ADMM 类算法的一些最新进展（II）

利用统一框架设计多种用途的分裂算法

统一处理线性等式和不等式约束

直接推广求解多块可分离凸优化问题

B. S. He, S. J. Xu and X. M. Yuan:
Extensions of ADMM for Separable Convex Optimization Problems with Linear Equality or Inequality Constraints, arXiv:2107.01897v2[math.OC]

何 炳 生

南京大学数学系

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

1 Mathematical Background

1.1 Optimization problem and VI

Let $\Omega \subset \Re^n$, we consider the convex minimization problem

$$\min\{f(x) \mid x \in \Omega\}. \tag{1.1}$$

What is the first-order optimal condition ?

$x^* \in \Omega^* \quad \Leftrightarrow \quad x^* \in \Omega$ and any feasible direction is not a descent one.

Optimal condition in variational inequality form

- $S_d(x^*) = \{s \in \Re^n \mid s^T \nabla f(x^*) < 0\} \;=\;$ Set of the descent directions.
- $S_f(x^*) = \{s \in \Re^n \mid s = x - x^*,\; x \in \Omega\} \;=\;$ Set of feasible directions.

$x^* \in \Omega^* \quad \Leftrightarrow \quad x^* \in \Omega$ **and** $S_f(x^*) \cap S_d(x^*) = \emptyset$.

瞎子爬山判定山顶的准则是: 所有可行方向都不再是上升方向

The optimal condition can be presented in a variational inequality (VI) form:

$$x^* \in \Omega, \quad (x - x^*)^T F(x^*) \geq 0, \quad \forall x \in \Omega, \tag{1.2}$$

where $F(x) = \nabla f(x)$. For general VI, F is an operator from $\Re^n$ to itself.

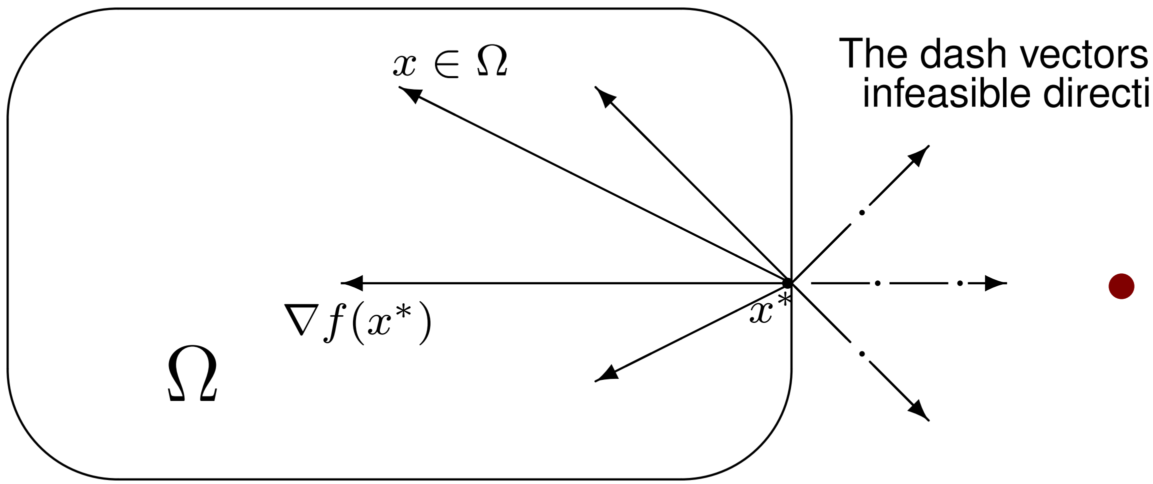


Fig. 1.1 Differential Convex Optimization and VI

Since $f(x)$ is a convex function, we have

$f(y) \geq f(x) + \nabla f(x)^T (y - x)$ and thus $(x - y)^T (\nabla f(x) - \nabla f(y)) \geq 0.$

We say the gradient ∇f of the convex function f is a monotone operator.

通篇我们需要用到的大学数学　主要是基于微积分学的一个引理

$$\min\{\theta(x)|x\in\mathcal{X}\},\quad x^*\in\mathcal{X},\qquad \theta(x)-\theta(x^*)\ge 0,\qquad \forall\, x\in\mathcal{X};$$

$$\min\{f(x)|x\in\mathcal{X}\},\quad x^*\in\mathcal{X},\quad (x-x^*)^T\nabla f(x^*)\ge 0,\quad \forall\, x\in\mathcal{X}.$$

上面的凸优化最优性条件是最基本的, 合在一起就是下面的引理:

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,*

$$x^*\in\arg\min\{\theta(x)+f(x)\,|\,x\in\mathcal{X}\} \tag{1.3a}$$

if and only if

$$x^*\in\mathcal{X},\;\; \theta(x)-\theta(x^*)+(x-x^*)^T\nabla f(x^*)\ge 0,\;\; \forall\, x\in\mathcal{X}. \tag{1.3b}$$

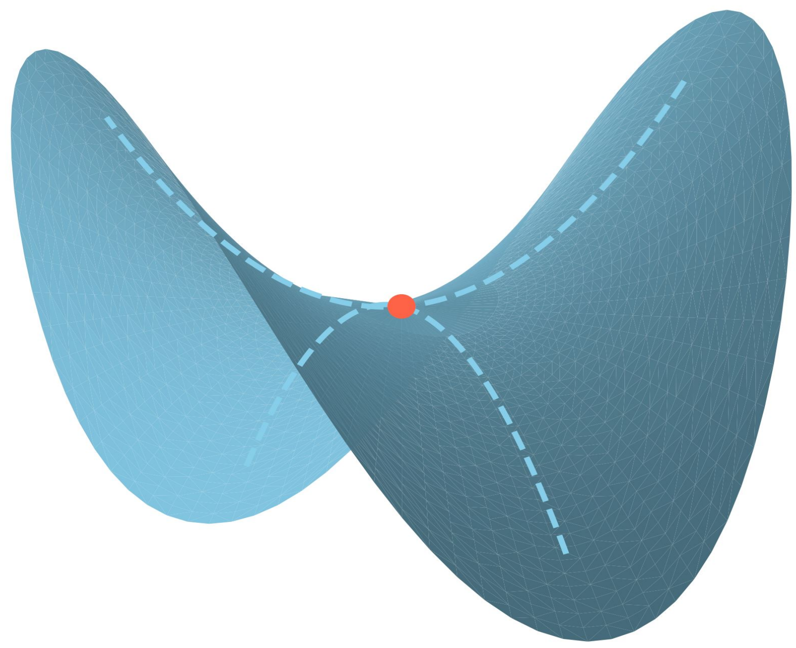
Linearly constrained Optimization in form of VI

