凸优化和单调变分不等式的收缩算法

第十一讲: 结构型优化的交替方向法

Alternating direction method of multipliers
for separable convex programming

南京大学数学系　　何炳生
hebma@nju.edu.cn

1 Structured constrained convex optimization

We consider the following structured constrained convex optimization problem

$$\min \{\theta_1(x) + \theta_2(y) \,|\, Ax + By = b,\ x \in \mathcal{X},\ y \in \mathcal{Y}\} \tag{1.1}$$

where $\theta_1(x) : \Re^{n_1} \to \Re,\ \theta_2(y) : \Re^{n_2} \to \Re$ are convex functions (but not necessary smooth), $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$ and $b \in \Re^m$, $\mathcal{X} \subset \Re^{n_1}$, $\mathcal{Y} \subset \Re^{n_2}$ are given closed convex sets. It is clear that the split feasibility problem:

$$\text{Find a point } \ x \in \mathcal{X} \ \text{ such that } \ Ax \in \mathcal{B},$$

can be formulated as a special problem of (1.1) with $\theta_1(x) = \theta_2(y) = 0$. Find (x, y) such that

$$\{Ax - y = 0,\ x \in \mathcal{X},\ y \in \mathcal{B}\}. \tag{1.2}$$

Let λ be the Lagrangian multiplier for the linear constraints $Ax + By = b$ in (1.1), the Lagrangian function of this problem is

$$L(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T (Ax + By - b),$$

which is defined on $\mathcal{X} \times \mathcal{Y} \times \Re^m$. Let (x^*, y^*, λ^*) be an saddle point of the Lagrangian function, then $(x^*, y^*, \lambda^*) \in \Omega$ and it satisfies

$$\left\{ \begin{array}{l} \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, \\ \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-B^T\lambda^*) \geq 0, \quad \forall\, (x, y, \lambda) \in \Omega, \\ (\lambda - \lambda^*)^T(Ax^* + By^* - b) \geq 0, \end{array} \right. \tag{1.3}$$

where

$$\Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m.$$

By denoting

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}$$

and

$$\theta(u) = \theta_1(x) + \theta_2(y),$$

the first order optimal condition (1.3) can be written in a compact form such as

$$w^* \in \Omega, \;\; \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \;\; \forall\, w \in \Omega. \tag{1.4}$$

Note that the mapping F is monotone. We use Ω^* to denote the solution set of the variational inequality (1.4). For convenience we use the notations

$$v = \begin{pmatrix} y \\ \lambda \end{pmatrix} \qquad \text{and} \qquad \mathcal{V}^* = \{(y^*, \lambda^*) \,|\, (x^*, y^*, \lambda^*) \in \Omega^*\}.$$

Augmented Lagrangian Method to structured COP

Augmented Lagrangian Method is one of the attractive methods for nonlinear optimization as demonstrated in Chapter 17 of [21]. We try to use the Augmented Lagrangian Method to solve (1.1) and set

$$M = (A, B) \qquad \text{and} \qquad \mathcal{U} = \mathcal{X} \times \mathcal{Y}.$$

Now, the problem (1.1) is rewritten as

$$\min\{\theta(u) \mid Mu = b, \;\; u \in \mathcal{U}\}. \tag{1.5}$$

For given $\beta > 0$, the augmented Lagrangian function of (1.5) is

$$\mathcal{L}_A(u,\lambda) = \theta(u) - \lambda^T(Mu-b) + \frac{\beta}{2}\|Mu-b\|^2,$$

which defined on $\Omega = \mathcal{U} \times \Re^m$. Directly applied Augmented Lagrangian Method to the problem (1.5), the k-th iteration begins with λ^k, obtain

$$u^{k+1} = \text{Argmin}\{\mathcal{L}_A(u,\lambda^k) \mid u \in \mathcal{U}\}, \tag{1.6}$$

and then update the new iterate by

$$\lambda^{k+1} = \lambda^k - \beta(Mu^{k+1} - b). \tag{1.7}$$

Note that $u^{k+1} \in \mathcal{U}$ generated by (1.6) satisfies

$$\theta(u) - \theta(u^{k+1}) + (u-u^{k+1})^T\{-M^T\lambda^k + \beta M^T(Mu^{k+1}-b)\} \geq 0, \;\; \forall u \in \mathcal{U}.$$

By using (1.7) in the last inequality, we obtain

$$u^{k+1} \in \mathcal{U}, \quad \theta(u) - \theta(u^{k+1}) + (u-u^{k+1})^T\big(-M^T\lambda^{k+1}\big) \geq 0, \quad \forall u \in \mathcal{U}. \tag{1.8}$$

