一类适用范围更广和便于推广的交替方向法

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

文章可见： 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 $F(x) = \nabla f(x)$.

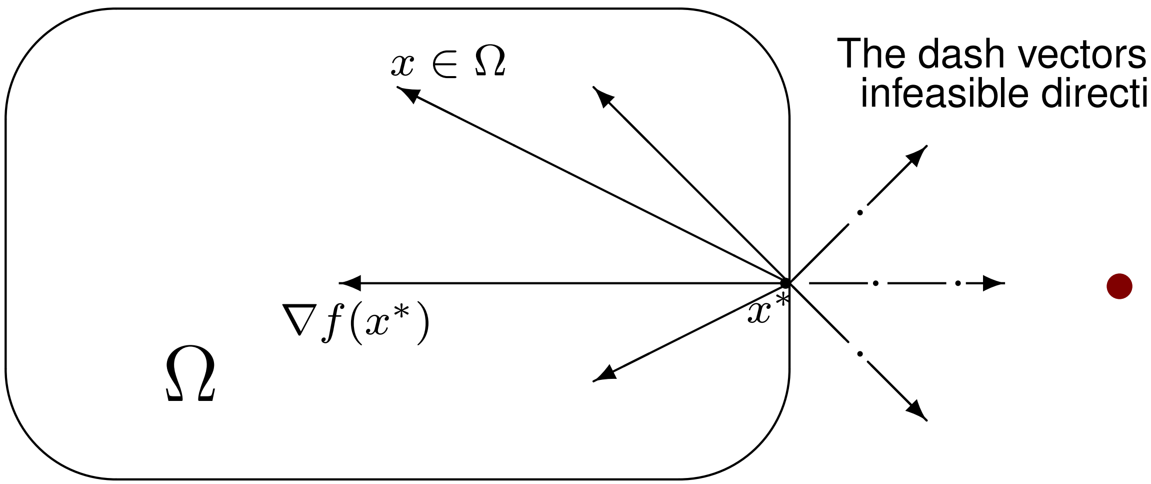

Fig. 1.1 Differential Convex Optimization and VI

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

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

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

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

2 ADMM with wider application $\&$ easy extensions

Let us consider the general separable convex optimization model

$$\min \left\{ \theta_1(x) + \theta_2(y) \mid Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y} \right\}. \tag{2.1}$$

The augmented Lagrangian function is

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

2.1 From ALM to ADMM

Augmented Lagrangian Method for (2.1). From λ^k to λ^{k+1}:

$$\left\{ \begin{array}{rcl} (x^{k+1}, y^{k+1}) & \in & \arg\min\{\mathcal{L}_\beta(x, y, \lambda^k) \,|\, x \in \mathcal{X}, y \in \mathcal{Y}\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array} \right. \tag{2.2}$$

ADMM for (2.1) From (y^k, λ^k) to (y^{k+1}, λ^{k+1})

$$\left\{\begin{array}{lcl} x^{k+1} & \in & \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ y^{k+1} & \in & \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array}\right. \tag{2.3}$$

From (2.2) to (2.3), 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 (2.3) as

$$\left\{\begin{array}{l} x^{k+1} \in \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2}\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\}, \\ y^{k+1} \in \operatorname{argmin}\left\{\begin{array}{c} \theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \\ \frac{\beta}{2}\|A(x^{k+1} - x^k) + B(y - y^k)\|^2 \end{array} \middle| \; y \in \mathcal{Y}\right\}, \\ \lambda^{k+1} = \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} - b\big) \end{array}\right. \tag{2.4}$$

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 (2.4).

2.2 ADMM with wider applications

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

$$\min\left\{\theta_1(x)+\theta_2(y)\mid Ax+By=b\ (\text{or} \geq b), x\in\mathcal{X}, y\in\mathcal{Y}\right\}. \tag{2.5}$$

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

We define

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

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

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

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

2.2.1 Primal-dual extension of ADMM with wider application

A Primal-Dual Extension of the ADMM for (2.5).

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

$$\left\{\begin{array}{l} \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \mid y \in \mathcal{Y}\big\}, \\ \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(A\tilde{x}^k + B\tilde{y}^k - b\big)\big]. \end{array}\right. \tag{2.6a}$$

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

$$\begin{pmatrix} Ax^{k+1} \\ By^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} Ax^k \\ By^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 \\ 0 & \nu I_m & 0 \\ -\nu\beta I_m & 0 & I_m \end{pmatrix} \begin{pmatrix} Ax^k - A\tilde{x}^k \\ By^k - B\tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{2.6b}$$

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.

