凸优化和单调变分不等式的收缩算法

第十九讲: 多块可分离凸优化问题部分平行正则化的交替方向类方法

Partially parallel and regularized Alternating direction method of multipliers 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 [11]

1 Convex optimization problem with m-blocks

We consider the linearly constrained convex optimization with m separable operators

$$\min\Big\{\sum_{i=1}^m \theta_i(x_i) \,\Big|\, \sum_{i=1}^m A_i x_i = b,\ x_i \in \mathcal{X}_i\Big\}. \tag{1.1}$$

Its Lagrange function is

$$L(x_1,\ldots,x_m,\lambda) = \sum_{i=1}^m \theta_i(x_i) - \lambda^T\big(\sum_{i=1}^m A_i x_i - b\big), \tag{1.2}$$

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 (1.2). Then we have

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

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

$$\left\{ \begin{array}{l} \theta_1(x_1) - \theta_1(x_1^*) + (x_1 - x_1^*)^T\{-A_1^T\lambda^*\} \geq 0, \\ \theta_2(x_2) - \theta_2(x_2^*) + (x_2 - x_2^*)^T\{-A_2^T\lambda^*\} \geq 0, \\ \qquad\qquad\vdots \\ \qquad\qquad\vdots \\ \theta_m(x_m) - \theta_m(x_m^*) + (x_m - x_m^*)^T\{-A_m^T\lambda^*\} \geq 0, \\ (\lambda - \lambda^*)^T(\sum_{i=1}^m A_i x_i^* - b) \geq 0, \end{array} \right. \tag{1.3}$$

for all $w = (x_1, x_2, \cdots, x_m, \lambda) \in \Omega$. More compactly, (1.3) can be written into the following VI:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{1.4a}$$

where

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

$$w = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ \lambda \end{pmatrix} \quad \text{and} \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ -A_2^T\lambda \\ \vdots \\ -A_m^T\lambda \\ \sum_{i=1}^m A_i x_i - b \end{pmatrix}. \tag{1.4b}$$

Note that the operator $F(w)$ defined in (1.4b) is an affine operator of and its matrix is skew-symmetric, thus, we have

$$(w - \tilde{w})^T \big(F(w) - F(\tilde{w})\big) \equiv 0, \quad \forall w, \tilde{w}. \tag{1.5}$$

Since we have assumed that the solution set of (1.1) is not empty, the solution set of (1.4), denoted by Ω^*, is also nonempty.

In addition to the notation of $w = (x_1, x_2, \cdots, x_m, \lambda)$, we also use the following notation:

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

Moreover, we define

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

The augmented Lagrangian function of the problem (1.1) is

$$\mathcal{L}_\beta(x_1,\ldots,x_m,\lambda) = L(x_1,\ldots,x_m,\lambda) + \frac{\beta}{2}\|\textstyle\sum_{i=1}^m A_i x_i - b\|^2. \tag{1.6}$$

Now, we are in the stage to describe the direct extension of ADMM to the problem (1.1).

Direct Extension of ADMM Start with given $(x_2^k,\ldots,x_m^k,\lambda^k)$,

$$\left\{\begin{array}{l} x_1^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1,x_2^k,x_3^k,\ldots,x_m^k,\lambda^k) \mid x_1\in\mathcal{X}_1\}; \\ x_2^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1^{k+1},x_2,x_3^k,\ldots,x_m^k,\lambda^k) \mid x_2\in\mathcal{X}_2\}; \\ \qquad\vdots \\ x_i^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1^{k+1},\ldots,x_{i-1}^{k+1},x_i,x_{i+1}^k,\ldots,x_m^k,\lambda^k) \mid x_i\in\mathcal{X}_i\}; \\ \qquad\vdots \\ x_m^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1^{k+1},\ldots,x_{m-1}^{k+1},x_m,\lambda^k) \mid x_m\in\mathcal{X}_m\}; \\ \lambda^{k+1} = \lambda^k - \beta\big(\sum_{i=1}^m A_i x_i^{k+1} - b\big). \end{array}\right. \tag{1.7}$$

There is counter example [3], it is not necessary convergent for the problem with $m \geq 3$.