Combining (1.8) and (1.7), we get $w^{k+1} \in \Omega$ and for any $w \in \Omega$, it holds that

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} u - u^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -M^T \lambda^{k+1} \\ Mu^{k+1} - b \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{\beta}(\lambda^{k+1} - \lambda^k) \end{pmatrix} \right\} \geq 0.$$

Substituting $w = w^*$ in the above variational inequality and using the notation of $F(w)$, we get

$$\begin{aligned} &(\lambda^{k+1} - \lambda^*)^T (\lambda^k - \lambda^{k+1}) \\ &\geq \quad \beta\{(w^{k+1} - w^*)^T F(w^{k+1}) + \theta(u^{k+1}) - \theta(u^*)\}. \end{aligned} \tag{1.9}$$

Using the monotonicity of F and the fact

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

we derive that the right hand side of (1.9) is non-negative. Therefore, we have

$$(\lambda^{k+1} - \lambda^*)^T (\lambda^k - \lambda^{k+1}) \geq 0, \quad \forall\, \lambda^* \in \Lambda^*. \tag{1.10}$$

It follows from (1.10) that

$$\|\lambda^k - \lambda^*\|^2 = \|(\lambda^{k+1} - \lambda^*) + (\lambda^k - \lambda^{k+1})\|^2 \geq \|\lambda^{k+1} - \lambda^*\|^2 + \|\lambda^k - \lambda^{k+1}\|^2.$$

We get the nice convergence property:

$$\|\lambda^{k+1} - \lambda^*\|^2 \le \|\lambda^k - \lambda^*\|^2 - \|\lambda^k - \lambda^{k+1}\|^2.$$

Summarize the Augmented Lagrangian Method to structured COP

For given λ^k, $u^{k+1} = (x^{k+1}, y^{k+1})$ is the solution of the following problem

$$\boxed{(x^{k+1}, y^{k+1}) = \text{Argmin}\left\{ \begin{array}{c} \theta_1(x) + \theta_2(y) - (\lambda^k)^T(Ax + By - b) \\ +\frac{\beta}{2}\|Ax + By - b\|^2 \end{array} \middle| \begin{array}{c} x \in \mathcal{X} \\ y \in \mathcal{Y} \end{array} \right\}}$$

The new iterate $\qquad \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b).$

Convergence $\qquad \|\lambda^{k+1} - \lambda^*\|^2 \le \|\lambda^k - \lambda^*\|^2 - \|\lambda^k - \lambda^{k+1}\|^2.$

Shortcoming $\qquad$ The structure property is not used !

By using the augmented Lagrangian method for the structured problem (1.5), the k-th iteration is from λ^k to λ^{k+1}. The variable $u = (x, y)$ is only an intermediate variable.

2 Alternating Direction Method of Multipliers

To overcome the shortcoming the ALM for the problem (1.1), we use the alternating direction method. The main idea is splitting the subproblem (1.6) in two parts and only the x-part is the intermediate variable. Thus the iteration begins with $v^0 = (y^0, \lambda^0)$.

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

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

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

Use λ^k and the obtained x^{k+1}, y^{k+1} is the solution of the following problem

$$y^{k+1} = \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(Ax^{k+1} + By - b) \\ + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \tag{2.1b}$$

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{2.1c}$$

Advantages

The sub-problems (2.1a) and (2.1b) are separately solved one by one.

Remark 2.1 *The sub-problems* (2.1a) *and* (2.1b) *is equivalent to*

$$x^{k+1} = \textit{Argmin}\Big\{\theta_1(x) + \tfrac{\beta}{2}\|(Ax + By^k - b) - \tfrac{1}{\beta}\lambda^k\|^2 \,\Big|\, x \in \mathcal{X}\Big\} \tag{2.2a}$$

and

$$y^{k+1} = \textit{Argmin}\Big\{\theta_2(y) + \tfrac{\beta}{2}\|(Ax^{k+1} + By - b) - \tfrac{1}{\beta}\lambda^k\|^2 \,\Big|\, y \in \mathcal{Y}\Big\} \tag{2.2b}$$

respectively. Note that the equation (2.1c) *can be written as*

$$(\lambda - \lambda^{k+1})\{(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \ge 0, \ \forall \lambda \in \Re^m. \tag{2.2c}$$

Analysis Note that the solution of (2.1a) and (2.1b) satisfies

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad & \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ & \Big\{-A^T\lambda^k + \beta A^T\big(Ax^{k+1} + By^k - b\big)\Big\} \ge 0, \ \forall\, x \in \mathcal{X} \end{aligned} \tag{2.3a}$$

and

$$\begin{aligned} y^{k+1} \in \mathcal{Y}, \quad & \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \\ & \Big\{-B^T\lambda^k + \beta B^T\big(Ax^{k+1} + By^{k+1} - b\big)\Big\} \ge 0, \ \forall\, y \in \mathcal{Y}, \end{aligned} \tag{2.3b}$$

respectively. Substituting λ^{k+1} (see (2.1c)) in (2.3) (eliminating λ^k in (2.3)), we get

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \Big\{ -A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1}) \Big\} \geq 0, \ \forall\, x \in \mathcal{X}, \tag{2.4a}$$

and

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \big\{ -B^T\lambda^{k+1} \big\} \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{2.4b}$$

For analysis convenience, we rewrite (2.4) as $u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$.