We consider the linearly constrained convex optimization problem

$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{1.4}$$

The Lagrange function of (1.4) is

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \qquad (u,\lambda) \in \mathcal{U} \times \Re^m. \tag{1.5}$$



Saddle Point of the Lagrange Function

A pair of $(u^*, \lambda^*) \in \mathcal{U} \times \Re^m$ is called a saddle point if

$$L_{\lambda \in \Re^m}(u^*, \lambda) \leq L(u^*, \lambda^*) \leq L_{u \in \mathcal{U}}(u, \lambda^*).$$

The above inequalities can be written as

$$\begin{cases} u^* \in \mathcal{U}, \quad L(u, \lambda^*) - L(u^*, \lambda^*) \geq 0, \ \ \forall\, u \in \mathcal{U}, & \text{(1.6a)} \\ \lambda^* \in \Re^m, \ \ L(u^*, \lambda^*) - L(u^*, \lambda) \geq 0, \ \ \forall\, \lambda \in \Re^m. & \text{(1.6b)} \end{cases}$$

According to the definition of $L(u, \lambda)$ (see(1.5)), we get the following variational inequality:

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

Using a more compact form, the saddle-point can be characterized as the solution of the following VI:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega. \tag{1.7}$$

where

$$w = \left(\begin{array}{c} u \\ \lambda \end{array} \right), \quad F(w) = \left(\begin{array}{c} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{array} \right) \text{ and } \quad \Omega = \mathcal{U} \times \Re^m. \tag{1.8}$$

Because F is a affine operator and

$$F(w) = \begin{pmatrix} 0 & -\mathcal{A}^T \\ \mathcal{A} & 0 \end{pmatrix} \begin{pmatrix} u \\ \lambda \end{pmatrix} - \begin{pmatrix} 0 \\ b \end{pmatrix},$$

the matrix is skew-symmetric, and we have

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

Convex optimization problem with two separable functions

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\, x\in\mathcal{X}, y\in\mathcal{Y}\}. \tag{1.9}$$

The same analysis tells us that the saddle point is a solution of the following VI:

$$w^*\in\Omega, \quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0, \quad \forall\, w\in\Omega. \tag{1.10a}$$

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

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix}, \quad u=\begin{pmatrix} x\\ y\end{pmatrix}, \quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}, \tag{1.10b}$$

and $\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m$. The affine operator $F(w)$ has the form

$$F(w)=\begin{pmatrix} 0 & 0 & -A^T\\ 0 & 0 & -B^T\\ A & B & 0\end{pmatrix}\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix}-\begin{pmatrix} 0\\ 0\\ b\end{pmatrix}.$$

Again, we have

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

Convex optimization problem with three separable functions

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

is a special problem of (1.4) with three blocks. The Lagrangian function is

$$L^3(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b).$$

The same analysis tells us that the saddle point is a solution of the following VI:

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega.$$

where

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

$$\theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z)\quad \text{and}\quad \Omega=\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\times\Re^m.$$

1.2 Proximal point algorithms and its Beyond

Lemma 2 *Let the vectors* $a, b \in \Re^n$, $H \in \Re^{n\times n}$ *be a positive definite matrix. If* $b^T H(a-b) \ge 0$, *then we have*

$$\|b\|_H^2 \le \|a\|_H^2 - \|a-b\|_H^2. \tag{1.11}$$

The assertion follows from $\|a\|_H^2 = \|b+(a-b)\|_H^2 \ \ge\ \|b\|_H^2 + \|a-b\|_H^2$.

1.2.1 Proximal point algorithms for convex optimization

$$\text{(Convex Optimization)} \qquad \min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}. \tag{1.12}$$

Basic formula of the PPA (not too far away from the last iterate):

$$\text{(PPA)} \qquad x^{k+1} = \text{Argmin}\{\theta(x) + f(x) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \tag{1.13}$$

Nice property of the Proximal Point Algorithm:

$$\|x^{k+1} - x^*\|^2 \le \|x^k - x^*\|^2 - \|x^k - x^{k+1}\|^2. \tag{1.14}$$

1.2.2 Preliminaries of PPA for Variational Inequalities

The optimal condition of the linearly constrained convex optimization is characterized as a mixed monotone variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{1.15}$$

PPA for VI (1.15) in H-norm For given w^k and $H \succ 0$, find w^{k+1},

$$\begin{aligned} w^{k+1} \in \Omega, \;\; \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w^{k+1}) \\ &\geq (w - w^{k+1})^T H(w^k - w^{k+1}), \;\; \forall\, w \in \Omega. \end{aligned} \tag{1.16}$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (1.15).

w^k is the solution of (1.15) if and only if $w^k = w^{k+1}$

Setting $w = w^*$ in (1.16), we obtain

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

Note that (see the structure of $F(w)$ in (1.8))

$$(w^{k+1} - w^*)^T F(w^{k+1}) = (w^{k+1} - w^*)^T F(w^*),$$

and consequently (by using (1.15)) we obtain

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

Thus, we have

$$(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \geq 0. \tag{1.17}$$

By setting $\boldsymbol{a = w^k - w^*}$ and $\boldsymbol{b = w^{k+1} - w^*}$,

the inequality (1.17) means that $b^T H(a - b) \geq 0$.

By using Lemma 2, we obtain

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \tag{1.18}$$

We get the nice convergence property of Proximal Point Algorithm for VI.

2 Splitting Methods in a Unified Framework

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:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{2.1}$$

2.1 Algorithms in a unified framework for VI (3.1)

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{2.2a}$$

where the matrix Q has the property: $Q^T + Q$ is positive definite.

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

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \tag{2.2b}$$

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

Convergence Conditions

For the matrices Q and M, there is a positive definite matrix H such that

$$HM = Q. \tag{2.3a}$$

For the given H, M and Q satisfied the condition (2.3a), and the step size α determined in (2.2), the matrix

$$G = Q^T + Q - \alpha M^T HM \succ 0. \tag{2.3b}$$

Convergence using the unified framework

Theorem 1 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (3.1) *and $\tilde{w}^k$ is obtained in the k-th iteration. If v^k, v^{k+1} and $\tilde{w}^k$ satisfy the conditions in the unified framework, then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha\|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*, \tag{2.4}$$

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

2.2 Convergence proof in the unified framework

In this section, assuming the conditions (2.3) in the unified framework are satisfied, we prove some convergence properties

Theorem 2 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (3.1) *and* $\tilde{w}^k$ *is obtained in the* k*-th iteration. If* v^k*,* v^{k+1} *and* $\tilde{w}^k$ *satisfy the conditions in the unified framework, then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha\|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.5}$$

