构造凸优化的一阶分裂收缩算法

— 基于变分不等式和邻近点算法的统一框架

何 炳 生

南京大学数学系　　南方科技大学数学系

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

中学的数理基础　必要的社会实践
普通的大学数学　一般的优化原理

一个科学家最大的本领就在于化复杂为简单, 用简单的方法去解决复杂的问题。　—— 冯康

计算方法的创新来源于计算实践.　—— 冯康

✠ 先贤名言, 铭记在心. ✠

越民义先生在【发展数学之我见】中就告诫我们：学生“很容易将老师所说作为数学的前沿作品来接受，工作后又将这些东西作为衣钵往下传。若因循守旧, 则易造成谬种流传, 误人不浅”。

✠ 前辈良言，置我案前. ✠

尺有所短, 寸有所长。听我宣讲, 读我文章, 是否真有道理, 敬请独立考量。找导师, 选方向, 扬长避短第一桩。 —— 何炳生

连续优化中一些代表性数学模型

1. 简单约束问题 $\min\{f(x) \mid x \in \mathcal{X}\}$, $\mathcal{X}$ is a closed convex set.
2. 变分不等式问题 $u \in \Omega, \quad (u' - u)^T F(u) \geq 0, \quad \forall u' \in \Omega$
3. $\min$-$\max$ 问题 $\min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} \{\mathcal{L}(x, y) = \theta_1(x) - y^T Ax - \theta_2(y)\}$
4. 线性约束的凸优化问题 $\min\{\theta(x) | Ax = b \,(\text{or} \geq b),\ x \in \mathcal{X}\}$
5. 结构型凸优化 $\min\{\theta_1(x) + \theta_2(y) | Ax + By = b,\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$
6. 三个算子的凸优化
   $\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz = b, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}$

变分不等式(VI) 是瞎子爬山的数学表达形式

邻近点算法(PPA) 是步步为营 稳扎稳打的求解方法.

变分不等式和邻近点算法是我们的两大法宝.

Text Outline

1 Preliminaries – Convex Optimization Variational Inequality and Proximal Point Algorithms

1.1 Differential convex optimization and monotone 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 descent direction.

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.

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

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

$$x \in \Omega, \quad (x' - x)^T F(x) \geq 0, \quad \forall x' \in \Omega, \tag{1.2}$$

where $F(x) = \nabla f(x)$.

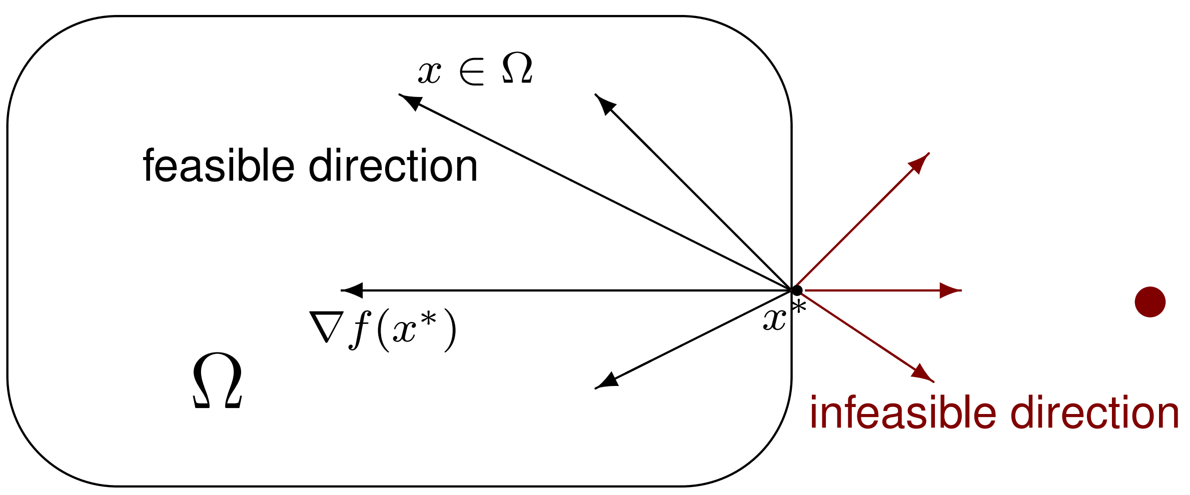


Fig. 1.1 Differential Convex Optimization and VI

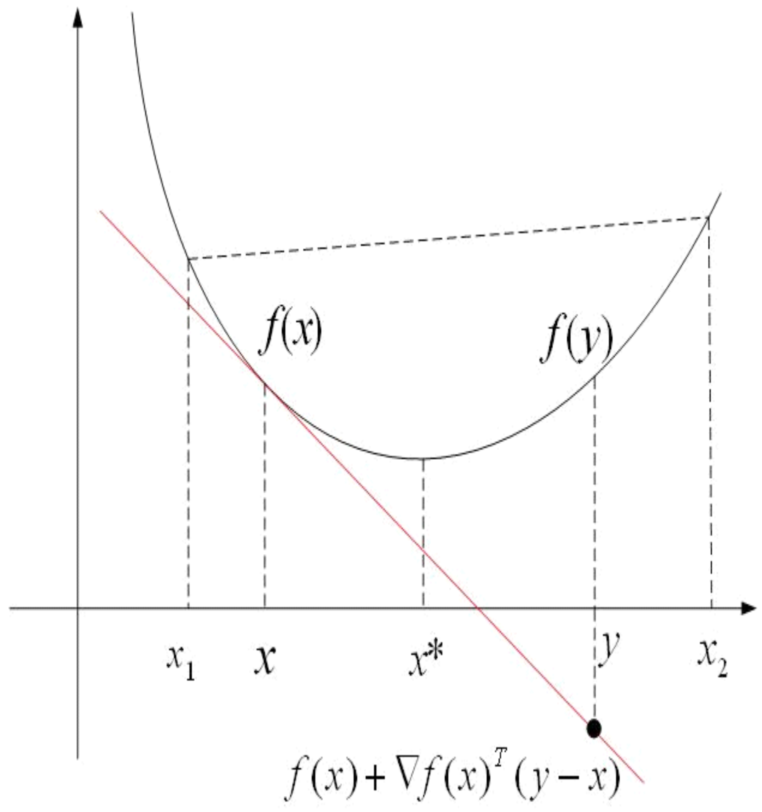
Since $f(x)$ is a convex function, we have

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

A function $f(x)$ is convex iff

$$f((1-\theta)x+\theta y) \leq (1-\theta)f(x)+\theta f(y)$$
$$\forall \theta \in [0,1].$$



Convex function

Properties of convex function

- $f \in \mathcal{C}^1$. f is convex iff

  $f(y) - f(x) \geq \nabla f(x)^T(y-x)$.

  Thus, we have also

  $f(x) - f(y) \geq \nabla f(y)^T(x-y)$.

- Adding above two inequalities, we get

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

- $f \in \mathcal{C}^1$, ∇f is monotone. $f \in \mathcal{C}^2$, $\nabla^2 f(x)$ is positive semi-definite.
- Any local minimum of a convex function is a global minimum.

For the analysis in this paper, we need only the basic property which is described in the following lemma.

Lemma 1.1 *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

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

Proof : First, if (1.3a) is true, then for any $x \in \mathcal{X}$, we have

$$\frac{\theta(x_\alpha) - \theta(x^*)}{\alpha} + \frac{f(x_\alpha) - f(x^*)}{\alpha} \geq 0, \tag{1.4}$$

where

$$x_\alpha = (1-\alpha)x^* + \alpha x, \quad \forall \alpha \in (0,1].$$

Because $\theta(\cdot)$ is convex, it follows that

$$\theta(x_\alpha) \le (1-\alpha)\theta(x^*) + \alpha\theta(x),$$

and thus

$$\theta(x) - \theta(x^*) \ge \frac{\theta(x_\alpha) - \theta(x^*)}{\alpha}, \quad \forall \alpha \in (0,1].$$

Substituting the last inequality in the left hand side of (1.4), we have

$$\theta(x) - \theta(x^*) + \frac{f(x_\alpha) - f(x^*)}{\alpha} \ge 0, \quad \forall \alpha \in (0,1].$$

Using $f(x_\alpha) = f(x^* + \alpha(x - x^*))$ and letting $\alpha \to 0_+$, from the above inequality we get

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

Thus (1.3b) follows from (1.3a). Conversely, since f is convex, it follow that

$$f(x_\alpha) \le (1-\alpha)f(x^*) + \alpha f(x)$$

and it can be rewritten as

$$f(x_\alpha) - f(x^*) \le \alpha(f(x) - f(x^*)).$$

Thus, we have

$$f(x) - f(x^*) \ge \frac{f(x_\alpha) - f(x^*)}{\alpha} = \frac{f(x^* + \alpha(x - x^*)) - f(x^*)}{\alpha},$$

for all $\alpha \in (0,1]$. Letting $\alpha \to 0_+$, we get

$$f(x) - f(x^*) \ge \nabla f(x^*)^T(x - x^*).$$

Substituting it in the left hand side of (1.3b), we get

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + f(x) - f(x^*) \ge 0, \quad \forall\, x \in \mathcal{X},$$

and (1.3a) is true. The proof is complete. $\square$

1.2 Linear constrained convex optimization and VI

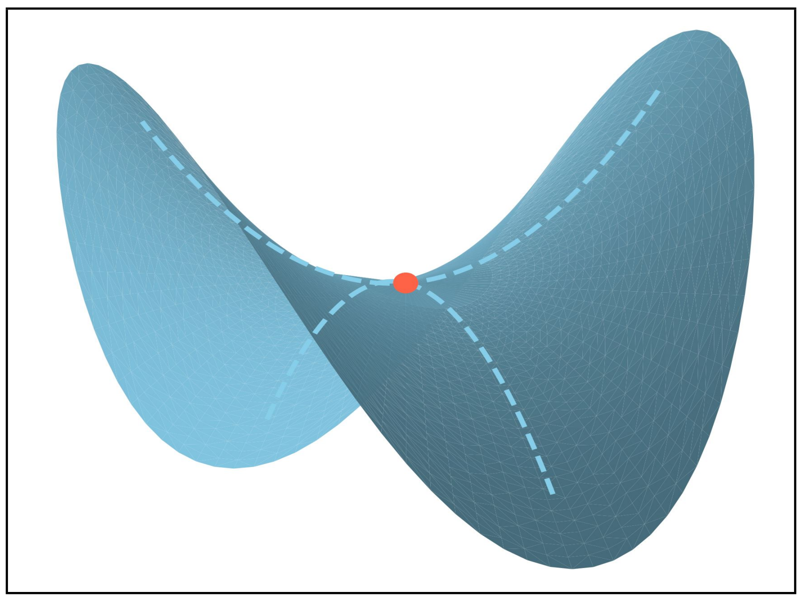
We consider the linearly constrained convex optimization problem

$$\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}. \quad (1.5)$$

The Lagrangian function of the problem (1.5) is

$$L(x,\lambda) = \theta(x) - \lambda^T(Ax - b),$$

which is defined on $\mathcal{X} \times \Re^m$.



A pair of (x^*, λ^*) is called a saddle point of the Lagrange function, if

$$L_{\lambda\in\Re^m}(x^*,\lambda) \le L(x^*,\lambda^*) \le L_{x\in\mathcal{X}}(x,\lambda^*).$$

An equivalent expression of the saddle point is the following variational inequality:

$$\begin{cases} x^* \in \mathcal{X}, & \theta(x) - \theta(x^*) + (x-x^*)^T(-A^T\lambda^*) \ge 0, \quad \forall\, x \in \mathcal{X}, \\ \lambda^* \in \Re^m, & (\lambda-\lambda^*)^T(Ax^* - b) \ge 0, \quad \forall\, \lambda \in \Re^m. \end{cases}$$

Thus, by denoting

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Re^m, \quad (1.6)$$

the optimal condition can be characterized as a monotone variational inequality:

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

Note that the operator F is monotone, because

$$(w-\tilde{w})^T(F(w)-F(\tilde{w})) \ge 0, \ \text{Here } (w-\tilde{w})^T(F(w)-F(\tilde{w})) = 0. \quad (1.8)$$

Example 1 of the problem (1.5): Finding the nearest correlation matrix

A positive semi-definite matrix, whose each diagonal element is equal 1, is called the correlation matrix. For given symmetric $n \times n$ matrix C, the mathematical form of finding the nearest correlation matrix X is

$$\min\{\tfrac{1}{2}\|X - C\|_F^2 \mid \text{diag}(X) = e,\ X \in S_+^n\}, \quad (1.9)$$

where S_+^n is the positive semi-definite cone and e is a n-vector whose each element is equal 1. The problem (1.9) is a concrete problem of type (1.5).

Example 2 of the problem (1.5): The matrix completion problem

Let M be a given $m \times n$ matrix, Π is the elements indices set of M,

$$\Pi \subset \{(ij) | i \in \{1, \ldots, m\},\ j \in \{1, \ldots, n\}\}.$$

The mathematical form of the matrix completion problem is relaxed to

$$\min\{\|X\|_* \mid X_{ij} = M_{ij},\ (ij) \in \Pi\}, \quad (1.10)$$

where $\|\cdot\|_*$ is the nuclear norm–the sum of the singular values of a given matrix. The problem (1.10) is a convex optimization of form (1.5). The matrix A in (1.5) for the linear constraints

$$X_{ij} = M_{ij},\ (ij) \in \Pi,$$

is a projection matrix, and thus $\|A^T A\| = 1$.

M is low Rank, only some elements of M are known.

$$\begin{array}{cccccccccc}
\ast & & & \ast & & \ast & & & \ast & \\
 & & \ast & & \ast & & & \ast & & \\
\ast & & & \ast & & \ast & & \ast & & \ast \\
 & & \ast & & & \ast & & & \ast & \\
 & \ast & & & \ast & & \ast & & & \ast \\
\ast & & & \ast & & \ast & & & \ast & \\
 & \ast & & & \ast & & \ast & & & \ast \\
\ast & & \ast & & \ast & \ast & & \ast & \ast & \\
 & & & & & \ast & & & & \ast \\
\ast & & & \ast & & & \ast & & \ast & \\
 & \ast & & \ast & & \ast & & \ast & & \\
 & & \ast & & \ast & & \ast & & \ast & \\
\ast & & \ast & & \ast & & & \ast & & \ast \\
 & \ast & & \ast & & & \ast & & \ast &
\end{array}$$

Convex optimization problem with two separable objective functions

Some convex optimization problems have the following separable structure:

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

The Lagrangian function of this problem is

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

The saddle point $((x^*, y^*), \lambda^*)$ of $L^{(2)}(x, y, \lambda)$ is a solution of the following VI:

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

where

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

and

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

Convex optimization problem with three separable objective functions

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

Its 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 saddle point $((x^*,y^*,z^*),\lambda^*)$ of $L^{(3)}(x,y,z,\lambda)$ is a solution of the VI:

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

where

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

and

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

1.3 Proximal point algorithms for convex optimization

Lemma 1.2 *Let the vectors $a,b\in\Re^n$, $H\in\Re^{n\times n}$ be a positive definite matrix. If $b^TH(a-b)\ge 0$, then we have*

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

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

Convex Optimization Now, let us consider the *simple* convex optimization

$$\min\{\theta(x)+f(x)\mid x\in\mathcal{X}\}, \tag{1.12}$$

where $\theta(x)$ and $f(x)$ are convex but $\theta(x)$ is not necessary smooth, $\mathcal{X}$ is a closed convex set.