2.2.2 Dual-Primal extension of ADMM with wider application

A Dual-Primal Extension of the ADMM for (2.5).

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

$$\left\{ \begin{array}{l} \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(Ax^k + By^k - b\big)\big], \\ \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^T A^T \tilde{\lambda}^k + \frac{1}{2}\beta\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^T B^T \tilde{\lambda}^k + \frac{1}{2}\beta\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \mid y \in \mathcal{Y}\big\}. \end{array} \right. \tag{2.7a}$$

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

$$\begin{pmatrix} Ax^{k+1} \\ By^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} Ax^k \\ By^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 \\ 0 & \nu I_m & 0 \\ -\beta I_m & -\beta I_m & I_m \end{pmatrix} \begin{pmatrix} Ax^k - A\tilde{x}^k \\ By^k - B\tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{2.7b}$$

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.

3 p-block separable convex optimization problems

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

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i)\ \Big|\ \sum_{i=1}^{p}A_ix_i = b\ (\text{or} \geq b),\quad x_i\in\mathcal{X}_i\Big\}. \tag{3.1}$$

The Lagrangian function is

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

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

$$\Lambda=\begin{cases}\Re^m, & \text{if } \sum_{i=1}^{p}A_ix_i=b,\\ \Re^m_+, & \text{if } \sum_{i=1}^{p}A_ix_i\geq b.\end{cases}$$

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

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

The optimality condition of (3.1) can be written as the following VI:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{3.2a}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_p^T\lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}, \tag{3.2b}$$

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 (3.2).

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

A Primal-Dual Extension of the ADMM for (3.1) **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

$$\begin{cases} \tilde{x}_1^k \in \arg\min\big\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\big\}; \\ \tilde{x}_2^k \in \arg\min\big\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\big\}; \\ \qquad\vdots \\ \tilde{x}_i^k \in \arg\min_{x_i \in \mathcal{X}_i}\big\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\big\}; \\ \qquad\vdots \\ \tilde{x}_p^k \in \arg\min_{x_p \in \mathcal{X}_p}\big\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\big\}; \\ \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^{p} A_j\tilde{x}_j^k - b\big)\big]. \end{cases} \tag{3.3}$$

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 (3.1) **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

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{3.4}$$

对照一下就可以发现, §3.1 中的方法, 就是 §2.2.1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:

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

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & 0 \\ 0 & \nu I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -\nu I_m \\ 0 & \cdots & 0 & \nu I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \end{pmatrix}, \tag{3.5}$$

λ^{k+1}

$$\lambda^{k+1} = \tilde{\lambda}^k + \nu\beta(A_1x_1^k - A_1\tilde{x}_1^k). \tag{3.6}$$

还能说校正不简单!?

3.2 Dual-primal extension of the ADMM for (3.1)

A Dual-Primal Extension of the ADMM for (3.1) **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

$$\left\{\begin{array}{l}
\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^p A_jx_j^k - b\big)\big] \\
\tilde{x}_1^k \in \arg\min\big\{\theta_1(x_1) - x_1^TA_1^T\tilde{\lambda}^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\big\}; \\
\tilde{x}_2^k \in \arg\min\big\{\theta_2(x_2) - x_2^TA_2^T\tilde{\lambda}^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\big\}; \\
\qquad\vdots \\
\tilde{x}_i^k \in \arg\min_{x_i\in\mathcal{X}_i}\{\theta_i(x_i) - x_i^TA_i^T\tilde{\lambda}^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1}A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\
\qquad\vdots \\
\tilde{x}_p^k \in \mathrm{argmin}_{x_p\in\mathcal{X}_p}\{\theta_p(x_p) - x_p^TA_p^T\tilde{\lambda}^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}.
\end{array}\right. \tag{3.7}$$

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 (3.1) **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

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\beta I_m & -\beta I_m & \cdots & -\beta I_m & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{3.8}$$

对照一下就可以发现, §3.2 中的方法, 就是 §2.2.2 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:

