^k \\
\vdots \\
x_m^k - \tilde{x}_m^k \\
\lambda^k - \tilde{\lambda}^k
\end{pmatrix}
\begin{pmatrix}
2\beta A_2^T A_2 & \cdots & \beta A_2^T A_m & A_2^T \\
\vdots & \ddots & \vdots & \vdots \\
\beta A_m^T A_2 & \cdots & 2\beta A_m^T A_m & A_m^T \\
A_2 & \cdots & A_m & \frac{2}{\beta} I_l
\end{pmatrix}
\begin{pmatrix}
x_2^k - \tilde{x}_2^k \\
\vdots \\
x_m^k - \tilde{x}_m^k \\
\lambda^k - \tilde{\lambda}^k
\end{pmatrix}
\]
Use the notation of the matrices $H$ and $Q$ to the right-hand side of the last equality, we obtain

$$(v^k - \tilde{v}^k)^T M (v^k - \tilde{v}^k) + (\lambda^k - \tilde{\lambda}^k)^T \left( \sum_{j=2}^{m} A_j (x^k_j - \tilde{x}^k_j) \right) = \frac{1}{2} \|v^k - \tilde{v}^k\|_H^2 + \frac{1}{2} \|v^k - \tilde{v}^k\|_Q^2.$$ 

Substituting the last equality in (4.15), this theorem is proved. $\square$

It follows from (4.14) that

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

In other words, by setting

$$G = M H^{-1} M^T,$$

(4.18)

$MH^{-1} M^T (v^k - v^*)$ is the gradient of the distance function $\frac{1}{2} \|v - v^*\|^2_G$, and $M^{-T} H (\tilde{v}^k - v^k)$ is a descent direction of $\frac{1}{2} \|v - v^*\|^2_G$ at the current point $v^k$ whenever $\tilde{v}^k \neq v^k$. 
4.2 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 $V^*$. Note that we follow the definition of contractive type methods. With this contractive property, the convergence of the proposed ADM with Gaussian back substitution can be easily derived with subroutine analysis.

**Theorem 4.2** 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 (3.4a) from the given vector $v^k = (x_2^k, \ldots, x_m^k, \lambda^k)$. Let the matrix $G$ be given by (4.18). For the new iterate $v^{k+1}$ produced by the Gaussian back substitution (3.5), there exists a constant $c_0 > 0$ such that

$$\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 - c_0 (\|v^k - \tilde{v}^k\|_H^2 + \|v^k - \tilde{v}^k\|_Q^2), \quad \forall \; v^* \in V^*,$$

(4.19)

where $H$ and $Q$ are defined in (3.2) and (3.8), respectively.
Proof For $G = MH^{-1}M^T$ and any $\alpha \geq 0$, we obtain

$$
\|v^k - v^*\|^2_G - \|v^{k+1} - v^*\|^2_G \\
= \|v^k - v^*\|^2_G - \|(v^k - v^*) - \alpha M^{-T}H(v^k - \tilde{v}^k)\|^2_G \\
= 2\alpha (v^k - v^*)^T M(v^k - \tilde{v}^k) - \alpha^2 \|v^k - \tilde{v}^k\|^2_H.
$$

(4.20)

Substituting the result of Theorem 4.1 into the right-hand side of the last equation, we get

$$
\|v^k - v^*\|^2_G - \|v^{k+1} - v^*\|^2_G \\
\geq \alpha(\|v^k - \tilde{v}^k\|^2_H + \|v^k - \tilde{v}^k\|^2_Q) - \alpha^2 \|v^k - \tilde{v}^k\|^2_H \\
= \alpha(1 - \alpha)\|v^k - \tilde{v}^k\|^2_H + \alpha\|v^k - \tilde{v}^k\|^2_Q,
$$

and thus

$$
\|v^{k+1} - v^*\|^2_G \leq \|v^k - v^*\|^2_G \\
- \alpha((1 - \alpha)\|v^k - \tilde{v}^k\|^2_H + \|v^k - \tilde{v}^k\|^2_Q), \quad \forall v^* \in V^*.
$$

(4.21)

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

Corollary 4.1 The assertion of Theorem 4.2 also holds if the Gaussian back substitution update form is (3.6) with the calculated step length by (3.7).
Proof Analogous to the proof of Theorem 4.2, we have that
\[
\|v^k - v^*\|_G^2 - \|v^{k+1} - v^*\|_G^2 \\
\geq 2\gamma \alpha_k^*(v^k - v^*)^T M(v^k - \tilde{v}^k) - (\gamma \alpha_k^*)^2 \|v^k - \tilde{v}^k\|_H^2,
\]
(4.22)

where \(\alpha_k^*\) is given by (3.7). According to (3.7), we have that
\[
\alpha_k^*(\|v^k - \tilde{v}^k\|_H^2) = \frac{1}{2} (\|v^k - \tilde{v}^k\|_H^2 + \|v^k - \tilde{v}^k\|_Q^2).
\]