Proof. Using $Q = HM$ (see (2.3a)), the prediction can be written as

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

By using relation (2.2b), $v^k - v^{k+1} = \alpha M(v^k - \tilde{v}^k)$, we get

$$\alpha\{\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\} \ge (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \ \forall w \in \Omega.$$

Setting $w = w^*$ in the above inequality, we get

$$(v^k - v^{k+1})^T H(\tilde{v}^k - v^*) \ge \alpha\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}, \ \forall w^* \in \Omega^*.$$

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

$$(v^k - v^{k+1})^T(\tilde{v}^k - v^*) \ge 0, \quad \forall v^* \in \mathcal{V}^*. \tag{2.6}$$

Setting $a = v^k$, $b = v^{k+1}$, $c = \tilde{v}^k$ and $d = v^*$, in the identity

$$2(a-b)^T H(c-d) = \{\|a-d\|_H^2 - \|b-d\|_H^2\} - \{\|a-c\|_H^2 - \|b-c\|_H^2\},$$

it follows from (2.6) that

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2. \tag{2.7}$$

For the right hand side of the last inequality, we have

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&\quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\
&\quad \overset{(2.2b)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - \alpha M(v^k - \tilde{v}^k)\|_H^2 \\
&\quad = \quad 2\alpha(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - \alpha^2(v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\
&\quad = \quad \alpha(v^k - \tilde{v}^k)^T(Q^T + Q - \alpha M^T HM)(v^k - \tilde{v}^k) \\
&\quad \overset{(2.3b)}{=} \quad \alpha\|v^k - \tilde{v}^k\|_G^2.
\end{aligned} \tag{2.8}$$

Substituting (2.8) in (2.7), the assertion of this theorem is proved. $\square$

3 Special prediction-correction methods

We study the optimization algorithms using the guidance of variational inequality.

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \qquad (3.1)$$

3.1 Algorithms $Q = H$, H is positive definite

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \ \forall w \in \Omega, \qquad (3.2a)$$

where the matrix H is symmetric and positive definite.

[Correction Step.] The new iterate v^{k+1} by

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0, 2) \qquad (3.2b)$$

H is a symmetric positive definite matrix.

Since $G=(2-\alpha)H, \quad \alpha\|v^k-\tilde{v}^k|_G^2=\alpha(2-\alpha)\|v^k-\tilde{v}^k|_H^2.$

The sequence $\{v^k\}$ generated by the prediction-correction method (3.2) satisfies

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

The above inequality is the Key for convergence analysis !

Set $\alpha=1$ in (2.2b), the prediction (3.2a) becomes: $w^{k+1} \in \Omega$ such that

$$\theta(u)-\theta(u^{k+1})+(w-w^{k+1})^T F(w^{k+1}) \ge (v-v^{k+1})^T H(v^k-v^{k+1}), \ \forall w \in \Omega.$$

The generated sequence $\{v^k\}$ satisfies

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2-\|v^k-v^{k+1}\|_H^2. \quad \forall v^* \in \mathcal{V}^*.$$

上式是跟 (1.18) 类似的不等式, 是关于核心变量 v 的 PPA 方法.

3.2 Applications for separable problems

This section presents various applications of the proposed algorithms for the separable convex optimization problem

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{3.1}$$

Its VI-form is

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{3.2}$$

where

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ \\ y \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \tag{3.3a}$$

and

$$\theta(u) = \theta_1(x) + \theta_2(y), \qquad\qquad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{3.3b}$$

The augmented Lagrangian Function of the problem (3.1) is

$$\mathcal{L}_\beta(x,y,\lambda) = \theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)+\frac{\beta}{2}\|Ax+By-b\|^2. \quad (3.4)$$

Solving the problem (3.1) by using ADMM, the k-th iteration begins with given (y^k,λ^k), it offers the new iterate (y^{k+1},λ^{k+1}) via

$$\text{(ADMM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x,y^k,\lambda^k) \;\big|\; x\in\mathcal{X}\big\}, & (3.5a) \\ y^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x^{k+1},y,\lambda^k) \;\big|\; y\in\mathcal{Y}\big\}, & (3.5b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1}+By^{k+1}-b). & (3.5c) \end{array} \right.$$

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

The main convergence result is

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - \|v^k-v^{k+1}\|_H^2, \quad \forall\, v^*\in\mathcal{V}^*$$

where

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

Ignoring some constant term in the objective function, ADMM (3.5) **is implemented by**

(ADMM)
$$\begin{cases} x^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T p^k \\ + \frac{\beta}{2}\|A(x - x^k)\|^2. \end{array} \,\middle|\, x \in \mathcal{X} \right\}, & (3.6a) \\ y^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T q^k \\ + \frac{\beta}{2}\|B(y - y^k)\|^2. \end{array} \,\middle|\, y \in \mathcal{Y} \right\}, & (3.6b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & (3.6c) \end{cases}$$

where

$$\begin{aligned} p^k &= \lambda^k - \beta(Ax^k + By^k - b), \\ q^k &= \lambda^k - \beta(Ax^{k+1} + By^k - b). \end{aligned}$$

3.3 ADMM in PPA-sense

根据 Special 算法的要求 设计的右端矩阵为对称正定.

In order to solve the separable convex optimization problem (3.1), we construct a method whose prediction-step is

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{3.7a}$$

where

$$H = \begin{pmatrix} \beta B^T B + \delta I_{n_2} & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad \text{(a small } \delta > 0\text{, say } \delta = 0.05\text{)}. \tag{3.7b}$$

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 (3.7) 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 $F(\tilde{w}^k)$:

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

In fact, the prediction can be arranged by

$$\begin{cases} \tilde{x}^k \in \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, & (3.8a) \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), & (3.8b) \\ \tilde{y}^k \in \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T [2\tilde{\lambda}^k - \lambda^k] \\ +\frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\}. & (3.8c) \end{cases}$$

这个预测与经典的交替方向法 (3.5) 相当, 采用(2.2b) 校正, 会加快速度.

According to Lemma 1, the solution of (3.8a), $\tilde{x}^k$ satisfies

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

By using (3.8b), $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, the above variational inequality can be written as

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

The equation (3.8b) 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 (3.7), namely,

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

can be achieved by

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

3.4 Linearized ADMM-Like Method

当子问题 (3.8c) 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 (3.8), the prediction step becomes

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\geq(v-\tilde{v}^k)^TH(v^k-\tilde{v}^k),\ \forall w\in\Omega, \quad (3.9)$$

where

$$H=\begin{bmatrix} sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}, \text{ 代替 (3.7) 中的 } \begin{bmatrix} (1+\delta)\beta B^TB & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}. \quad (3.10)$$

The concrete formula of (3.9) is

$$\begin{cases} \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k}\}\geq 0, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k}+\ \boldsymbol{s}(\tilde{y}^k-y^k)\boldsymbol{-B^T}(\tilde{\lambda}^k-\lambda^k)\}\geq 0, \\ (\underline{A\tilde{x}^k+B\tilde{y}^k-b})\boldsymbol{-B}(\tilde{y}^k-y^k)+\boldsymbol{(1/\beta)}(\tilde{\lambda}^k-\lambda^k)=0. \end{cases} \quad (3.11)$$

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

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

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}.