For solving (1.12), the k-th iteration of the proximal point algorithm (abbreviated to PPA) [31, 34] begins with a given x^k, offers the new iterate x^{k+1} via the recursion

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

Since x^{k+1} is the optimal solution of (1.13), it follows from Lemma 1.1 that

$$\theta(x)-\theta(x^{k+1})+(x-x^{k+1})^T\{\nabla f(x^{k+1})+r(x^{k+1}-x^k)\}\geq 0,\ \forall x\in\mathcal{X}. \tag{1.14}$$

Setting $x=x^*$ in the above inequality, it follows that

$$(x^{k+1}-x^*)^T r(x^k-x^{k+1})\geq\theta(x^{k+1})-\theta(x^*)+(x^{k+1}-x^*)^T\nabla f(x^{k+1}). \tag{1.15}$$

Since

$$(x^{k+1}-x^*)^T\nabla f(x^{k+1})\geq(x^{k+1}-x^*)^T\nabla f(x^*),$$

it follows that the right hand side of (1.15) is nonnegative. And consequently,

$$(x^{k+1}-x^*)^T(x^k-x^{k+1})\geq 0. \tag{1.16}$$

Let $a=x^k-x^*$ and $b=x^{k+1}-x^*$ and using Lemma 1.2, we obtain

$$\|\boldsymbol{x^{k+1}-x^*}\|^2\leq\|\boldsymbol{x^k-x^*}\|^2-\|\boldsymbol{x^k-x^{k+1}}\|^2, \tag{1.17}$$

which is the nice convergence property of Proximal Point Algorithm.

The residue sequence $\{\|x^k-x^{k+1}\|\}$ is also monotonically no-increasing.

Proof. Replacing $k+1$ in (1.14) with k, we get

$$\theta(x)-\theta(x^k)+(x-x^k)^T\{\nabla f(x^k)+r(x^k-x^{k-1})\}\geq 0,\quad\forall x\in\mathcal{X}.$$

Let $x=x^{k+1}$ in the above inequality, it follows that

$$\theta(x^{k+1})-\theta(x^k)+(x^{k+1}-x^k)^T\{\nabla f(x^k)+r(x^k-x^{k-1})\}\geq 0. \tag{1.18}$$

Setting $x=x^k$ in (1.14), we become

$$\theta(x^k)-\theta(x^{k+1})+(x^k-x^{k+1})^T\{\nabla f(x^{k+1})+r(x^{k+1}-x^k)\}\geq 0. \tag{1.19}$$

Adding (1.18) and (1.19) and using $(x^k-x^{k+1})^T[\nabla f(x^k)-\nabla f(x^{k+1})]\geq 0$,

$$(x^k-x^{k+1})^T\{(x^{k-1}-x^k)-(x^k-x^{k+1})\}\geq 0. \tag{1.20}$$

Setting $a=x^{k-1}-x^k$ and $b=x^k-x^{k+1}$ in (1.20) and using (1.11), we obtain

$$\boxed{\|\boldsymbol{x^k-x^{k+1}}\|^2\leq\|\boldsymbol{x^{k-1}-x^k}\|^2-\|\boldsymbol{(x^{k-1}-x^k)-(x^k-x^{k+1})}\|^2.} \tag{1.21}$$

We write the problem (1.12) and its PPA (1.13) in VI form

The equivalent variational inequality form of the optimization problem (1.12) is

$$x^* \in \mathcal{X}, \;\; \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \;\; \forall\, x \in \mathcal{X}. \qquad (1.22a)$$

For solving the problem (1.12), the variational inequality form of the k-th iteration of the PPA (see (1.14)) is:

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \;\; & \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \nabla f(x^{k+1}) \\ & \geq (x - x^{k+1})^T r(x^k - x^{k+1}), \;\; \forall\, x \in \mathcal{X}. \end{aligned} \qquad (1.22b)$$

PPA reaches the solution of (1.22a) via solving a series of subproblems (1.22b).
采用的是步步为营的策略, 稳扎稳打 !

Using (1.22), we study PPA for VI arising from the constrained optimization

1.4 Preliminaries of PPA for Variational Inequalities

The optimal condition of the problem (1.5) is characterized as a mixed monotone variational inequality:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \qquad (1.23)$$

PPA for VI in Euclidean-norm For given w^k and $r > 0$, find w^{k+1}

$$\begin{aligned} w^{k+1} \in \Omega, \quad & \theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ & \geq (w - w^{k+1})^T r(w^k - w^{k+1}), \; \forall\, w \in \Omega. \end{aligned} \qquad (1.24)$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (1.23).

✠ **w^k is the solution of (1.23) if and only if $w^k = w^{k+1}$** ✠

Setting $w = w^*$ in (1.24), we obtain

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

Note that (see the structure of $F(w)$ in (1.6))

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

and consequently (by using (1.23)) we obtain

$$(w^{k+1}-w^*)^Tr(w^k-w^{k+1}) \geq \theta(x^{k+1})-\theta(x^*)+(w^{k+1}-w^*)^TF(w^*) \geq 0,$$

and thus

$$(w^{k+1}-w^*)^T(w^k-w^{k+1}) \geq 0. \tag{1.25}$$

Now, by setting $a = w^k - w^*$ and $b = w^{k+1} - w^*$ in the inequality (1.25), it is $b^T(a-b) \geq 0$. Using Lemma1.2, we obtain

$$\|w^{k+1}-w^*\|^2 \leq \|w^k-w^*\|^2 - \|w^k-w^{k+1}\|^2. \tag{1.26}$$

We get the nice convergence property of Proximal Point Algorithm:

The sequence $\{w^k\}$ generated by PPA is Fejér monotone.

PPA for monotone mixed VI in H-norm

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

$$\begin{aligned} w^{k+1}\in\Omega, \quad \theta(x)-\theta(x^{k+1})+(w-w^{k+1})^TF(w^{k+1}) \\ \geq (w-w^{k+1})^TH(w^k-w^{k+1}), \ \forall\, w\in\Omega, \end{aligned} \tag{1.27}$$

where H is a symmetric positive definite matrix.

✠ Again, w^k is the solution of (1.23) if and only if $w^k = w^{k+1}$ ✠

Convergence Property of Proximal Point Algorithm in H-norm

$$\|w^{k+1}-w^*\|_H^2 \leq \|w^k-w^*\|_H^2 - \|w^k-w^{k+1}\|_H^2. \tag{1.28}$$

The sequence $\{w^k\}$ is Fejér monotone in H-norm. In customized PPA, via choosing a proper positive definite matrix H, the solution of the subproblem (1.27) has a closed form. An iterative algorithm is called the contraction method, if its generated sequence $\{w^k\}$ satisfies $\|w^{k+1}-w^*\|_H^2 < \|w^k-w^*\|_H^2$.

2 From PDHG to CP-PPA and Customized-PPA

We consider the $\min-\max$ problem

$$\min_x \max_y\{\Phi(x,y)=\theta_1(x)-y^TAx-\theta_2(y)\,|\,x\in\mathcal{X},y\in\mathcal{Y}\}. \tag{2.1}$$

Let (x^*,y^*) be the solution of (2.1), then we have

$$\left\{\begin{array}{ll} x^*\in\mathcal{X}, \ \ \Phi(x,y^*)-\Phi(x^*,y^*)\ge 0, \ \ \forall\, x\in\mathcal{X}, & \text{(2.2a)}\\ y^*\in\mathcal{Y}, \ \ \Phi(x^*,y^*)-\Phi(x^*,y)\ge 0, \ \ \forall\, y\in\mathcal{Y}. & \text{(2.2b)}\end{array}\right.$$

Using the notation of $\Phi(x,y)$, it can be written as

$$\left\{\begin{array}{lll} x^*\in\mathcal{X}, & \theta_1(x)-\theta_1(x^*)+(x-x^*)^T(-A^Ty^*)\ge 0, & \forall\, x\in\mathcal{X},\\ y^*\in\mathcal{Y}, & \theta_2(y)-\theta_2(y^*)\;+\;(y-y^*)^T(\;Ax^*)\;\;\ge 0, & \forall\, y\in\mathcal{Y}.\end{array}\right.$$

Furthermore, it can be written as a variational inequality in the compact form:

$$u\in\Omega, \quad \theta(u)-\theta(u^*)+(u-u^*)^TF(u^*)\ge 0, \ \forall\, u\in\Omega, \tag{2.3}$$

where

$$u=\begin{pmatrix}x\\y\end{pmatrix}, \quad \theta(u)=\theta_1(x)+\theta_2(y), \quad F(u)=\begin{pmatrix}-A^Ty\\Ax\end{pmatrix}, \quad \Omega=\mathcal{X}\times\mathcal{Y}.$$

Since $F(u)=\begin{pmatrix}-A^Ty\\Ax\end{pmatrix}=\begin{pmatrix}0 & -A^T\\A & 0\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$, we have

$$(u-v)^T(F(u)-F(v))\equiv 0.$$

2.1 Original primal-dual hybrid gradient algorithm [38]

For given (x^k,y^k), PDHG [38] produces a pair of (x^{k+1},y^{k+1}). First,

$$x^{k+1}=\text{argmin}\{\Phi(x,y^k)+\frac{r}{2}\|x-x^k\|^2\,|\,x\in\mathcal{X}\}, \tag{2.4a}$$

and then we obtain y^{k+1} via

$$y^{k+1}=\text{argmax}\{\Phi(x^{k+1},y)-\frac{s}{2}\|y-y^k\|^2\,|\,y\in\mathcal{Y}\}. \tag{2.4b}$$

Ignoring the constant term in the objective function, the subproblems (2.4) are reduced to

$$\left\{\begin{array}{ll} x^{k+1}=\text{argmin}\{\theta_1(x)-x^TA^Ty^k+\frac{r}{2}\|x-x^k\|^2\,|\,x\in\mathcal{X}\}, & \text{(2.5a)}\\ y^{k+1}=\text{argmin}\{\theta_2(y)+y^TAx^{k+1}+\frac{s}{2}\|y-y^k\|^2\,|\,y\in\mathcal{Y}\}. & \text{(2.5b)}\end{array}\right.$$

According to Lemma 1.1, the optimality condition of (2.5a) is $x^{k+1}\in\mathcal{X}$ and

$$\theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{-A^Ty^k+r(x^{k+1}-x^k)\}\ge 0, \ \forall\, x\in\mathcal{X}. \tag{2.6}$$

Similarly, from (2.5b) we get $y \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{Ax^{k+1} + s(y^{k+1} - y^k)\} \geq 0, \ \forall\, y \in \mathcal{Y}. \quad (2.7)$$

Combining (2.6) and (2.7), we have

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T y^{k+1} \\ Ax^{k+1} \end{pmatrix} + \begin{pmatrix} r(x^{k+1} - x^k) + A^T(y^{k+1} - y^k) \\ s(y^{k+1} - y^k) \end{pmatrix} \right\} \geq 0, \quad \forall (x, y) \in \Omega.$$

The compact form is $u^{k+1} \in \Omega$,

$$\theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{F(u^{k+1}) + Q(u^{k+1} - u^k)\} \geq 0, \ \forall u \in \Omega, \quad (2.8)$$

where

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \qquad \text{is not symmetric.}$$

It does not be the PPA form (1.27), and we can not expect its convergence.

The following example of linear programming indicates the original PDHG (2.4) is not necessary convergent.

Consider a pair of the primal-dual linear programming:

$$\text{(Primal)} \quad \begin{array}{ll} \min & c^T x \\ \text{s. t.} & Ax = b \\ & x \geq 0. \end{array} \qquad \text{(Dual)} \quad \begin{array}{ll} \max & b^T y \\ \text{s. t.} & A^T y \leq c. \end{array}$$

We take the following example

$$\text{(P)} \quad \begin{array}{ll} \min & x_1 + 2x_2 \\ \text{s. t.} & x_1 + x_2 = 1 \\ & x_1, \ x_2 \geq 0. \end{array} \qquad \text{(D)} \quad \begin{array}{ll} \max & y \\ \text{s. t.} & \begin{bmatrix} 1 \\ 1 \end{bmatrix} y \leq \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{array}$$

where $A = [1, 1], \ b = 1, c = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and the vector $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

The optimal solutions of this pair of linear programming are $x^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $y^* = 1$.

Note that its Lagrange function is

$$L(x, y) = c^T x - y^T (Ax - b) \tag{2.9}$$

which defined on $\Re^2_+ \times \Re$. (x^*, y^*) is the unique saddle point of the Lagrange function.

For the convex optimization problem $\min\{\theta(x) \mid Ax = b, x \in \mathcal{X}\}$, its Lagrangian function is $$L(x, y) = \theta(x) - y^T (Ax - b),$$ which defined on $\mathcal{X} \times \Re^m$. Find the saddle point of the Lagrangian function is a special $\min - \max$ problem (2.1) whose $\theta_1(x) = \theta(x)$, $\theta_2(y) = -b^T y$ and $\mathcal{Y} = \Re^m$.

For solving the $\min$-$\max$ problem (2.9), by using (2.4), the iterative formula is

$$\begin{cases} x^{k+1} = \max\{(x^k + \frac{1}{r}(A^T y^k - c)), 0\}, \\ y^{k+1} = y^k - \frac{1}{s}(Ax^{k+1} - b). \end{cases}$$

We use $(x_1^0, x_2^0; y^0) = (0, 0; 0)$ as the start point. For this example, the method is not convergent.

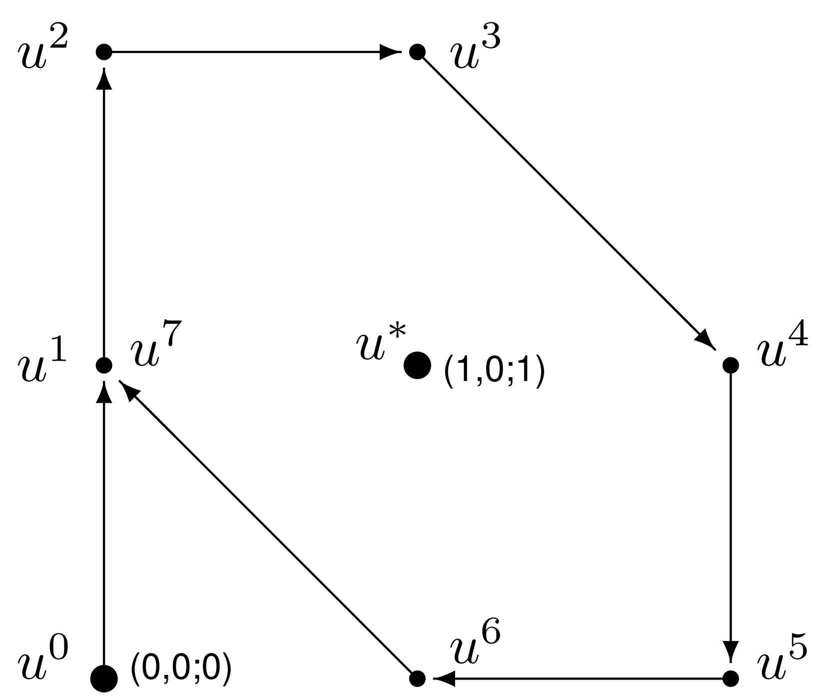
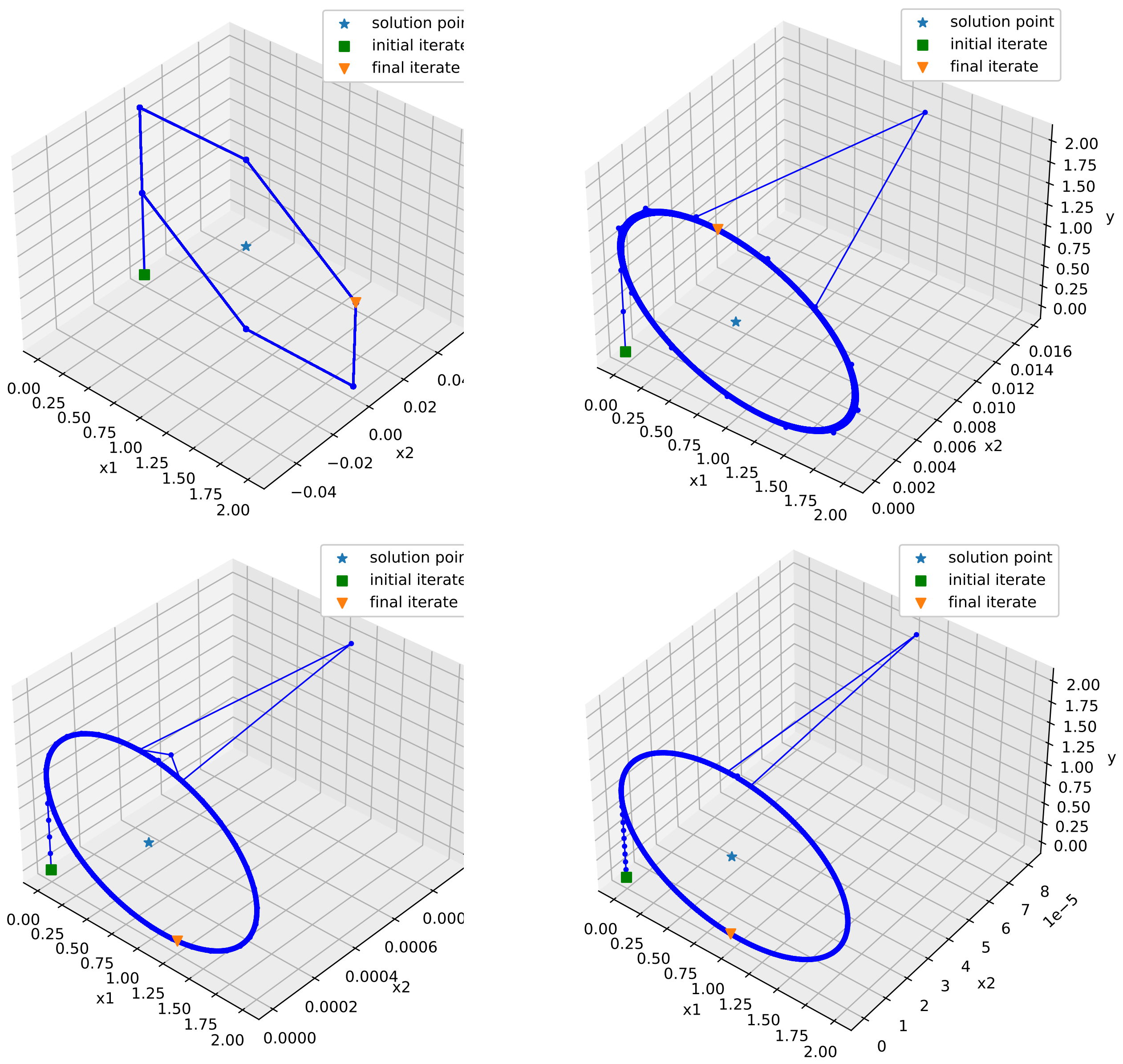


