凸优化和单调变分不等式的收缩算法

第十八讲: 多块可分离凸优化问题
带回代的交替方向收缩算法

Alternating direction method of multipliers with back substitution for convex optimization containing more separable blocks

南京大学数学系　　何炳生

hebma@nju.edu.cn

The context of this lecture is based on the publication [8]

1 Introduction

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

$$\begin{array}{ll} \min & \theta_1(x_1) + \theta_2(x_2) \\ \text{s. t.} & A_1x_1 + A_2x_2 = b, \\ & x_1 \in \mathcal{X}_1 \quad \text{and} \quad x_2 \in \mathcal{X}_2, \end{array} \tag{1.1}$$

where $\theta_1 : \Re^{n_1} \to \Re$ and $\theta_2 : \Re^{n_2} \to \Re$ are closed proper convex functions (not necessarily smooth); $\mathcal{X}_1 \subset \Re^{n_1}$ and $\mathcal{X}_2 \subset \Re^{n_2}$ are closed convex sets; $A_1 \in \Re^{l\times n_1}$ and $A_2 \in \Re^{l\times n_2}$ are given matrices; and $b \in \Re^l$ is a given vector.

The augmented Lagrangian function of the problem (1.1) is

$$\mathcal{L}_\beta(x_1, x_2, \lambda) = \theta_1(x_1)+\theta_2(x_2)-\lambda^T(A_1x_1+A_2x_2-b)+\frac{\beta}{2}\|A_1x_1+A_2x_2-b\|^2,$$

where $\lambda^k \in \Re^l$ is the Lagrange multiplier associated with the linear constraint and $\beta > 0$ is the penalty parameter for the violation of the linear constraint.

The iterative scheme of ADMM for solving (1.1) is as follows:

$$\left\{\begin{array}{l} x_1^{k+1} \in \arg\min\{\mathcal{L}_\beta(x_1, x_2^k, \lambda^k)\} \\ \qquad = \arg\min\{\theta_1(x_1) + \frac{\beta}{2}\|(A_1x_1 + A_2x_2^k - b) - \frac{1}{\beta}\lambda^k\|^2 \;\big|\; x_1 \in \mathcal{X}_1\}, \\ x_2^{k+1} \in \arg\min\{\mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^k)\} \\ \qquad = \arg\min\{\theta_2(x_2) + \frac{\beta}{2}\|(A_1x_1^{k+1} + A_2x_2 - b) - \frac{1}{\beta}\lambda^k\|^2 \;\big|\; x_2 \in \mathcal{X}_2\}, \\ \lambda^{k+1} = \lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} - b), \end{array}\right. \tag{1.2}$$

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

$$\begin{array}{ll} \min & \sum_{i=1}^m \theta_i(x_i) \\ & \sum_{i=1}^m A_ix_i = b; \\ & x_i \in \mathcal{X}_i, \quad i = 1, \cdots, m; \end{array} \tag{1.3}$$

where $\theta_i : \Re^{n_i} \to \Re \ (i = 1, \ldots, m)$ 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.

Because of the efficiency of ADMM for (1.1), a natural idea for solving (1.3) is to extend the ADMM (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 extended ADMM is not necessarily convergent [2].

In this paper, we provide a novel approach towards the extension of ADMM for the problem (1.3). More specifically, we show that if a new iterate is generated by correcting the output of the ADMM 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 ADMM for (1.3). The resulting method is called the ADMM with Gaussian back substitution from now on. Alternatively, the ADMM with Gaussian back substitution can be regarded as a prediction-correction type method whose predictor is generated by the ADMM procedure and the correction is completed by a Gaussian back substitution procedure. We prove the convergence of the ADMM with Gaussian back substitution under the analytic framework of contractive type methods

Throughout, we assume that the solution set of (1.3) is nonempty [a].

---

[a] In our published paper [8], we assume that $A_i^T A_i$ is nonsingular, this assumption is not necessary.