2 ADMM + Prox-Parallel Splitting ALM

The following splitting method does not need correction. Its k-th iteration begins with given $v^k = (x_2^k, \ldots, x_m^k, \lambda^k)$, and obtain v^{k+1} via the following procedure:

$$\left\{\begin{array}{l} x_1^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \ldots, x_m^k, \lambda^k) \mid x_1 \in \mathcal{X}_1\}; \\ \text{for} \quad i = 2, \ldots, m, \quad \text{do :} \\ x_i^{k+1} = \arg\min\limits_{x_i \in \mathcal{X}_i} \left\{\begin{array}{c} \mathcal{L}_\beta(x_1^{k+1}, x_2^k \ldots, x_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k) \\ + \frac{\tau\beta}{2}\|A_i(x_i - x_i^k)\|^2 \end{array}\right\}; \\ \lambda^{k+1} = \lambda^k - \beta\big(\sum_{i=1}^m A_i x_i^{k+1} - b\big) \end{array}\right. \tag{2.1}$$

- **The $x_2 \ldots x_m$-subproblems are solved in a parallel manner.**
- **To ensure the convergence, in the x_i-subproblem, $i = 2, \ldots, m$, an extra proximal term $\frac{\tau\beta}{2}\|A_i(x_i - x_i^k)\|^2$ is necessary.**

An equivalent recursion of (2.1)

$\mu = \tau + 1$ and τ is given in (2.1).

$$\left\{\begin{array}{l} x_1^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \ldots, x_m^k, \lambda^k) \;\big|\; x_1 \in \mathcal{X}_1\big\}; \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta\big(A_1 x_1^{k+1} + \sum_{i=2}^m A_i x_i^k - b\big); \\ \text{for} \quad i = 2, \ldots, m, \;\; \text{do}: \\ x_i^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_i(x_i) - (\lambda^{k+\frac{1}{2}})^T A_i x_i \\ + \frac{\mu\beta}{2}\|A_i(x_i - x_i^k)\|^2 \end{array} \middle| \; x_i \in \mathcal{X}_i \right\}; \\ \lambda^{k+1} = \lambda^k - \beta\big(\sum_{i=1}^m A_i x_i^{k+1} - b\big) \end{array}\right. \tag{2.2}$$

The method (2.2) is proposed in IMA Numerical Analysis [11]:

- **B. He, M. Tao and X. Yuan, A splitting method for separable convex programming. IMA J. Numerical Analysis, 31(2015), 394-426.**

Equivalence of (2.1) and (2.2)

It needs only to check the optimization conditions of their x_i-subproblems for $i = 2, \ldots, m$. Note that the optimal condition of the x_i-subproblem of (2.1) is

$$x_i^{k+1} \in \mathcal{X}_i, \quad \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T \big\{ -A_i^T \lambda^k + \\ +\beta A_i^T \big[(A_1 x_1^{k+1} + \textstyle\sum_{j=2}^m A_j x_j^k - b) + A_i(x_i^{k+1} - x_i^k) \big] \\ +\tau\beta A_i^T A_i (x_i^{k+1} - x_i^k) \big\} \ \ge \ 0.$$

for all $x_i \in \mathcal{X}_i$. By using

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \beta\big(A_1 x_1^{k+1} + \textstyle\sum_{j=2}^m A_j x_j^k - b\big); \qquad (2.3)$$

it can be written as

$$x_i^{k+1} \in \mathcal{X}_i, \quad \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T \big\{ -A_i^T \lambda^{k+\frac{1}{2}} \\ +\beta A_i^T A_i (x_i^{k+1} - x_i^k) + \tau\beta A_i^T A_i (x_i^{k+1} - x_i^k) \big\} \ \ge \ 0.$$

and consequently

$$x_i^{k+1} \in \mathcal{X}_i, \quad \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T \{-A_i^T \lambda^{k+\frac{1}{2}} \\ +(1+\tau)\beta A_i^T A_i(x_i^{k+1} - x_i^k)\} \geq 0, \ \forall\, x_i \in \mathcal{X}_i. \quad (2.4)$$

Setting $\mu = 1 + \tau$, (2.4) is the optimal condition of the x_i-subproblem of (2.2) !

Notice that the subproblems in (2.2) are

$$x_1^{k+1} = \arg\min\{\theta_1(x_1) + \frac{\beta}{2}\|A_1 x_1 + (\sum_{i=2}^m A_i x_i^k - b) - \frac{1}{\beta}\lambda^k \,|\, x_1 \in \mathcal{X}_1\}$$

and

$$x_i^{k+1} = \arg\min\{\theta_i(x_i) + \frac{\mu\beta}{2}\|A_i(x_i - x_i^k) - \frac{1}{\mu\beta}\lambda^{k+\frac{1}{2}}\|^2 \,|\, x_i \in \mathcal{X}_i\},$$

for $i = 2, \ldots, m$. We assume that the above problems are not difficult to solve.