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) \right.$$

$$\left. + \begin{pmatrix} 0 & 0 \\ 0 & \beta B^T B \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\} \geq 0, \ \forall\, (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

Combining the last inequality with (2.2c), we have $w^{k+1} \in \Omega$ and

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \\ Ax^{k+1} + By^{k+1} - b \end{pmatrix} \right.$$

$$\left. +\beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}) + \begin{pmatrix} 0 & 0 \\ \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} \right\} \geq 0, \quad (2.5)$$

for any $w \in \Omega$. The above inequality can be rewritten as $w^{k+1} \in \Omega$ and

$$\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) + \beta \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1})$$

$$\geq \begin{pmatrix} y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix}, \ \forall\, w \in \Omega. \quad (2.6)$$

3 Convergence of ADMM

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

Lemma 3.1 *Let* $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by* (2.1) *from the given* $v^k = (y^k, \lambda^k)$*. Then, we have*

$$(v^{k+1} - v^*)^T H (v^k - v^{k+1}) \geq (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}), \tag{3.1}$$

where

$$\eta(y^k, y^{k+1}) = \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}) \tag{3.2}$$

and

$$H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{3.3}$$

Proof. Setting $w = w^*$ in (2.6), and using H and $\eta(y^k, y^{k+1})$, we get

$$\begin{aligned} (v^{k+1} - v^*)^T H(v^k - v^{k+1}) \\ \geq \quad & (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) \\ & + \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}). \end{aligned} \tag{3.4}$$

Since F is monotone, it follows that

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

The last inequality is due to $w^{k+1} \in \Omega$ and $w^* \in \Omega^*$ (see (1.4)). Substituting it in (3.4), the lemma is proved. $\Box$

Lemma 3.2 *Let* $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by* (2.1) *from the given* $v^k = (y^k, \lambda^k)$*. Then, we have*

$$(w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) = (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}), \tag{3.5}$$

and

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0. \tag{3.6}$$

Proof. By using $\eta(y^k, y^{k+1})$ (see (3.2)), $Ax^* + By^* = b$ and (2.1c), we have

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

Because (2.4b) is true for the k-th iteration and the previous iteration, we have

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T\lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{3.7}$$

and

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T \{-B^T\lambda^k\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{3.8}$$

Setting $y = y^k$ in (3.7) and $y = y^{k+1}$ in (3.8), respectively, and then adding the two resulting inequalities, we get

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

The assertion of this lemma is proved. $\square$

Even though H is positive semi-definite (see (3.3) when B is not full column rank), in this lecture we use $\|v - \tilde{v}\|_H$ to denote that

$$\|v - \tilde{v}\|_H^2 = (v - \tilde{v})^T H (v - \tilde{v}) = \beta \|B(y - \tilde{y})\|^2 + \frac{1}{\beta}\|\lambda - \tilde{\lambda}\|^2.$$

Lemma 3.3 *Let* $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by* (2.1) *from the given* $v^k = (y^k, \lambda^k)$*. Then, we have*

$$(v^{k+1} - v^*)^T H (v^k - v^{k+1}) \geq 0, \quad \forall\, v^* \in \mathcal{V}^*. \tag{3.9}$$

Proof. The assertion follows (3.1), (3.5) and (3.6) directly. $\square$

Theorem 3.1 *Let* $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by* (2.1) *from the given* $v^k = (y^k, \lambda^k)$*. Then, we have*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \ \forall\, v^* \in \mathcal{V}^*. \tag{3.10}$$

