典型凸优化问题的分裂收缩算法讲座

III. 交替方向法 (ADMM) 及 PPA 意义下的 ADMM

何 炳 生　　南京大学数学系

Homepage: maths.nju.edu.cn/~hebma

江苏省研究生视觉计算与可信人工智能暑期学校

2022 年 7 月 12 日

1 Two blocks separable convex optimization

We consider the following separable convex optimization

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

Example: Best matrix approximation under some conditions

$$\min_X \{\tfrac{1}{2}\|X-C\|_F^2 \mid X\in S_\Lambda^n \cap S_B\},$$

where

$$S_\Lambda^n = \{H\in\mathcal{S}^n \mid \lambda_{\min} I \preceq H \preceq \lambda_{\max} I\}$$

and

$$S_B = \{H\in\mathcal{S}^n \mid H_L \le H \le H_U\}.$$

It can be translated to the following equivalent problem:

$$\begin{array}{rl} \min_{X,Y} & \frac{1}{2}\|X-C\|^2+\frac{1}{2}\|Y-C\|^2 \\ \text{s.t} & X-Y=0,\ X\in S_\Lambda^n,\ Y\in S_B. \end{array} \tag{1.2}$$

The problem (1.2) is a concrete problem of type (1.1).

Smooth Optimization Approach for Covariance Selection — Statistics

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

where C is a given symmetric matrix, $e^T|X|e = \sum_{i=1}^n \sum_{j=1}^n |X_{ij}|$. Its equivalent optimization problem is

$$\begin{aligned} \min_{X,Y} \quad & \mathbf{Tr}(CX) - \log(\det(X)) + \rho e^T |Y| e \\ \text{s.t} \quad & X - Y = 0, \\ & X \in S_+^n,\ Y \in R^{n\times n}. \end{aligned}$$

Low rank and sparse optimization problem in statistics

$$\begin{aligned} \min_{X,Y} \quad & \|X\|_* + \rho e^T |Y| e \\ \text{s.t} \quad & X + Y = H \\ & X,\ Y \in R^{n\times n}. \end{aligned} \tag{1.3}$$

这些矩阵优化的数学模型本身就是一个形如 (1.1) 的结构型优化问题.

2 Mathematical Background

两大基本概念：变分不等式　和　邻近点 (PPA) 算法

Lemma 1 *Let $\mathcal{X} \subset \Re^n$ be a closed convex set, $\theta(x)$ and $f(x)$ be convex functions and $f(x)$ is differentiable. Assume that the solution set of the minimization problem $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ is nonempty. Then,*

$$x^* \in \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{2.1a}$$

if and only if

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.1b}$$

2.1 Linearly constrained convex optimization and VI

The Lagrangian function of the problem (1.1) is

$$L^2(x,y,\lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b).$$

According to Lemma 1, the saddle point is a solution of the following variational inequality:

$$\left\{ \begin{array}{lrl} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, & \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-B^T\lambda^*) \geq 0, & \forall\, y \in \mathcal{Y}, \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T(Ax^* + By^* - b) \geq 0, & \forall\, \lambda \in \Re^m. \end{array} \right.$$

Its compact form is the following variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \qquad (2.2)$$

where

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

and

$$\theta(u) = \theta_1(x) + \theta_2(y), \qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m.$$

Note that the operator F is monotone, because

$$(w - \tilde{w})^T(F(w) - F(\tilde{w})) \geq 0, \text{ Here } (w - \tilde{w})^T(F(w) - F(\tilde{w})) = 0. \quad (2.3)$$

2.2 Preliminaries of PPA for Variational Inequalities

The optimal condition of the problem (1.1) is characterized as a mixed monotone variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \qquad (2.4)$$

PPA for monotone mixed VI in H-norm

For given w^k, find the proximal point w^{k+1} in H-norm which satisfies

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T \{F(w^{k+1}) + H(w^{k+1} - w^k)\} \ge 0, \ \forall\, w \in \Omega, \tag{2.5}$$

where H is a symmetric positive definite matrix.

✠ Again, w^k is the solution of (2.4) if and only if $w^k = w^{k+1}$ ✠

Convergence Property of Proximal Point Algorithm in H-norm

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \tag{2.6}$$

The sequence $\{w^k\}$ is Fejér monotone in H-norm. In customized PPA, via choosing a proper positive definite matrix H, the solution of the subproblem (2.5) has a closed form. An iterative algorithm is called the contraction method, if its generated sequence $\{w^k\}$ satisfies $\|w^{k+1} - w^*\|_H^2 < \|w^k - w^*\|_H^2$.

2.3 Augmented Lagrangian Method (ALM)

We consider the convex optimization, namely

$$\min\{\theta(u) \,|\, \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{2.7}$$

The related variational inequality of the saddle point of the Lagrangian function is

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{2.8a}$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \ \text{and} \quad \Omega = \mathcal{U} \times \Re^m. \tag{2.8b}$$

Augmented Lagrangian Method

The augmented Lagrangian function of the problem (2.7) is

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

The k-th iteration of the **Augmented Lagrangian Method** [15, 18] begins with a given λ^k, obtain $w^{k+1} = (u^{k+1}, \lambda^{k+1})$ via

$$\text{(ALM)} \quad \left\{ \begin{array}{l} u^{k+1} = \arg\min\{\mathcal{L}_\beta(u, \lambda^k) \mid u \in \mathcal{U}\}, \qquad \text{(2.9a)} \\ \lambda^{k+1} = \lambda^k - \beta(\mathcal{A}u^{k+1} - b). \qquad \text{(2.9b)} \end{array} \right.$$

In (2.9), u^{k+1} is only a computational result of (2.9a) from given λ^k, it is called the intermediate variable. In order to start the k-th iteration of ALM, we need only to have λ^k and thus we call it as the essential variable.