Fig. 4.1 The sequence generated by PDHG Method with $r = s = 1$

$$\begin{aligned} u^0 &= (0, 0; 0) \\ u^1 &= (0, 0; 1) \\ u^2 &= (0, 0; 2) \\ u^3 &= (1, 0; 2) \\ u^4 &= (2, 0; 1) \\ u^5 &= (2, 0; 0) \\ u^6 &= (1, 0; 0) \\ u^7 &= (0, 0; 1) \end{aligned}$$

$$\boldsymbol{u^{k+6} = u^k}$$



对 $r = s = 1, 2, 5, 10$, PDHG 方法都不收敛

2.2 Proximal Point Algorithm – CP-PPA

If we change the non-symmetric matrix Q to a symmetric matrix H such that

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \quad \Rightarrow \quad H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix},$$

then the variational inequality (2.8) will become the following desirable form:

$$\theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{F(u^{k+1}) + H(u^{k+1} - u^k)\} \geq 0, \ \forall u \in \Omega.$$

For this purpose, we need only to change (2.7) in PDHG, namely,

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

to

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{A[2x^{k+1} - x^k] + s(y^{k+1} - y^k)\} \geq 0, \ \forall y \in \mathcal{Y}. \tag{2.10}$$

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{Ax^{k+1} + A(x^{k+1} - x^k) + s(y^{k+1} - y^k)\} \geq 0.$$

Thus, for given (x^k, y^k), producing a proximal point (x^{k+1}, y^{k+1}) via (2.4a) and (2.10) can be summarized as:

$$x^{k+1} = \text{argmin}\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}. \tag{2.11a}$$

$$y^{k+1} = \text{argmax}\{\Phi\big([2x^{k+1} - x^k], y\big) - \frac{s}{2}\|y - y^k\|^2\} \tag{2.11b}$$

By ignoring the constant term in the objective function, getting x^{k+1} from (2.11a) is equivalent to obtaining x^{k+1} from

$$x^{k+1} = \text{argmin}\{\theta_1(x) + \frac{r}{2}\|x - \big[x^k + \frac{1}{r}A^T y^k\big]\|^2 \,|\, x \in \mathcal{X}\}.$$

The solution of (2.11b) is given by

$$y^{k+1} = \text{argmin}\{\theta_2(y) + \frac{s}{2}\|y - \big[y^k + \frac{1}{s}A(2x^{k+1} - x^k)\big]\|^2 \,|\, y \in \mathcal{Y}\}.$$

According to the assumption, there is no difficulty to solve (2.11a)-(2.11b).

In the case that $rs > \|A^T A\|$, the matrix

$$H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \quad \text{is positive definite.}$$

The method (2.11) was first suggested by Chambolle and Pock [4]. A few months later, by using the PPA interpretation, we have proved the convergence as mentioned here [26]. In this sense, we call it **CP-PPA**.

Theorem 2.1 *The sequence $\{u^k = (x^k, \lambda^k)\}$ generated by the CP-PPA (2.11) satisfies*

$$\|u^{k+1} - u^*\|_H^2 \le \|u^k - u^*\|_H^2 - \|u^k - u^{k+1}\|_H^2. \tag{2.12}$$

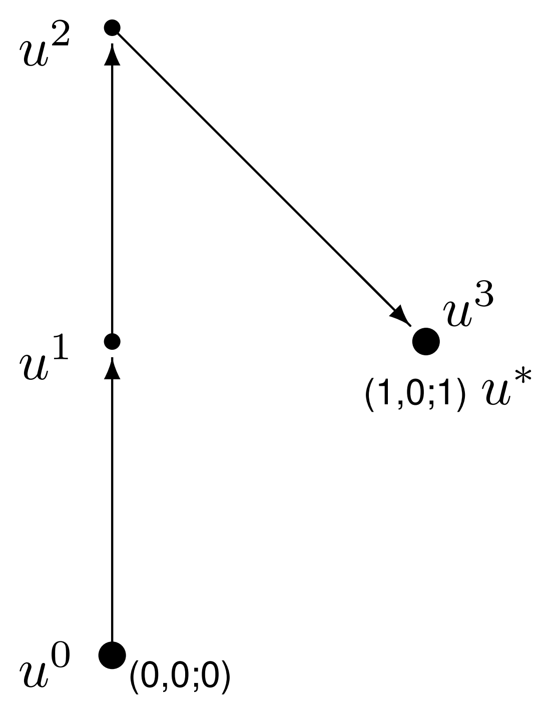
For the minimization problem $\min\{\theta(x) \,|\, Ax = b, x \in \mathcal{X}\}$,
the iterative scheme is

$$x^{k+1} = \text{argmin}\{\theta(x) + \frac{r}{2}\|x - \big[x^k + \frac{1}{r}A^T y^k\big]\|^2 \,|\, x \in \mathcal{X}\}. \tag{2.13a}$$

$$y^{k+1} = y^k - \frac{1}{s}\big[A(2x^{k+1} - x^k) - b\big]. \tag{2.13b}$$

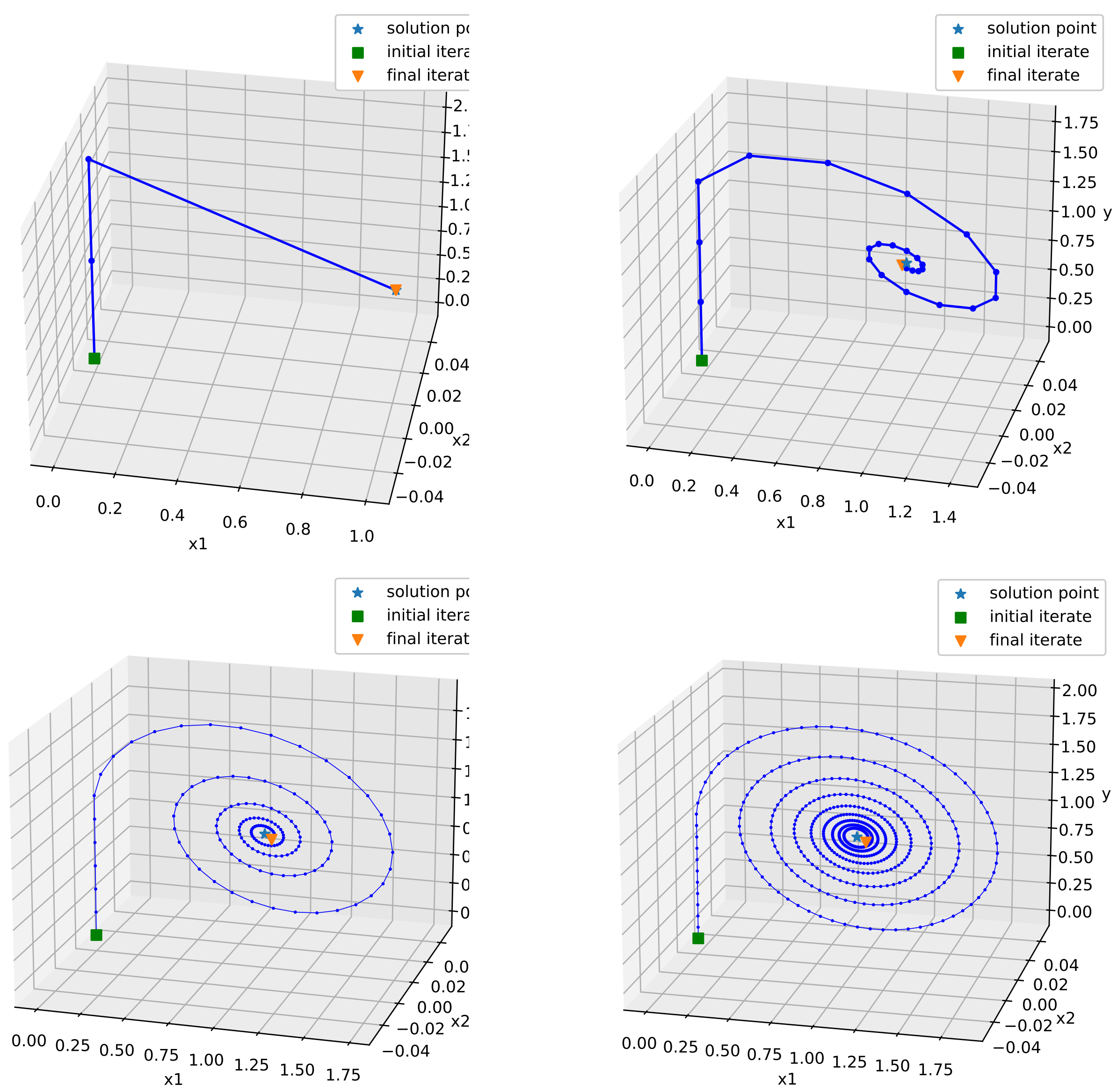
For solving the $\min\text{-}\max$ problem (2.9), by using (2.11), the iterative formula is

$$\begin{cases} x^{k+1} = \max\{(x^k + \frac{1}{r}(A^T y^k - c)), 0\}, \\ y^{k+1} = y^k - \frac{1}{s}[A(2x^{k+1} - x^k) - b]. \end{cases}$$



$$u^0 = (0,0;0)$$
$$u^1 = (0,0;1)$$
$$u^2 = (0,0;2)$$
$$u^3 = (1,0;1)$$
$$\boldsymbol{u^3 = u^*}.$$

Fig. 4.2 The sequence generated by
C-PPA Method with $r = s = 1$



对 $r = s = 1, 2, 5, 10$, C-PPA 方法都收敛. 参数越大, 收敛越慢

Besides (2.11), (x^{k+1}, y^{k+1}) can be produced by using the dual-primal order:

$$y^{k+1} = \text{argmax}\big\{\Phi\big(x^k, y\big) - \frac{s}{2}\big\|y - y^k\big\|^2\big\} \qquad \text{(2.14a)}$$

$$x^{k+1} = \text{argmin}\big\{\Phi(x, (2y^{k+1} - y^k)) + \frac{r}{2}\big\|x - x^k\big\|^2 \,\big|\, x \in \mathcal{X}\big\}. \qquad \text{(2.14b)}$$

By using the notation of u, $F(u)$ and Ω in (2.3), we get $u^{k+1} \in \Omega$ and

$$\theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T\{F(u^{k+1}) + H(u^{k+1} - u^k)\} \geq 0, \;\; \forall u \in \Omega,$$

where

$$H = \begin{pmatrix} rI_n & -A^T \\ -A & sI_m \end{pmatrix}.$$

Note that in the primal-dual order,

$$H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}.$$

In the both cases, $rs > \|A^TA\|$, the matrix H is positive definite.

Remark We use CP-PPA to solve linearly constrained convex optimization.

If the equality constraints $Ax = b$ is changed to $Ax \geq b$, namely,

$$\boxed{\min\{\theta(x) \mid Ax = b, \; x \in \mathcal{X}\}} \quad \Rightarrow \quad \boxed{\min\{\theta(x) \mid Ax \geq b, \; x \in \mathcal{X}\}.}$$

In this case, the Lagrange multiplier y should be nonnegative. $\Omega = \mathcal{X} \times \Re^m_+$.
We need only to make a slight change in the algorithms.

In the primal-dual order (2.11b), it needs to change the update dual update form

$$\boxed{y^{k+1} = \lambda^k - \tfrac{1}{s}\big(A(2x^{k+1} - x^k) - b\big) \;\Rightarrow\; y^{k+1} = \big[y^k - \tfrac{1}{s}\big(A(2x^{k+1} - x^k) - b\big]_+}$$

In the dual-primal order (2.14a), it needs to change the update dual update form

$$\boxed{y^{k+1} = y^k - \tfrac{1}{s}\big(Ax^{k+1} - b\big) \qquad \Rightarrow \qquad y^{k+1} = \big[y^k - \tfrac{1}{s}\big(Ax^{k+1} - b\big)\big]_+}$$

2.3 Simplicity recognition

Frame of VI is recognized by some Researcher in Image Science

Diagonal preconditioning for first order primal-dual algorithms in convex optimization*

Thomas Pock
Institute for Computer Graphics and Vision
Graz University of Technology
pock@icg.tugraz.at

Antonin Chambolle
CMAP & CNRS
École Polytechnique
antonin.chambolle@cmap.polytechnique.fr

- T. Pock and A. Chambolle, IEEE ICCV, 1762-1769, 2011
- A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vison, 40, 120-145, 2011.

preconditioned algorithm. In very recent work [10], it has been shown that the iterates (2) can be written in form of a proximal point algorithm [14], which greatly simplifies the convergence analysis.

From the optimality conditions of the iterates (4) and the convexity of G and F^* it follows that for any $(x, y) \in X \times Y$ the iterates x^{k+1} and y^{k+1} satisfy

$$\left\langle \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}, F\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} + M\begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\rangle \geq 0\,, \tag{5}$$

where

$$\boxed{F\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} \partial G(x^{k+1}) + K^T y^{k+1} \\ \partial F^*(y^{k+1}) - Kx^{k+1} \end{pmatrix}}$$

and

$$M = \begin{bmatrix} \mathrm{T}^{-1} & -K^T \\ -\theta K & \Sigma^{-1} \end{bmatrix}\,. \tag{6}$$

It is easy to check, that the variational inequality (5) now takes the form of a proximal point algorithm [10, 14, 16].