Proof. By using (3.9), we have

$$\begin{array}{rcl}\|v^k - v^*\|_H^2 & = & \|(v^{k+1} - v^*) + (v^k - v^{k+1})\|_H^2 \\ & = & \|v^{k+1} - v^*\|_H^2 + 2(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \\ & & \quad + \|v^k - v^{k+1}\|_H^2 \\ & \geq & \|v^{k+1} - v^*\|_H^2 + \|v^k - v^{k+1}\|_H^2,\end{array}$$

and thus (3.10) is proved. $\square$

The inequality (3.10) is essential for the convergence of the alternating direction method. It tells us that the alternating direction method is a contraction method. Multiplying a factor $1/\beta$, it can be written as

$$\left\| \begin{array}{c} B(y^{k+1} - y^*) \\ \frac{1}{\beta}(\lambda^{k+1} - \lambda^*) \end{array} \right\|^2 \leq \left\| \begin{array}{c} B(y^k - y^*) \\ \frac{1}{\beta}(\lambda^k - \lambda^*) \end{array} \right\|^2 - \left\| \begin{array}{c} B(y^k - y^{k+1}) \\ \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) \end{array} \right\|^2, \ \forall\, v^* \in \mathcal{V}^*.$$

This result is included in Theorem 1 of [14] as a special case for fixed β and $\gamma \equiv 1$.

4 The extended Alternating Direction Method

In the extended ADM, the k-th iteration begins with (y^k, λ^k). However, we take the solution of the classical ADM as a predictor, and denote it by $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$.

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

$$\tilde{x}^k = \text{Argmin}\left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k - b) \\ +\frac{\beta}{2}\|Ax + By^k - b\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \tag{4.1a}$$

2. Then, use λ^k and the obtained $\tilde{x}^k$, $\tilde{y}^k$ is the solution of the following problem

$$\tilde{y}^k = \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(A\tilde{x}^k + By - b) \\ +\frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \tag{4.1b}$$

3. Finally,

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b). \tag{4.1c}$$

Based on the predictor $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$, we consider how to produce the new iterate $v^{k+1} = (y^{k+1}, \lambda^{k+1})$ and drive it more close to the set $\mathcal{V}^*$.

According to the same analysis in the last section (see (2.5)) we have $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ and

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - \tilde{y}^k) \right.$$

$$\left. + \begin{pmatrix} 0 & 0 \\ \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \right\} \ge 0, \ \forall w \in \Omega. \tag{4.2}$$

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

Lemma 4.1 *Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ be generated by* (4.1) *from the given $v^k = (y^k, \lambda^k)$. Then, we have*

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \ge (\tilde{w}^k - w^*)^T \eta(y^k, \tilde{y}^k), \tag{4.3}$$

where

$$\eta(y^k, \tilde{y}^k) = \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - \tilde{y}^k) \tag{4.4}$$

and H is the same as defined in (3.3)*.*

Proof. The prof is similar as those for Lemma 3.1 and thus omitted. $\square$

Similarly as in (3.5), by using $\eta(y^k, \tilde{y}^k)$ (see (4.4)) and $Ax^* + By^* = b$, we have

$$(\tilde{w}^k - w^*)^T \eta(y^k, \tilde{y}^k) = (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k). \tag{4.5}$$

Lemma 4.2 *Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ be generated by* (4.1) *from the given $v^k = (y^k, \lambda^k)$. Then, we have*

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \varphi(v^k, \tilde{v}^k), \quad \forall\, v^* \in \mathcal{V}^*. \tag{4.6}$$

and

$$\varphi(v^k, \tilde{v}^k) = \|v^k - \tilde{v}^k\|_H^2 + (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k). \tag{4.7}$$

Proof. It follows from (4.3) and (4.5) that

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k).$$

Assertion (4.6) follows from the last inequality and the definition of $\varphi(v^k, \tilde{v}^k)$ and the lemma is proved. $\square$

Now, we observe the right hand side of (4.6). Note that

$$\begin{aligned} \varphi(v^k, \tilde{v}^k) &= \|v^k - \tilde{v}^k\|_H^2 + (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k) \\ &= \frac{1}{2}\|v^k - \tilde{v}^k\|_H^2 + \frac{1}{2\beta}\|\beta B(y^k - \tilde{y}^k) + (\lambda^k - \tilde{\lambda}^k)\|^2 \\ &\geq \frac{1}{2}\|v^k - \tilde{v}^k\|_H^2. \end{aligned} \tag{4.8}$$

Ye-Yuan's Alternating Direction Method of Multipliers

Ye-Yuan's alternating direction method is a prediction-correction method. The predictor is generated by (4.1). The correction step is to update the new iterate.

Correction Using the $\tilde{v}^k$ produced by (4.1), update the new iterate v^{k+1} by

$$v^{k+1} = v^k - \alpha_k(v^k - \tilde{v}^k), \qquad \alpha_k = \gamma\alpha_k^*, \qquad \gamma \in (0,2) \tag{4.9a}$$

where

$$\alpha_k^* = \frac{\varphi(v^k, \tilde{v}^k)}{\|v^k - \tilde{v}^k\|_H^2} \tag{4.9b}$$

Usually, in comparison with the computational load for obtaining $(\tilde{x}^k, \tilde{y}^k)$ in (4.1), the calculation cost for step-size α_k^* is slight.