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

$$\begin{pmatrix} A_1 x_1^{k+1} \\ A_2 x_2^{k+1} \\ \vdots \\ A_p x_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1 x_1^{k} \\ A_2 x_2^{k} \\ \vdots \\ A_p x_p^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & 0 \\ 0 & \nu I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -\nu I_m \\ 0 & \cdots & 0 & \nu I_m \end{pmatrix} \begin{pmatrix} A_1 x_1^k - A_1 \tilde{x}_1^k \\ A_2 x_2^k - A_2 \tilde{x}_2^k \\ \vdots \\ A_p x_p^k - A_p \tilde{x}_p^k \end{pmatrix}, \tag{3.9}$$

λ^{k+1} **The correction form of the dual parts are slightly different.**

$$\lambda^{k+1} = \tilde{\lambda}^k + \beta \sum_{i=1}^{p} (A_i x_i^k - A_i \tilde{x}_i^k). \tag{3.10}$$

两种不同方法的

$$\lambda^{k+1} = \tilde{\lambda}^k + \nu\beta(A_1 x_1^k - A_1 \tilde{x}_1^k) \ \Rightarrow \ \lambda^{k+1} = \tilde{\lambda}^k + \beta \textstyle\sum_{i=1}^{p} (A_i x_i^k - A_i \tilde{x}_i^k).$$

4 Convergence

The optimization problem (3.1) has been translated to VI (3.2), namely,

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

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

$$P = \begin{pmatrix} \sqrt{\beta}A_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta}A_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta}A_p & 0 \\ 0 & \cdots & \cdots & 0 & (1/\sqrt{\beta})I_m \end{pmatrix}, \quad \xi = Pw = \begin{pmatrix} \sqrt{\beta}A_1x_1 \\ \sqrt{\beta}A_2x_2 \\ \vdots \\ \sqrt{\beta}A_px_p \\ (1/\sqrt{\beta})\lambda \end{pmatrix}. \tag{4.1}$$

Accordingly, we define

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

and

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

We will prove that both the primal-dual algorithm (3.3)-(3.4) and the dual-primal algorithm (3.7)-(3.8) belong to the following prototypical algorithmic framework.

A Prototypical Algorithmic Framework for VI (3.2).

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

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k), \quad \forall\, w \in \Omega, \tag{4.2a}$$

   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 (4.2a) and $\tilde{\xi}^k = P\tilde{w}^k$, the new iterate ξ^{k+1} is updated by

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k), \tag{4.2b}$$

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

Theorem 1 *For the matrices $\mathcal{Q}$ and $\mathcal{M}$ in the algorithm* (4.2), *if there is a positive definite matrix $\mathcal{H} \in \Re^{(p+1)m\times(p+1)m}$ such that*

$$\mathcal{H}\mathcal{M} = \mathcal{Q} \tag{4.3a}$$

and

$$\mathcal{G} := \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0, \tag{4.3b}$$

then we have

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall\, \xi^* \in \Xi^*. \tag{4.4}$$

Proof. Setting w in (4.2a) 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

$$(\xi^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \ge (\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k), \quad \forall\, \xi^* \in \Xi^*. \tag{4.5}$$

Then, by simple manipulations, we obtain

$$\begin{aligned}
&\|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \\
&\overset{(4.2b)}{=} \quad \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|(\xi^k - \xi^*) - \mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&\overset{(4.3a)}{=} \quad 2(\xi^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) - \|\mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&\overset{(4.5)}{\ge} \quad 2(\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) - \|\mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&= \quad (\xi^k - \tilde{\xi}^k)^T [(\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T \mathcal{H} \mathcal{M}](\xi^k - \tilde{\xi}^k) \\
&\overset{(4.3b)}{=} \quad \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2.
\end{aligned}$$

The assertion of this theorem is proved. □

We call (4.3) the convergence conditions for the algorithm framework (4.2).

The inequality (4.4) is the key for the convergence proofs, for details, see [5]

5 Convergence of the Primal-Dual Algorithm in §3.1

In order to prove the convergence of the algorithm (3.3)-(3.4), we need only to show that it belongs to the algorithmic framework (4.2) and to verify the convergence conditions (4.3)

5.1 The algorithm (3.3)-(3.4) belongs to the framework (4.2)

Prediction First, for the primal part of the predictor,

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

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

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

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

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{\underline{-A_i^T\tilde{\lambda}^k} + \beta A_i^T\big(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)\big) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \tag{5.1a}$$