The subproblem (2.9a) is a problem of mathematical form

$$\min\{\theta(u) + \tfrac{\beta}{2}\|\mathcal{A}u - p^k\|^2 \mid u \in \mathcal{U}\} \qquad (2.10)$$

where $\beta > 0$ is a given scalar and $p^k = b + \frac{1}{\beta}\lambda^k$.

Assumption: The solution of problem (2.10) has closed-form solution or can be efficiently computed with a high precision.

The optimal condition of (2.9) can be written as $w^{k+1} \in \Omega = \mathcal{U} \times \Re^m$ and

$$\begin{cases} \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T\{-\mathcal{A}^T\lambda^k + \beta\mathcal{A}^T(\mathcal{A}u^{k+1} - b)\} \geq 0, \ \forall u \in \mathcal{U}, \\ \qquad (\lambda - \lambda^{k+1})^T\{(\mathcal{A}u^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0, \quad \forall \lambda \in \Re^m. \end{cases}$$

The above relations can be written as

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} u - u^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} -\mathcal{A}^T\lambda^{k+1} \\ \mathcal{A}u^{k+1} - b \end{pmatrix} \geq (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}),$$

for all $w \in \Omega$. Using the notations in (2.8), we get the compact form

$$\begin{aligned} \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq \quad (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}), \ \forall w \in \Omega. \end{aligned} \tag{2.11}$$

Setting $w = w^*$ in (2.11), we get

$$(\lambda^{k+1} - \lambda^*)^T(\lambda^k - \lambda^{k+1}) \geq \beta\{\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1})\}.$$

By using the monotonicity of F and the optimality of w^*, it follows that

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

Thus, we have

$$(\lambda^{k+1} - \lambda^*)^T (\lambda^k - \lambda^{k+1}) \geq 0. \tag{2.12}$$

By using the above inequality, we obtain

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

It means that

$$\|\lambda^{k+1} - \lambda^*\|^2 \leq \|\lambda^k - \lambda^*\|^2 - \|\lambda^k - \lambda^{k+1}\|^2. \tag{2.13}$$

The above inequality is the key for the convergence proof of the Augmented Lagrangian Method.

3 ADMM for two-block problems

Recall the separable convex optimization problem

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

The augmented Lagrangian function

$$\mathcal{L}_\beta(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2.$$

Applied ALM to solve the problem (1.1), the k-th iteration begins with given λ^k,

$$\left\{ \begin{array}{l} (x^{k+1}, y^{k+1}) = \arg\min\big\{\mathcal{L}_\beta(x, y, \lambda^k) \;\big|\; x \in \mathcal{X}, y \in \mathcal{Y}\big\}, \qquad (3.1a) \\ \qquad \lambda^{k+1} \in \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \qquad (3.1b) \end{array} \right.$$

ADMM is a relaxed ALM for the problem (1.1),the k-th iteration begins with given (y^k, λ^k),

$$\left\{ \begin{array}{ll} x^{k+1} \in \arg\min\big\{\mathcal{L}_\beta(x, y^k, \lambda^k) \;\big|\; x \in \mathcal{X}\big\}, & (3.2a) \\ y^{k+1} \in \arg\min\big\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \;\big|\; y \in \mathcal{Y}\big\}, & (3.2b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & (3.2c) \end{array} \right.$$

两个可分离目标函数问题的 ADMM 方法 [3, 4]

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

$$x^{k+1} = \text{Argmin}\left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k - b) \\ +\frac{\beta}{2}\|Ax + By^k - b\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \quad (3.3a)$$

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

$$y^{k+1} = \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(Ax^{k+1} + By - b) \\ +\frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \quad (3.3b)$$

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \quad (3.3c)$$

Advantages

The x and y sub-problems are separately solved one by one.

Remark Ignoring the constant term in the objective function, the sub-problems (3.22a) and (3.22b) is equivalent to

$$x^{k+1} = \text{Argmin}\big\{\theta_1(x) + \tfrac{\beta}{2}\|(Ax + By^k - b) - \tfrac{1}{\beta}\lambda^k\|^2 \big| x \in \mathcal{X}\big\} \qquad \text{(3.4a)}$$

and

$$y^{k+1} = \text{Argmin}\big\{\theta_2(y) + \tfrac{\beta}{2}\|(Ax^{k+1} + By - b) - \tfrac{1}{\beta}\lambda^k\|^2 \big| y \in \mathcal{Y}\big\} \qquad \text{(3.4b)}$$

respectively. Note that the equation (3.3c) can be written as

$$(\lambda - \lambda^{k+1})\{(Ax^{k+1} + By^{k+1} - b) + \tfrac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0, \ \forall \lambda \in \Re^m. \qquad \text{(3.4c)}$$

Notice that the sub-problems (3.4a) and (3.4b) are the type of

$$x^{k+1} = \text{Argmin}\big\{\theta_1(x) + \tfrac{\beta}{2}\|Ax - p^k\|^2 \big| x \in \mathcal{X}\big\}$$

and

$$y^{k+1} = \text{Argmin}\big\{\theta_2(y) + \tfrac{\beta}{2}\|By - q^k\|^2 \big| y \in \mathcal{Y}\big\},$$

respectively. (子问题求解有困难怎么处理放在后面讲)

Analysis According to Lemma 1, the solution of (3.22a) and (3.22b) satisfies

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \Big\{-A^T\lambda^k + \beta A^T\big(Ax^{k+1} + By^k - b\big)\Big\} \geq 0, \ \forall\, x \in \mathcal{X} \tag{3.5a}$$

and

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \Big\{-B^T\lambda^k + \beta B^T\big(Ax^{k+1} + By^{k+1} - b\big)\Big\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{3.5b}$$

respectively. Substituting $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$ (see (3.3c)) in (3.5) (eliminating λ^k in (3.5)), we get

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \Big\{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1})\Big\} \geq 0, \ \forall\, x \in \mathcal{X}, \tag{3.6a}$$

and

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{3.6b}$$

The compact form of (3.6) is $u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$ and

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T \lambda^{k+1} \\ -B^T \lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^T B \\ 0 \end{pmatrix} \big(y^k - y^{k+1}\big) \right\} \geq 0, \tag{3.7}$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

By adding and subtracting the term $\beta B^T B (y^k - y^{k+1})$, we rewrite the about variational inequality in our desirable form

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

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

Combining the last inequality with (3.4c), we have the following lemma.