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (3.11), we need only use the following procedure:

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

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 $s > \beta\|B^TB\|$. 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.

4 Wider Application & Easy Extensions

Let us consider the general separable convex optimization model

$$\min \left\{\theta_1(x) + \theta_2(y) \mid Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\right\}. \tag{4.1}$$

The augmented Lagrangian function is

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

4.1 From ALM to ADMM

Augmented Lagrangian Method for (4.1). From λ^k to λ^{k+1}:

$$\left\{\begin{array}{rcl} (x^{k+1}, y^{k+1}) & \in & \arg\min\{\mathcal{L}_\beta(x, y, \lambda^k) \,|\, x \in \mathcal{X}, y \in \mathcal{Y}\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array}\right. \tag{4.2}$$

ADMM for (4.1) From (y^k, λ^k) to (y^{k+1}, λ^{k+1})

$$\left\{ \begin{array}{lcl} x^{k+1} & \in & \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} & \in & \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array} \right. \tag{4.3}$$

From (4.2) to (4.3), ADMM is a relaxed ALM.

ADMM is designed for equality constraints problems.

The direct extension of ADMM is not necessarily convergent !

Ignoring some constant terms in the objective functions of the corresponding subproblems, we can rewrite the ADMM (4.3) as

$$\left\{\begin{array}{l} x^{k+1} \in \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2}\|A(x-x^k)\|^2 \mid x \in \mathcal{X}\}, \\ y^{k+1} \in \operatorname{argmin}\left\{\begin{array}{c} \theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \\ \frac{\beta}{2}\|A(x^{k+1}-x^k) + B(y-y^k)\|^2 \end{array} \;\middle|\; y \in \mathcal{Y}\right\}, \\ \lambda^{k+1} = \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} - b\big) \end{array}\right. \tag{4.4}$$

where

$$\lambda^{k+\frac{1}{2}} := \lambda^k - \beta(Ax^k + By^k - b).$$

The λ update form can be also denoted by

$$\lambda^{k+1} = P_{\Re^m}\big[\lambda^k - \beta\big(Ax^{k+1} + By^{k+1} - b\big)\big].$$

为了说明我们后面提出的方法和 ADMM 的关系,
我们把经典的 ADMM 改写成等价的 (4.4).

4.2 ADMM with wider applications

Let us consider the general two-block separable convex optimization model

$$\min \left\{\theta_1(x) + \theta_2(y) \mid Ax + By = b \text{ (or } \ge b), x \in \mathcal{X}, y \in \mathcal{Y}\right\}. \tag{4.5}$$

The linear constraints can be a system of linear equations or linear inequalities. We define

$$\Lambda = \begin{cases} \Re^m, & \text{if } Ax + By = b, \\ \Re^m_+, & \text{if } Ax + By \ge b. \end{cases}$$

The projection on Λ is denoted by $P_\Lambda[\cdot]$.

For such special Λ, the projection on Λ is clear !

The only difference: $\quad P_{\Re^m}(\lambda) = \lambda, \quad P_{\Re^m_+}(\lambda) = \max\{\lambda, 0\}.$

4.2.1 Primal-dual extension of ADMM with wider application

A Primal-Dual Extension of the ADMM for (4.5).

From (Ax^k, By^k, λ^k) to $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$:

1. (Prediction Step) With given (Ax^k, By^k, λ^k), find $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via

$$\begin{cases} \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \mid y \in \mathcal{Y}\big\}, \\ \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(A\tilde{x}^k + B\tilde{y}^k - b\big)\big]. \end{cases} \tag{4.6a}$$

2. (Correction Step) Generate the new iterate $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} Ax^{k+1} \\ By^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} Ax^k \\ By^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 \\ 0 & \nu I_m & 0 \\ -\nu\beta I_m & 0 & I_m \end{pmatrix} \begin{pmatrix} Ax^k - A\tilde{x}^k \\ By^k - B\tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{4.6b}$$

这是一类预测–校正方法. 需要额外的校正, 但校正花费很小！

预测先做 Primal 部分, 再做 Dual 部分, 顺序也可以倒过来.

4.2.2 Dual-Primal extension of ADMM with wider application

A Dual-Primal Extension of the ADMM for (4.5).

From (Ax^k, By^k, λ^k) to $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$:

1. (Prediction Step) With given (Ax^k, By^k, λ^k), find $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via

$$\left\{\begin{array}{l} \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(Ax^k + By^k - b\big)\big], \\[2mm] \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^T A^T \tilde{\lambda}^k + \frac{1}{2}\beta\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\[2mm] \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^T B^T \tilde{\lambda}^k + \frac{1}{2}\beta\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \mid y \in \mathcal{Y}\big\}. \end{array}\right. \tag{4.7a}$$

2. (Correction Step) Generate the new iterate $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} Ax^{k+1} \\ By^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} Ax^k \\ By^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 \\ 0 & \nu I_m & 0 \\ -\beta I_m & -\beta I_m & I_m \end{pmatrix} \begin{pmatrix} Ax^k - A\tilde{x}^k \\ By^k - B\tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{4.7b}$$

预测采用不同顺序, 校正公式也略有不同. 校正同样是花费很小的.

无论是 primal-dual, 还是 dual-primal 方法, 都可以向多块问题直接推广.

5 p-block separable convex optimization problems

In the following we consider the multiple-block convex optimization:

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i) \;\Big|\; \sum_{i=1}^{p}A_i x_i = b \;(\text{or} \ge b), \;\; x_i\in\mathcal{X}_i\Big\}. \tag{5.1}$$

The Lagrangian function is

$$L(x_1,\ldots,x_p,\lambda)=\sum_{i=1}^{p}\theta_i(x_i)-\lambda^T\Big(\sum_{i=1}^{p}A_i x_i-b\Big),$$

which is defined on $\Omega=\prod_{i=1}^{p}\mathcal{X}_i\times\Lambda$, where

$$\Lambda=\begin{cases}\Re^m, & \text{if } \sum_{i=1}^{p}A_i x_i=b,\\[2ex] \Re^m_+, & \text{if } \sum_{i=1}^{p}A_i x_i\ge b.\end{cases}$$

Let $(x_1^*, \ldots, x_p^*, \lambda^*) \in \Omega$ be a saddle point of the Lagrangian function, then

$$L_{\lambda\in\Lambda}(x_1^*, \ldots, x_p^*, \lambda) \le L(x_1^*, \ldots, x_p^*, \lambda^*) \le L_{x_i\in\mathcal{X}_i}(x_1, \ldots, x_p, \lambda^*).$$

The optimality condition of (5.1) can be written as the following VI:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \qquad \text{(5.2a)}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_p^T\lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}, \qquad \text{(5.2b)}$$

and

$$\theta(x) = \sum_{i=1}^p \theta_i(x_i), \qquad \Omega = \prod_{i=1}^p \mathcal{X}_i \times \Lambda.$$

Again, we denote by Ω^* the solution set of the VI (5.2).