The authors said that the PPA explanation greatly simplifies the convergence analysis

We think that only when the matrix M in (6) is symmetric and positive definite, the related method is convergent.

Otherwise, as we have shown in the previous example, the method is not necessarily convergent

In the matrix M, the parameter $\theta = 0$ can not guarantee the convergence.

For $\theta \in (0, 1)$, there is not proof for the convergence, it is still an Open Problem.

[9] L. Ford and D. Fulkerson. *Flows in Networks*. Princeton University Press, Princeton, New Jersey, 1962.

[10] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for total variation image restoration. Technical report, Nanjing University, China, 2010.

Later, the Reference [10] is published in SIAM J. Imaging Science [26].

Chambolle and Pock's Math. Progr. Paper only uses the PPA form $\theta = 1$.

Math. Program., Ser. A
DOI 10.1007/s10107-015-0957-3

CrossMark

FULL LENGTH PAPER

On the ergodic convergence rates of a first-order primal–dual algorithm

Antonin Chambolle[1] · Thomas Pock[2,3]

The paper published by Chambolle and Pock in Math. Progr. uses the VI framework

1 Introduction

In this work we revisit a first-order primal–dual algorithm which was introduced in [15, 26] and its accelerated variants which were studied in [5]. We derive new estimates for the rate of convergence. In particular, exploiting a proximal-point interpretation due to [16], we are able to give a very elementary proof of an ergodic $O(1/N)$ rate of convergence (where N is the number of iterations), which also generalizes to non-

Algorithm 1: $O(1/N)$ Non-linear primal–dual algorithm

- Input: Operator norm $L := \|K\|$, Lipschitz constant L_f of ∇f, and Bregman distance functions D_x and D_y.
- Initialization: Choose $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$, $\tau, \sigma > 0$
- Iterations: For each $n \geq 0$ let

$$(x^{n+1}, y^{n+1}) = \mathcal{PD}_{\tau,\sigma}(x^n, y^n, 2x^{n+1} - x^n, y^n) \qquad (11)$$

The elegant interpretation in [16] shows that by writing the algorithm in this form

♣ The authors mentioned, the elegant explanation in [16] (our paper) shows that by writing the algorithm in (our suggested form), it can be regarded as Proximal Point Algorithm, $\cdots$, A proof of convergence is easily deduced.

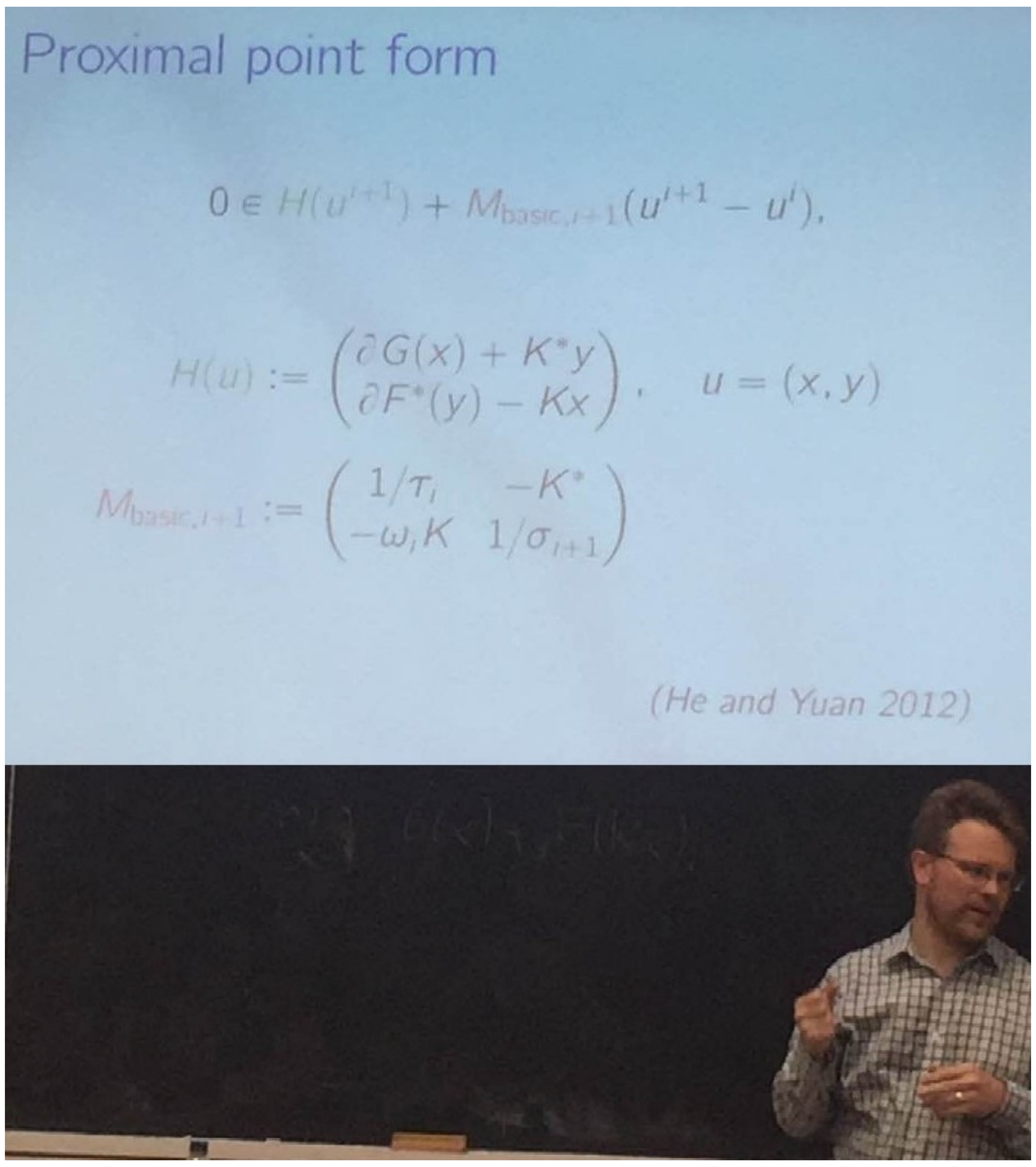
15. Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal–dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. **3**(4), 1015–1046 (2010)
16. He, B., Yuan, X.: Convergence analysis of primal–dual algorithms for a saddle-point problem: from contraction perspective. SIAM J. Imaging Sci. **5**(1), 119–149 (2012)
17. He, B., Yuan, X.: On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method. SIAM J. Numer. Anal. **50**(2), 700–709 (2012)

♣ Reference [16] is our paper published on SIAM J. Imaging Science, 2012. Reference [17] shows the $O(1/t)$ convergence rate of the alternating direction method of multipliers, it published on SIAM J. Numerical Analysis, 2012.

We are showing these quotations to illustrate:

♣ By using the framework of proximal point algorithm for variational inequality, the proof of convergence can be greatly simplified. This approach is regarded a very simple yet powerful technique for analysing the optimization methods (S. Becker, 2011, 2019).

♣ Only a simple and clear point of view can be quickly adopted by scholars in the field of application and have some impact and valuable utility.



In July 2017, one of my colleagues from Mathematics Department of Southern University of Science and Technology visited the UK. At an academic conference he attended, the first reporter mentioned that the work is based on the Proximal point form proposed by us (He and Yuan, 2012).

Seeing a slide show about our contribution, my colleague snapped a picture and sent it to me.

It shows that only simple and powerful ideas can be easily spread and accepted.

2.4 Customized PPA – an extended version of PPA

In practical computation, instead of $w^{k+1} = (x^{k+1}, y^{k+1})$, we denote the output of (2.11) and (2.14) by $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k)$. Then, similarly as (1.25), for such $\tilde{w}^k$, we have

$$(\tilde{w}^k - w^*)^T H(w^k - \tilde{w}^k) \geq 0,$$

and thus

$$(w^k - w^*)^T H(w^k - \tilde{w}^k) \geq \|w^k - \tilde{w}^k\|_H^2. \tag{2.15}$$

In the extended PPA, the new iterate is given by

$$w^{k+1} = w^k - \gamma(w^k - \tilde{w}^k), \quad \gamma \in (0, 2). \tag{2.16}$$

The method in this section is called Extended customized PPA **(E-C-PPA)**. From (2.15) and (2.16) follows immediately the following contraction inequality:

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \gamma(2-\gamma)\|w^k - \tilde{w}^k\|_H^2. \tag{2.17}$$

In order to see how to take a relaxed factor $\gamma \in (0, 2)$ in (2.16), we define

$$w^{k+1}(\alpha) = w^k - \alpha(w^k - \tilde{w}^k), \tag{2.18}$$

as a step-size α dependent new iterate. It is natural to consider maximizing the α-dependent profit function

$$\vartheta_k(\alpha) = \|w^k - w^*\|_H^2 - \|w^{k+1}(\alpha) - w^*\|_H^2. \tag{2.19}$$

Using (8.12), we get

$$\begin{aligned} \vartheta_k(\alpha) &= \|w^k - w^*\|_H^2 - \|w^k - w^* - \alpha(w^k - \tilde{w}^k)\|_H^2 \\ &= 2\alpha(w^k - w^*)^T H(w^k - \tilde{w}^k) - \alpha^2\|w^k - \tilde{w}^k\|_H^2. \end{aligned} \tag{2.20}$$

For any fixed solution point w^*, (A1.12) tell us that $\vartheta_k(\alpha)$ is a quadratic function of α. Because w^* is unknown, it is impossible to get the maxmum point of $\vartheta_k(\alpha)$. Fortunately, using (2.15), we have

$$\vartheta_k(\alpha) \geq 2\alpha\|w^k - \tilde{w}^k\|_H^2 - \alpha^2\|w^k - \tilde{w}^k\|_H^2 = q_k(\alpha). \tag{2.21}$$

The right hand side of the last inequality is defined by $q_k(\alpha)$, which is a lower bound function of $\vartheta_k(\alpha)$ and quadratic. It is clear that $q_k(\alpha)$ reaches its maximum at $\alpha_k^* = 1$.

Recall that ideal is to maximize the unknown quadratic profit function $\vartheta_k(\alpha)$ (see (A1.12)), and $q_k(\alpha)$ is a lower bound function of $\vartheta_k(\alpha)$.

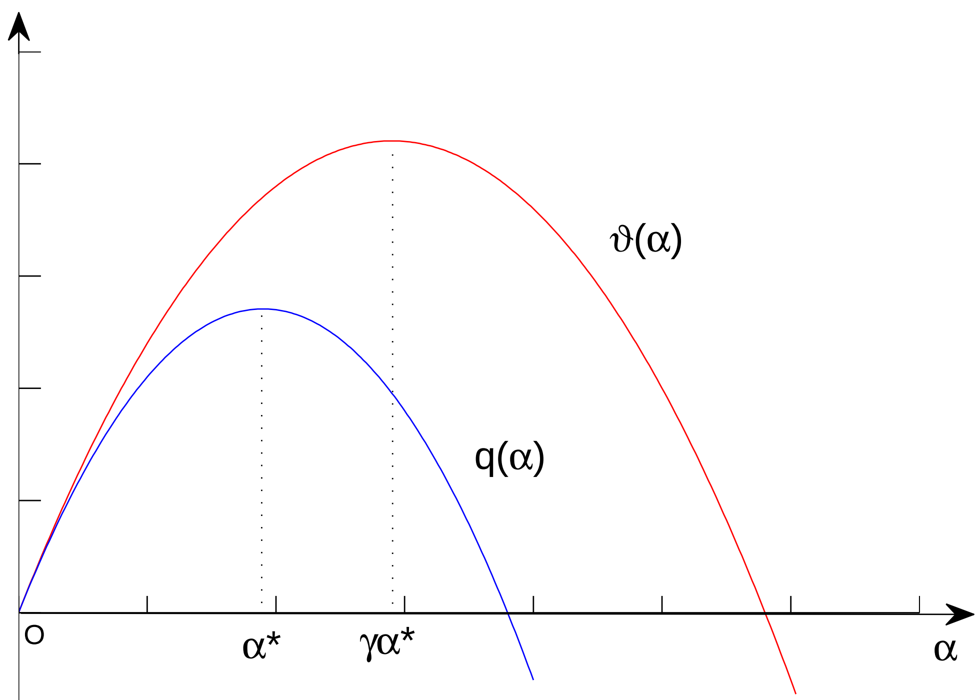


Fig. 3.3 Instruction for $\gamma \in [1, 2)$

We take a relaxed factor $\gamma \in (0, 2)$, and use (2.16) to produce the new iterate. From (2.19) and (8.1.1), the contraction inequality (2.17) follows immediately.

Thus, the Chambolle-Pock method is a special algorithm of (2.16) with $\gamma = 1$. In other words, CP method is a classical customized PPA. In practical computation, we suggest to use the extended customized PPA with $\gamma \in [1.2, 1.8]$.

2.5 Applications in scientific computation

2.5.1 Finding the nearest correlation matrix

$$\min\{\frac{1}{2}\|X - C\|_F^2 \,|\, \text{diag}(X) = e,\ X \in S_+^n\}, \tag{2.22}$$

where e is a n-vector whose each element is equal 1.

The problem has the mathematical form (1.5) with $\|A^T A\| = 1$.

We use $y \in \Re^n$ as the Lagrange multiplier for the linear equality constraint.

Applied Customized PPA to the problem (2.22)

For given (X^k, y^k), produce the predictor (X^{k+1}, y^{k+1}) by using (2.14):

1. Producing y^{k+1} by

$$y^{k+1} = y^k - \frac{1}{s}(\text{diag}(X^k) - e).$$

2. Finding X^{k+1} which is the solution of the following minimization problem

$$\min\{\frac{1}{2}\|X-C\|_F^2+\frac{r}{2}\|X-[X^k+\frac{1}{r}\operatorname{diag}(2y^{k+1}-y^k)]\|_F^2 | X\in S_+^n\}. \tag{2.23}$$

How to solve the subproblem (2.23) The problem (2.23) is equivalent to

$$\min\{\frac{1}{2}\|X-\frac{1}{1+r}[rX^k+\operatorname{diag}(2y^{k+1}-y^k)+C]\|_F^2 | X\in S_+^n\}.$$

The main computational load of each iteration is a SVD decomposition.

Numerical Tests To construct the test examples, we give the matrix C via:

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

In this way, C is symmetric, $C_{jj}\in(0,2)$, and $C_{ij}\in(-1,1)$, for $i\neq j$.

Matlab code for Creating the test examples

```
clear;  close all; n = 1000;      tol=1e-5;     r=2.0;   s=1.01/r;
gamma=1.5; rand('state',0);  C=rand(n,n);   C=(C'+C)-ones(n,n) +
eye(n);
```


Matlab code of the classical Customized PPA

```
%%%     Classical PPA for calibrating correlation matrix           %(1)
function PPAC(n,C,r,s,tol)                                         %(2)
X=eye(n);      y=zeros(n,1);    tic;      %%  The initial iterate  %(3)
stopc=1;                       k=0;                                %(4)
while (stopc>tol && k<=100)         %% Beginning of an Iteration   %(5)
 if mod(k,20)==0 fprintf(' k=%4d   epsm=%9.3e  \n',k,stopc); end;  %(6)
    X0=X;     y0=y;     k=k+1;                                     %(7)
    yt=y0 - (diag(X0)-ones(n,1))/s;               EY=y0-yt;        %(8)
    A=(X0*r + C + diag(yt*2-y0))/(1+r);                            %(9)
    [V,D]=eig(A);   D=max(0,D);  XT=(V*D)*V';     EX=X0-XT;       %(10)
    ex=max(max(abs(EX)));  ey=max(abs(EY));   stopc=max(ex,ey);   %(11)
    X=XT;                  y=yt;                                  %(12)
end;                                      % End of an Iteration   %(13)
toc;                       TB = max(abs(diag(X-eye(n))));         %(14)
fprintf(' k=%4d  epsm=%9.3e   max|X_jj - 1|=%8.5f \n',k,stopc,TB); %%
```

The SVD decomposition is performed by [V,D]=eig(A) in the line (10) of the above code.