Lemma 2 *Let* $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\in\Omega$ *be generated by* (3.3) *with given* (y^k,λ^k)*, then we have*

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

$$+\begin{pmatrix} 0 & 0 \\ \beta B^TB & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix}\begin{pmatrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix}\Biggr\}\geq 0, \ \forall\, w\in\Omega. \quad (3.8)$$

For convenience we use the notations

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

Then, we get the following lemma:

Lemma 3 *Let the sequence* $\{w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})\} \in \Omega$ *be generated by* (3.3)*. Then, we have*

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq (y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}), \ \forall\, w^* \in \Omega^*, \tag{3.9}$$

where

$$H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{3.10}$$

Proof. Setting $w = w^*$ in (3.8), we get

$$
\begin{aligned}
&(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \\
\geq \quad & \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} \beta B(y^k - y^{k+1}) \\
& + \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}), \ \forall\, w^* \in \Omega^*. \quad (3.11)
\end{aligned}
$$

Observe the first part of the right hand side of (3.11),

$$
\begin{aligned}
& \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} \beta B(y^k - y^{k+1}) \\
= \quad & (y^k - y^{k+1})^T B^T \beta \big(A, B\big) \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix} \\
= \quad & (y^k - y^{k+1})^T B^T \beta \big(Ax^{k+1} + By^{k+1} - (Ax^* + By^*)\big) \\
= \quad & (y^k - y^{k+1}) B^T \underline{\beta \big(Ax^{k+1} + By^{k+1} - b\big)} \\
= \quad & (y^k - y^{k+1}) B^T \underline{(\lambda^k - \lambda^{k+1})}. \qquad (3.12)
\end{aligned}
$$

To the second part, since $(w^{k+1}-w^*)^T F(w^{k+1}) = (w^{k+1}-w^*)^T F(w^*)$ and w^* is the optimal solution, it follows that

$$\begin{aligned} & \theta(u^{k+1}) - \theta(u^*) + (w^{k+1}-w^*)^T F(w^{k+1}) \\ = \; & \theta(u^{k+1}) - \theta(u^*) + (w^{k+1}-w^*)^T F(w^*) \ge 0. \end{aligned} \tag{3.13}$$

The assertion (3.11) immediately. □

Lemma 4 *Let the sequence* $\{w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})\} \in \Omega$ *be generated by* (3.3). *Then, we have*

$$(y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}) \ge 0. \tag{3.14}$$

Proof. Because (3.6b) is true for the k-th iteration and the previous iteration, we have

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \ge 0, \ \forall\, y \in \mathcal{Y}, \tag{3.15}$$

and

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T\{-B^T\lambda^k\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{3.16}$$

Setting $y = y^k$ in (3.15) and $y = y^{k+1}$ in (3.16), respectively, and then adding the two resulting inequalities, we get the assertion (3.14) immediately. □

Substituting (3.14) in (3.9), we get

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0, \ \ \forall\, v^* \in \mathcal{V}^*. \tag{3.17}$$

Using the above inequality, as in the last lecture, we have the following theorem, which is the key for the proof of the convergence of ADMM.

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

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \ \forall\, v^* \in \mathcal{V}^*. \tag{3.18}$$

交替方向法收敛性证明的　再阐述

交替方向法处理的是两个可分离块的凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\}. \tag{3.19}$$

将其拉格朗日函数 $L(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)$ 的鞍点归结为等价的变分不等式的解点：

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{3.20a}$$

其中

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix},\quad \Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m. \tag{3.20b}$$

ADMM 的 k 步迭代从给定的核心变量 $v^k=(y^k,\lambda^k)$ 出发

$$\left\{\begin{array}{rcl} x^{k+1} &=& \arg\min\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k-b\|^2 \mid x\in\mathcal{X}\}, \quad (3.21a)\\ y^{k+1} &=& \arg\min\{\theta_2(y)-y^TB^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2 \mid y\in\mathcal{Y}\}, \quad (3.21b)\\ \lambda^{k+1} &=& \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). \quad (3.21c)\end{array}\right.$$

根据最优性引理 1, ADMM k-步迭代满足

$$\left\{\begin{array}{ll} x^{k+1}\in\mathcal{X}, & \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{-A^T\lambda^k+\beta A^T(Ax^{k+1}+By^k-b)\}\ge 0,\ \forall x\in\mathcal{X},\\ y^{k+1}\in\mathcal{Y}, & \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{-B^T\lambda^k+\beta B^T(Ax^{k+1}+By^{k+1}-b)\}\ge 0,\ \forall y\in\mathcal{Y},\\ \lambda^{k+1}\in\Re^m, & (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0,\ \forall\lambda\in\Re^m.\end{array}\right.$$

利用 $\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b)$ 上面的式子可以整理改写成

$$\left\{\begin{array}{lll} x^{k+1}\in\mathcal{X}, & \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{-A^T\lambda^{k+1}+\beta A^TB(y^k-y^{k+1})\}\ge 0,\ \forall x\in\mathcal{X}, & \text{(3.22a)}\\ y^{k+1}\in\mathcal{Y}, & \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{-B^T\lambda^{k+1}\}\ge 0,\ \forall y\in\mathcal{Y}, & \text{(3.22b)}\\ \lambda^{k+1}\in\Re^m, & (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0,\ \forall\lambda\in\Re^m. & \text{(3.22c)}\end{array}\right.$$