The convergence analysis is under the assumption $\mu \geq m - 1$. In the next section, Section 3, we prove the convergence of Algorithm (2.2). In Section 4, we present a prediction-correction method which uses the output of (2.2) as the predictor, and the new iterate is updated by a simple correction.

3 Convergence Analysis for the algorithm (2.2)

We use (2.2) to analyze the convergence and assume that $\boldsymbol{\mu \geq m-1}$.

The optimal condition of the x_i-subproblems of (2.2) can be written as

$$\begin{cases} \theta_1(x_1) - \theta_1(x_1^{k+1}) + (x_1 - x_1^{k+1})^T(-A_1^T\lambda^{k+\frac{1}{2}}) \geq 0, \;\forall x_1 \in \mathcal{X}_1; \\ \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T(-A_i^T\lambda^{k+\frac{1}{2}}) \\ \quad \geq (x_i - x_i^{k+1})^T \mu\beta A_i^T A_i(x_i^k - x_i^{k+1}), \;\forall x_i \in \mathcal{X}_i; \;\; i = 2, \ldots, m. \end{cases} \tag{3.1}$$

Since

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \beta\big(A_1 x_1^{k+1} + \sum_{j=2}^m A_j x_j^k - b\big)$$

and

$$\lambda^{k+1} = \lambda^k - \beta\big(\sum_{i=1}^m A_i x_i^{k+1} - b\big),$$

we have

$$-\lambda^{k+\frac{1}{2}} = -\lambda^{k+1} + \beta\sum_{j=2}^m A_j(x_j^k - x_j^{k+1}). \tag{3.2}$$

Substituting (3.2) in (3.1), we get

$$\begin{cases} \theta_1(x_1) - \theta_1(x_1^{k+1}) + (x_1 - x_1^{k+1})^T(-A_1^T\lambda^{k+1} + A_1^T p_k) \\ \qquad\qquad \geq 0, \;\forall x_1 \in \mathcal{X}_1; \\ \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T(-A_i^T\lambda^{k+1} + A_i^T p_k) \\ \qquad\qquad \geq (x_i - x_i^{k+1})^T \mu\beta A_i^T A_i(x_i^k - x_i^{k+1}), \forall x_i \in \mathcal{X}_i; \\ \qquad\qquad\quad i = 2, \ldots, m. \end{cases} \tag{3.3}$$

where

$$p_k = \beta \sum_{j=2}^{m} A_j(x_j^k - x_j^{k+1}). \tag{3.4}$$

Since

$$\big(\sum_{i=1}^{m} A_i x_i^{k+1} - b\big) = (1/\beta)(\lambda^k - \lambda^{k+1}),$$

It can be written as

$$(\lambda - \lambda^{k+1})^T \Big(\sum_{i=1}^{m} A_i x_i^{k+1} - b\Big) \ge (\lambda - \lambda^{k+1})^T (1/\beta)(\lambda^k - \lambda^{k+1}), \quad (3.5)$$

for all $\lambda \in \Re^l$.

Combining (3.3) and (3.5)

$$\left\{\begin{array}{l} \theta_1(x_1) - \theta_1(x_1^{k+1}) + (x_1 - x_1^{k+1})^T(-A_1^T\lambda^{k+1} + A_1^T p_k) \\ \qquad\qquad \ge 0, \ \forall x_1 \in \mathcal{X}_1; \\ \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T(-A_i^T\lambda^{k+1} + A_i^T p_k) \\ \qquad\qquad \ge (x_i - x_i^{k+1})^T \mu\beta A_i^T A_i(x_i^k - x_i^{k+1}), \forall x_i \in \mathcal{X}_i; \\ \qquad\qquad\quad i = 2, \ldots, m. \\ (\lambda - \lambda^{k+1})^T \big(\sum_{i=1}^m A_i x_i^{k+1} - b\big) \\ \qquad \ge (\lambda - \lambda^{k+1})^T (1/\beta)(\lambda^k - \lambda^{k+1}). \ \ \forall \lambda \in \Re^l. \end{array}\right. \quad (3.6)$$

By using the notations $\theta(x)$ and $F(w)$, it follows from (3.6) that

$$\theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) + p_k^T \big(\sum_{i=1}^m A_i(x_i - x_i^{k+1})\big)$$

$$\geq \sum_{i=2}^m (x_i - x_i^{k+1})^T \mu\beta A_i^T A_i (x_i^k - x_i^{k+1})$$

$$+(\lambda - \lambda^{k+1})^T (1/\beta)(\lambda^k - \lambda^{k+1}), \quad \forall w \in \Omega.$$