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

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T\big\{\big(\underline{\textstyle\sum_{j=1}^p A_j \tilde{x}_j^k - b}\big) + \tfrac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\big\} \geq 0, \quad \forall \lambda \in \Lambda. \tag{5.1b}$$

Summating (5.1a) and (5.1b), for the predictor $\tilde{w}^k$ generated by (3.3), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \geq (w - \tilde{w}^k)^T Q_{PD}(w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \tag{5.2a}$$

where

$$Q_{PD} = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{5.2b}$$

Using the notation P in (4.1), for the the matrix Q_{PD} in (5.2b), we have

$$Q_{PD} = P^T \mathcal{Q}_{PD} P, \qquad \text{where} \qquad \mathcal{Q}_{PD} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{5.3}$$

Thus, for the right hand side of (5.2a), 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 (5.2) that we have the following inequality:

$$\begin{aligned} \tilde{w}^k \in \Omega, \;\; \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) & \\ \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q}_{PD} (\xi^k - \tilde{\xi}^k), \;\; \forall\, w \in \Omega. & \end{aligned} \tag{5.4}$$

where $\mathcal{Q}_{PD}$ is given in (5.3).

Correction Left-multiplying the matrix 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, (3.4), 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(4.1)).

The correction step of the primal-dual algorithm, (3.4), can be written as

$$\xi^{k+1} = \xi^k - \mathcal{M}_{PD}(\xi^k - \tilde{\xi}^k), \tag{5.5a}$$

where

$$\mathcal{M}_{PD} = \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu I_m & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{5.5b}$$

5.2 Verifying the convergence conditions of the algorithm

In the algorithm (5.4)-(5.5), 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:

$$\mathcal{L} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}, \qquad \mathcal{I} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{5.6}$$

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

$$\mathcal{E} = \begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix}. \tag{5.7}$$

Recall the respective definitions $\mathcal{L}$ and $\mathcal{E}$ in (5.6) and (5.7). 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 (5.3) and (5.5b), we have

$$\mathcal{Q}_{PD} = \begin{pmatrix} \mathcal{L} & \mathcal{E}^T \\ 0 & I_m \end{pmatrix} \quad \text{and} \quad \mathcal{M}_{PD} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}\mathcal{L}^{-T} & I_m \end{pmatrix} \tag{5.8}$$

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 (4.3)are satisfied.

Lemma 2 *For the matrices $\mathcal{Q}_{PD}$ and $\mathcal{M}_{PD}$ given by* (5.3) *and* (5.5b), *respectively, the matrix*

$$\mathcal{H}_{PD} = \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T + \mathcal{E}^T\mathcal{E} & \mathcal{E}^T \\ \mathcal{E} & I_m \end{pmatrix} \quad \textit{with} \quad \nu \in (0,1) \tag{5.9}$$

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 (5.3), 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 3 *Let* $\mathcal{Q}_{PD}$, $\mathcal{M}_{PD}$ *and* $\mathcal{H}_{PD}$ *be defined in* (5.3), (5.5b) *and* (5.9), *respectively. Then the matrix*

$$\mathcal{G}_{PD} := (\mathcal{Q}_{PD}^T + \mathcal{Q}_{PD}) - \mathcal{M}_{PD}^T \mathcal{H}_{PD} \mathcal{M}_{PD} \tag{5.10}$$

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 (5.6)-(5.7)) 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 2 and Lemma 3 have verified the convergence conditions (4.3) and thus the key convergence inequality (4.4) holds. The primal-dual algorithm (3.3)-(3.4) is convergent.

6 Convergence of the Dual-Primal Algorithm in §3.2

In order to prove the convergence of the algorithm (3.7)-(3.8), we need only to show that it belongs to the algorithmic framework (4.2) and to verify the convergence conditions (4.3).

6.1 The algorithm (3.7)-(3.8) belongs to the framework (4.2)

Prediction For the dual part of the predictor, $\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^p A_j x_j^k - b\big)\big]$,

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

The optimal condition is

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T\big\{\big(\textstyle\sum_{j=1}^p A_j x_j^k - b\big) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\big\} \ge 0, \quad \forall \lambda \in \Lambda.$$

It can be rewritten as

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T\Big\{\big(\underline{\textstyle\sum_{j=1}^p A_j \tilde{x}_j^k - b}\big) - \textstyle\sum_{j=1}^p A_j(\tilde{x}_j^k - x_j^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\Big\} \ge 0, \quad \forall \lambda \in \Lambda. \tag{6.1a}$$