在 (3.22b) 的后半部加上和为零的两项, 得到

$$\left\{\begin{array}{l} \underline{\theta_1(x)-\theta_1(x^{k+1})}+(x-x^{k+1})^T\{-A^T\lambda^{k+1}+\beta A^TB(y^k-y^{k+1})\}\ge 0,\\ \underline{\theta_2(y)-\theta_2(y^{k+1})}+(y-y^{k+1})^T\{-B^T\lambda^{k+1}+\underbrace{\beta B^TB(y^k-y^{k+1})+\beta B^TB(y^{k+1}-y^k)}\}\ge 0,\\ \qquad(\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0.\end{array}\right.$$

利用变分不等式 (3.20), 进行合理整合, 得到

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

将上式中那个任意的 w, 设成解点 w^* 便有

$$\underline{\theta(u^*)-\theta(u^{k+1})}+(w^*-w^{k+1})^T F(w^{k+1})$$
$$+\begin{pmatrix} x^*-x^{k+1} \\ y^*-y^{k+1}\end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k-y^{k+1}) + \begin{pmatrix} y^*-y^{k+1} \\ \lambda^*-\lambda^{k+1}\end{pmatrix}^T \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix}\begin{pmatrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} \ge 0.$$

经转换, 得到

$$\begin{pmatrix} y^{k+1}-y^* \\ \lambda^{k+1}-\lambda^*\end{pmatrix}^T \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix}\begin{pmatrix} y^k-y^{k+1} \\ \lambda^k-\lambda^{k+1} \end{pmatrix}$$

后面记 $v=\begin{pmatrix} y \\ \lambda \end{pmatrix}$, $H=\begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix}$

$$\ge \begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^*\end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k-y^{k+1}) + \underbrace{[\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^T F(w^{k+1})]}. \quad (3.23)$$

假如 (3.23) 式右端非负, 证明就基本上完成了. 由于

$$\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1})=\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^*)\geq 0.$$

(3.23) 式右端下划线部分非负. 因此从 (3.23) 式得到

$$(v^{k+1}-v^*)^TH(v^k-v^{k+1}) \;\geq\; \begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^* \end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k-y^{k+1}). \tag{3.24}$$

对 (3.24) 式的右端进行处理, 有

$$\begin{aligned}
\begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^* \end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k-y^{k+1}) &= (y^k-y^{k+1})^TB^T\beta(A,B)\begin{pmatrix} x^{k+1}-x^* \\ y^{k+1}-y^* \end{pmatrix} \\
&= (y^k-y^{k+1})^TB^T\beta\big(Ax^{k+1}+By^{k+1}-(Ax^*+By^*)\big) \qquad \text{利用}(Ax^*+By^*=b) \\
&= (y^k-y^{k+1})B^T\beta\big(Ax^{k+1}+By^{k+1}-b\big) \\
&= (y^k-y^{k+1})B^T(\lambda^k-\lambda^{k+1}).
\end{aligned} \tag{3.25}$$

后面我们要证明 $(y^k-y^{k+1})B^T(\lambda^k-\lambda^{k+1})\geq 0.$

利用 (3.22b) 有

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

和

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

（将任意的 y 分别设成 y^k 和 y^{k+1}）

$$\theta_2(y^k) - \theta_2(y^{k+1}) + (y^k - y^{k+1})^T\{-B^T\lambda^{k+1}\} \geq 0.$$

$$\theta_2(y^{k+1}) - \theta_2(y^{k}) + (y^{k+1} - y^{k})^T\{-B^T\lambda^{k}\} \geq 0.$$

(将上面两式相加, 就有) $(y^k - y^{k+1})B^T(\lambda^k - \lambda^{k+1}) \geq 0.$ ((3.25) 式右端非负)

证明了(3.25) 式右端非负, 进而得到 (3.24) 式右端非负. 所以

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0. \tag{3.26}$$

Lemma 2 告诉我们:

$$b^T H(a-b) \geq 0 \quad \Rightarrow \quad \|b\|_H^2 \leq \|a\|_H^2 - \|a-b\|_H^2. \tag{3.27}$$

在 (3.27) 中置 $a = (v^k - v^*)$ 和 $b = (v^{k+1} - v^*)$, 根据 (3.26) 就得到收敛的关键不等式

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2.$$

由 $\|v^k - v^{k+1}\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2$ 得 $\sum_{k=0}^{\infty} \|v^k - v^{k+1}\|_H^2 \leq \|v^0 - v^*\|_H^2.$

How to choose the parameter β. The efficiency of ADMM is heavily dependent on the parameter β in (3.3). We discuss how to choose a suitable β in the practical computation.

Note that if $\boldsymbol{\beta A^T B(y^k - y^{k+1}) = 0}$, then it follows from (3.7)

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} \ge 0, \ \forall (x,y) \in \mathcal{X} \times \mathcal{Y}.$$

In this case, if additionally $\boldsymbol{Ax^{k+1} + By^{k+1} - b = 0}$, then we have

$$\left\{ \begin{array}{rl} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T(-A^T\lambda^{k+1}) \ge 0, & \forall x \in \mathcal{X} \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T(-B^T\lambda^{k+1}) \ge 0, & \forall y \in \mathcal{Y} \\ (\lambda - \lambda^{k+1})^T(Ax^{k+1} + By^{k+1} - b) \ge 0, & \forall \lambda \in \Re^m \end{array} \right.$$

and consequently $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is a solution of the VI (2.2).