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 $\lambda \in \Re^l$ to the linear constraint, the Lagrange function of (1.3) is:

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

which is defined on

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

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

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

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

$w^* = (x_1^*, x_2^*, ..., x_m^*, \lambda^*) \in \Omega$, such that

$$\left\{ \begin{array}{l} \theta_1(x_1) - \theta_1(x_1^*) + (x_1 - x_1^*)^T \{-A_1^T \lambda^*\} \geq 0, \\ \theta_2(x_2) - \theta_2(x_2^*) + (x_2 - x_2^*)^T \{-A_2^T \lambda^*\} \geq 0, \\ \qquad\qquad\qquad \vdots \\ \qquad\qquad\qquad \vdots \\ \theta_m(x_m) - \theta_m(x_m^*) + (x_m - x_m^*)^T \{-A_m^T \lambda^*\} \geq 0, \\ (\lambda - \lambda^*)^T (\sum_{i=1}^m A_i x_i^* - b) \geq 0, \end{array} \right. \tag{2.2}$$

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

$$\theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{2.3a}$$

where

$$\theta(x) = \sum_{i=1}^m \theta_i(x_i),$$

$$w = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ \lambda \end{pmatrix} \quad \text{and} \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ -A_2^T\lambda \\ \vdots \\ -A_m^T\lambda \\ \sum_{i=1}^m A_i x_i - b \end{pmatrix}. \tag{2.3b}$$

Note that the operator $F(w)$ defined in (2.3b) is monotone. Especially, we have

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \equiv 0. \tag{2.4}$$

the fact that θ_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 Ω^*, is also nonempty. In addition to the notation of $w = (x_1, x_2, \cdots, x_m, \lambda)$, for any integer number k we also use the following notation:

$$v = (x_2, \cdots, x_m, \lambda).$$

Moreover, we define

$$\mathcal{V}^* = \{(x_2^*, \ldots, x_m^*, \lambda^*) \,|\, (x_1^*, x_2^*, \ldots, x_m^*, \lambda^*) \in \Omega^*\}.$$

3 ADMM with Gaussian back substitution

In this section, we show the combination of the extended ADMM scheme (1.2) with a Gaussian back substitution procedure, and derive the resulting ADMM 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:

$$L = \begin{pmatrix} I_l & 0 & \cdots & \cdots & \cdots & 0 \\ I_l & I_l & \ddots & & & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ I_l & I_l & \cdots & \cdots & I_l & 0 \\ 0 & 0 & \cdots & \cdots & 0 & I_l \end{pmatrix}_{(m+1)\text{Blocks}}, \tag{3.1}$$

and

$$P = \text{diag}\big(\sqrt{\beta}A_2, \sqrt{\beta}A_3, \ldots, \sqrt{\beta}A_m, \sqrt{\frac{1}{\beta}}I_l\big), \qquad H = P^T P. \tag{3.2}$$

Note that for any $\beta > 0$, from Pv^k, we get $(A_2 x_2^k, \cdots, A_m x_m^k, \lambda^k)$ very easily. The matrix L 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:

$$L^{-T} = \begin{pmatrix} I_l & -I_l & 0 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \ddots & \ddots & -I_l & & 0 \\ 0 & \cdots & 0 & I_l & -I_l & 0 \\ 0 & \cdots & 0 & 0 & I_l & 0 \\ 0 & \cdots & 0 & 0 & 0 & I_l \end{pmatrix}_{(m+1)\text{Blocks}} . \tag{3.3}$$

which is a upper-triangular block matrix whose diagonal components are identity matrices.

Algorithm: ADMM with Gaussian back substitution for (1.3):

Let $\beta > 0$ and $\alpha \in [0.5, 1)$, With the given iterate $Pv^k = P(x_2^k, \cdots, x_m^k, \lambda^k)$.