The computational load of each decomposition [V,D]=eig(A) is about $9n^3$ flops.

Modifying the Classical PPA to Extended PPA, it needs only change the line (12)

Matlab Code of the Extended Customized PPA

```
%%%     Extended PPA for calibrating correlation matrix            %(1)
function PPAE(n,C,r,s,tol,gamma)                                   %(2)
X=eye(n);      y=zeros(n,1);    tic;     %%  The initial iterate  %(3)
stopc=1;                      k=0;                                 %(4)
while (stopc>tol && k<=100)          %% Beginning of an Iteration  %(5)
 if mod(k,20)==0 fprintf(' k=%4d   epsm=%9.3e  \n',k,stopc); end;  %(6)
    X0=X;     y0=y;     k=k+1;                                     %(7)
    yt=y0 - (diag(X0)-ones(n,1))/s;               EY=y0-yt;        %(8)
    A=(X0*r + C + diag(yt*2-y0))/(1+r);                            %(9)
    [V,D]=eig(A);   D=max(0,D);  XT=(V*D)*V';     EX=X0-XT;       %(10)
    ex=max(max(abs(EX)));  ey=max(abs(EY));   stopc=max(ex,ey);   %(11)
    X=X0-EX*gamma;          y=y0-EY*gamma;                        %(12)
end;                                         % End of an Iteration %(13)
toc;                       TB = max(abs(diag(X-eye(n))));         %(14)
fprintf(' k=%4d  epsm=%9.3e   max|X_jj - 1|=%8.5f \n',k,stopc,TB); %%
```


The difference of the above mentioned codes only in the line 12, the method is much efficient by taking the relaxed factor $\gamma = 1.5$.

Numerical results of (2.22)–SVD by using Mexeig

$n \times n$ Matrix	Classical PPA		Extended PPA	
$n =$	No. It	CPU Sec.	No. It	CPU Sec.
100	30	0.12	23	0.10
200	33	0.54	25	0.40
500	38	7.99	26	6.25
800	38	37.44	28	27.04
1000	45	94.32	30	55.32
2000	62	723.40	38	482.18

The extended PPA converges faster than the classical PPA.

$$\frac{\text{It. No. of Extended PPA}}{\text{It. No. of Classical PPA}} \approx 65\%.$$

2.5.2 Application in matrix completion problem

$$(\textbf{Problem}) \qquad \min\{\|X\|_* \mid X_\Omega = M_\Omega\}. \tag{2.24}$$

We let $Y \in \Re^{n\times n}$ as the Lagrangian multiplier to the constraints $X_\Omega = M_\Omega$.

For given (X^k, Y^k), applying (2.14) to produce (X^{k+1}, Y^{k+1}):

1. Producing Y^{k+1} by

$$Y_\Omega^{k+1} = Y_\Omega^k - \frac{1}{s}(X_\Omega^k - M_\Omega). \tag{2.25}$$

2. Finding X^{k+1} by

$$X^{k+1} = \arg\min \left\{\|X\|_* + \frac{r}{2}\left\|X - \left[X^k + \frac{1}{r}(2Y_\Omega^{k+1} - Y_\Omega^k)\right]\right\|_F^2\right\}. \tag{2.26}$$

Then, the new iterate is given by

$$X^{k+1} := X^k - \gamma(X^k - X^{k+1}), \qquad Y^{k+1} := Y^k - \gamma(Y^k - Y^{k+1}).$$

Test examples The test examples is taken from

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

Code for Creating the test examples of Matrix Completion

```
%%  Creating the test examples of the matrix Completion problem     %(1)
clear all;  clc                                                      %(2)
maxIt=100;        tol = 1e-4;                                        %(3)
r=0.005;          s=1.01/r;        gamma=1.5;                        %(4)
  n=200;            ra = 10;         oversampling = 5;               %(5)
% n=1000;    ra=100;     oversampling = 3; %%  Iteration No. 31      %(6)
% n=1000;    ra=50;      oversampling = 4; %%  Iteration No. 36      %(7)
% n=1000;    ra=10;      oversampling = 6; %%  Iteration No. 78      %(8)
%% Generating the test problem                                       %(9)
rs = randseed;         randn('state',rs);                           %(10)
M=randn(n,ra)*randn(ra,n);            %% The matrix will be completed %(11)
df =ra*(n*2-ra);                      %% The freedom of the matrix   %(12)
mo=oversampling;                                                    %(13)
m =min(mo*df,round(.99*n*n));          %% No. of the known elements  %(14)
Omega= randsample(n^2,m);              %% Define the subset Omega    %(15)
fprintf('Matrix: n=%4d  Rank(M)=%3d  Oversampling=%2d \n',n,ra,mo);%(16)
```

Code: Extended Customized PPA for Matrix Completion Problem

```
function PPAE(n,r,s,M,Omega,maxIt,tol,gamma)    %  Ititial Process %%(1)
X=zeros(n);     Y=zeros(n);      YT=zeros(n);                        %(2)
  nM0=norm(M(Omega),'fro');    eps=1;  VioKKT=1;   k=0;   tic;       %(3)
%% Minimum nuclear norm solution by PPA method                       %(4)
while  (eps > tol && k<= maxIt)                                      %(5)
 if mod(k,5)==0                                                      %(6)
 fprintf('It=%3d |X-M|/|M|=%9.2e VioKKT=%9.2e\n',k,eps,VioKKT); end;%(7)
  k=k+1;      X0=X;      Y0=Y;                                       %(8)
  YT(Omega)=Y0(Omega)-(X0(Omega)-M(Omega))/s;            EY=Y-YT;    %(9)
  A = X0 + (YT*2-Y0)/r;                       [U,D,V]=svd(A,0);     %(10)
  D=D-eye(n)/r;     D=max(D,0);      XT=(U*D)*V';         EX=X-XT;  %(11)
  DXM=XT(Omega)-M(Omega);           eps = norm(DXM,'fro')/nM0;      %(12)
  VioKKT = max( max(max(abs(EX)))*r, max(max(abs(EY))) );           %(13)
   if (eps <= tol)   gamma=1;   end;                                %(14)
  X = X0 - EX*gamma;                                                %(15)
  Y(Omega) = Y0(Omega) - EY(Omega)*gamma;                           %(16)
end;                                                                %(17)
 fprintf('It=%3d  |X-M|/|M|=%9.2e  Vi0KKT=%9.2e \n',k,eps,VioKKT); %(18)
   RelEr=norm((X-M),'fro')/norm(M,'fro');   toc;                    %(19)
 fprintf(' Relative error = %9.2e  Rank(X)=%3d \n',RelEr,rank(X)); %(20)
 fprintf(' Violation of KKT Condition = %9.2e  \n',VioKKT);         %(21)
```


Numerical Results: using SVD in Matlab

Unknown $n \times n$ matrix M				Computational Results		
n	rank(ra)	m/d_{ra}	m/n^2	$\#$iters	times(Sec.)	relative error
1000	10	6	0.12	76	841.59	9.38E-5
1000	50	4	0.39	37	406.24	1.21E-4
1000	100	3	0.58	31	362.58	1.50E-4

Numerical Results: Using SVD in PROPACK

Unknown $n \times n$ matrix M				Computational Results		
n	rank(ra)	m/d_{ra}	m/n^2	$\#$iters	times(Sec.)	relative error
1000	10	6	0.12	**76**	30.99	9.30E-5
1000	50	4	0.39	**36**	40.25	1.29E-4
1000	100	3	0.58	**30**	42.45	1.50E-4

♣ The paper by Cai *et. al* is the first publication in SIAM J. Opti. for matrix completion problem. For the same accuracy, the iteration numbers are listed in the last column of the above table (See the first 3 examples in Table 5.1 of [2], Page. 1974).

♣ The reader may find, for the two examples in in §2.4, the constrained matrix A is a projection matrix and thus $\|A^TA\| = 1$, thus we take $rs = 1.01$. However, we take $r = 2$ an $r = 1/200$ in §2.4.1 and §2.4.2, respectively. r is problems-dependent.

3 From augmented Lagrangian method to ADMM

For the primal-dual methods and customized PPA in the last section, we assume that the subproblem $\min\{\theta(x) + \frac{r}{2}\|x - a\|^2 \,|\, x \in \mathcal{X}\}$ is simple. Since this section, the mathematical form of the sub-problems of the proposed methods is

$$\min\{\theta(x) + \frac{\beta}{2}\|Ax - p\|^2 \,|\, x \in \mathcal{X}\}, \tag{3.1}$$

where $\beta > 0$, and p is a given vector. In comparison with the subproblem in the last section, the subproblem (3.1) is a little bit difficult. However, we still assume its solution has a closed-form representation or it can be efficiently solved up to a high precision.

3.1 Augmented Lagrangian Method

We consider the convex optimization (1.5), namely

$$\min\{\theta(x) \,|\, Ax = b,\, x \in \mathcal{X}\}.$$

Its augmented Lagrangian function is

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

where the additional quadratic term is the penalty for the linear constraints $Ax = b$. The k-th iteration of the **Augmented Lagrangian Method** [28, 32] begins with a given λ^k, obtain $w^{k+1} = (x^{k+1}, \lambda^{k+1})$ via

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

This is equivalent to

$$\left\{ \begin{array}{l} x^{k+1} = \arg\min\big\{L(x, \lambda^k) + \frac{\beta}{2}\|Ax - b\|^2 \;\big|\; x \in \mathcal{X}\big\}, \\ \lambda^{k+1} = \arg\max\big\{L(x^{k+1}, \lambda) - \frac{1}{2\beta}\|\lambda - \lambda^k\|^2 \;\big|\; \lambda \in \Re^m\big\}. \end{array} \right.$$

where $L(x, \lambda) = \theta(x) - \lambda^T(Ax - b)$ is the usual Lagrangian function. In (3.2), x^{k+1} is only a computational result of (3.2a) from given λ^k, it is called the intermediate variable. In order to start the k-th iteration of ALM, we need only to have λ^k and thus we call it as the essential variable. The optimal condition can be written as $w^{k+1} \in \Omega$ and

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

The above relations can be written as

$$\theta(x) - \theta(x^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} -A^T\lambda^{k+1} \\ Ax^{k+1} - b \end{pmatrix} \geq (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}), \quad \forall\, w \in \Omega. \tag{3.3}$$

Setting $w = w^*$ in (3.3) and using the notations in (1.6), we get

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

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

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

Thus, we have

$$(\lambda^{k+1} - \lambda^*)^T(\lambda^k - \lambda^{k+1}) \geq 0. \tag{3.4}$$

By using Lemma 1.2, we obtain

$$\|\lambda^{k+1} - \lambda^*\|^2 \leq \|\lambda^k - \lambda^*\|^2 - \|\lambda^k - \lambda^{k+1}\|^2. \tag{3.5}$$

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

3.2 Alternating direction method of multipliers

We consider the following structured constrained convex optimization problem

$$\boxed{\min\{\theta_1(x) + \theta_2(y)\,|\, Ax + By = b,\ x \in \mathcal{X},\ y \in \mathcal{Y}\} \tag{3.6}}$$

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

Let λ be the Lagrangian multiplier for the linear constraints $Ax + By = b$ in (3.6), the Lagrangian function of this problem is

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

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

$$\left\{ \begin{array}{r} \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, \quad \forall x \in \mathcal{X} \\ \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-B^T\lambda^*) \geq 0, \quad \forall y \in \mathcal{Y} \\ (\lambda - \lambda^*)^T(Ax^* + By^* - b) \geq 0, \quad \forall \lambda \in \Re^m \end{array} \right. \tag{3.7}$$

Note that these first order optimal conditions (3.7) can be written in a compact form such as

$$w^* \in \Omega, \ \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \ \forall\, w \in \Omega. \tag{3.8a}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix} \tag{3.8b}$$

and

$$\theta(u) = \theta_1(x) + \theta_2(y) \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m, \tag{3.8c}$$

Note that the mapping F is monotone. We use Ω^* to denote the solution set of the variational inequality (3.8).

The augmented Lagrange Function of (3.6) is

$$\begin{aligned} \mathcal{L}_\beta^{[2]}(x, y, \lambda) &= L(x, y, \lambda) + \frac{\beta}{2}\|Ax + By - b\|^2 \\ &= \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2. \end{aligned} \tag{3.9}$$

Applying ALM to the structured Convex Optimization problem (3.6)

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

$$\boxed{\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \text{Argmin} \left\{ \begin{array}{c|c} \theta_1(x) + \theta_2(y) - (\lambda^k)^T(Ax + By - b) & x \in \mathcal{X} \\ +\frac{\beta}{2}\|Ax + By - b\|^2 & y \in \mathcal{Y} \end{array} \right\} \tag{3.10}}$$

The new iterate

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{3.11}$$

Convergence

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

Shortcoming

The structure property is not used !

To overcome the shortcoming of the ALM for the problem (3.6), we use the alternating direction method of multipliers. The main idea is splitting the subproblem (3.10) in two parts and only the x-part is the intermediate variable. The iteration begins with $v^0 = (y^0, \lambda^0)$.

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

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

$$x^{k+1} = \text{Argmin}\left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k - b) \\ + \frac{\beta}{2}\|Ax + By^k - b\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \tag{3.12a}$$

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

$$y^{k+1} = \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(Ax^{k+1} + By - b) \\ + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \tag{3.12b}$$

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{3.12c}$$

Advantages The sub-problems (3.12a) and (3.12b) are separately solved one by one.

Remark Ignoring the constant term in the objective function, the sub-problems (3.12a) and (3.12b) is equivalent to

$$x^{k+1} = \text{Argmin}\{\theta_1(x) + \tfrac{\beta}{2}\|(Ax + By^k - b) - \tfrac{1}{\beta}\lambda^k\|^2 \,|\, x \in \mathcal{X}\} \tag{3.13a}$$

and

$$y^{k+1} = \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|(Ax^{k+1} + By - b) - \tfrac{1}{\beta}\lambda^k\|^2 \,|\, y \in \mathcal{Y}\} \tag{3.13b}$$

respectively. Note that the equation (3.12c) can be written as

$$(\lambda - \lambda^{k+1})\{(Ax^{k+1} + By^{k+1} - b) + \tfrac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \ge 0, \ \forall \lambda \in \Re^m. \tag{3.13c}$$

Notice that the sub-problems (3.13a) and (3.13b) are the type of

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

and

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

respectively.