We obtain an essential inequality for convergence in the following theorem which was proved by Ye and Yuan in [29].

Theorem 4.1 *Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ be generated by* (4.1) *from the given $v^k = (y^k, \lambda^k)$ and the new iterate v^{k+1} be given by* (4.9). *Then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \frac{\gamma(2-\gamma)}{4}\|v^k - \tilde{v}^k\|_H^2, \ \forall\, v^* \in \mathcal{V}^*. \tag{4.10}$$

Proof. By using (4.6) and (4.9), we obtain

$$
\begin{aligned}
&\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\
&\quad = \quad \|v^k - v^*\|_H^2 - \|(v^k - v^*) - \gamma\alpha_k^*(v^k - \tilde{v}^k)\|_H^2 \\
&\quad \geq \quad 2\gamma\alpha_k^*\varphi(v^k, \tilde{v}^k) - \gamma^2(\alpha_k^*)^2\|v^k - \tilde{v}^k\|_H^2 \\
&\quad = \quad \gamma(2-\gamma)(\alpha_k^*)^2\|v^k - \tilde{v}^k\|_H^2. \qquad (4.11)
\end{aligned}
$$

In addition, it follows from (4.8) and (4.9b) that $\alpha_k^* \geq \frac{1}{2}$. Substituting this fact in (4.11), the theorem is proved. $\square$

Convergence

Both (3.10) and (4.10) can be written as

$$
\left\| \begin{matrix} B(y^{k+1} - y^*) \\ \frac{1}{\beta}(\lambda^{k+1} - \lambda^*) \end{matrix} \right\|^2 \leq \left\| \begin{matrix} B(y^k - y^*) \\ \frac{1}{\beta}(\lambda^k - \lambda^*) \end{matrix} \right\|^2 - c_0 \left\| \begin{matrix} B(y^k - \tilde{y}^k) \\ \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \end{matrix} \right\|^2, \ \forall\, v^* \in \mathcal{V}^*.
$$

It leads to that

$$
\lim_{k\to\infty} By^k = By^* \qquad \text{and} \qquad \lim_{k\to\infty} \lambda^k = \lambda^*.
$$

5 Application and Numerical Experiments

5.1 Calibrating the correlation matrices

We consider to solve the following problem:

$$\min\{\frac{1}{2}\|X - C\|_F^2 \mid X \in S_+^n \cap S_B\}, \tag{5.1}$$

where

$$S_+^n = \{H \in R^{n\times n} \mid H^T = H,\ H \succeq 0\},$$

and

$$S_B = \{H \in R^{n\times n} \mid H^T = H,\ H_L \leq H \leq H_U\}.$$

H_L and H_U are given symmetric matrices.

Use the following Matlab Code to produce the matrices C, H_L and H_U

```
rand('state',0);     C=rand(n,n);  C=(C'+C)-ones(n,n) + eye(n);
%%% C is symmetric and C_{ij} is in (-1,1),  C_{jj} is in (0,2)  %%
HU=ones(n,n)*0.1;  HL=-HU; for i=1:n   HU(i,i)=1;  HL(i,i)=1; end;
```

The problem is converted to the following equivalent one:

$$
\begin{aligned}
\min \quad & \tfrac{1}{2}\|X-C\|^2 + \tfrac{1}{2}\|Y-C\|^2 \\
\text{s.t} \quad & X - Y = 0, \\
& X \in S_+^n, \; Y \in S_B.
\end{aligned}
\tag{5.2}
$$

The basic sub-problems in alternating direction methods for the problem (5.2)

- For fixed Y^k and Z^k,

$$\tilde{X}^k = \text{Argmin}\{\frac{1}{2}\|X-C\|_F^2 - \mathbf{Tr}(Z^k X) + \frac{\beta}{2}\|X - Y^k\|_F^2 \mid X \in S_+^n\}$$

- With fixed $\tilde{X}^k$ and Z^k,

$$\tilde{Y}^k = \text{Argmin}\{\frac{1}{2}\|Y-C\|_F^2 + \mathbf{Tr}(Z^k Y) + \frac{\beta}{2}\|\tilde{X}^k - Y\|_F^2 \mid Y \in S_B\}$$

$\tilde{X}^k$ can be directly obtained via

$$\tilde{X}^k = P_{S_+^n}\{\frac{1}{1+\beta}(\beta Y^k + Z^k + C)\}. \tag{5.3}$$

Note that

$$P_{S_+^n}(A) = U\Lambda^+U^T, \text{ where } \Lambda^+ = \max(\Lambda, 0) \text{ and } [U,\Lambda] = \text{eig}(A).$$