The primal part of the predictor, $\tilde{x}_i^k$ is given by

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

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

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \{\underline{-A_i^T \tilde{\lambda}^k} + \beta A_i^T \big(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)\big)\} \ge 0, \quad \text{(6.1b)}$$

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

Summating (6.1b) and (6.1a), for the predictor $\tilde{w}^k$ generated by (3.7), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \ge (w - \tilde{w}^k)^T Q_{DP}(w^k - \tilde{w}^k), \quad \forall\, w \in \Omega, \quad \text{(6.2a)}$$

where

$$Q_{DP} = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & 0 \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & 0 \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & 0 \\ -A_1 & -A_2 & \cdots & -A_p & \frac{1}{\beta} I_m \end{pmatrix}. \quad \text{(6.2b)}$$

Using the notation P in (4.1), for the matrix Q_{DP} in (6.2b), we have

$$Q_{DP} = P^T \mathcal{Q}_{DP} P, \qquad \text{where} \qquad \mathcal{Q}_{DP} = \begin{pmatrix} I_m & 0 & \cdots & 0 & 0 \\ I_m & I_m & \ddots & \vdots & 0 \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & 0 \\ -I_m & -I_m & \cdots & -I_m & I_m \end{pmatrix}. \tag{6.3}$$

Thus, for the right hand side of (6.2a), 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 (6.2) that we have the following inequality:

$$\begin{aligned} \tilde{w}^k \in \Omega, \ \ \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) & \\ \ge (\xi - \tilde{\xi}^k)^T \mathcal{Q}_{PD} (\xi^k - \tilde{\xi}^k), \ \ \forall\, w \in \Omega. & \end{aligned} \tag{6.4}$$

where $\mathcal{Q}_{PD}$ is given in (6.3).

Correction Left-multiplying the matrix diag$(\sqrt{\beta}I_m, \ldots, \sqrt{\beta}I_m, (1/\sqrt{\beta})I_m)$ to both sides of the correction step of the dual-primal algorithm, (3.8), we get

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

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

Recall the definitions of the matrix P and $Pw = \xi$ (see(4.1)).

The correction step of the dual-primal algorithm, (3.8), can be written as

$$\xi^{k+1} = \xi^k - \mathcal{M}_{DP}(\xi^k - \tilde{\xi}^k), \tag{6.5a}$$

where

$$\mathcal{M}_{DP} = \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -I_m & -I_m & \cdots & -I_m & I_m \end{pmatrix}. \tag{6.5b}$$

6.2 Verify the convergence conditions of the D-P algorithm

In the algorithm (4.2), 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 (5.6). 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 (5.7))

$$\mathcal{Q}_{DP} = \begin{pmatrix} \mathcal{L} & 0 \\ -\mathcal{E} & I_m \end{pmatrix} \quad \text{and} \quad \mathcal{M}_{DP} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\mathcal{E} & I_m \end{pmatrix} \tag{6.6}$$

Lemma 4 *For the matrices* $\mathcal{Q}_{DP}$ *and* $\mathcal{M}_{DP}$ *given by* (6.3) *and* (6.5b), *respectively, the matrix*

$$\mathcal{H}_{DP} = \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T & 0 \\ 0 & I_m \end{pmatrix} \quad \textit{with} \quad \nu \in (0,1) \tag{6.7}$$

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 (6.3), 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 5 *Let* $\mathcal{Q}_{DP}$, $\mathcal{M}_{DP}$ *and* $\mathcal{H}_{DP}$ *be defined in* (6.3), (6.5b) *and* (6.7), *respectively. Then the matrix*

$$\mathcal{G}_{DP} := (\mathcal{Q}_{DP}^T + \mathcal{Q}_{DP}) - \mathcal{M}_{DP}^T \mathcal{H}_{DP} \mathcal{M}_{DP} \tag{6.8}$$

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 (5.6)-(5.7)) 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 $\nu \in (0, 1)$. □

Lemma 4 and Lemma 5 have verified the convergence conditions (4.3) and thus the key convergence inequality (4.4) holds. The dual-primal algorithm (3.7)-(3.8) is convergent.

7 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 (2.6) 和(2.7) 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(4.4)

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

  关于收敛性的进一步的细节可以参考文献 [5].

- 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] 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.

[3] 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.

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

[5] 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]

Thank you very much for reading ! !