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

$$\min\Big\{\theta_i(x) + \frac{\beta}{2}\|(\textstyle\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + A_i x_i + \sum_{j=i+1}^{m} A_j x_j^k - b) - \frac{1}{\beta}\lambda^k\|^2 \mid x_i \in \mathcal{X}_i\Big\} \tag{3.4a}$$

and set

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

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

$$L^T P(v^{k+1} - v^k) = \alpha P\big(\tilde{v}^k - v^k\big). \tag{3.4c}$$

where the matrices L and P are defined by (3.1) and (3.2), respectively.

Note that the k-th iteration starts with a given $(A_2x_2^k, \cdots, A_mx_m^k, \lambda^k)$, and offers the new iterate $(A_2x_2^{k+1}, \cdots, A_mx_m^{k+1}, \lambda^{k+1})$. Only for analysis convenience, we use Pv^k and Pv^{k+1}. In any way, the variable x_1 is only an intermediate variable.

Recall that the matrix L^T in (3.4c) is a upper-triangular block matrix. Thus, we call the correction step (3.4c) is a Gaussian back substitution, and it is very easy to execute. In fact, as we mentioned, after the predictor is generated by the ADMM 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 ADMM with Gaussian back substitution is dominated by the ADMM 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

$$Pv^{k+1} = Pv^k - \alpha L^{-T} P(v^k - \tilde{v}^k). \tag{3.5}$$

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 α can be modified accordingly into the following variant with a dynamical step size:

$$Pv^{k+1} = Pv^k - \gamma \alpha_k^* L^{-T} P(v^k - \tilde{v}^k). \tag{3.6}$$

where

$$\alpha_k^* = \frac{\|v^k - \tilde{v}^k\|_H^2 + \|v^k - \tilde{v}^k\|_Q^2}{2\|v^k - \tilde{v}^k\|_H^2}; \tag{3.7}$$

$$Q = \begin{pmatrix} \beta A_2^T A_2 & \beta A_2^T A_3 & \cdots & \beta A_2^T A_m & A_2^T \\ \beta A_3^T A_2 & \beta A_3^T A_3 & \cdots & \beta A_3^T A_m & A_3^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta A_m^T A_2 & \beta A_m^T A_3 & \cdots & \beta A_m^T A_m & A_m^T \\ A_2 & A_3 & \cdots & A_m & \frac{1}{\beta} I_l \end{pmatrix}; \tag{3.8}$$

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 ADMM procedure (3.4a), we have that

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

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 α_k^* defined in (3.7) satisfies $\frac{1}{2} \le \alpha_k^* \le \frac{m+1}{2}$.

4 Convergence of the proposed method

In this section, we prove the convergence of the proposed ADMM 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 the sequence generated by the proposed ADMM with Gaussian back substitution is contractive with respect to $\mathcal{V}^*$.

2.) Derive the convergence based on the Fej$\acute{e}$r monotonicity of the sequence generated by the proposed ADMM 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

For this purpose, we first prove two lemmas.

Lemma 4.1 *Let* $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k)$ *be generated by the ADMM step* (3.4a) *from the given vector* $Pv^k = P(x_2^k, \ldots, x_m^k, \lambda^k)$*. Then, we have*

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad & \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) + (x - \tilde{x}^k)^T d_2(x^k, \tilde{x}^k) \\ \geq \quad & (v - \tilde{v}^k)^T d_1(v^k, \tilde{v}^k), \quad \forall\, w \in \Omega, \end{aligned} \tag{4.1}$$

where

$$\begin{aligned} d_1(v^k, \tilde{v}^k) \quad &= \begin{pmatrix} \beta A_2^T A_2 & 0 & \cdots & \cdots & 0 \\ \beta A_3^T A_2 & \beta A_3^T A_3 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \beta A_m^T A_2 & \beta A_m^T A_3 & \cdots & \beta A_m^T A_m & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_l \end{pmatrix} \begin{pmatrix} x_2^k - \tilde{x}_2^k \\ x_3^k - \tilde{x}_3^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}, \\ &= \quad P^T L P (v^k - \tilde{v}^k). \end{aligned} \tag{4.2}$$

and

$$d_2(x^k, \tilde{x}^k) = \beta \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \end{pmatrix} \big(\sum_{j=2}^{m} A_j(x_j^k - \tilde{x}_j^k)\big). \tag{4.3}$$