In other words, $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is not a solution of (2.2) because

$$\beta A^T B(y^k - y^{k+1}) \neq 0 \qquad \text{and/or} \qquad Ax^{k+1} + By^{k+1} - b \neq 0.$$

We call

$$\|\boldsymbol{\beta A^T B(y^k - y^{k+1})}\| \qquad \text{and} \qquad \|\boldsymbol{Ax^{k+1} + By^{k+1} - b}\|$$

the primal-residual and the dual-residual, respectively. It seems that we should balance the primal and the dual residuals dynamically. If

$$\mu\|\beta A^T B(y^k - y^{k+1})\| < \|Ax^{k+1} + By^{k+1} - b\| \quad \text{with a} \quad \mu > 1,$$

it means that the dual residual is too large and thus we should enlarge the parameter β in the augmented Lagrangian function. Otherwise, we should reduce the parameter β. A simple scheme that often works well is (see, e.g., [10]):

$$\beta_{k+1} = \begin{cases} \beta_k * \tau, & \text{if } \mu\|\beta A^T B(y^k - y^{k+1})\| < \|Ax^{k+1} + By^{k+1} - b\|; \\ \beta_k/\tau, & \text{if } \|\beta A^T B(y^k - y^{k+1})\| > \mu\|Ax^{k+1} + By^{k+1} - b\|; \\ \beta_k, & \text{otherwise.} \end{cases}$$

where $\mu > 1, \tau > 1$ are parameters. Typical choices might be $\mu = 10$ and $\tau = 2$. The idea behind this penalty parameter update is to try to keep the primal and dual residual norms within a factor of μ of one another as they both converge to zero. This self adaptive adjusting rule has been used by S. Boyd's group [1] and in their Optimization Solver [5].

4 Customized PPA for Variational Inequality

We study the algorithms using the guidance of variational inequality. The optimal condition of the linearly constrained convex optimization is resulted in a variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \qquad (4.1)$$

4.1 Customized PPA for VI (4.1)

[Prediction Step.] With given v^k, find a vector $\tilde{w}^k \in \Omega$ which satisfying

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \ \forall w \in \Omega, \qquad (4.2a)$$

where the matrix H is positive definite.

[Correction Step.] Determine a nonsingular matrix M and a scalar $\alpha > 0$, let

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0, 2). \qquad (4.2b)$$

v is a part of the elements of the vector w, $\quad v = w$ is also possible.

Problem: $w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \ \forall\, w \in \Omega.$

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H (v^k - \tilde{v}^k), \ \forall w \in \Omega.$$

Prediction-Correction $\qquad v_\alpha^{k+1} = v^k - \alpha(v^k - \tilde{v}^k).$

H 范数矩阵 ***G*** 效益矩阵

$$H \succ 0, \qquad G = (2 - \alpha) H \succ 0.$$

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \alpha \|v^k - \tilde{v}^k\|_G^2.$$

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \alpha(2 - \alpha) \|v^k - \tilde{v}^k\|_H^2.$$

4.2 Convergence proof

We prove the following main convergence property.

Theorem 1 *Let* $\{v^k\}$ *be the sequence generated by* (4.2) *for the problem* (4.1) *and* $\tilde{w}^k$ *is obtained from* (4.2a). *Then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha(2-\alpha)\|v^k - \tilde{v}^k\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \qquad (4.3)$$

where $\mathcal{V}^* = \{v^* \mid v^* \text{is a part of } w^*, w^* \in \Omega^*\}$.

Proof. Setting $w = w^*$ in (4.2a), we get

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k), \ \forall w^* \in \Omega^*.$$

By using $(\tilde{w}^k - w^*)^T F(\tilde{w}^k) = (\tilde{w}^k - w^*)^T F(w^*)$ and the optimality of w^*, we have

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \ge 0, \quad \forall v^* \in \mathcal{V}^*.$$

It can be written as

$$\{(v^k - v^*) - (v^k - \tilde{v}^k)\}^T H(v^k - \tilde{v}^k) \ge 0, \quad \forall v^* \in \mathcal{V}^*,$$

and thus

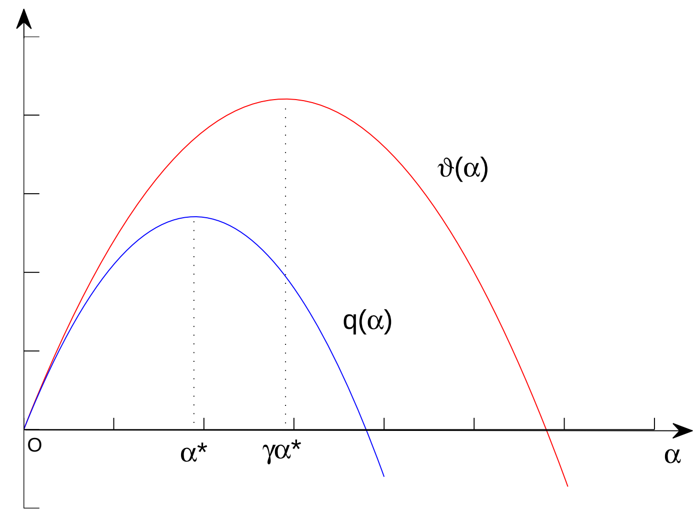
$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \ge \|v^k - \tilde{v}^k\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \qquad (4.4)$$

Let

$$\vartheta(\alpha) = \|v^k - v^*\|_H^2 - \|v_\alpha^{k+1} - v^*\|_H^2.$$

It follows that

$$\begin{aligned}\vartheta(\alpha) &= \|v^k - v^*\|_H^2 - \|v_\alpha^{k+1} - v^*\|_H^2 \\ &= \|v^k - v^*\|_H^2 \\ &\quad -\|(v^k - v^*) - \alpha(v^k - \tilde{v}^k)\|_H^2 \\ &= 2\alpha(v^k - v^*)^T H(v^k - \tilde{v}^k) \\ &\quad -\alpha^2\|v^k - \tilde{v}^k\|_H^2. \end{aligned} \tag{4.5}$$



取 $\gamma \in [1, 2)$ 的示意图

Using (4.4), we get

$$\begin{aligned}\vartheta(\alpha) &\geq 2\alpha\|v^k - \tilde{v}^k\|_H^2 - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \\ &:= q(\alpha) \end{aligned} \tag{4.6}$$

The assertion (4.3) follows from (4.5) and (4.6) immediately. □

我们本想极大化 $\vartheta(\alpha)$, 虽然 $\vartheta(\alpha)$ 是 α 的二次函数, 但线性项系数 $2(v^k - v^*)^T H(v^k - \tilde{v}^k)$ 中含有未知的 v^*, 利用 (4.4), 得到 $\vartheta(\alpha)$ 的下界函数 $q(\alpha)$. 极大化 $q(\alpha)$, $\alpha_k^* \equiv 1$. 可以松弛延拓.