Setting $w = w^*$ in the above inequality and by a manipulation, we get

$$\sum_{i=2}^m (x_i^{k+1} - x^*)^T \mu\beta A_i^T A_i (x_i^k - x_i^{k+1}) + (\lambda^{k+1} - \lambda^*)^T \frac{1}{\beta} (\lambda^k - \lambda^{k+1})$$

$$\geq \quad \theta(x^{k+1}) - \theta(x^*) + (w^{k+1} - w^*)^T F(w^{k+1})$$

$$+ p_k^T \big(\sum_{i=1}^m A_i (x_i^{k+1} - x_i^*)\big) \tag{3.7}$$

Now, we treat the right hand side of (3.7). Using (1.5) and the optimality, we have

$$\begin{aligned} &\theta(x^{k+1}) - \theta(x^*) + (w^{k+1} - w^*)^T F(w^{k+1}) \\ = \quad & \theta(x^{k+1}) - \theta(x^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0. \end{aligned}$$

Because $\sum_{i=1}^m A_i x_i^* = b$ and $\sum_{i=1}^m A_i x_i^{k+1} - b = \frac{1}{\beta}(\lambda^k - \lambda^{k+1})$, using the definition of p_k (see (3.4)), we get

$$p_k^T \big(\sum_{i=1}^m A_i (x_i^{k+1} - x^*) \big) = (\lambda^k - \lambda^{k+1})^T \sum_{j=2}^m A_j (x_j^k - x_j^{k+1}).$$

Substituting it in (3.7), we get

$$\begin{aligned} &\sum_{i=2}^m (x_i^{k+1} - x^*)^T \mu\beta A_i^T A_i (x_i^k - x_i^{k+1}) + (\lambda^{k+1} - \lambda^*)^T \frac{1}{\beta} (\lambda^k - \lambda^{k+1}) \\ \geq \quad & (\lambda^k - \lambda^{k+1})^T \sum_{j=2}^m A_j (x_j^k - x_j^{k+1}). \end{aligned} \tag{3.8}$$

By denoting

$$v^k = \begin{pmatrix} x_2^k \\ \vdots \\ x_m^k \\ \lambda^k \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} \mu\beta A_2^T A_2 & & & \\ & \ddots & & \\ & & \mu\beta A_m^T A_m & \\ & & & \frac{1}{\beta} I \end{pmatrix}, \qquad (3.9)$$

we get the following assertion:

Lemma 3.1 *Let* v^{k+1} *be generated by* (2.2) *from the given vector* v^k*, then we have*

$$(v^k - v^*)^T G(v^k - v^{k+1}) \geq \varphi(v^k, v^{k+1}), \qquad (3.10)$$

where

$$\varphi(v^k, v^{k+1}) = \|v^k - v^{k+1}\|_G^2 + (\lambda^k - \lambda^{k+1})^T \big(\textstyle\sum_{i=2}^m A_i(x_i^k - x_i^{k+1})\big). \qquad (3.11)$$

Proof. Using the notations (3.9), the inequality (3.8) can be written as

$$(v^{k+1} - v^*)^T G(v^k - v^{k+1}) \geq (\lambda^k - \lambda^{k+1})^T \sum_{j=2}^{m} A_j(x_j^k - x_j^{k+1}).$$

Adding $(v^k - v^{k+1})^T G(v^k - v^{k+1})$ to the both sides of the above inequality, we get the assertion directly. $\square$

Now we consider the profit of the k-th iteration. Using (3.10) and (3.11), we have

$$\begin{aligned}
&\|v^k - v^*\|_G^2 - \|v^{k+1} - v^*\|_G^2 \\
&= \|v^k - v^*\|_G^2 - \|(v^k - v^*) - (v^k - v^{k+1})\|_G^2 \\
&= 2(v^k - v^*)^T G(v^k - v^{k+1}) - \|v^k - v^{k+1}\|_G^2 \\
&\geq \|v^k - v^{k+1}\|_G^2 + 2(\lambda^k - \lambda^{k+1})^T \sum_{j=2}^{m} A_j(x_j^k - x_j^{k+1}). \quad (3.12)
\end{aligned}$$

In the following is to show that, **when $\mu > m - 1$, there is a constant $\sigma > 0$, such that the right hand side of (3.12) is greater than $\sigma\|v^k - v^{k+1}\|_G^2$** .