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

$$\begin{aligned} \tilde{x}_i^k \in \mathcal{X}_i, \quad & \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \big\{-A_i^T \lambda^k \\ & +\beta A_i^T (\textstyle\sum_{j=1}^{i} A\tilde{x}_j^k + \sum_{j=i+1}^{m} A_j x_j^k - b)\big\} \geq 0, \quad \forall\, x_i \in \mathcal{X}_i. \end{aligned} \tag{4.4}$$

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

$$\tilde{x}_i^k \in \mathcal{X}_i, \quad \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \big\{ -A_i^T \tilde{\lambda}^k + \beta A_i^T \big(\textstyle\sum_{j=i+1}^m A_j (x_j^k - \tilde{x}_j^k) \big) \big\} \ge 0, \;\; \forall \, x_i \in \mathcal{X}_i. \tag{4.5}$$

It follows from (4.5) that $\tilde{x}^k \in \mathcal{X}$ and

$$\tilde{x}^k \in \mathcal{X}, \quad \theta(x) - \theta(\tilde{x}^k) +$$

$$\begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ x_3 - \tilde{x}_3^k \\ \vdots \\ x_{m-1} - \tilde{x}_{m-1} \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A_1^T \tilde{\lambda}^k \\ -A_2^T \tilde{\lambda}^k \\ -A_3^T \tilde{\lambda}^k \\ \vdots \\ -A_{m-1}^T \tilde{\lambda}^k \\ -A_m^T \tilde{\lambda}^k \end{pmatrix} + \beta \begin{pmatrix} A_1^T \big(\sum_{j=2}^m A_j (x_j^k - \tilde{x}_j^k) \big) \\ A_2^T \big(\sum_{j=3}^m A_j (x_j^k - \tilde{x}_j^k) \big) \\ A_3^T \big(\sum_{j=4}^m A_j (x_j^k - \tilde{x}_j^k) \big) \\ \vdots \\ A_{m-1}^T \big(A_m (x_m^k - \tilde{x}_m^k) \big) \\ 0 \end{pmatrix} \right\} \ge 0, \tag{4.6}$$

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

$$\begin{pmatrix} x_2 - \tilde{x}_2^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \beta \begin{pmatrix} A_2^T\big(A_2(x_2^k - \tilde{x}_2^k)\big) \\ A_3^T\big(\sum_{j=2}^3 A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ A_m^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \end{pmatrix}$$

to the both sides of (4.6), we get $\tilde{x}^k \in \mathcal{X}$ and

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

$$\geq \begin{pmatrix} x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \begin{pmatrix} \beta A_2^T\big(\sum_{j=2}^2 A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ \beta A_m^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \end{pmatrix}, \quad \forall\, x \in \mathcal{X}. \qquad (4.7)$$

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

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

Adding (4.7) and the last equality together, we get $\tilde{w}^k \in \mathcal{W}$ and

$$\theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} -A_1^T\tilde{\lambda}^k + \beta A_1^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ -A_2^T\tilde{\lambda}^k + \beta A_2^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ -A_m^T\tilde{\lambda}^k + \beta A_m^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ \sum_{i=1}^m A_i\tilde{x}_i^k - b \end{pmatrix}$$

$$\geq \quad \begin{pmatrix} x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} \beta A_2^T\big(\sum_{j=2}^2 A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ \beta A_m^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}, \quad \forall\, w \in \Omega.$$