Analysis Note that the solution of (3.12a) and (3.12b) satisfies

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \big\{-A^T\lambda^k + \beta A^T\big(Ax^{k+1} + By^k - b\big)\big\} \geq 0, \ \forall\, x \in \mathcal{X} \tag{3.14a}$$

and

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \\ \big\{-B^T\lambda^k + \beta B^T\big(Ax^{k+1} + By^{k+1} - b\big)\big\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{3.14b}$$

respectively. Substituting λ^{k+1} (see (3.12c)) in (3.14) (eliminating λ^k in (3.14)), we get

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \big\{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1})\big\} \geq 0, \ \forall\, x \in \mathcal{X}, \tag{3.15a}$$

and

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

The compact form of (3.15) is $u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$ and

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^T B \\ 0 \end{pmatrix} (y^k - y^{k+1}) \right\} \geq 0, \tag{3.16}$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. We rewrite the about variational inequality in our desirable form

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

Notice that (3.13c) can be written as

$$\lambda^{k+1} \in \Re^m, \quad (\lambda - \lambda^{k+1})^T\{(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0, \quad \forall \lambda \in \Re^m.$$

Combining the last two inequalities, we have $w^{k+1} \in \Omega$ and

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

$$\left. + \begin{pmatrix} 0 & 0 \\ \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} \right\} \ge 0, \ \forall\, w \in \Omega. \quad (3.17)$$

For convenience we use the notations

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

Then, we get the following lemma:

Lemma 3.1 *Let the sequence* $\{w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})\} \in \Omega$ *be generated by* (3.12). *Then, we have*

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \ge (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}). \quad (3.18)$$

where

$$H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \quad (3.19)$$

Proof. First, using the notation of the matrix H, (3.17) can be rewritten as

$$w^{k+1} \in \Omega, \ \ \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w)$$

$$+\,\beta \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1})$$

$$\ge (v - v^{k+1})^T H(v^k - v^{k+1}), \ \ \forall w \in \Omega. \quad (3.20)$$

Setting $w = w^*$ in (3.20), we get

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1})$$

$$\ge \ \ \beta \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1})$$

$$+\{\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1})\}, \ \forall\, w^* \in \Omega^*. \quad (3.21)$$

By using $Ax^* + By^* = b$ and $\beta(Ax^{k+1} + By^{k+1} - b) = \lambda^k - \lambda^{k+1}$ (see (3.12c)), we have

$$\begin{aligned} & \beta\begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) \\ = & \ \beta\{(Ax^{k+1} + By^{k+1}) - (Ax^* + By^*)\}^T B(y^k - y^{k+1}) \\ = & \ (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}). \end{aligned} \quad (3.22)$$

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

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

Substituting (3.22) and the last inequality in (3.20), the assertion of this lemma follows immediately. $\square$

Lemma 3.2 *Let the sequence* $\{w^k = (x^k, y^k, \lambda^k)\} \in \Omega$ *be generated by* (3.12). *Then, we have*

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0. \quad (3.23)$$

Proof. Because (3.15b) is true for the k-th iteration and the previous iteration, we have

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0, \ \forall \, y \in \mathcal{Y}, \quad (3.24)$$

and

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T \{-B^T \lambda^k\} \geq 0, \ \forall \, y \in \mathcal{Y}, \quad (3.25)$$

Setting $y = y^k$ in (3.24) and $y = y^{k+1}$ in (3.25), respectively, and then adding the two resulting inequalities, we get the assertion (3.23) immediately. $\square$

Substituting (3.23) in (3.18), we get

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0, \quad \forall \, v^* \in \mathcal{V}^*. \quad (3.26)$$

Using the above inequality, as from (3.4) to (3.5) in Section 3.1, we have the following theorem, which is the key for the proof of the convergence of ADMM.

Theorem 3.1 *Let the sequence* $\{w^k = (x^k, y^k, \lambda^k)\} \in \Omega$ *be generated by* (3.12). *Then, we have*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \ \forall \, v^* \in \mathcal{V}^*. \quad (3.27)$$

How to choose the parameter β. The efficiency of ADMM is heavily dependent on the parameter β in (3.12). We discuss how to choose a suitable β in the practical computation.

Note that if $\beta A^T B(y^k - y^{k+1}) = 0$, then it follows from (3.16)

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} \geq 0, \ \forall (x,y) \in \mathcal{X} \times \mathcal{Y}. \quad (3.28)$$

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

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

and consequently $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is a solution of the variational inequality (3.8).

In other words, $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is not a solution of (3.8) because

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

We call

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

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

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

it means that the dual residual is too large and thus we should enlarge the parameter β in the augmented Lagrangian function (3.9). Otherwise, we should reduce the parameter β. A simple scheme that often works well is (see, e.g., [24]):

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

where $\mu > 1, \tau > 1$ are parameters. Typical choices might be $\mu = 10$ and $\tau = 2$. The idea behind this penalty parameter update is to try to keep the primal and dual residual norms within a factor of μ of one another as they both converge to zero. This self adaptive adjusting rule has been used by S. Boyd's group [1] and in their Optimization Solver [12].

3.3 Linearized ADMM

The augmented Lagrangian Function of the problem (3.6) is

$$\mathcal{L}_\beta^{[2]}(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)+\frac{\beta}{2}\|Ax+By-b\|^2. \quad (3.29)$$

Solving the problem (3.6) 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

$$\text{(ADMM)}\quad \left\{\begin{array}{ll} x^{k+1}=\arg\min\{\mathcal{L}_\beta^{[2]}(x,y^k,\lambda^k)\mid x\in\mathcal{X}\}, & (3.30a)\\ y^{k+1}=\arg\min\{\mathcal{L}_\beta^{[2]}(x^{k+1},y,\lambda^k)\mid y\in\mathcal{Y}\}, & (3.30b)\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). & (3.30c)\end{array}\right.$$

In optimization problem, the solution is invariant by changing the constant term in the objective function. Thus, by using the augmented Lagrangian function,

$$\begin{aligned} y^{k+1} &= \arg\min\{\mathcal{L}_\beta^{[2]}(x^{k+1},y,\lambda^k)\mid y\in\mathcal{Y}\}\\ &= \arg\min\{\theta_2(y)-y^TB^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2\mid y\in\mathcal{Y}\}.\end{aligned}$$

Thus, by denoting $q^k=b-Ax^{k+1}+\frac{1}{\beta}\lambda^k$, the solution of (3.12b) is given by

$$\min\{\theta_2(y)+\frac{\beta}{2}\|By-q^k\|^2\mid y\in\mathcal{Y}\}. \quad (3.31)$$

In some practical applications, because of the structure of the matrix B, the subproblem (3.31) is not so easy to be solved. In this case, it is necessary to use the linearized version of the ADMM.

Notice that the Taylor expansion of the quadratic term of (3.30b), namely,

$$\begin{aligned}\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2 = & \frac{\beta}{2}\|Ax^{k+1}+By^k-b\|^2\\ & +\beta(y-y^k)^TB^T(Ax^{k+1}+By^k-b)\\ & +\frac{\beta}{2}\|B(y-y^k)\|^2.\end{aligned}$$

Changing the constant term in the objective function of (3.30b) accordingly, we have

$$
\begin{aligned}
y^{k+1} &= \arg\min\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\} \\
&= \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k + \beta y^T B^T (Ax^{k+1} + By^k - b) \\ + \frac{\beta}{2}\|B(y - y^k)\|^2 \end{array} \middle| \; y \in \mathcal{Y} \right\}.
\end{aligned}
$$

So-called linearized version of ADMM, we remove the unfavorable term $\frac{\beta}{2}\|B(y - y^k)\|^2$ in the objective function, and add the term $\frac{s}{2}\|y - y^k\|^2$.

Strictly speaking, it should be a "linearization" plus "regularization" method. It can also be interpreted as:

The term $\frac{\beta}{2}\|B(y - y^k)\|^2$ is replaced with $\frac{s}{2}\|y - y^k\|^2$.

In other words, it is equivalent to adding the term

$$
\frac{1}{2}\|y - y^k\|_{D_B}^2 \qquad (\text{with } D_B = sI_{n_2} - \beta B^T B) \tag{3.32}
$$

to the objective function of (3.30b), we get

$$
\begin{aligned}
y^{k+1} &= \arg\min\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_{D_B}^2 \mid y \in \mathcal{Y}\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k + \beta y^T B^T (Ax^{k+1} + By^k - b) \\ + \frac{s}{2}\|y - y^k\|^2 \end{array} \middle| \; y \in \mathcal{Y} \right\} \\
&= \arg\min\Big\{\theta_2(y) + \frac{s}{2}\|y - d^k\|^2 \mid y \in \mathcal{Y}\Big\},
\end{aligned} \tag{3.33}
$$

where

$$
d^k = y^k - \frac{1}{s}B^T\big[\beta(Ax^{k+1} + By^k - b) - \lambda^k\big].
$$

By using such strategy, the sub-problems of ADMM is simplified. The linearized version of ADMM are applied successfully in scientific computing [29, 33, 36, 37]. The following analysis is based on the fact that the sub-problems (3.12a) and

$$
\min\{\theta_2(y) + \frac{s}{2}\|y - d^k\|^2 \mid y \in \mathcal{Y}\}
$$

are easy to be solved.

Linearized ADMM. For solving the problem (3.6), the k-th iteration of the linearized ADMM begins with given $v^k = (y^k, \lambda^k)$, produces the $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$

via the following procedure:

$$\left\{\begin{array}{ll} x^{k+1} = \arg\min\{\mathcal{L}_{\beta}^{[2]}(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(3.34a)} \\ y^{k+1} = \arg\min\{\mathcal{L}_{\beta}^{[2]}(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_{D_B}^2 \mid y \in \mathcal{Y}\}, & \text{(3.34b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(3.34c)} \end{array}\right.$$

where D_B is defined by (3.32).

First, using the optimality of the sub-problems of (3.34), we prove the following lemma as the base of convergence.

Lemma 3.3 *Let* $\{w^k\}$ *be the sequence generated by Linearized ADMM* (3.34) *for the problem* (3.6). *Then, we have*

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w) \\ &+ \beta(x - x^{k+1})^T A^T (By^k - By^{k+1}) \\ &\geq (y - y^{k+1})^T D_B (y^k - y^{k+1}) \\ &+ \frac{1}{\beta}(\lambda - \lambda^{k+1})^T (\lambda^k - \lambda^{k+1}), \quad \forall w \in \Omega. \end{aligned} \tag{3.35}$$

Proof. For the x-subproblem in (3.34a), by using Lemma 1.1, we have

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) & \\ + (x - x^{k+1})^T \{-A^T\lambda^k + \beta A^T(Ax^{k+1} + By^k - b)\} & \\ \geq 0, \quad \forall x \in \mathcal{X}. & \end{aligned}$$

By using the multipliers update form in (3.34), $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$, the above inequality can be written as

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) & \\ + (x - x^{k+1})^T \{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1})\} & \\ \geq 0, \quad \forall x \in \mathcal{X}. & \end{aligned} \tag{3.36}$$

For the y-subproblem in (3.34b), by using Lemma 1.1, we have

$$\begin{aligned} y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) & \\ + (y - y^{k+1})^T \{-B^T\lambda^k + \beta B^T(Ax^{k+1} + By^{k+1} - b)\} & \\ + (y - y^{k+1})^T D_B (y^{k+1} - y^k) \geq 0, \ \forall y \in \mathcal{Y}. & \end{aligned}$$

Again, by using the update form $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$, the above inequality can be written as

$$\begin{aligned} y^{k+1} \in \mathcal{Y},\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \\ \geq (y - y^{k+1})^T D_B(y^k - y^{k+1})\},\ \forall\, y \in \mathcal{Y}. \end{aligned} \tag{3.37}$$

Notice that the update form for the multipliers, $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$, can be written as $\lambda^{k+1} \in \Re^m$ and

$$(\lambda - \lambda^{k+1})^T\{(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0,\ \forall\, \lambda \in \Re^m. \tag{3.38}$$

Adding (3.36), (3.37) and (3.38), and using the notation in (3.8), we get

$$\begin{aligned} w^{k+1} \in \Omega,\quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ + \beta(x - x^{k+1})^T A^T(By^k - By^{k+1}) \\ \geq (y - y^{k+1})^T D_B(y^k - y^{k+1}) \\ + \frac{1}{\beta}(\lambda - \lambda^{k+1})^T(\lambda^k - \lambda^{k+1}),\quad \forall w \in \Omega. \end{aligned} \tag{3.39}$$

For the term $(w - w^{k+1})^T F(w^{k+1})$ in the left side of (3.39), by using (1.8), we have

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

The assertion (3.35) is proved. $\square$

This lemma is the base for the convergence analysis of the linearized ADMM.

The contractive property of the sequence $\{w^k\}$ by Linearized ADMM (3.34)

In the following we will prove, for any $w^* \in \Omega^*$, the sequence

$$\{\|v^{k+1} - v^*\|_G + \|y^k - y^{k+1}\|_{D_B}^2\}$$

is monotonically decreasing. For this purpose， we prove some lemmas.

Lemma 3.4 *Let $\{w^k\}$ be the sequence generated by Linearized ADMM* (3.34) *for the* *problem* (3.6)*. Then, we have*

$$
\begin{aligned}
w^{k+1} \in \Omega, \;\; \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w) \\
&+ \beta \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) \\
&\geq (v - v^{k+1})^T G(v^k - v^{k+1}), \;\; \forall w \in \Omega,
\end{aligned} \tag{3.40}
$$

where G *is given by*

$$
G = \begin{pmatrix} D_B + \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{3.41}
$$

.

Proof. Adding $(y - y^{k+1})^T \beta B^T B(y^k - y^{k+1})$ to the both sides of (3.35) in Lemma 3.3, and using the notation of the matrix G, we obtain (3.40) immediately and the lemma is proved. □

Lemma 3.5 *Let* $\{w^k\}$ *be the sequence generated by Linearized ADMM* (3.34) *for the*

problem (3.6)*. Then, we have*

$$
(v^{k+1} - v^*)^T G(v^k - v^{k+1}) \geq (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}), \;\; \forall w^* \in \Omega^*. \tag{3.42}
$$

Proof. Setting the $w \in \Omega$ in (3.40) by any $w^* \in \Omega^*$, we obtain

$$
\begin{aligned}
&(v^{k+1} - v^*)^T G(v^k - v^{k+1}) \\
&\quad \geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \\
&\quad + \beta \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}).
\end{aligned} \tag{3.43}
$$

According to the optimality, a part of the terms in the right hand side of the above inequality,

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

Using $Ax^* + By^* = b$ and $\lambda^k - \lambda^{k+1} = \beta(Ax^{k+1} + By^{k+1} - b)$ (see (3.34c)) to deal the last term in the right hand side of (3.43) , it follows that

$$\begin{aligned}
&\beta \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) \\
&= \beta[(Ax^{k+1} - Ax^*) + (By^{k+1} - By^*)]^T B(y^k - y^{k+1}) \\
&= (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}).
\end{aligned}$$

The lemma is proved. □

Lemma 3.6 *Let* $\{w^k\}$ *be the sequence generated by Linearized ADMM* (3.34) *for the problem* (3.6)*. Then, we have*

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq \frac{1}{2}\|y^k - y^{k+1}\|_{D_B}^2 - \frac{1}{2}\|y^{k-1} - y^k\|_{D_B}^2. \quad (3.44)$$

Proof. First, (3.37) represents

$$\begin{aligned}
&y^{k+1} \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \\
&\qquad \{-B^T\lambda^{k+1} + D_B(y^{k+1} - y^k)\} \geq 0, \; \forall\, y \in \mathcal{Y}. \quad (3.45)
\end{aligned}$$

Setting k in (3.45) by $k-1$, we have

$$\begin{aligned}
&y^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(y^k) + (y - y^k)^T \\
&\qquad \{-B^T\lambda^k + D_B(y^k - y^{k-1})\} \geq 0, \;\; \forall\, y \in \mathcal{Y}. \quad (3.46)
\end{aligned}$$

Setting the y in (3.45) and (3.46) by y^k and y^{k+1}, respectively, and adding them, we get

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

From the above inequality we get

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

Using the Cauchy-Schwarz inequality for the right hand side term of the above inequality, we get (3.44) and the lemma is proved. □

By using Lemma 3.5 and Lemma 3.6, we can prove the following convergence theorem.

Theorem 3.2 *Let* $\{w^k\}$ *be the sequence generated by Linearized ADMM* (3.34) *for the problem* (3.6). *Then, we have*

$$\begin{aligned}&\big(\|v^{k+1}-v^*\|_G^2+\|y^k-y^{k+1}\|_{D_B}^2\big)\\ &\quad\le\big(\|v^k-v^*\|_G^2+\|y^{k-1}-y^k\|_{D_B}^2\big)-\|v^k-v^{k+1}\|_G^2,\ \forall w^*\in\Omega^*, \end{aligned}\tag{3.47}$$

where G *is given by* (3.41).