According to the definition of the matrix G, we have

$$\|v^k - v^{k+1}\|_G^2 + 2(\lambda^k - \lambda^{k+1})^T\big(\sum_{i=2}^{m} A_i(x_i^k - x_i^{k+1})\big)$$

$$= \begin{pmatrix} \sqrt{\beta}A_2(x_2^k - x_2^{k+1}) \\ \sqrt{\beta}A_3(x_3^k - x_3^{k+1}) \\ \vdots \\ \sqrt{\beta}A_m(x_m^k - x_m^{k+1}) \\ (1/\sqrt{\beta})(\lambda^k - \lambda^{k+1}) \end{pmatrix}^T \begin{pmatrix} \mu I_l & 0 & \cdots & 0 & I_l \\ 0 & \mu I_l & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & I_l \\ 0 & \cdots & 0 & \mu I_l & I_l \\ I_l & \cdots & I_l & I_l & I_l \end{pmatrix} \begin{pmatrix} \sqrt{\beta}A_2(x_2^k - x_2^{k+1}) \\ \sqrt{\beta}A_3(x_3^k - x_3^{k+1}) \\ \vdots \\ \sqrt{\beta}A_m(x_m^k - x_m^{k+1}) \\ (1/\sqrt{\beta})(\lambda^k - \lambda^{k+1}) \end{pmatrix}. \tag{3.13}$$

Notice that the block-wise matrix

$$\begin{pmatrix} \mu I_l & 0 & \cdots & 0 & I_l \\ 0 & \mu I_l & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & I_l \\ 0 & \cdots & 0 & \mu I_l & I_l \\ I_l & \cdots & I_l & I_l & I_l \end{pmatrix}$$

in (3.13) have the same largest (resp. smallest) eigenvalues as the $m \times m$ symmetric matrix

$$M = \begin{pmatrix} \mu & 0 & \cdots & 0 & 1 \\ 0 & \mu & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & 0 & \mu & 1 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}_{m \times m} . \quad (3.14)$$

Lemma 3.2 *For* $m \geq 2$*, the* $m \times m$ *symmetric matrix* M *defined in* (3.14) *has* $(m-2)$ *multiple eigenvalues*

$$\nu_1 = \nu_2 = \cdots = \nu_{m-2} = \mu,$$

and another two eigenvalues

$$\nu_{m-1}, \nu_m = \frac{1}{2}\left[(\mu+1) \pm \sqrt{(\mu+1)^2 + 4((m-1)-\mu)}\right].$$

Proof. Let e be a $(m-1)$-vector whose each element equals 1. Thus

$$M = \begin{pmatrix} \mu I_{m-1} & e \\ e^T & 1 \end{pmatrix}.$$

Without loss of generality, we assume that the eigenvectors of M have forms

$$z = \begin{pmatrix} y \\ 0 \end{pmatrix} \quad \text{or} \quad z = \begin{pmatrix} y \\ 1 \end{pmatrix}, \quad \text{where } \; y \in \Re^{m-1}.$$

In the first case, we have

$$\begin{cases} \mu y = \nu y, \\ e^T y = 0. \end{cases} \tag{3.15}$$

It is clear that the $(m-1)$-vectors

$$y^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}, \quad y^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix} \quad \cdots\cdots \quad y^{m-2} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ -1 \end{pmatrix}$$

are linear independent and satisfy (3.15) with $\nu = \mu$. Thus,

$$z^i = \begin{pmatrix} y^i \\ 0 \end{pmatrix}, \quad i = 1, \ldots, m-2,$$

are eigenvectors of M and the related eigenvalue

$$\nu_1 = \nu_2 = \cdots = \nu_{m-2} = \mu.$$

In the second case, $z^T = (y^T, 1)$, we have

$$\begin{cases} \mu y + e = \nu y, \\ e^T y + 1 = \nu. \end{cases} \tag{3.16}$$

It follows from (3.16) that

$$(\nu - \mu)(\nu - 1) - (m-1) = 0,$$

and thus

$$\nu_{m-1}, \nu_m = \frac{1}{2}\left[(\mu+1) \pm \sqrt{(\mu+1)^2 + 4((m-1)-\mu)}\right].$$