Use the notations of $F(w)$, $d_1(v^k, \tilde{v}^k)$ and $d_2(x^k, \tilde{x}^k)$, the left hand side and the right hand side of the above inequality can be written as

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) + (x - \tilde{x}^k)^T d_2(x^k, \tilde{x}^k)$$

and

$$(v - \tilde{v}^k)^T d_1(v^k, \tilde{v}^k),$$

respectively. Thus, the assertion of this lemma is proved. $\square$

Note that the $d_1(v^k, \tilde{v}^k)$ depends only on v^k and $\tilde{v}^k$, while $d_2(x^k, \tilde{x}^k)$ is determined by both x^k and $\tilde{x}^k$.

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 ADMM step (3.4a) from the given vector $Pv^k = P(x_2^k, \ldots, x_m^k, \lambda^k)$. Then, we have*

$$(\tilde{v}^k - v^*)^T d_1(v^k, \tilde{v}^k) \ge (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=2}^{m} A_j(x_j^k - \tilde{x}_j^k)\Big), \quad \forall\, v^* \in \mathcal{V}^*, \tag{4.8}$$

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

Proof. Since $w^* \in \Omega$, it follows from (4.1) that

$$\begin{aligned}&(\tilde{v}^k - v^*)^T d_1(v^k, \tilde{v}^k) \\ &\quad \geq \quad \theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}) + (\tilde{x}^k - x^*)^T d_2(x^k, \tilde{x}^k). \end{aligned} \tag{4.9}$$

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

$$\begin{aligned}&\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}) + (\tilde{x}^k - x^*)^T d_2(x^k, \tilde{x}^k) \\ &\quad = \quad \theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \\ &\qquad\quad +\Big(\sum_{j=2}^{m} A_j(x_j^k - \tilde{x}_j^k)\Big)^T \beta\Big(\sum_{j=1}^{m} A_j(\tilde{x}_j^k - x_j^*)\Big). \end{aligned} \tag{4.10}$$

Then, we look at the right-hand side of (4.10). First, by using the skew-symmetry of F (2.4) and the optimality of w^*, we have

$$\begin{aligned}&\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \\ &\quad = \quad \theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(w^*) \\ &\quad \geq \quad 0.\end{aligned}$$

For the last part of the right hand side of (4.10), because

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

it follows

$$\Big(\sum_{j=2}^{m} A_j (x_j^k - \tilde{x}_j^k)\Big)^T \beta\Big(\sum_{j=1}^{m} A_j (\tilde{x}_j^k - x_j^*)\Big) = \Big(\sum_{j=2}^{m} A_j (x_j^k - \tilde{x}_j^k)\Big)^T (\lambda^k - \tilde{\lambda}^k).$$

Finally, from (4.10) that

$$\begin{aligned} &\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}) + (\tilde{x}^k - x^*)^T d_2(x^k, \tilde{x}^k) \\ \geq \quad & (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=2}^{m} A_j (x_j^k - \tilde{x}_j^k)\Big). \end{aligned} \tag{4.11}$$

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

Since (see (3.1) and (4.2))

$$d_1(v^k, \tilde{v}^k) = P^T L P (v^k - \tilde{v}^k), \tag{4.12}$$

consequently from (4.8) follows that

$$(\tilde{v}^k - v^*)^T P^T LP(v^k - \tilde{v}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=2}^{m} A_j (x_j^k - \tilde{x}_j^k) \Big), \ \forall\, v^* \in \mathcal{V}^*. \quad (4.13)$$

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 ADMM step* (3.4a) *from the given vector* $Pv^k = P(x_2^k, \ldots, x_m^k, \lambda^k)$. *Then, we have*

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

where L, P, 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

$$\begin{aligned} &(v^k - v^*)^T P^T LP(v^k - \tilde{v}^k) \\ &\quad \geq \ (v^k - \tilde{v}^k)^T P^T LP(v^k - \tilde{v}^k) + (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=2}^{m} A_j (x_j^k - \tilde{x}_j^k) \Big). \end{aligned} \quad (4.15)$$