5 Applications for separable problems

5.1 ADMM in PPA-sense

根据 PPA 算法的要求　设计的右端矩阵为对称正定. 具体算法可参阅 [19]

In order to solve the separable convex optimization problem (1.1), we construct a method whose prediction-step is

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{5.1a}$$

where

$$H = \begin{pmatrix} \beta B^T B + \delta I_{n_2} & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad \text{(a small } \delta > 0\text{).} \tag{5.1b}$$

Since H is positive definite, we can use the update form of Algorithm I to produce the new iterate $v^{k+1} = (y^{k+1}, \lambda^{k+1})$. (In the algorithm [2], we took $\delta = 0$).

The concrete form of (5.1) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + (\boldsymbol{\beta B^T B + \delta I_{n_2}})(\tilde{y}^k - y^k) \boldsymbol{- B^T}(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b}) \quad \boldsymbol{-B}(\tilde{y}^k - y^k) \quad + \quad \boldsymbol{(1/\beta)}\,(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

The underline part is $F(\tilde{w}^k)$:

$$\boldsymbol{F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}}$$

In fact, the prediction can be arranged by

$$\begin{cases} \tilde{x}^k \in \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(5.2a)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), & \text{(5.2b)} \\ \tilde{y}^k \in \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T [2\tilde{\lambda}^k - \lambda^k] \\ +\frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\}. & \text{(5.2c)} \end{cases}$$

这个预测与经典的交替方向法 (3.3) 相当, 采用(4.2b) 校正, 会加快速度.

According to Lemma 1, the solution of (5.2a), $\tilde{x}^k$ satisfies

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) \\ + (x - \tilde{x}^k)^T\{-A^T\lambda^k + \beta A^T(A\tilde{x}^k + By^k - b)\} \geq 0, \quad \forall x \in \mathcal{X}.$$

By using (5.2b), $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, the above variational inequality can be written as

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\} \geq 0, \quad \forall x \in \mathcal{X}.$$

The equation (5.2b) can written as

$$(\underline{A\tilde{x}^k + B\tilde{y}^k - b}) \boldsymbol{- B}(\tilde{y}^k - y^k) + \boldsymbol{(1/\beta)}(\tilde{\lambda}^k - \lambda^k) = 0.$$

The remainder part of the prediction (5.2c), namely,

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \{\underline{-B^T\tilde{\lambda}^k} + \boldsymbol{(\beta B^T B + \delta I_{n_2})}(\tilde{y}^k - y^k) \boldsymbol{- B^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0$$

can be achieved by

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) - y^T B^T[2\tilde{\lambda}^k - \lambda^k] + \frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \mid y \in \mathcal{Y}\}.$$

如果把 (5.2) 中取 $\delta = 0$, 并将其输出记为 $(x^{k+1}, \lambda^{k+1}, y^{k+1})$, 则迭代式为

$$\left\{\begin{array}{ll} x^{k+1} \in \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, & \text{(5.3a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), & \text{(5.3b)} \\ y^{k+1} \in \text{Argmin}\big\{\theta_2(y) - y^T B^T [2\lambda^{k+1} - \lambda^k] + \frac{\beta}{2}\|B(y - y^k)\|^2 \bigm| y \in \mathcal{Y}\big\}. & \text{(5.3c)} \end{array}\right.$$

注意在 (5.3c) 中,

$$\begin{aligned} y^{k+1} &\in \text{Argmin}\big\{\theta_2(y) - y^T B^T [2\lambda^{k+1} - \lambda^k] + \frac{\beta}{2}\|B(y - y^k)\|^2 \bigm| y \in \mathcal{Y}\big\} \\ &= \text{Argmin}\big\{\theta_2(y) - y^T B^T \lambda^{k+1} - y^T B^T (\lambda^{k+1} - \lambda^k) + \frac{\beta}{2}\|B(y - y^k)\|^2 \bigm| y \in \mathcal{Y}\big\} \\ &= \text{Argmin}\big\{\theta_2(y) - y^T B^T \lambda^{k+1} + \frac{\beta}{2}\|B(y - y^k) - \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\|^2 \bigm| y \in \mathcal{Y}\big\} \\ &= \text{Argmin}\big\{\theta_2(y) - y^T B^T \lambda^{k+1} + \frac{\beta}{2}\|B(y - y^k) + \frac{1}{\beta}(\lambda^k - \lambda^{k+1})\|^2 \bigm| y \in \mathcal{Y}\big\} \\ &= \text{Argmin}\big\{\theta_2(y) - y^T B^T \lambda^{k+1} + \frac{\beta}{2}\|B(y - y^k) + (Ax^{k+1} + By^k - b)\|^2 \bigm| y \in \mathcal{Y}\big\} \\ &= \text{Argmin}\big\{\theta_2(y) - y^T B^T \lambda^{k+1} + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \bigm| y \in \mathcal{Y}\big\}. \end{aligned}$$

所以, (5.3) 就是

$$
\left\{
\begin{array}{ll}
x^{k+1} \in \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(5.4a)} \\
\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), & \text{(5.4b)} \\
y^{k+1} \in \text{Argmin}\{\theta_2(y) - y^T B^T \lambda^{k+1} + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\}. & \text{(5.4c)}
\end{array}
\right.
$$

请注意, 经典的 ADMM 是

$$
\begin{array}{rcl}
x^{k+1} & \in & \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, \\
y^{k+1} & \in & \text{Argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\}, \\
\lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b).
\end{array}
$$

所以, (5.3), 就是交换了 y, λ 顺序的交替方向法. 由于可以采用

$$
v^{k+1} := v^k - \alpha(v^k - v^{k+1}), \quad \alpha \in (0, 2).
$$

通常取 $\alpha = 1.5$, 收敛更快.