5.1 Primal-dual extension of the ADMM for p-block Problems

A Primal-Dual Extension of the ADMM for (5.1) **Prediction Step** .

From $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ to $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$:

With given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, find $\tilde{w}^k \in \Omega$ via

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\big\{\theta_1(x_1) - x_1^TA_1^T\lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\big\};\\
\tilde{x}_2^k \in \arg\min\big\{\theta_2(x_2) - x_2^TA_2^T\lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\big\};\\
\qquad\vdots\\
\tilde{x}_i^k \in \arg\min_{x_i\in\mathcal{X}_i}\big\{\theta_i(x_i) - x_i^TA_i^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1}A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\big\};\\
\qquad\vdots\\
\tilde{x}_p^k \in \arg\min_{x_p\in\mathcal{X}_p}\big\{\theta_p(x_p) - x_p^TA_p^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\big\};\\
\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^{p}A_j\tilde{x}_j^k - b\big)\big].
\end{array}\right. \tag{5.3}$$

预测先原始再对偶. 对可分离的原始变量子问题逐一按序求解.

A Primal-Dual Extension of the ADMM for (5.1) **Correction Step** .

From $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ to $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$:

Generate the new iterate $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \quad (5.4)$$

对照一下就可以发现, §5.1 中的方法, 就是 §4.2.1方法的直接推广.

校正非常简单, 工作量也很小. 把校正公式分开来写就是:

$Ax_i^{k+1}, i = 1, \ldots, p$

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \end{pmatrix} - \nu \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \end{pmatrix}, \tag{5.5}$$

λ^{k+1}

$$\lambda^{k+1} = \tilde{\lambda}^k + \nu\beta(A_1x_1^k - A_1\tilde{x}_1^k). \tag{5.6}$$

还能说校正不简单！？

5.2 Dual-primal extension of the ADMM for (5.1)

A Dual-Primal Extension of the ADMM for (5.1) **Prediction Step** .

From $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ to $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$:

With given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, find $\tilde{w}^k \in \Omega$ via

$$\left\{\begin{array}{l}
\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^p A_jx_j^k - b\big)\big] \\
\tilde{x}_1^k \in \arg\min\big\{\theta_1(x_1) - x_1^TA_1^T\tilde{\lambda}^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\big\}; \\
\tilde{x}_2^k \in \arg\min\big\{\theta_2(x_2) - x_2^TA_2^T\tilde{\lambda}^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\big\}; \\
\qquad\vdots \\
\tilde{x}_i^k \in \arg\min_{x_i\in\mathcal{X}_i}\{\theta_i(x_i) - x_i^TA_i^T\tilde{\lambda}^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1}A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\
\qquad\vdots \\
\tilde{x}_p^k \in \mathrm{argmin}_{x_p\in\mathcal{X}_p}\{\theta_p(x_p) - x_p^TA_p^T\tilde{\lambda}^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}.
\end{array}\right. \tag{5.7}$$

预测先对偶再原始. 对可分离的原始变量子问题逐一按序求解.

A Dual-Primal Extension of the ADMM for (5.1) **Correction Step** .

From $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ to $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$:

Generate the new iterate $(A_1x_1^{k+1}, A_2x_2^{k+1}, \cdots, A_px_p^{k+1}, \lambda^{k+1})$ with $\nu \in (0,1)$ by

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\beta I_m & -\beta I_m & \cdots & -\beta I_m & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{5.8}$$

对照一下就可以发现, §5.2 中的方法, 就是 §4.2.2 方法的直接推广.

校正工作量很小. 把校正公式分开来写就是:

$\boldsymbol{Ax_i^{k+1}\ (i=1,\ldots,p)}$ **The correction form of the primal parts are equal.**

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \end{pmatrix} - \nu \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \end{pmatrix}, \tag{5.9}$$

$\boldsymbol{\lambda^{k+1}}$ **The correction form of the dual parts are slightly different.**

$$\lambda^{k+1} = \tilde{\lambda}^k + \beta \sum_{i=1}^{p} (A_ix_i^k - A_i\tilde{x}_i^k). \tag{5.10}$$

两种不同方法的

$\lambda^{k+1} = \tilde{\lambda}^k + \nu\beta(A_1x_1^k - A_1\tilde{x}_1^k) \;\; \Rightarrow \;\; \lambda^{k+1} = \tilde{\lambda}^k + \beta\sum_{i=1}^p (A_ix_i^k - A_i\tilde{x}_i^k).$

6 Convergence

The optimization problem (5.1) has been translated to VI (5.2), namely,

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega.$$

For the easy analysis, we need to denote the following notations:

$$P = \begin{pmatrix} \sqrt{\beta}A_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta}A_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta}A_p & 0 \\ 0 & \cdots & \cdots & 0 & (1/\sqrt{\beta})I_m \end{pmatrix}, \quad \xi = Pw = \begin{pmatrix} \sqrt{\beta}A_1x_1 \\ \sqrt{\beta}A_2x_2 \\ \vdots \\ \sqrt{\beta}A_px_p \\ (1/\sqrt{\beta})\lambda \end{pmatrix}. \tag{6.1}$$

Accordingly, we define

$$\Xi = \{\xi \mid \xi = Pw, \ w \in \Omega\},$$

and

$$\Xi^* = \{\xi^* \mid \xi^* = Pw^*, \ w^* \in \Omega^*\}.$$

We will prove that both the primal-dual algorithm (5.3)-(5.4) and the dual-primal algorithm (5.7)-(5.8) belong to the following prototypical algorithmic framework.

A Prototypical Algorithmic Framework for VI (5.2).

1. (Prediction Step) With given $\xi^k = Pw^k$, find $\tilde{w}^k \in \Omega$ such that

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k), \quad \forall\, w \in \Omega, \tag{6.2a}$$

   with $\mathcal{Q} \in \Re^{(p+1)m \times (p+1)m}$, and the matrix $\mathcal{Q}^T + \mathcal{Q}$ is positive definite.