The lemma is proved. $\square$

For $\mu \geq 1$, it is easy to verify that

$$\nu_m = \frac{(\mu+1) - \sqrt{(\mu+1)^2 + 4((m-1)-\mu)}}{2} \qquad (3.17)$$

is the smallest eigenvalue of M. For fixed $\mu > (m-1)$, there is a σ such that $\nu_m > \sigma$. Together with (3.12) and (3.13), we have the following assetions:

Lemma 3.3 *Let $\mu > m-1$, then there is a $\sigma > 0$ such that*

$$\|v^k - v^{k+1}\|_G^2 + 2(\lambda^k - \lambda^{k+1})^T \Big(\sum_{i=2}^{m} A_i(x_i^k - x_i^{k+1})\Big) \geq \sigma \|v^k - v^{k+1}\|_G^2, \qquad (3.18)$$

where G is defined in (3.9).

Theorem 3.1 *Let $\mu > m-1$ and $\{v^k\}$ be the sequence generated by* (2.2), *then there is a $\sigma > 0$ such that*

$$\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 - \sigma \|v^k - v^{k+1}\|_G^2, \quad \forall\, v^* \in \mathcal{V}^*. \qquad (3.19)$$

where G is defined in (3.9).

The inequality (3.19) is is the key for convergence of the method (2.1) and (2.2).

Implementation of the method for three block problems

For the problem with three separable operators

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\},$$

we have

$$\begin{array}{rcl}\mathcal{L}^3_\beta(x,y,z,\lambda) & = & \theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b)\\ & & +\frac{\beta}{2}\|Ax+By+Cz-b\|^2.\end{array}$$

For given $v^k=(y^k,z^k,\lambda^k)$, by using the method proposed in this subsection, the new iterate $v^{k+1}=(y^{k+1},z^{k+1},\lambda^{k+1})$ is obtained via ($\tau\geq 1$) :

$$\left\{\begin{array}{l} x^{k+1}=\text{Argmin}\{\mathcal{L}^3_\beta(x,y^k,z^k,\lambda^k)\,|\,x\in\mathcal{X}\},\\ y^{k+1}=\text{Argmin}\{\mathcal{L}^3_\beta(x^{k+1},y,z^k,\lambda^k)+\frac{\tau\beta}{2}\|B(y-y^k)\|^2\,|\,y\in\mathcal{Y}\},\\ z^{k+1}=\text{Argmin}\{\mathcal{L}^3_\beta(x^{k+1},y^k,z,\lambda^k)+\frac{\tau\beta}{2}\|C(z-z^k)\|^2\,|\,z\in\mathcal{Z}\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b),\end{array}\right. \tag{3.20}$$

An equivalent recursion of (3.20) is

$$\left\{\begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}^3_\beta(x, y^k, z^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\ y^{k+1} = \text{Argmin}\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, \\ z^{k+1} = \text{Argmin}\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T Cz + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{array}\right. \tag{3.21}$$

where $\mu = \tau + 1 \geq 2$. Implementation of (3.21) is via

$$\left\{\begin{array}{l} x^{k+1} = \text{Argmin}\{\theta_1(x) + \frac{\beta}{2}\|Ax + [By^k + Cz^k - b - \frac{1}{\beta}\lambda^k]\|^2 \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\ y^{k+1} = \text{Argmin}\{\theta_2(y) + \frac{\mu\beta}{2}\|By - [By^k + \frac{1}{\mu\beta}\lambda^{k+\frac{1}{2}}]\|^2 \,|\, y \in \mathcal{Y}\}, \\ z^{k+1} = \text{Argmin}\{\theta_3(z) + \frac{\mu\beta}{2}\|Cz - [Cz^k + \frac{1}{\mu\beta}\lambda^{k+\frac{1}{2}}]\|^2 \,|\, z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{array}\right.$$

4 Method with the calculated stepsize

The iteration of the method (2.1) and/or (2.2) begin with $v^k = (x_2^k, \cdots, \lambda^k)$ and finish with $v^{k+1} = (x_2^{k+1}, \cdots x_m^{k+1}, \lambda^{k+1})$. In this section, we consider the method with the calculated step-size. In practice, we use the output of (2.2) as a predictor.