Similarly, $\tilde{Y}^k$ in is given by

$$\tilde{Y}^k = P_{S_B}\{\frac{1}{1+\beta}(\beta\tilde{X}^k - Z^k + C)\}. \tag{5.4}$$

$$S_B = \{H \mid H_L \le H \le H_U\}, \qquad P_{S_B}(A) = \min(\max(H_L, A), H_U)$$

The most time consuming calculation is $[U,\Lambda] = \text{eig}(A)$, $9n^3$

The main Matlab Code of an iteration in the Classical ADM

```
Y0=Y;                          Z0=Z;              k=k+1;
X=(Y0*beta+Z0+C)/(1+beta);
   [V,D]=eig(X);               D=max(0,D);        X=(V*D)*V';
Y=min(max((X*beta-Z0+C)/(1+beta),HL),HU);
Z=Z0-(X-Y)*beta;
```

The main Matlab Code of an iteration in Ye-Yuan's ADM

```
Y0=Y;                              Z0=Z;            k=k+1;
X=(Y0*beta+Z0+C)/(1+beta);
[V,D]=eig(X);                      D=max(0,D);      X=(V*D)*V';
Y=min(max((X*beta-Z0+C)/(1+beta),HL),HU);
%%%%%%%%%    Calculating the step size    %%%%%%%%%%%%%%
EY=Y0-Y;        EZ=(X-Y)*beta;
T1 = EY(:)'*EY(:);  T2 = EZ(:)'*EZ(:);    TA=T1*beta + T2/beta
T2 = (EY(:)'*EZ(:));
alpha=(TA-T2)*gammaY/TA;
Y=Y0-EY*alpha;                                  Z=Z0-EZ*alpha;
```

Numerical results for problem (5.1)

C=rand(n,n); C=(C'+C)-ones(n,n) + eye(n)

H_U=ones(n,n)*0.2; $H_L = -H_U$; $H_U(jj) = H_L(jj)$=1.

Numerical Results for calibrating correlation matrix (Using Matlab EIG)

$n \times n$ Matrix	Classical ADM		Glowinski's ADM		Ye-Yuan's ADM	
$n =$	No. It	CPU Sec.	No. It	CPU Sec.	No. It	CPU Sec.
500	40	18.03	39	17.68	32	14.99
800	41	73.28	39	70.00	33	60.80
1000	43	141.69	42	138.30	34	114.67
1500	47	471.77	45	452.22	41	419.70
2000	55	1254.01	53	1206.94	45	1035.38

Numerical Results for calibrating correlation matrix (Using MeXEIG)

$n \times n$ Matrix	Classical ADM		Glowinski's ADM		Ye-Yuan's ADM	
$n =$	No. It	CPU Sec.	No. It	CPU Sec.	No. It	CPU Sec.
500	40	5.57	39	5.38	32	4.67
800	41	18.13	39	17.15	33	15.13
1000	43	34.75	42	34.00	34	28.50
1500	47	123.77	45	117.87	41	110.17
2000	55	306.32	53	294.75	45	255.72

$$\frac{\text{It. No. of Ye-Yuan's ADM}}{\text{It. No. of Classical ADM}} \approx \frac{5}{6}.$$

It seems that Ye-Yuan's Algorithm converges faster than primary ADM.

5.2 Application for Sparse Covariance Selection

For the details of applications in this subsection, please see the reference [30].

The problem

$$\min_{X} \left\{ \mathbf{Tr}(\Sigma X) - \log(\det(X)) + \rho e^T |X| e \;\middle|\; X \in S_+^n \right\}, \tag{5.5}$$

The equivalent problem:

$$\begin{array}{ll} \min & \mathbf{Tr}(\Sigma X) - \log(\det(X)) + \rho e^T |Y| e \\ \text{s.t} & X - Y = 0, \\ & X \in S_+^n. \end{array} \tag{5.6}$$

For given Y^k and Z^k, get $(\tilde{X}^k, \tilde{Y}^k, \tilde{Z}^k)$ in the following procedure:

1. For fixed Y^k and Z^k, $\tilde{X}^k$ is the solution of the following problem

$$\min\{\mathbf{Tr}(\Sigma X) - \log(\det(X)) - \mathbf{Tr}(Z^k X) + \frac{\beta}{2}\|X - Y^k\|_F^2 \mid X \in S_+^n\}$$