Now, we treat the first term of the right hand side of (4.15). Using the matrix L and P (see

(3.1) and (3.2)) , we have

$$
\begin{aligned}
&(v^k-\tilde v^k)^T P^T L P (v^k-\tilde v^k)\\
&= \begin{pmatrix} x_2^k-\tilde x_2^k\\ \vdots\\ x_m^k-\tilde x_m^k\\ \lambda^k-\tilde\lambda^k \end{pmatrix}
\begin{pmatrix} \beta A_2^T A_2 & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \vdots\\ \beta A_m^T A_2 & \cdots & \beta A_m^T A_m & 0\\ 0 & \cdots & 0 & \frac{1}{\beta} I_l \end{pmatrix}^T
\begin{pmatrix} x_2^k-\tilde x_2^k\\ \vdots\\ x_m^k-\tilde x_m^k\\ \lambda^k-\tilde\lambda^k \end{pmatrix}
\end{aligned}
\tag{4.16}
$$

Let us deal with the second term of the right-hand side of (4.15). We have

$$
\begin{aligned}
&(\lambda^k-\tilde\lambda^k)^T\Big(\sum_{j=2}^m A_j(x_j^k-\tilde x_j^k)\Big)\\
&= \begin{pmatrix} x_2^k-\tilde x_2^k\\ \vdots\\ x_m^k-\tilde x_m^k\\ \lambda^k-\tilde\lambda^k \end{pmatrix}^T
\begin{pmatrix} 0 & \dots & 0 & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \dots & 0 & 0\\ A_2 & \dots & A_m & 0 \end{pmatrix}
\begin{pmatrix} x_2^k-\tilde x_2^k\\ \vdots\\ x_m^k-\tilde x_m^k\\ \lambda^k-\tilde\lambda^k \end{pmatrix}.
\end{aligned}
\tag{4.17}
$$

Adding (4.16) and (4.17) together, it follows that

$$
\begin{aligned}
&(v^k-\tilde v^k)^T P^T L P(v^k-\tilde v^k) + (\lambda^k-\tilde\lambda^k)^T\Big(\sum_{j=2}^m A_j(x_j^k-\tilde x_j^k)\Big)\\
= \quad & \begin{pmatrix} x_2^k-\tilde x_2^k \\ \vdots \\ x_m^k-\tilde x_m^k \\ \lambda^k-\tilde\lambda^k \end{pmatrix} \begin{pmatrix} \beta A_2^T A_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \beta A_m^T A_2 & \cdots & \beta A_m^T A_m & 0 \\ A_2 & \cdots & A_m & \frac{1}{\beta} I_l \end{pmatrix}^T \begin{pmatrix} x_2^k-\tilde x_2^k \\ \vdots \\ x_m^k-\tilde x_m^k \\ \lambda^k-\tilde\lambda^k \end{pmatrix}\\
= \quad & \frac{1}{2}\begin{pmatrix} x_2^k-\tilde x_2^k \\ \vdots \\ x_m^k-\tilde x_m^k \\ \lambda^k-\tilde\lambda^k \end{pmatrix}^T \begin{pmatrix} 2\beta A_2^T A_2 & \cdots & \beta A_2^T A_m & A_2^T \\ \vdots & & \vdots & \vdots \\ \beta A_m^T A_2 & \cdots & 2\beta A_m^T A_m & A_m^T \\ A_2 & \cdots & A_m & \frac{2}{\beta} I_l \end{pmatrix} \begin{pmatrix} x_2^k-\tilde x_2^k \\ \vdots \\ x_m^k-\tilde x_m^k \\ \lambda^k-\tilde\lambda^k \end{pmatrix}.
\end{aligned}
$$