2. (Correction Step) With the predictor $\tilde{w}^k$ by (6.2a) and $\tilde{\xi}^k = P\tilde{w}^k$, the new iterate ξ^{k+1} is updated by

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k), \tag{6.2b}$$

   where $\mathcal{M} \in \Re^{(p+1)m \times (p+1)m}$ is a non-singular matrix.

Theorem 3 *For the matrices* $\mathcal{Q}$ *and* $\mathcal{M}$ *in the algorithm* (6.2)*, if there is a positive definite matrix* $\mathcal{H} \in \Re^{(p+1)m\times(p+1)m}$ *such that*

$$\mathcal{H}\mathcal{M} = \mathcal{Q} \tag{6.3a}$$

and

$$\mathcal{G} := \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0, \tag{6.3b}$$

then we have

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall\, \xi^* \in \Xi^*. \tag{6.4}$$

Proof. Setting w in (6.2a) as any fixed $w^* \in \Omega^*$, and using

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) \equiv (\tilde{w}^k - w^*)^T F(w^*),$$

we get

$$(\tilde{\xi}^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*), \;\; \forall\, w^* \in \Omega^*.$$

The right-hand side of the last inequality is non-negative. Thus, we have

$$(\xi^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \geq (\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k), \quad \forall\, \xi^* \in \Xi^*. \qquad (6.5)$$

Then, by simple manipulations, we obtain

$$\begin{aligned}
&\|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \\
&\overset{(6.2b)}{=} \quad \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|(\xi^k - \xi^*) - \mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&\overset{(6.3a)}{=} \quad 2(\xi^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) - \|\mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&\overset{(6.5)}{\geq} \quad 2(\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) - \|\mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&= \quad (\xi^k - \tilde{\xi}^k)^T [(\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T \mathcal{H} \mathcal{M}](\xi^k - \tilde{\xi}^k) \\
&\overset{(6.3b)}{=} \quad \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2.
\end{aligned}$$

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

We call (6.3) the convergence conditions for the algorithm framework (6.2).

The inequality (6.4) is the key for the convergence proofs, for details, see [6]

7 Convergence of the Primal-Dual Algorithm in §5.1

In order to prove the convergence of the algorithm (5.3)-(5.4), we need only to show that it belongs to the algorithmic framework (6.2) and to verify the convergence conditions (6.3)

7.1 The algorithm (5.3)-(5.4) belongs to the framework (6.2)

Prediction First, for the primal part of the predictor,

$$\tilde{x}_i^k \in \arg\min\Big\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 \,|\, x_i \in \mathcal{X}_i\Big\}.$$

According to Lemma 1, the optimal condition is $\tilde{x}_i^k \in \mathcal{X}_i$ and

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{-A_i^T\lambda^k + \beta A_i^T\big(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)\big)\} \geq 0,$$

for all $x_i \in \mathcal{X}_i$. It can be written as $\tilde{x}_i^k \in \mathcal{X}_i$ and

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{\underline{-A_i^T\tilde{\lambda}^k} + \beta A_i^T\big(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)\big) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \tag{7.1a}$$

for all $x_i \in \mathcal{X}_i$. The dual part of the predictor, $\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^p A_j\tilde{x}_j^k - b\big)\big]$,

$$\tilde{\lambda}^k = \arg\min\Big\{\big\|\lambda - \big[\lambda^k - \beta\big(\textstyle\sum_{j=1}^p A_j\tilde{x}_j^k - b\big)\big]\big\|^2 \,\big|\, \lambda \in \Lambda\Big\}.$$

The optimal condition is

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T\Big\{\big(\underline{\textstyle\sum_{j=1}^p A_j\tilde{x}_j^k - b}\big) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\Big\} \geq 0, \quad \forall \lambda \in \Lambda. \tag{7.1b}$$

Summating (7.1a) and (7.1b), for the predictor $\tilde{w}^k$ generated by (5.3), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \geq (w - \tilde{w}^k)^T Q_{PD}(w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \tag{7.2a}$$

where

$$Q_{PD} = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{7.2b}$$

Using the notation P in (6.1), for the the matrix Q_{PD} in (7.2b), we have

$$Q_{PD} = P^T \mathcal{Q}_{PD} P, \qquad \text{where} \qquad \mathcal{Q}_{PD} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{7.3}$$

Thus, for the right hand side of (7.2a), we have

$$\begin{aligned} (w - \tilde{w}^k)^T Q_{PD} (w^k - \tilde{w}^k) &= (w - \tilde{w}^k)^T P^T \mathcal{Q}_{PD} P (w^k - \tilde{w}^k) \\ &= (\xi - \tilde{\xi}^k)^T \mathcal{Q}_{PD} (\xi^k - \tilde{\xi}^k). \end{aligned}$$

Then, it follows from (7.2) that we have the following inequality:

$$\begin{aligned} \tilde{w}^k \in \Omega, \ \ \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) & \\ \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q}_{PD} (\xi^k - \tilde{\xi}^k), \ \ \forall\, w \in \Omega. & \end{aligned} \tag{7.4}$$

where $\mathcal{Q}_{PD}$ is given in (7.3).

Correction Left-multiplying the matrix $\mathrm{diag}(\sqrt{\beta}I_m, \ldots, \sqrt{\beta}I_m, \frac{1}{\sqrt{\beta}}I_m)$ to both sides of the correction step of the primal-dual algorithm, (5.4), we get

$$\begin{pmatrix} \sqrt{\beta}A_1x_1^{k+1} \\ \sqrt{\beta}A_2x_2^{k+1} \\ \vdots \\ \sqrt{\beta}A_px_p^{k+1} \\ (1/\sqrt{\beta})\lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \sqrt{\beta}A_1x_1^{k} \\ \sqrt{\beta}A_2x_2^{k} \\ \vdots \\ \sqrt{\beta}A_px_p^{k} \\ (1/\sqrt{\beta})\lambda^{k} \end{pmatrix}$$

$$- \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu I_m & 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} \sqrt{\beta}(A_1x_1^k - A_1\tilde{x}_1^k) \\ \sqrt{\beta}(A_2x_2^k - A_2\tilde{x}_2^k) \\ \vdots \\ \sqrt{\beta}(A_px_p^k - A_p\tilde{x}_p^k) \\ (1/\sqrt{\beta})(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}.$$

Recall the definitions of the matrix P and $Pw = \xi$ (see(6.1)).