$$\left\{\begin{array}{l} x_1^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \ldots, x_m^k, \lambda^k) \;\big|\; x_1 \in \mathcal{X}_1\big\}; \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta\big(A_1 x_1^{k+1} + \sum_{i=2}^m A_i x_i^k - b\big); \\ \text{for} \quad i = 2, \ldots, m, \;\; \text{do}: \\ \tilde{x}_i^k = \arg\min\left\{ \begin{array}{c} \theta_i(x_i) - (\lambda^{k+\frac{1}{2}})^T A_i x_i \\ + \frac{\mu\beta}{2}\|A_i(x_i - x_i^k)\|^2 \end{array} \right| \left. x_i \in \mathcal{X}_i \right\}; \\ \tilde{\lambda}^k = \lambda^k - \beta\big(A_1 x^{k+1} + \sum_{i=2}^m A_i \tilde{x}_i^k - b\big) \end{array}\right. \tag{4.1}$$

We only denote the output $v^{k+1} = (x_2^{k+1}, \cdots x_m^{k+1}, \lambda^{k+1})$ generated from (2.2) by using the new notations $\tilde{v}^k = (\tilde{x}_2^k, \cdots, \tilde{x}_m^k, \tilde{\lambda}^k)$. After getting $\tilde{v}^k$, we offer thenew iterate v^{k+1} by $v^{k+1} = v^k - \alpha_k(v^k - \tilde{v}^k)$.

Algorithm 2: a prediction-correction splitting method for solving (1.1)

Step 1. Prediction step. From the given $v^k = (x_2^k, \cdots, x_m^k, \lambda^k)$, using (4.1) to produce the predictor $\tilde{v}^k = (\tilde{x}_2^k, \cdots, \tilde{x}_m^k, \tilde{\lambda}^k)$.

Step 2. Correction step. The new iterate $v^{k+1} = (x_2^{k+1}, \cdots x_m^{k+1}, \lambda^{k+1})$ is updated via:

$$v^{k+1} = v^k - \alpha_k(v^k - \tilde{v}^k), \tag{4.2}$$

where

$$\alpha_k = \gamma \alpha_k^*, \qquad \gamma_k \in (0, 2), \qquad \alpha_k^* = \frac{\varphi(v^k, \tilde{v}^k)}{\|v^k - \tilde{v}^k\|_G^2} \tag{4.3}$$

and

$$\varphi(v^k, \tilde{v}^k) = \|v^k - \tilde{v}^k\|_G^2 + (\lambda^k - \tilde{\lambda}^k)^T \left(\textstyle\sum_{i=2}^m A_i(x_i^k - \tilde{x}_i^k)\right). \tag{4.4}$$

As we can see easily, Algorithm 1 (2.2) turns out to be a special case of Algorithm

2 where $\gamma_k \equiv 1/\alpha_k$ in (4.3). Thus, in the following, we prove the convergence for Algorithm 2, from which the convergence of Algorithm 1 becomes trivial.

Since the $\tilde{v}^k$ in (4.1) is the same of v^{k+1} in Algorithm (2.2), similarly as in Lemma 3.1, we have the following assertion directly.

Lemma 4.1 *Let $\tilde{v}^k$ be generated by* (4.1) *from the given vector v^k, then we have*

$$(v^k - v^*)^T G(v^k - \tilde{v}^k) \geq \varphi(v^k, \tilde{v}^k), \tag{4.5}$$

where $\varphi(v^k, \tilde{v}^k)$ is defined in (4.4)*.*

Lemma 4.2 *Under the assumption $\mu > m - 1$, it holds that*

$$\varphi(v^k, \tilde{v}^k) \geq \frac{1+\sigma}{2} \|v^k - \tilde{v}^k\|_G^2. \tag{4.6}$$

Proof. According to the definition of $\varphi(v^k, \tilde{v}^k)$ (see (4.4)) and the inequality

(3.18) in Lemma 3.3, we have

$$\begin{aligned} 2\varphi(v^k, \tilde{v}^k) &= 2\|v^k - \tilde{v}^k\|_G^2 + 2(\lambda^k - \tilde{\lambda}^k)^T \big(\sum_{i=2}^{m} A_i(x_i^k - \tilde{x}_i^k)\big) \\ &\geq (1+\sigma)\|v^k - \tilde{v}^k\|_G^2, \end{aligned}$$

and the assertion follows from the definitions of $\varphi(v^k, \tilde{v}^k)$ and α_k^* (see (4.3) and (3.11)) directly. □

For determinate the step size α_k in (4.2), we define the step-size dependent new iterate by

$$v^{k+1}(\alpha) = v^k - \alpha(v^k - \tilde{v}^k), \tag{4.7}$$

In this way,

$$\vartheta(\alpha) = \|v^k - v^*\|_G^2 - \|v^{k+1}(\alpha) - v^*\|_G^2 \tag{4.8}$$

is the distance decrease functions in the k-th iteration by using updating form (4.7). By defining

$$q(\alpha) = 2\alpha\varphi(v^k, \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_G^2. \tag{4.9}$$