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

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

Substituting the last equality in (4.15), this theorem is proved. □

It follows from (4.14) that

$$
\big\langle L^T P(v^k-v^*), P(v^k-\tilde{v}^k)\big\rangle \ge \frac{1}{2}\|v^k-\tilde{v}^k\|_{(H+Q)}^2. \tag{4.18}
$$

Based on (4.18), in order to minimize $\|L^T P(v-v^*)\|^2$, we need only to let

$$
L^T P(v^{k+1}-v^*) = L^T P(v^k-v^*) - \alpha P(v^k-\tilde{v}^k) \tag{4.19}
$$

and by selecting a suitable step-size α.

$$
Pv^{k+1} = Pv^k - \alpha L^{-T} P(v^k-\tilde{v}^k).
$$

By setting

$$
G = P^T L L^T P, \tag{4.20}
$$

$\frac{1}{2}\|v - v^*\|_G^2$, and $L^{-T}P(\tilde{v}^k - v^k)$ is a descent direction of $\frac{1}{2}\|v - v^*\|_G^2$ 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 ADMM with Gaussian back substitution is contractive with respect to the set $\mathcal{V}^*$. Note that we follow the definition of contractive type methods. With this contractive property, the convergence of the proposed ADMM 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 ADMM step* (3.4a) *from the given vector* $Pv^k = P(x_2^k, \ldots, x_m^k, \lambda^k)$*. Let the matrix* G *be given by* (4.20)*. For the new iterate* v^{k+1} *produced by the Gaussian back substitution* (3.5)*, there exists a constant* $c_0 > 0$ *such that*

$$\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 - c_0\big(\|v^k - \tilde{v}^k\|_H^2 + \|v^k - \tilde{v}^k\|_Q^2\big), \quad \forall\, v^* \in \mathcal{V}^*, \tag{4.21}$$

where H *and* Q *are defined in* (3.2) *and* (3.8)*, respectively.*

Proof By using (3.5) and $G = P^T LL^T P$, we obtain

$$
\begin{aligned}
&\|v^k - v^*\|_G^2 - \|v^{k+1} - v^*\|_G^2 \\
&\quad = \quad \|P(v^k - v^*)\|_{(LL^T)}^2 - \|P(v^{k+1} - v^*)\|_{(LL^T)}^2 \\
&\quad = \quad \|P(v^k - v^*)\|_{(LL^T)}^2 - \|P(v^k - v^*) - \alpha L^{-T} P(v^k - \tilde{v}^k)\|_{(LL^T)}^2 \\
&\quad = \quad 2\alpha(v^k - v^*)^T P^T LP(v^k - \tilde{v}^k) - \alpha^2 \|v^k - \tilde{v}^k\|_H^2. \qquad (4.22)
\end{aligned}
$$

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

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

and thus

$$
\begin{aligned}
&\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 \\
&\qquad -\alpha\big((1-\alpha)\|v^k - \tilde{v}^k\|_H^2 + \|v^k - \tilde{v}^k\|_Q^2\big), \quad \forall\, v^* \in \mathcal{V}^*. \qquad (4.23)
\end{aligned}
$$

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

$$\begin{aligned} &\|v^k - v^*\|_G^2 - \|v^{k+1} - v^*\|_G^2 \\ &\quad \geq \quad 2\gamma\alpha_k^*(v^k - v^*)^T P^T L P(v^k - \tilde{v}^k) - (\gamma\alpha_k^*)^2 \|v^k - \tilde{v}^k\|_H^2, \end{aligned} \tag{4.24}$$

where α_k^* is given by (3.7). According to (3.7), we have that

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

Then, it follows from the above equality and (4.14) that

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

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

$$\begin{aligned} &\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 \\ &\qquad - \tfrac{1}{4}\gamma(2-\gamma)\big(\|v^k - \tilde{v}^k\|_H^2 + \|v^k - \tilde{v}^k\|_Q^2\big), \quad \forall\, v^* \in \mathcal{V}^*. \end{aligned} \tag{4.25}$$