5.2 Linearized ADMM-Like Method

当子问题 (5.2c) 求解有困难时, 用 $\frac{s}{2}\|y-y^k\|^2$ 代替 $\frac{1+\delta}{2}\beta\|B(y-y^k)\|^2$.

By using the linearized version of (5.2c), the prediction step becomes

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T H(v^k-\tilde{v}^k), \ \forall w\in\Omega, \tag{5.5}$$

where

$$H=\begin{bmatrix} sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}, \text{ 代替 (5.1) 中的 } \begin{bmatrix} (1+\delta)\beta B^TB & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}. \tag{5.6}$$

The concrete formula of (5.5) is

$$\begin{cases} \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k}\}\ge 0, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k}+\ \boldsymbol{s}(\tilde{y}^k-y^k)\boldsymbol{-B^T}(\tilde{\lambda}^k-\lambda^k)\}\ge 0, \\ (\underline{A\tilde{x}^k+B\tilde{y}^k-b})\boldsymbol{-B}(\tilde{y}^k-y^k)+\boldsymbol{(1/\beta)}(\tilde{\lambda}^k-\lambda^k)=0. \end{cases} \tag{5.7}$$

The underline part is $F(\tilde{w}^k)$:

$$\boldsymbol{F(w)}=\begin{pmatrix} -\boldsymbol{A^T\lambda} \\ -\boldsymbol{B^T\lambda} \\ \boldsymbol{Ax+By-b} \end{pmatrix}$$

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (5.7),

we need only use the following procedure:

$$\left\{\begin{array}{l} \tilde{x}^k \in \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\theta_2(y) - y^T B^T[\mathbf{2\tilde{\lambda}^k - \lambda^k}] + \dfrac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \end{array}\right. \tag{5.8}$$

用 $\frac{s}{2}\|y-y^k\|^2$ 代替 $\frac{1}{2}(\beta\|B(y-y^k)\|^2 + \delta\|y-y^k\|^2)$, 为保证收敛,

需要 $s > \beta\|B^T B\|$. 对给定的 $\beta > 0$, 太大的 s 会影响收敛速度.

只有当由二次项 $\frac{1}{2}\beta\|B(y-y^k)\|^2$ 引发求解困难, 才用线性化方法.

Then, we use the form

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0, 2)$$

to update the new iterate v^{k+1}.

6 Solving the primal subproblem in parallel

根据 PPA 算法的要求 设计的右端矩阵为对称正定.

Primal-Dual Order

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \;\; \forall w \in \Omega, \tag{6.1a}$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & 0 & A^T \\ 0 & \beta B^T B + \delta I_{n_2} & B^T \\ A & B & \frac{2}{\beta} I_m \end{pmatrix}. \tag{6.1b}$$

The both matrices

$$\begin{pmatrix} \beta A^T A + \delta I_{n_1} & A^T \\ A & \frac{1}{\beta} I_m \end{pmatrix} \succ 0, \qquad \begin{pmatrix} \beta B^T B + \delta I_{n_2} & B^T \\ B & \frac{1}{\beta} I_m \end{pmatrix} \succ 0.$$

The concrete form of (6.1) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_{n_1})}(\tilde{x}^k - x^k) \boldsymbol{+ A^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + \boldsymbol{(\beta B^T B + \delta I_{n_2})}(\tilde{y}^k - y^k) \boldsymbol{+ B^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b}) \boldsymbol{+ A}(\tilde{x}^k - x^k) \boldsymbol{+ B}(\tilde{y}^k - y^k) + \boldsymbol{(2/\beta)}\, (\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

整理一下得到

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T\lambda^k + (\beta A^T A + \delta I_{n_1})(\tilde{x}^k - x^k)\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{-B^T\lambda^k + (\beta B^T B + \delta I_{n_2})(\tilde{y}^k - y^k)\} \geq 0, \\ [2(A\tilde{x}^k + B\tilde{y}^k - b) - (Ax^k + By^k - b)] + (2/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

In fact, the prediction can be arranged by

$$\left\{\begin{array}{ll} \tilde{x}^k = \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T \lambda^k \\ +\frac{1}{2}\beta\|A(x-x^k)\|^2 + \frac{1}{2}\delta\|x-x^k\|^2 \end{array} \middle| x \in \mathcal{X} \right\} & \text{(6.2a)} \\ \\ \tilde{y}^k = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k \\ +\frac{1}{2}\beta\|B(y-y^k)\|^2 + \frac{1}{2}\delta\|y-y^k\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} & \text{(6.2b)} \\ \tilde{\lambda}^k = \lambda^k - \frac{1}{2}\beta\big[2(A\tilde{x}^k + B\tilde{y}^k - b) - (Ax^k + By^k - b)\big] & \text{(6.2c)} \end{array}\right.$$

$$\left\{\begin{array}{l} \tilde{x}^k = \arg\min\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}(x-x^k)^T(\beta A^T A + \delta I_{n_1})(x-x^k) | x \in \mathcal{X}\big\} \\ \tilde{y}^k = \arg\min\big\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}(y-y^k)^T(\beta B^T B + \delta I_{n_2})(y-y^k) | y \in \mathcal{Y}\big\} \\ \tilde{\lambda}^k = \lambda^k - \frac{1}{2}\beta\big[2(A\tilde{x}^k + B\tilde{y}^k - b) - (Ax^k + By^k - b)\big] \end{array}\right.$$

$$w^{k+1} = w^k - \alpha(w^k - \tilde{w}^k), \qquad \alpha \in (0,2).$$

Dual-Primal Order

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \;\; \forall w \in \Omega, \tag{6.3a}$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & 0 & -A^T \\ 0 & \beta B^T B + \delta I_{n_2} & -B^T \\ -A & -B & \frac{2}{\beta} I_m \end{pmatrix}. \tag{6.3b}$$