2. Then, with fixed $(\tilde{X}^k, Z^k)$, $\tilde{Y}^k$ is a solution of

$$\min\{\rho e^T|Y|e + \mathbf{Tr}(Z^k Y) + \frac{\beta}{2}\|\tilde{X}^k - Y\|_F^2\}$$

3. Finally, update $\tilde{Z}^k$ by

$$\tilde{Z}^k = Z^k - \beta(\tilde{X}^k - \tilde{Y}^k). \tag{5.7}$$

Solving the X subproblem for getting $\tilde{X}^k$:

$$\tilde{X}^k = \text{Argmin}\Big\{\frac{1}{2}\|X - [Y^k - \frac{1}{\beta}(\Sigma - Z^k)]\|_F^2 - \frac{1}{\beta}\log(\det(X)) \;\Big|\; X \in S_+^n\Big\}. \tag{5.8}$$

It should hold that $\tilde{X} \succ 0$ and thus $\tilde{X}$ is the solution of matrix equation

$$X - \left(Y^k - \frac{1}{\beta}(\Sigma - Z^k)\right) - \frac{1}{\beta}X^{-1} = 0. \tag{5.9}$$

By setting

$$A = Y^k - \frac{1}{\beta}(\Sigma - Z^k), \tag{5.10}$$

and using

$$[V, \Lambda] = \text{eig}(A) \tag{5.11}$$

in Matlab, we get

$$A = V\Lambda V^T, \qquad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n).$$

In fact, the solution of matrix equation (5.9) should have the same eigenvectors as matrix A.

$$\tilde{X} = V\tilde{\Lambda}V^T, \qquad \tilde{\Lambda} = \text{diag}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_n). \tag{5.12}$$

It follows from (5.9) that

$$\tilde{\Lambda} - \Lambda - \frac{1}{\beta}\tilde{\Lambda}^{-1} = 0$$

and thus

$$\tilde{\lambda}_j = \frac{\lambda_j + \sqrt{\lambda_j^2 + (4/\beta)}}{2}, \quad j = 1, \ldots, n.$$

Indeed, $\tilde{\lambda}_j > 0$ and thus $\tilde{X} \succ 0$ (see (5.12)).

The main computational load for getting $\tilde{X}^k$ is the eigenvalues-vectors decomposition in (5.11).

Solving the Y subproblem for getting $\tilde{Y}^k$:

The first order condition for minimization problem

$$\min\{\rho e^T |Y| e + \mathbf{Tr}(Z^k Y) + \frac{\beta}{2} \|\tilde{X}^k - Y\|_F^2\}$$

is

$$0 \in \frac{\rho}{\beta} \partial(|Y|) + Y - (\tilde{X}^k - \frac{1}{\beta} Z^k).$$

In fact,

$$\tilde{Y}^k = (\tilde{X}^k - \frac{1}{\beta} Z^k) - P_{B_\infty^{\rho/\beta}}[\tilde{X}^k - \frac{1}{\beta} Z^k],$$

where $B_{\infty}^{\rho/\beta} = \{X \in \mathbf{R}^{n\times n} \mid -\frac{\rho}{\beta} \le X_{ij} \le \frac{\rho}{\beta}\}$. **The projection on a 'box' is very easy to be carried out !**

5.3 Split feasibility problem and Matrix completion

Applying ADMM to the reformulated split feasibility problem (1.2), the k-th iteration begins with $(y^k, \lambda^k) \in \mathcal{B} \times \Re^m$, and the new iterate is generated by the following procedure:

$$\begin{cases} x^{k+1} = \text{Argmin}\{\frac{1}{2}\|Ax - (y^k + \lambda^k/\beta)\|^2 \mid x \in \mathcal{X}\}, \\ y^{k+1} = P_{\mathcal{B}}[Ax^{k+1} - \lambda^k/\beta], \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - y^{k+1}). \end{cases}$$

The x-subproblem is a problem of form $\min\{\frac{1}{2}\|Ax - a\|^2 \mid x \in \mathcal{X}\}$. After getting x^{k+1}, (y^{k+1}, λ^{k+1}) is obtained by a projection and an evaluation.

Matrix completion is to recover an unknown matrix from a sampling of its entries. For an $m \times n$ matrix M, Ω denotes the indices subset of the matrix

$$\Omega = \{(ij) \mid i \in \{1, 2, \ldots, m\},\ j \in \{1, 2, \ldots, n\}\}.$$

The mathematical form of the considered problem is

$$\boxed{\min\{\|X\|_* \mid X_{ij} = M_{ij},\ (ij) \in \Omega\}}$$

where $\|X\|_*$ is the nuclear norm of X [2]. It can convert to the following problem

$$\begin{array}{rl} \min_{X,Y} & \|X\|_* \\ \text{s. t} & X - Y = 0, \\ & Y_{ij} = M_{ij}, \forall\ (ij) \in \Omega. \end{array}$$

It belongs to the problem (1.1) and was successfully solved by the alternating direction methods [4].