Since $\gamma \in (0,2)$, the assertion of this corollary follows from (4.25) directly. $\square$

4.3 Convergence

The proposed lemmas and theorems are adequate to establish the global convergence of the proposed ADMM 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 ADMM with Gaussian back substitution. Then we have*

1. $\lim_{k\to\infty} \|A_i(x^k - \tilde{x}_i^k)\| = 0, i = 2, \ldots, m,$ *and* $\lim_{k\to\infty} \|\lambda^k - \tilde{\lambda}^k\| = 0.$
2. *If each A_i is full column rank matrix, then*
   (a) *any cluster point of $\{\tilde{w}^k\}$ is a solution point of* (2.3).
   (b) *The sequence $\{\tilde{v}^k\}$ converges to some $v^\infty \in \mathcal{V}^*$.*

Proof. From (4.21) we get

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

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

$$\lim_{k\to\infty}\|A_i(x^k-\tilde{x}_i^k)\|=0, \quad i=2,\ldots,m, \tag{4.26}$$

and

$$\lim_{k\to\infty}\|\lambda^k-\tilde{\lambda}^k\|=0. \tag{4.27}$$

The first assertion is proved.

Now, we assume that each A_i is full column rank matrix. Substituting (4.26) into (4.5), for $i=1,2,\ldots,m$, we have

$$\tilde{x}_i^k\in\mathcal{X}_i, \ \lim_{k\to\infty}\{\theta_i(x_i)-\theta_i(\tilde{x}_i^k)+(x_i-\tilde{x}_i^k)^T(-A_i^T\tilde{\lambda}^k)\}\geq 0, \quad \forall\, x_i\in\mathcal{X}_i. \tag{4.28}$$

It follows from (3.4a) and (4.27) that

$$\lim_{k\to\infty}(\sum_{j=1}^m A_j\tilde{x}_j^k-b)=0. \tag{4.29}$$

Combining (4.28) and (4.29) we get

$$\tilde{w}^k\in\Omega, \quad \lim_{k\to\infty}\{\theta(x)-\theta_i(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\}\geq 0, \quad \forall\, w\in\Omega, \tag{4.30}$$

and thus any cluster point of $\{\tilde{w}^k\}$ is a solution point of (2.3). The part (a) of the second assertion is proved.

It follows from the full column rank assumption, the matrix H and G are positive definite. Since $\lim_{k\to\infty}\|v^k-\tilde{v}^k\|_H^2=0$, then $\{\tilde{v}^k\}$ is bounded. Let v^∞ be a cluster point of $\{\tilde{v}^k\}$ and the subsequence $\{\tilde{v}^{k_j}\}$ converges to v^∞. It follows from (4.30)

$$\tilde{w}^{k_j}\in\Omega,\quad \lim_{k\to\infty}\{\theta(x)-\theta(\tilde{x}^{k_j})+(w-\tilde{w}^{k_j})^TF(\tilde{w}^{k_j})\}\ge 0,\quad \forall\, w\in\Omega \quad (4.31)$$

and consequently

$$\left\{\begin{array}{l} \theta_i(x_i)-\theta(x_i^\infty)+(x_i-x_i^\infty)^T\{-A_i^T\lambda^\infty\}\ge 0, \quad \forall x_i\in\mathcal{X}_i,\ \ i=1,\ldots,m, \\ \\ \sum_{j=1}^m A_jx_j^\infty-b=0. \end{array}\right. \quad (4.32)$$

This means that $v^\infty\in\mathcal{V}^*$. Since $\|v^{k+1}-v^*\|_G^2\le\|v^k-v^*\|_G^2$ 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 $v^\infty\in\mathcal{V}^*$. $\Box$

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

[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] C.H. Chen, B.S. He, Y.Y. Ye and X. M. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessary convergent, Mathematical Programming, 155 (2016) 57-79.

[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, M. Tao and X.M. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, *SIAM J. Optim. 22, 313-340, 2012*

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

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