Proof. From Lemma 3.5 and Lemma 3.6, it follows that

$$(v^{k+1}-v^*)^TG(v^k-v^{k+1})\ge\frac{1}{2}\|y^k-y^{k+1}\|_{D_B}^2-\frac{1}{2}\|y^{k-1}-y^k\|_{D_B}^2,\quad\forall w^*\in\Omega^*.$$

Using the above inequality, for any $w^*\in\Omega^*$, we get

$$\begin{aligned}\|v^k-v^*\|_G^2&=\|(v^{k+1}-v^*)+(v^k-v^{k+1})\|_G^2\\ &\ge\|v^{k+1}-v^*\|_G^2+\|v^k-v^{k+1}\|_G^2+2(v^{k+1}-v^*)^TG(v^k-v^{k+1})\\ &\ge\|v^{k+1}-v^*\|_G^2+\|v^k-v^{k+1}\|_G^2\\ &\qquad+\|y^k-y^{k+1}\|_{D_B}^2-\|y^{k-1}-y^k\|_{D_B}^2.\end{aligned}$$

The assertion of the Theorem 3.2 is proved. $\Box$

Optimal linearized ADMM – Main result in OO6228

In the subproblem of the Linearized ADMM, namely (3.34b), in order to ensure the convergence, it was required that

$$D_B=sI_{n_2}-\beta B^TB\quad\text{and}\quad s>\beta\|B^TB\|.\tag{3.48}$$

It is well known that the large parameter s will lead a slow convergence.

Recent Advance in : Bingsheng He, Feng Ma, Xiaoming Yuan:
Optimal Linearized Alternating Direction Method of Multipliers for Convex Programming. http://www.optimization-online.org/DB_HTML/2017/09/6228.html

We have proved: For the matrix D_B in (3.34b) with form (3.48)

- if $s>\frac{3}{4}\beta\|B^TB\|$, the method is still convergent;
- if $s<\frac{3}{4}\beta\|B^TB\|$, there is divergent example.

4 Splitting Methods in a Unified Framework

We study the algorithms using the guidance of variational inequality. Similarly as described in (1.7), together with the Lagrangian multipliers, the optimal condition of the linearly constrained convex optimization is resulted in a variational inequality:

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

The analysis can be fund in [17] (Sections 4 and 5 therin). In order to illustrate the unified framework, let us restudy the augmented Lagrangian method.

4.1 Extended Augmented Lagrangian Method

For the convex optimization (1.5), namely

$$\min\{\theta(x) \,|\, Ax = b,\ x \in \mathcal{X}\}.$$

If we denote the output of (3.2) by $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$, then the optimal condition can be

written as $\tilde{w}^k \in \Omega$ and

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

The above relation can be written as

$$\theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{x}^k - b \end{pmatrix} \geq (\lambda - \tilde{\lambda}^k)^T \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k), \ \forall\, w \in \Omega. \tag{4.2a}$$

In the classical augmented Lagrangian method, $\lambda^{k+1} = \tilde{\lambda}^k$. In practice, we can use relaxation techniques and offer the new iterate by

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

Setting $w = w^*$ in (4.2a), we get $(\tilde{\lambda}^k - \lambda^*)^T \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \geq 0$ and thus

$$(\lambda^k - \lambda^*)^T(\lambda^k - \tilde{\lambda}^k) \geq \|\lambda^k - \tilde{\lambda}^k\|^2. \tag{4.3}$$

Similar as (2.15) in Section 2.4, by using (4.2b) and (4.3), we get

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

In practical computation, we take $\alpha \in (1,2)$ and (4.2) is called the extended augmented Lagrangian method. Usually, it will accelerate the convergence significantly if we take an enlarged $\alpha \in [1.2, 1.8]$. The reason is the same as the one illustrated in Section 2.4.

In order to describe the algorithm prototype, we give the following definition.

Definition (Intermediate variables and Essential Variables)

For an iterative algorithm solving VI(Ω, F, θ), if some coordinates of w are not involved in the beginning of each iteration, then these coordinates are called intermediate variables and those required by the iteration are called essential variables (denoted by v).

- The sub-vector $w \backslash v$ is called intermediate variables.
- In some Algorithms, v is a proper sub-vector of w; however, $v = w$ is also possible.

According to the above mentioned definition, in the the augmented Lagrangian method, x is an intermediate variable and λ is the essential variable.

4.2 Algorithms in a unified framework

A Prototype Algorithm for (4.1)

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \quad \text{(4.4a)}$$

where Q is not necessarily symmetric, but $Q^T + Q$ is essentially positive definite.

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

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \quad \text{(4.4b)}$$

- Usually, we do not take the output of (4.4a), $\tilde{v}^k$, as the new iterate. Thus, $\tilde{v}^k$ is called a predictor. The new iterate v^{k+1} given by (4.4b) is called the corrector.
- We say a matrix G is essentially positive definite, when $G = R^T H R$, H is positive definite, and $\tilde{w}^k$ is a solution of (4.1) when $\|R(v^k - \tilde{v}^k)\| = 0$.
- We use the extended ALM in Section 4.1 as an example. In (4.2a) we have $v = \lambda$, $Q = \frac{1}{\beta} I$, while in the correction step (4.2b), $M = I$ and $\alpha \in (0,2)$.
- When $v^k = \tilde{v}^k$, it follows from (4.4a) directly that $\tilde{w}^k$ is a solution of (4.1). Thus, one can use $\|v^k - \tilde{v}^k\| < \epsilon$ as the stopping criterion in (4.4).

Convergence Conditions

For the matrices Q and M, and the step size α determined in (4.4), the matrices

$$H = QM^{-1} \tag{4.5a}$$

and

$$G = Q^T + Q - \alpha M^T HM. \tag{4.5b}$$

are positive definite (or $H \succ 0$ and $G \succeq 0$).

- We use the extended ALM in Section 4.1 as an example. Since $Q = \frac{1}{\beta}I$ in the prediction step, and $M = I$ and $\alpha \in (0, 2)$ in the correction step, it follows that
  $H = QM^{-1} = \frac{1}{\beta}I$ and $G = Q^T + Q - \alpha M^T HM = \frac{2-\alpha}{\beta}I.$
  Therefore, the convergence conditions are satisfied.
- For $G \succeq 0$, it has the $O(1/t)$ convergence rate in a ergodic sense. If $G \succ 0$, the sequence $\{v^k\}$ is Fèjer monotone and converges to a $v^* \in \mathcal{V}^*$ in H-norm.
- Using the unified framework, the convergence proof is very simple. In addition, it will help us to construct more efficient splitting contraction method for convex optimization with different structures.

Given a positive definite matrix Q in (4.4a) ($Q^T + Q \succ 0$), for satisfying the convergence conditions (4.5), how to choose the matrix M and $\alpha > 0$ in the correction step (4.4b) ?

There are many possibilities, the principle is simplicity and efficiency. See an example:

- In order to ensure the symmetry and positivity of $H = QM^{-1}$, we take
  $$H = QD^{-1}Q^T,$$
  where D is a symmetric investable block diagonal matrix. Because
  $$H = QD^{-1}Q^T \qquad \text{and} \qquad H = QM^{-1},$$
  we only need to set $M^{-1} = D^{-1}Q^T$ and thus
  $$M = Q^{-T}D \quad \text{satisfies the condition (4.5a).}$$
  In this case, $M^T HM = Q^T M = Q^T Q^{-T} D = D.$
- After choosing the matrix M, let
  $$\alpha_{\max} = \arg\max\{\alpha \,|\, Q^T + Q - \alpha M^T HM \succeq 0\},$$
  the condition (4.5b) is satisfied for any $\alpha \in (0, \alpha_{\max})$.

4.3 Customized PPA satisfies the Convergence Condition

Recall the convex optimization problem discussed in Section **??**, namely,

$$\min\{\theta(x) \mid Ax = b, \ x \in \mathcal{X}\}.$$

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

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

where

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

For given $v^k = w^k = (x^k, \lambda^k)$, let the output of the (2.11) as a predictor and denote it as $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$. Then, we have

$$\begin{cases} \tilde{x}^k = \arg\min\{\theta(x) - (\lambda^k)^T(Ax - b) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \frac{1}{s}[A(2\tilde{x}^k - x^k) - b]. \end{cases} \tag{4.6}$$

Similar as (2.2), the output $\tilde{w}^k \in \Omega$ of the iteration (4.6) satisfies

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

It is a form of (4.4a) where

$$Q = H = \begin{pmatrix} rI & A^T \\ A & sI \end{pmatrix}.$$

This matrix is positive definite when $rs > \|A^TA\|$. We take $M = I$ in the correction (4.4b) and the new iterate is updated by

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

Then, we have and

$$H = QM^{-1} = Q \succ 0 \quad \text{and} \quad G = Q^T + Q - \alpha M^T H M = (2 - \alpha)H \succ 0.$$

The convergence conditions (4.5) are satisfied. More about customized PPA, please see

♣ G.Y. Gu, B.S. He and X.M. Yuan, Customized Proximal point algorithms for linearly constrained convex minimization and saddle-point problem: a unified Approach, Comput. Optim. Appl., 59(2014), 135-161.

4.4 Primal-Dual relaxed PPA-based Contraction Methods

For given $v^k = w^k = (x^k, \lambda^k)$, denote the output of (2.4) by $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$, it leads

$$\tilde{x}^k = \arg\min\{\theta(x) - (\lambda^k)^T(Ax-b) + \tfrac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\} \tag{4.7a}$$

and (according to equality constraints $Ax = b$ or inequality constraints $Ax \geq b$)

$$\tilde{\lambda}^k = \lambda^k - \frac{1}{s}(A\tilde{x}^k - b) \quad \text{or} \quad \tilde{\lambda}^k = [\lambda^k - \frac{1}{s}(A\tilde{x}^k - b)]_+. \tag{4.7b}$$

Similar as in (2.8), the predictor $\tilde{w}^k \in \Omega$ generated by (4.7) satisfies

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{4.8}$$

where the matrix

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}, \tag{4.9}$$

is not symmetric. However, (4.8) can be viewed as (4.4a). In this subsection, all the mentioned matrix Q is (4.9). The example in Subsection 2.1 shows that the method is not necessary convergent if we directly take $w^{k+1} = \tilde{w}^k$.

Corrector—the new iterate For given v^k and the predictor $\tilde{v}^k$ by (4.7), we use

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{4.10}$$

to produce the new iterate, where

$$M = \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix}$$

is a upper triangular block matrix whose diagonal part is unit matrix. Note that

$$H = QM^{-1} = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & -\frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix} \succ 0.$$

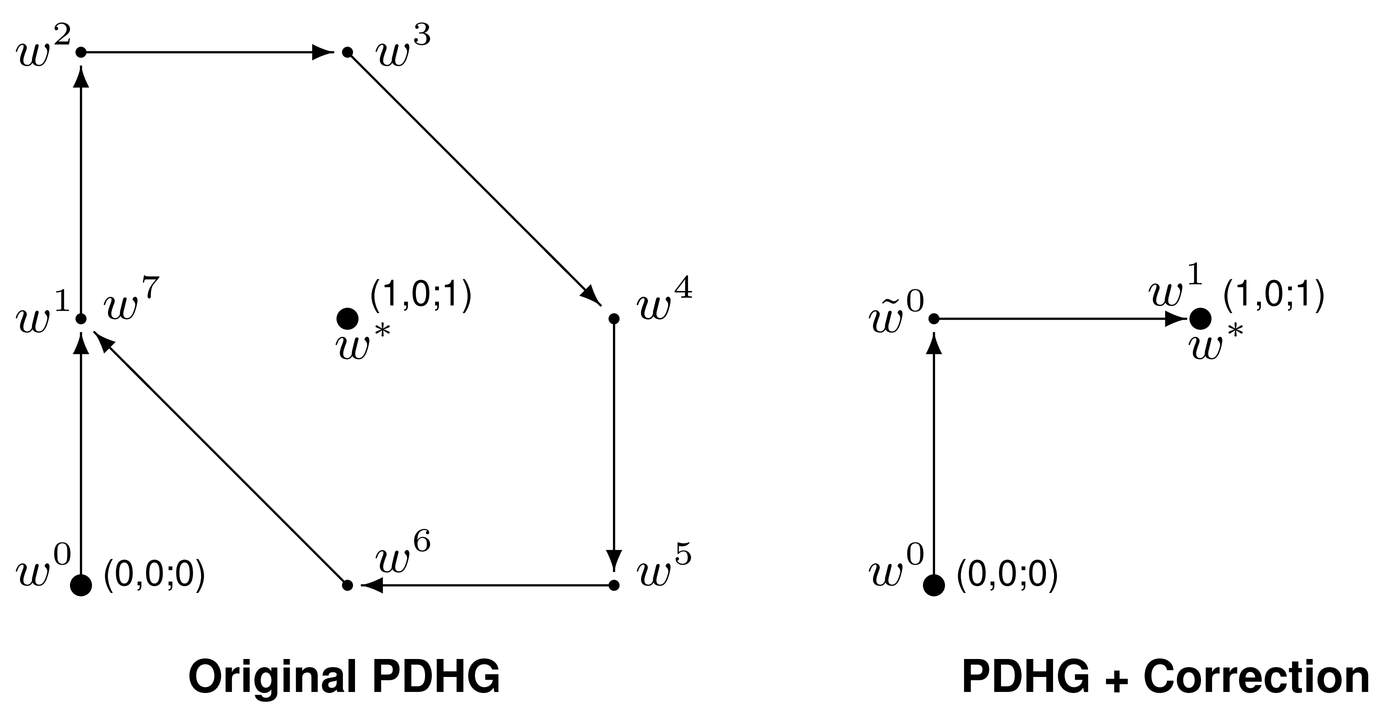
In addition,

$$\begin{aligned} G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} rI_n & 0 \\ 0 & sI_m - \frac{1}{r}AA^T \end{pmatrix}. \end{aligned}$$

G is positive definite when $rs > \|A^T A\|$. The convergence conditions (4.5) are satisfied.

Convergence behaviors for LP **Same toy example as in Section 3**

$$\min\{x_1 + 2x_2 \,|\, x_1 + x_2 = 1,\ x \geq 0\}, \quad (x^*, y^*) = (1, 0; 1).$$



This example shows, sometimes the correction has surprising effectiveness.

In the correction step (4.10), the matrix M is a upper-triangular matrix. We can also use the lower-triangular matrix

$$M = \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix}$$

According to (4.5a), $H = QM^{-1}$, by a simple computation, we have

$$H = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ \frac{1}{s}A & I_m \end{pmatrix} = \begin{pmatrix} rI_n + \frac{1}{s}A^T A & A^T \\ A & sI_m \end{pmatrix}.$$

H is positive definite for any $r, s > 0$. In addition,

$$\begin{aligned} G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix} = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}. \end{aligned}$$

G is positive definite when $rs > \|A^T A\|$. The convergence conditions (4.5) are satisfied.