Then, it follows from the above equality and (4.14) that
\[
\|v^k - v^*\|_G^2 - \|v^{k+1} - v^*\|_G^2 \\
\geq \gamma \alpha_k^*(\|v^k - \tilde{v}^k\|_H^2 + \|v^k - \tilde{v}^k\|_Q^2) - \frac{1}{2} \gamma^2 \alpha_k^*(\|v^k - \tilde{v}^k\|_H^2 + \|v^k - \tilde{v}^k\|_Q^2) \\
= \frac{1}{2} \gamma (2 - \gamma) \alpha_k^*(\|v^k - \tilde{v}^k\|_H^2 + \|v^k - \tilde{v}^k\|_Q^2).
\]

Because \(\alpha_k^* \geq \frac{1}{2}\), it follows from the last inequality that
\[
\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 \\
- \frac{1}{4} \gamma (2 - \gamma) (\|v^k - \tilde{v}^k\|_H^2 + \|v^k - \tilde{v}^k\|_Q^2), \quad \forall v^* \in V^*.
\]
(4.23)
Since $\gamma \in (0, 2)$, the assertion of this corollary follows from (4.23) directly. □

### 4.3 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 4.3** Let $\{v^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the proposed ADM with Gaussian back substitution. Then we have

1. The sequence $\{v^k\}$ is bounded.
2. $\lim_{k \to \infty} \|A_i(x^k - \tilde{x}_i^k)\| = 0$, $i = 2, \ldots, m$, and $\lim_{k \to \infty} \|\lambda^k - \tilde{\lambda}^k\| = 0$.
3. Any cluster point of $\{\tilde{w}^k\}$ is a solution point of (2.3).
4. The sequence $\{\tilde{v}^k\}$ converges to some $v^\infty \in \mathcal{V}^*$.

**Proof.** The first assertion follows from (4.19) directly. From (4.19) we get

$$
\sum_{k=0}^{\infty} c_0 \|v^k - \tilde{v}^k\|_H^2 \leq \|v^0 - v^*\|_G^2
$$
and thus we get $\lim_{k \to \infty} \|v^k - \tilde{v}^k\|_H^2 = 0$, and consequently
\[
\lim_{k \to \infty} \|A_i(x^k - \tilde{x}_i^k)\| = 0, \quad i = 2, \ldots, m, \tag{4.24}
\]
and
\[
\lim_{k \to \infty} \|\lambda^k - \tilde{\lambda}^k\| = 0. \tag{4.25}
\]
The second assertion is proved.

Substituting (4.24) into (4.5), for $i = 1, 2, \ldots, m$, we have
\[
\tilde{x}_i^k \in \mathcal{X}_i, \quad \lim_{k \to \infty} (x_i - \tilde{x}_i^k)^T \left\{ f_i(\tilde{x}_i^k) - A_i^T \tilde{\lambda}^k \right\} \geq 0, \quad \forall \ x_i \in \mathcal{X}_i. \tag{4.26}
\]
It follows from (3.4a) and (4.25) that
\[
\lim_{k \to \infty} \left( \sum_{j=1}^{m} A_j \tilde{x}_j^k - b \right) = 0. \tag{4.27}
\]
Combining (4.26) and (4.27) we get
\[
\tilde{w}^k \in \mathcal{W}, \quad \lim_{k \to \infty} (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq 0, \quad \forall \ w \in \mathcal{W}, \tag{4.28}
\]
and thus any cluster point of \( \{ \tilde{w}^k \} \) is a solution point of (2.3). The third assertion is proved.

It follows from the first assertion and \( \lim_{k \to \infty} \| v^k - \tilde{v}^k \|^2_H = 0 \) that \( \{ \tilde{v}^k \} \) is also bounded. Let \( v^\infty \) be a cluster point of \( \{ \tilde{v}^k \} \) and the subsequence \( \{ \tilde{v}^{k_j} \} \) converges to \( v^\infty \). It follows from (4.28) that

\[
\tilde{w}^{k_j} \in \mathcal{W}, \quad \lim_{k \to \infty} (w - \tilde{w}^{k_j})^T F(\tilde{w}^{k_j}) \geq 0, \quad \forall w \in \mathcal{W} \tag{4.29}
\]

and consequently

\[
\begin{align*}
(x_i - x_i^\infty)^T \{ f_i(x_i^\infty) - A_i^T \lambda^\infty \} & \geq 0, \quad \forall x_i \in \mathcal{X}_i, \ i = 1, \ldots, m, \\
\sum_{j=1}^m A_j x_j^\infty - b & = 0.
\end{align*} \tag{4.30}
\]

This means that \( v^\infty \in \mathcal{V}^* \). Since \( \{ v^k \} \) is Fejér monotone and \( \lim_{k \to \infty} \| v^k - \tilde{v}^k \| = 0 \), the sequence \( \{ \tilde{v}^k \} \) cannot have other cluster point and \( \{ \tilde{v}^k \} \) converges to \( v^\infty \in \mathcal{V}^* \). \( \square \)

If we take \( \alpha \equiv 1 \) in the correction form (3.5), similarly as in the last lecture, the resulting method is convergent in the ergodic sense with the convergence rate \( O(1/t) \).
References