It follows from (4.7), (4.8) and (4.5) that

$$
\begin{aligned}
\vartheta(\alpha) &= \|v^k - v^*\|_G^2 - \|v^k - v^* - \alpha(v^k - \tilde{v}^k)\|_G^2 \\
&= 2\alpha(v^k - v^*)^T G(v^k - \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_G^2 \\
&\geq 2\alpha\varphi(v^k, \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_G^2 \\
&= q(\alpha).
\end{aligned} \tag{4.10}
$$

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

$$
\alpha_k^* = \frac{\varphi(v^k, \tilde{v}^k)}{\|v^k - \tilde{v}^k\|_G^2}, \tag{4.11}
$$

and this is just the same as defined in (4.3). Usually, in practical computation, taking a relaxed factor $\gamma > 1$ is useful for fast convergence.

Theorem 4.1 *Let $\{v^k\}$ be the sequence generated by Algorithm 2. We have*

$$
\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 - \frac{\gamma(2-\gamma)}{4}\|v^k - \tilde{v}^k\|_G^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{4.12}
$$

Proof. It follows from (4.8) and (4.10) that

$$\|v^{k+1} - v^*\|_G^2 \le \|v^k - v^*\|_G^2 - q(\gamma\alpha_k^*), \quad \forall v^* \in \mathcal{V}^*. \tag{4.13}$$

By using (4.9) and (4.11) we obtain

$$\begin{aligned} q(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*\varphi(v^k, \tilde{v}^k) - (\gamma\alpha_k^*)^2\|v^k - \tilde{v}^k\|_G^2 \\ &= \gamma(2-\gamma)\alpha_k^*\varphi(v^k, \tilde{v}^k). \end{aligned} \tag{4.14}$$

Since (see (4.6))

$$\varphi(v^k, \tilde{v}^k) > \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2$$

and consequently (see (4.3)),

$$\alpha_k^* > \frac{1}{2}.$$

Thus, we have

$$\alpha_k^*\varphi(v^k, \tilde{v}^k) \ge \frac{1}{4}\|v^k - \tilde{v}^k\|_G^2.$$

Substituting it in (4.14), the proof of this theorem is complete. □

Theorem 4.1 offers the key inequality for the convergence !

References

[1] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A proximal point algorithms revisit on the alternating direction method of multipliers, Science China Mathematics, 56 (2013), 2179-2186.

[2] A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vison, 40, 120-145, 2011.

[3] 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 necessarily convergent*, to appear in *Mathematical Programming, Series A*.

[4] E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

[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**, 69-76, 1997.

[7] B. S. He, Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities, *Computational Optimization and Applications* **42**(2009), 195–212.

[8] B. S. He, PPA-like contraction methods for convex optimization: a framework using variational inequality approach. J. Oper. Res. Soc. China 3 (2015) 391 − 420.

[9] B. S. He, L. Z. Liao, D. Han, and H. Yang, A new inexact alternating directions method for

monontone variational inequalities, *Mathematical Programming* **92**(2002), 103–118.

[10] B. S. He, M. Tao and X.M. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, *SIAM Journal on Optimization* **22**(2012), 313-340.

[11] B.S. He, M. Tao and X.M. Yuan, A splitting method for separable convex programming, *IMA J. Numerical Analysis* **31**(2015), 394-426.

[12] B. S. He, M. H. Xu, and X. M. Yuan, Solving large-scale least squares covariance matrix problems by alternating direction methods, *SIAM Journal on Matrix Analysis and Applications* **32**(2011), 136-152.

[13] B.S. He, H. Yang, and S.L. Wang, Alternating directions method with self-adaptive penalty parameters for monotone variational inequalities, *JOTA* **23**(2000), 349–368.

[14] B. S. He and X. M. Yuan, On the $O(1/t)$ convergence rate of the alternating direction method, *SIAM J. Numerical Analysis* **50**(2012), 700-709.

[15] B.S. He and X.M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, *SIAM J. Imag. Science* **5**(2012), 119-149.

[16] B.S. He and X.M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers, Numerische Mathematik, 130 (2015) 567-577.

[17] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, *Rev. Francaise d'Inform. Recherche Oper.*, **4**, 154-159, 1970.

[18] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Cont. Optim.*, **14**, 877-898, 1976.