The correction step of the primal-dual algorithm, (5.4), can be written as

$$\xi^{k+1} = \xi^k - \mathcal{M}_{PD}(\xi^k - \tilde{\xi}^k), \tag{7.5a}$$

where

$$\mathcal{M}_{PD} = \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu I_m & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{7.5b}$$

7.2 Verifying the convergence conditions of the algorithm

In the algorithm (7.4)-(7.5), the matrices $\mathcal{Q}$ and $\mathcal{M}$ have the following forms:

$$\mathcal{Q}_{PD} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}, \quad \mathcal{M}_{PD} = \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu I_m & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

In order to simplify the notations to be used, we define the following $p \times p$ block matrices:

$$\mathcal{L} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}, \qquad \mathcal{I} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{7.6}$$

We also define the $1 \times p$ block matrix

$$\mathcal{E} = \begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix}. \tag{7.7}$$

Recall the respective definitions $\mathcal{L}$ and $\mathcal{E}$ in (7.6) and (7.7). We have

$$\begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} = \mathcal{L}^{-T}$$

and

$$\begin{pmatrix} I_m & 0 & \cdots & 0 \end{pmatrix} = \mathcal{E}\mathcal{L}^{-T}.$$

Thus, see (7.3) and (7.5b), we have

$$\mathcal{Q}_{PD} = \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} \quad \text{and} \quad \mathcal{M}_{PD} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} \tag{7.8}$$

For the above matrices $\mathcal{Q}_{PD}$ and $\mathcal{M}_{PD}$, the remaining tasks is to find a positive definite matrix $\mathcal{H}_{PD}$, such that the convergence conditions (6.3)are satisfied.

Lemma 3 *For the matrices* $\mathcal{Q}_{PD}$ *and* $\mathcal{M}_{PD}$ *given by* (7.3) *and* (7.5b), *respectively, the matrix*

$$\mathcal{H}_{PD} = \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix} \quad \textit{with} \quad \nu \in (0,1) \tag{7.9}$$

is positive definite, and it satisfies $\mathcal{H}_{PD}\mathcal{M}_{PD} = \mathcal{Q}_{PD}$.

Proof. It is easy to check the positive definiteness of $\mathcal{H}$. In addition, for the block matrix $\mathcal{Q}$ in (7.3), we have

$$\begin{aligned} \mathcal{H}_{PD}\mathcal{M}_{PD} &= \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix} \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} = \mathcal{Q}_{PD}. \end{aligned}$$

The assertions of this lemma are proved. $\square$

Lemma 4 *Let* $\mathcal{Q}_{PD}$, $\mathcal{M}_{PD}$ *and* $\mathcal{H}_{PD}$ *be defined in* (7.3), (7.5b) *and* (7.9)*, respectively. Then the matrix*

$$\mathcal{G}_{PD} := (\mathcal{Q}_{PD}^T + \mathcal{Q}_{PD}) - \mathcal{M}_{PD}^T \mathcal{H}_{PD} \mathcal{M}_{PD} \tag{7.10}$$

is positive definite.

Proof. By elementary matrix multiplications, we know that

$$\mathcal{M}_{PD}^T \mathcal{H}_{PD} \mathcal{M}_{PD} = \mathcal{Q}_{PD}^T \mathcal{M}_{PD} = \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E} & I_m \end{pmatrix} \begin{pmatrix} \nu \mathcal{L}^{-T} & 0 \\ -\nu \mathcal{E} \mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} \nu \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}.$$

Then, it follows from $\mathcal{L}^T + \mathcal{L} = \mathcal{I} + \mathcal{E}^T \mathcal{E}$ (see (7.6)-(7.7)) that

$$\begin{aligned} \mathcal{G}_{PD} &= (\mathcal{Q}_{PD}^T + \mathcal{Q}_{PD}) - \mathcal{M}_{PD}^T \mathcal{H}_{PD} \mathcal{M}_{PD} \\ &= \begin{pmatrix} \mathcal{L}^T + \mathcal{L} & \mathcal{E}^T \\ \mathcal{E} & 2I_m \end{pmatrix} - \begin{pmatrix} \nu \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}^T \mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix}. \end{aligned}$$

Thus, the matrix $\mathcal{G}_{PD}$ is positive definite for any $\nu \in (0,1)$. $\square$

Lemma 3 and Lemma 4 have verified the convergence conditions (6.3) and thus the key convergence inequality (6.4) holds. The primal-dual algorithm (5.3)-(5.4) is convergent.

8 Convergence of the Dual-Primal Algorithm in §5.2

In order to prove the convergence of the algorithm (5.7)-(5.8), we need only to show that it belongs to the algorithmic framework (6.2) and to verify the convergence conditions (6.3).

8.1 The algorithm (5.7)-(5.8) belongs to the framework (6.2)

The predictor $\tilde{w}^k$ generated by (5.7), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \geq (w - \tilde{w}^k)^T Q_{DP} (w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \tag{8.1a}$$

where

$$Q_{DP} = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & 0 \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & 0 \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & 0 \\ -A_1 & -A_2 & \cdots & -A_p & \frac{1}{\beta} I_m \end{pmatrix}. \tag{8.1b}$$

Using the notation P in (6.1), for the matrix Q_{DP} in (8.1b), we have

$$Q_{DP} = P^T \mathcal{Q}_{DP} P, \qquad \text{where} \qquad \mathcal{Q}_{DP} = \begin{pmatrix} I_m & 0 & \cdots & 0 & 0 \\ I_m & I_m & \ddots & \vdots & 0 \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & 0 \\ -I_m & -I_m & \cdots & -I_m & I_m \end{pmatrix}. \tag{8.2}$$

Thus, for the right hand side of (8.1a), we have

$$\begin{aligned} (w - \tilde{w}^k)^T Q_{DP} (w^k - \tilde{w}^k) &= (w - \tilde{w}^k)^T P^T \mathcal{Q}_{DP} P (w^k - \tilde{w}^k) \\ &= (\xi - \tilde{\xi}^k)^T \mathcal{Q}_{DP} (\xi^k - \tilde{\xi}^k). \end{aligned}$$

Then, it follows from (8.1) that we have the following inequality:

$$\begin{aligned} \tilde{w}^k \in \Omega, \;\; \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) & \\ \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q}_{DP} (\xi^k - \tilde{\xi}^k), \;\; \forall\, w \in \Omega. & \end{aligned} \tag{8.3}$$

where $\mathcal{Q}_{DP}$ is given in (8.2).

Correction Recall the definitions of the matrix P and $Pw = \xi$ (see(6.1)).