The both matrices

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & -A^T \\ -A & \frac{1}{\beta} I_m \end{pmatrix} \succ 0, \quad \begin{pmatrix} \beta B^T B + \delta I_{n_2} & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix} \succ 0.$$

The concrete form of (6.3) is

$$\left\{\begin{array}{l}\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T \\ \qquad\qquad \{\underline{-A^T\tilde{\lambda}^k}+\boldsymbol{(\beta A^T A+\delta I_{n_1})}(\tilde{x}^k-x^k)\boldsymbol{-A^T}(\tilde{\lambda}^k-\lambda^k)\}\ge 0, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T \\ \qquad\qquad \{\underline{-B^T\tilde{\lambda}^k}+\boldsymbol{(\beta B^T B+\delta I_{n_2})}(\tilde{y}^k-y^k)\boldsymbol{-B^T}(\tilde{\lambda}^k-\lambda^k)\}\ge 0, \\ (\underline{A\tilde{x}^k+B\tilde{y}^k-b})\boldsymbol{-A}(\tilde{x}^k-x^k)\boldsymbol{-B}(\tilde{y}^k-y^k)+\boldsymbol{(2/\beta)}\,(\tilde{\lambda}^k-\lambda^k)=0.\end{array}\right.$$

经整理归并一下得到

$$\left\{\begin{array}{l}\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\{-A^T(2\tilde{\lambda}^k-\lambda^k) \\ \qquad\qquad\qquad\qquad +(\beta A^T A+\delta I_{n_1})(\tilde{x}^k-x^k)\}\ge 0, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T\{-B^T(2\tilde{\lambda}^k-\lambda^k) \\ \qquad\qquad\qquad\qquad +(\beta B^T B+\delta I_{n_2})(\tilde{y}^k-y^k)\}\ge 0, \\ (Ax^k+By^k-b)+(2/\beta)(\tilde{\lambda}^k-\lambda^k)=0.\end{array}\right.$$

In fact, the prediction can be arranged by

$$
\left\{
\begin{array}{ll}
\tilde{\lambda}^k = \lambda^k - \frac{1}{2}\beta(Ax^k + By^k - b), & \text{(6.4a)} \\[2ex]
\tilde{x}^k \in \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T[2\tilde{\lambda}^k - \lambda^k] \\ +\frac{1}{2}\beta\|A(x - x^k)\|^2 + \frac{1}{2}\delta\|x - x^k\|^2 \end{array} \;\middle|\; x \in \mathcal{X} \right\} & \text{(6.4b)} \\[2ex]
\tilde{y}^k \in \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T[2\tilde{\lambda}^k - \lambda^k] \\ +\frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \end{array} \;\middle|\; y \in \mathcal{Y} \right\}. & \text{(6.4c)}
\end{array}
\right.
$$

$$
w^{k+1} = w^k - \alpha(w^k - \tilde{w}^k), \qquad \alpha \in (0, 2).
$$

我们关于 ADMM 的研究, 始于 1997 年, 第一篇 ADMM 方面的论文发表于 1998 年. 这一讲中 §4-§6 介绍的 ADMM 类方法, 可以从 [19] 中找到.

利用变分不等式 (VI) 和邻近点算法 (PPA), 更自由地设计 ADMM 类分裂收缩算法

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 Vol. 3, No. 1 (2010) 1 – 122.

[2] 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.

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

[4] R. Glowinski and A. Marrocco, Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problémes non linéaires. RAIRO Anal. Numer. R2, 41–76 (1975)

[5] D. Hallac, Ch. Wong, S. Diamond, A. Sharang, R. Sosič, S. Boyd and J. Leskovec, SnapVX: A Network-Based Convex Optimization Solver, Journal of Machine Learning Research 18 (2017) 1-5.

[6] 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.

[7] B.S. He, L.Z. Liao and M.J. Qian, Alternating projection based prediction-correction methods for structured variational inequalities, Journal of Computational Mathematics, 24(6), 693-710, 2006.

[8] B. S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman-Rachford splitting method for convex programming, *SIAM Journal on Optimization* **24**(2014), 1011-1040.

[9] B. S. He and H. Yang, Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, *Operations Research Letters* **23**(1998), 151–161.

[10] 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.

[11] B. S. He and X. M. Yuan, On the $O(1/t)$ convergence rate of the Douglas-Rachford Alternating Direction Method, *SIAM J. Numerical Analysis* **50**(2012), 700-709.

[12] 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.

[13] 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.

[14] B. S. He and X. M. Yuan, Balanced Augmented Lagrangian Method for Convex Programming, arXiv 2108.08554 [math.OC]

[15] M. R. Hestenes, Multiplier and gradient methods, *JOTA* **4**, 303-320, 1969.

[16] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.* **15** (2004), 229–251.

[17] J. Nocedal and S. J. Wright, Numerical Optimization, Springer Verlag, New York, 1999.

[18] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in Optimization, R. Fletcher, ed., Academic Press, New York, NY, pp. 283-298, 1969.

[19] 利用统一框架设计凸优化的分裂收缩算法, 高等学校计算数学学报, 2022, 44: 1-35.