Simple iterative scheme $+$ nice convergence properties
$\Rightarrow$ wide applications of ADMM in large scale optimization

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, Foundations and Trends in Machine Learning, **3**, 1-122, 2010.

[2] J. F. Cai, E. J. Candès and Z. W. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. **20**, 1956-1982, 2010.

[3] R. H. Chan, J. F. Yang, and X. M. Yuan. Alternating direction method for image inpainting in wavelet domain. *SIAM J. Imag. Sci.*, 4:807–826, 2011.

[4] C. H. Chen, B.S. He and X. M. Yuan, Matrix completion via alternating direction methods, IMA Journal of Numerical Analysis, **3**, 227-245, 2012.

[5] E. Esser. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. *CAM Report*, 09-31, UCLA, 2009.

[6] 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.

[7] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximations, Computers and Mathematics with Applications, **2**, 17-40, 1976.

[8] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[9] D. Goldfarb, S.Q. Ma, and K. Scheinberg. Fast alternating linearization methods for minimizing the sum of two convex functions. *Math. Program.*, (2012). DOI 10.1007/s10107-012-0530-2

[10] B.S. He, Inexact implicit methods for monotone general variational inequalities, Math. Program. Series A, **86**, 199-217, 1999.

[11] B.S. He, L-Z Liao, D.R. Han and H. Yang, A new inexact alternating directions method for monotone variational inequalities, Math. Progr., **92**, 103-118, 2002.

[12] B. S. He, M. Tao, and X. M. Yuan. Alternating direction method with Gaussian back substitution for separable convex programming. *SIAM J. Opt.*, 22(2): 313–340, 2012.

[13] B. S. He, M. H. Xu and X. M. Yuan, Solving large-scale least squares covariance matrix problems by alternating direction methods, SIAM J. Matrix Analy. Appli., 32: 136-152, 2011.

[14] B. S. He and H. Yang, Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, Oper. Res. Let., **23**, 151-161, 1998.

[15] B.S. He, H. Yang and S.L. Wang, Alternating directions method with self-adaptive penalty parameters for monotone variational inequalities, JOTA **106**, 349-368, 2000.

[16] B.S. He and X.M. Yuan, On the $O(1/n)$ Convergence Rate of the Douglas-Rachford Alternating Direction Method， SIAM J. Numer. Anal. 50, 700-709, 2012.

[17] Z. Lu, T. K. Pong, and Y. Zhang. An alternating direction method for finding Dantzig selectors. Preprint, 2010.

[18] M. Ng, F. Wang, and X. M. Yuan. Inexact alternating direction methods for image recovery. *SIAM J. Sci. Comput.*, 33:1643–1668, 2011.

[19] M. Ng, P. A. Weiss, and X. M. Yuan. Solving constrained total-variation problems via alternating direction methods. *SIAM J. Sci. Comput.*, 32:2710–2736, 2010.

[20] S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. *Scale space and variational methods in computer vision*, 5567:464–476, 2009.

[21] J. Nocedal and S.J. Wright, Numerical Optimization, Springer Verlag, 1999.

[22] S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. *Scale space and variational methods in computer vision*, 5567:464–476, 2009.

[23] J. Sun and S. Zhang. A modified alternating direction method for convex quadratically constrained quadratic semidefinite programs. *European J. Oper. Res*, 207:1210–1220, 2010.

[24] M. Tao and X. M. Yuan. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J. Opt.*, 21(1):57–81, 2011.

[25] Z. W. Wen, D. Goldfarb, and W. T. Yin. Alternating direction augmented Lagrangian methods for semidefinite programming. *Math. Program. Comput.*, 2:203–230, 2010.

[26] J. F. Yang and Y. Zhang. Alternating direction algorithms for L_1-problems in compressive sensing, *SIAM J. Sci. Comput.,* 38:250–278, 2011

[27] J. F. Yang, Y. Zhang, and W. Yin. A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data. *IEEE J. Selected Topics in Signal Processing*, 4:288–297, 2010

[28] J. F. Yang and X. M. Yuan. Sparese and low-rank matrix decomposition via alternating direction methods. *manuscript*, 2012.

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

[30] X. M. Yuan. Alternating direction method of multipliers for covariance selection models. *J. Sci. Comput.*, 51:261-273, 2012.

[31] X. Q. Zhang, M. Burger, and S. Osher. A unified primal-dual algorithm framework based on Bregman iteration. *J. Sci. Comput.*, 46(1):20–46, 2011.