The correction step of the dual-primal algorithm, (5.8), can be written as

$$\xi^{k+1} = \xi^k - \mathcal{M}_{DP}(\xi^k - \tilde{\xi}^k), \tag{8.4a}$$

where

$$\mathcal{M}_{DP} = \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -I_m & -I_m & \cdots & -I_m & I_m \end{pmatrix}. \tag{8.4b}$$

8.2 Verify the convergence conditions of the D-P algorithm

In the algorithm (6.2), the matrices $\mathcal{Q}$ and $\mathcal{M}$ have the following forms:

$$\mathcal{Q}_{DP} = \begin{pmatrix} I_m & 0 & \cdots & 0 & 0 \\ I_m & I_m & \ddots & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & 0 \\ -I_m & -I_m & \cdots & -I_m & I_m \end{pmatrix}, \quad \mathcal{M}_{DP} = \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -I_m & -I_m & \cdots & -I_m & I_m \end{pmatrix}.$$

Recall the respective definition $\mathcal{L}$ in (7.6). We have

$$\begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} = \mathcal{L}^{-T}.$$

Thus, we have (see $\mathcal{E}$ in (7.7))

$$\mathcal{Q}_{DP} = \begin{pmatrix} \mathcal{L} & 0 \\ -\mathcal{E} & I_m \end{pmatrix} \quad \text{and} \quad \mathcal{M}_{DP} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\mathcal{E} & I_m \end{pmatrix} \tag{8.5}$$

Lemma 5 *For the matrices* $\mathcal{Q}_{DP}$ *and* $\mathcal{M}_{DP}$ *given by* (8.2) *and* (8.4b), *respectively, the matrix*

$$\mathcal{H}_{DP} = \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T & 0 \\ 0 & I_m \end{pmatrix} \quad \textit{with} \quad \nu \in (0,1) \tag{8.6}$$

is positive definite, and it satisfies $\mathcal{H}_{DP}\mathcal{M}_{DP} = \mathcal{Q}_{DP}$.

Proof. It is easy to check the positive definiteness of $\mathcal{H}$. In addition, for the block matrix $\mathcal{Q}$ in (8.2), we have

$$\begin{aligned} \mathcal{H}_{DP}\mathcal{M}_{DP} &= \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T & 0 \\ 0 & I_m \end{pmatrix} \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\mathcal{E} & I_m \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{L} & 0 \\ -\mathcal{E} & I_m \end{pmatrix} = \mathcal{Q}_{DP}. \end{aligned}$$

The assertions of this lemma are proved. $\square$

Lemma 6 *Let* $\mathcal{Q}_{DP}$, $\mathcal{M}_{DP}$ *and* $\mathcal{H}_{DP}$ *be defined in* (8.2), (8.4b) *and* (8.6), *respectively. Then the matrix*

$$\mathcal{G}_{DP} := (\mathcal{Q}_{DP}^T + \mathcal{Q}_{DP}) - \mathcal{M}_{DP}^T \mathcal{H}_{DP} \mathcal{M}_{DP} \tag{8.7}$$

is positive definite.

Proof. By elementary matrix multiplications, we know that

$$\mathcal{M}_{DP}^T \mathcal{H}_{DP} \mathcal{M}_{DP} = \mathcal{Q}_{DP}^T \mathcal{M}_{DP} = \begin{pmatrix} \mathcal{L}^T & -\mathcal{E}^T \\ 0 & I_m \end{pmatrix} \begin{pmatrix} \nu \mathcal{L}^{-T} & 0 \\ -\mathcal{E} & I_m \end{pmatrix} = \begin{pmatrix} \nu \mathcal{I} + \mathcal{E}^T \mathcal{E} & -\mathcal{E}^T \\ -\mathcal{E} & I_m \end{pmatrix}.$$

Then, it follows from $\mathcal{L}^T + \mathcal{L} = \mathcal{I} + \mathcal{E}^T \mathcal{E}$ (see (7.6)-(7.7)) that

$$\begin{aligned} \mathcal{G}_{DP} &= (\mathcal{Q}_{DP}^T + \mathcal{Q}_{DP}) - \mathcal{M}_{DP}^T \mathcal{H}_{DP} \mathcal{M}_{DP} \\ &= \begin{pmatrix} \mathcal{L}^T + \mathcal{L} & -\mathcal{E}^T \\ -\mathcal{E} & 2I_m \end{pmatrix} - \begin{pmatrix} \nu \mathcal{I} + \mathcal{E}^T \mathcal{E} & -\mathcal{E}^T \\ -\mathcal{E} & I_m \end{pmatrix} = \begin{pmatrix} (1-\nu)\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}. \end{aligned}$$

Thus, the matrix $\mathcal{G}_{DP}$ is positive definite for any $\nu \in (0, 1)$. $\square$

> Lemma 5 and Lemma 6 have verified the convergence conditions (6.3) and thus the key convergence inequality (6.4) holds. The dual-primal algorithm (5.7)-(5.8) is convergent.

9 Conclusions

- 通常所说的交替方向法, 是从增广拉格朗日乘子法松弛而来的, 用来处理等式约束的可分离凸优化问题. 从 ALM 到 ADMM, 是把可分离的问题分开来求解. 这种思想继续推广到三块和三块以上的可分离问题, 我们 2016 年的 MP 文章证明了其收敛性无法保证.

- 这篇文章里给出的两类交替方向法, 不管是 primal-dual, 还是 dual-primal, 都可以推广到任意整数块可分离凸优化问题的求解. 是的, 它需要额外的校正. 可喜的是, 校正特别简单！

- 我们特别推崇"预测-校正", 尤其是那种代价很小的校正. 生机勃勃的果树, 修剪就是校正. 社会治理也是一种校正！ 交替按序预测, 降低了问题难度; 全局整体校正, 把握了收敛方向.

- 带校正的交替方向法既可以用来求解等式约束的问题, 又可以用来求解不等式约束的问题. 适用从一块到任意多块的可分离问题, 算法结构和收敛性证明完全统一.

- 适用范围广的算法会不会影响效率？对经典 ADMM 擅长的两块可分离的等式约束凸优化问题, 我们也用本文提到的带校正的交替方向法 (4.6) 和(4.7) 去求解, 与网上他人提供的 ADMM 代码比较, 发现这种担心是多余的.

- 在这个报告中, 我们只证明了收敛的关键不等式(6.4)

$$\|\xi^{k+1}-\xi^*\|_{\mathcal{H}}^2 \le \|\xi^k-\xi^*\|_{\mathcal{H}}^2 - \|\xi^k-\tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall\, \xi^* \in \Xi^*.$$

  关于收敛性的进一步的细节可以参考文献 [6].

- 我们相信, 由于应用范围广又便于向多块问题推广, 新方法将会更受用户欢迎 !

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] 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*, *Mathematical Programming, Series A* 2016.

[4] D. Gabay, Applications of the method of multipliers to variational inequalities, *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, edited by M. Fortin and R. Glowinski, North Holland, Amsterdam, The Netherlands, 1983, pp. 299–331.

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

[6] B. S. He, S. J. Xu and X. M. Yuan, Extensions of ADMM for Separable Convex Optimization Problems with Linear Equality or Inequality Constraints, arXiv:2107.01897v2[math.OC]

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

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

[9] 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.

Thank you very much for your attention！！