For a given prediction, there are different corrections which satisfy the convergence conditions (4.5). For example, we can take a convex combination of the above mentioned matrices. Namely, for $\tau \in [0,1]$

$$\begin{aligned} M &= (1-\tau)\begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} + \tau \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix} \\ &= \begin{pmatrix} I_n & \frac{1-\tau}{r}A^T \\ -\frac{\tau}{s}A & I_m \end{pmatrix}. \end{aligned}$$

For this matrix M, we denote

$$\Pi = I + \frac{\tau(1-\tau)}{rs}AA^T.$$

Clearly, Π is positive definite. Let

$$H = \begin{pmatrix} rI_n + \frac{\tau^2}{s}A^T\Pi^{-1}A & \tau A^T\Pi^{-1} \\ \tau\Pi^{-1}A & s\Pi^{-1} \end{pmatrix}.$$

It is easy to verify that H is positive definite for any $r, s > 0$ and

$$HM = Q.$$

Now, we turn to observe the matrix G, it leads that

$$\begin{aligned} G &= Q^T + Q - M^THM = Q^T + Q - Q^TM \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ A & sI_m \end{pmatrix}\begin{pmatrix} I_n & \frac{1-\tau}{r}A^T \\ \frac{-\tau}{s}A & I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n & \tau A^T \\ \tau A & s(I_m - \frac{1-\tau}{rs}AA^T) \end{pmatrix} \\ &= \begin{pmatrix} rI_n & \tau A^T \\ \tau A & \tau^2 sI \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & s(1-\tau)[(1+\tau)I_m - \frac{1}{rs}AA^T] \end{pmatrix}. \end{aligned}$$

For $\tau \in [0,1]$, G is positive definite when $rs > \|A^TA\|$. The convergence conditions (4.5) are satisfied. Especially, in the case $\tau = 1/2$, when $rs > \frac{3}{4}\|A^TA\|$,

$$G = \begin{pmatrix} rI_n & \frac{1}{2}A^T \\ \frac{1}{2}A & s(I_m - \frac{1}{2rs}AA^T) \end{pmatrix} \succ 0.$$

We do not need to calculate H and G, only verifying their positivity is necessary.

5 Convergence proof in the unified framework

In this section, assuming the conditions (4.5) in the unified framework are satisfied, we prove some convergence properties.

Theorem 5.1 *Let* $\{v^k\}$ *be the sequence generated by a method for the problem* (4.1) *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*

$$\begin{aligned}&\alpha\big(\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k)\big)\\ &\quad\ge\ \frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{\alpha}{2}\|v^k-\tilde{v}^k\|_G^2,\quad \forall w\in\Omega. \qquad (5.1)\end{aligned}$$

Proof. Using $Q=HM$ (see (4.5a)) and the relation (4.4b), the right hand side of (4.4a) can be written as $(v-\tilde{v}^k)^T\frac{1}{\alpha}H(v^k-v^{k+1})$ and hence

$$\alpha\{\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k)\}\ge(v-\tilde{v}^k)^T H(v^k-v^{k+1}),\ \forall w\in\Omega. \qquad (5.2)$$

Applying the identity

$$(a-b)^T H(c-d)=\frac{1}{2}\{\|a-d\|_H^2-\|a-c\|_H^2\}+\frac{1}{2}\{\|c-b\|_H^2-\|d-b\|_H^2\},$$

to the right hand side of (5.2) with

$$a=v,\quad b=\tilde{v}^k,\quad c=v^k,\quad \text{and}\quad d=v^{k+1},$$

we thus obtain

$$\begin{aligned}&(v-\tilde{v}^k)^T H(v^k-v^{k+1})\\ &\quad=\frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{1}{2}\big(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\big). \qquad (5.3)\end{aligned}$$

For the last term of (5.3), we have

$$\begin{aligned}&\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\\ &\quad=\ \|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-(v^k-v^{k+1})\|_H^2\\ &\quad\overset{(4.4b)}{=}\ \|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-\alpha M(v^k-\tilde{v}^k)\|_H^2\\ &\quad=\ 2\alpha(v^k-\tilde{v}^k)^T HM(v^k-\tilde{v}^k)-\alpha^2(v^k-\tilde{v}^k)^T M^T HM(v^k-\tilde{v}^k)\\ &\quad=\ \alpha(v^k-\tilde{v}^k)^T(Q^T+Q-\alpha M^T HM)(v^k-\tilde{v}^k)\\ &\quad\overset{(4.5b)}{=}\ \alpha\|v^k-\tilde{v}^k\|_G^2. \qquad (5.4)\end{aligned}$$

Substituting (5.3), (5.4) in (5.2), the assertion of this theorem is proved. □

5.1 Convergence in a strictly contraction sense

Theorem 5.2 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (4.1) *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*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha\|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.5}$$

Proof. Setting $w = w^*$ in (5.1), we get

$$\begin{aligned} & \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ \ge \quad & \alpha\|v^k - \tilde{v}^k\|_G^2 + 2\alpha\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \end{aligned} \tag{5.6}$$

By using the optimality of w^* and the monotonicity of $F(w)$, we have

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \ge 0$$

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \alpha\|v^k - \tilde{v}^k\|_G^2. \tag{5.7}$$

The assertion (5.5) follows directly. $\square$

For the convergence in a strictly contraction, the matrix G should be positive definite.

5.2 Convergence rate in an ergodic sense

Equivalent Characterization of the Solution Set of VI

For the convergence rate analysis, we need another characterization of the solution set of VI (4.1). It can be described the following theorem and the proof can be found in [9] (Theorem 2.3.5) or [25] (Theorem 2.1).

Theorem 5.3 *The solution set of VI(Ω, F, θ) is convex and it can be characterized as*

$$\Omega^* = \bigcap_{w\in\Omega} \{\tilde{w} \in \Omega : \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \ge 0\big\}. \tag{5.8}$$

Proof. Indeed, if $\tilde{w} \in \Omega^*$, we have

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

By using the monotonicity of F on Ω, this implies that

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

Thus, $\tilde{w}$ belongs to the right-hand set in (5.8). Conversely, suppose $\tilde{w}$ belongs to the latter set of (5.8). Let $w \in \Omega$ be arbitrary. The vector

$$\bar{w} = \alpha\tilde{w} + (1-\alpha)w$$

belongs to Ω for all $\alpha \in (0,1)$. Thus we have

$$\theta(\bar{u}) - \theta(\tilde{u}) + (\bar{w} - \tilde{w})^T F(\bar{w}) \geq 0. \tag{5.9}$$

Because $\theta(\cdot)$ is convex, we have

$$\theta(\bar{u}) \leq \alpha\theta(\tilde{u}) + (1-\alpha)\theta(u) \quad \Rightarrow \quad (1-\alpha)(\theta(u) - \theta(\tilde{u})) \geq \theta(u) - \theta(\bar{u}).$$

Substituting it in (5.9) and using $\bar{w} - \tilde{w} = (1-\alpha)(w - \tilde{w})$, we get

$$(\theta(u) - \theta(\tilde{u})) + (w - \tilde{w})^T F(\alpha\tilde{w} + (1-\alpha)w) \geq 0$$

for all $\alpha \in (0,1)$. Letting $\alpha \to 1$, it yields

$$(\theta(u) - \theta(\tilde{u})) + (w - \tilde{w})^T F(\tilde{w}) \geq 0.$$

Thus $\tilde{w} \in \Omega^*$. Now, we turn to prove the convexity of Ω^*. For each fixed but arbitrary $w \in \Omega$, the set

$$\{\tilde{w} \in \Omega : \; \theta(\tilde{u}) + \tilde{w}^T F(w) \leq \theta(u) + w^T F(w)\}$$

and its equivalent expression

$$\{\tilde{w} \in \Omega : \; \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \geq 0\}$$

is convex. Since the intersection of any number of convex sets is convex, it follows that the solution set of VI(Ω, F, θ) is convex. $\Box$

In Theorem 5.3, we have proved the equivalence of

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

and

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

We use the late one to define the approximate solution of VI (4.1). Namely, for given $\epsilon > 0$, $\tilde{w} \in \Omega$ is called an ϵ-approximate solution of VI(Ω, F, θ), if it satisfies

$$\tilde{w} \in \Omega, \;\; \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq -\epsilon, \;\; \forall\; w \in \mathcal{D}_{(\tilde{w})},$$

where

$$\mathcal{D}_{(\tilde{w})} = \{w \in \Omega \,|\, \|w - \tilde{w}\| \leq 1\}.$$

We need to show that for given $\epsilon > 0$, after t iterations, it can offer a $\tilde{w} \in \mathcal{W}$, such that

$$\tilde{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \big\{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\big\} \leq \epsilon. \tag{5.10}$$

Theorem 5.1 is also the base for the convergence rate proof. Using the monotonicity of F, we have

$$(w - \tilde{w}^k)^T F(w) \geq (w - \tilde{w}^k)^T F(\tilde{w}^k).$$

Substituting it in (5.1), we obtain

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2\alpha}\|v - v^k\|_H^2 \geq \frac{1}{2\alpha}\|v - v^{k+1}\|_H^2, \ \forall w \in \Omega. \tag{5.11}$$

Note that the above assertion is hold for $G \succeq 0$.

Theorem 5.4 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (4.1) *and $\tilde{w}^k$ is obtained in the k-th iteration. Assume that v^k, v^{k+1} and $\tilde{w}^k$ satisfy the conditions in the unified framework and let $\tilde{w}_t$ be defined by*

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{w}^k. \tag{5.12}$$

Then, for any integer number $t > 0$, $\tilde{w}_t \in \Omega$ and

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2\alpha(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{5.13}$$

Proof. First, it holds that $\tilde{w}^k \in \Omega$ for all $k \geq 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (5.12) implies that $\tilde{w}_t \in \Omega$. Summing the inequality (5.11) over $k = 0, 1, \ldots, t$, we obtain

$$(t+1)\theta(u) - \sum_{k=0}^{t} \theta(\tilde{u}^k) + \Big((t+1)w - \sum_{k=0}^{t} \tilde{w}^k\Big)^T F(w) + \frac{1}{2\alpha}\|v - v^0\|_H^2 \geq 0, \quad \forall w \in \Omega.$$

Use the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2\alpha(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{5.14}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{u}^k,$$

we have that

$$\theta(\tilde{u}_t) \le \frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k).$$

Substituting it in (5.14), the assertion of this theorem follows directly. $\Box$

Recall (5.10). The conclusion (5.13) thus indicates obviously that the method is able to generate an approximate solution (i.e., $\tilde{w}_t$) with the accuracy $O(1/t)$ after t iterations. That is, in the case $G \succeq 0$, the convergence rate $O(1/t)$ of the method is established.

- **For the unified framework and the convergence proof, the reader can consult: B.S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman-Rachford splitting method for convex programming, *SIAM Journal on Optimization* 24(2014), 1011-1040.**
- **B. S. He and X. M. Yuan, On the $O(1/n)$ convergence rate of the alternating direction method, *SIAM J. Numerical Analysis* 50(2012), 700-709.**

5.3 Convergence rate in pointwise iteration-complexity

In this subsection, we show that if the matrix G defined in (4.5b) is positive definite, a worst-case $O(1/t)$ convergence rate in a nonergodic sense can also be established for the prototype algorithm (4.4). Note in general a nonergodic convergence rate is stronger than the ergodic convergence rate.

We first need to prove the following lemma.

Lemma 5.1 *For the sequence generated by the prototype algorithm (4.4) where the Convergence Condition is satisfied, we have*

$$\begin{aligned} & (v^k-\tilde{v}^k)^T M^T H M\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\} \\ \ge \;\; & \frac{1}{2\alpha}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \end{aligned} \tag{5.15}$$

Proof. First, set $w=\tilde{w}^{k+1}$ in (4.4a), we have

$$\theta(\tilde{u}^{k+1})-\theta(\tilde{u}^k)+(\tilde{w}^{k+1}-\tilde{w}^k)^T F(\tilde{w}^k) \ge (\tilde{v}^{k+1}-\tilde{v}^k)^T Q(v^k-\tilde{v}^k). \tag{5.16}$$

Note that (4.4a) is also true for $k := k+1$ and thus we have

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

Set $w = \tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}). \tag{5.17}$$

Combining (5.16) and (5.17) and using the monotonicity of F, we get

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq 0. \tag{5.18}$$

Adding the term

$$\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}$$

to the both sides of (5.18), and using $v^T Q v = \frac{1}{2} v^T (Q^T + Q) v$, we obtain

$$(v^k-v^{k+1})^T Q\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\} \geq \frac{1}{2}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}.$$

Substituting $(v^k - v^{k+1}) = \alpha M(v^k - \tilde{v}^k)$ in the left-hand side of the last inequality and using $Q = HM$, we obtain (5.15) and the lemma is proved. $\square$

Now, we are ready to prove (5.19), the key inequality in this section.

Theorem 5.5 *For the sequence generated by the prototype algorithm (4.4) where the Convergence Condition is satisfied, we have*

$$\|M(v^{k+1} - \tilde{v}^{k+1})\|_H \leq \|M(v^k - \tilde{v}^k)\|_H, \quad \forall\, k > 0. \tag{5.19}$$

Proof. Setting $a = M(v^k - \tilde{v}^k)$ and $b = M(v^{k+1} - \tilde{v}^{k+1})$ in the identity

$$\|a\|^2_H - \|b\|^2_H = 2a^T H(a-b) - \|a-b\|^2_H,$$

we obtain

$$\begin{aligned}
&\|M(v^k - \tilde{v}^k)\|^2_H - \|M(v^{k+1} - \tilde{v}^{k+1})\|^2_H \\
&\quad = \quad 2(v^k - \tilde{v}^k)^T M^T H M[(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})] \\
&\qquad\qquad -\|M[(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})]\|^2_H.
\end{aligned}$$

Inserting (5.15) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned}
&\|M(v^k - \tilde{v}^k)\|_H^2 - \|M(v^{k+1} - \tilde{v}^{k+1})\|_H^2 \\
&\quad \geq \quad \frac{1}{\alpha}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2 - \|M[(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})]\|_H^2 \\
&\quad = \quad \frac{1}{\alpha}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2 \geq 0,
\end{aligned}$$

where the last inequality is because of the positive definiteness of the matrix $(Q^T + Q) - \alpha M^T H M \succeq 0$. The assertion (5.19) follows immediately. $\square$

Note that it follows from $G \succ 0$ and Theorem 5.2 there is a constant $c_0 > 0$ such that

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - c_0\|M(v^k - \tilde{v}^k)\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.20}$$

Now, with (5.20) and (5.19), we can establish the worst-case $O(1/t)$ convergence rate in a nonergodic sense for the prototype algorithm (4.4).

Theorem 5.6 *Let $\{v^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prototype algorithm (4.4) under the Convergence Condition. For any integer $t > 0$, we have*

$$\|M(v^t - \tilde{v}^t)\|_H^2 \leq \frac{1}{(t+1)c_0}\|v^0 - v^*\|_H^2. \tag{5.21}$$

Proof. First, it follows from (5.20) that

$$\sum_{k=0}^{\infty} c_0\|M(v^k - \tilde{v}^k)\|_H^2 \leq \|v^0 - v^*\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{5.22}$$

According to Theorem 5.5, the sequence $\{\|M(v^k - \tilde{v}^k)\|_H^2\}$ is monotonically non-increasing. Therefore, we have

$$(t+1)\|M(v^t - \tilde{v}^t)\|_H^2 \leq \sum_{k=0}^{t} \|M(v^k - \tilde{v}^k)\|_H^2. \tag{5.23}$$

The assertion (5.21) follows from (5.22) and (5.23) immediately. $\square$

Let $d := \inf\{\|v^0 - v^*\|_H \,|\, v^* \in \mathcal{V}^*\}$. Then, for any given $\epsilon > 0$, Theorem 5.6 shows that it needs at most $\lfloor d^2/c_0\epsilon \rfloor$ iterations to ensure that $\|M(v^k - \tilde{v}^k)\|_H^2 \leq \epsilon$. Recall that v^k is a solution of VI(Ω, F, θ) if $\|M(v^k - \tilde{v}^k)\|_H^2 = 0$ (see (4.4a) and due to $Q = HM$). A worst-case $O(1/t)$ convergence rate in pointwise iteration-complexity is thus established for the prototype algorithm (4.4).

Notice that, for a differentiable unconstrained convex optimization $\min f(x)$, it holds that

$$f(x) - f(x^*) = \nabla f(x^*)^T(x - x^*) + O(\|x - x^*\|^2) = O(\|x - x^*\|^2).$$

6 ADMM for problems with two separable blocks

This section concern the structured convex optimization problem (3.6) in Section 3.2, namely,

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

The augmented Lagrange Function of (3.6) is

$$\mathcal{L}_\beta^{[2]}(x,y,\lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax+By-b) + \frac{\beta}{2}\|Ax+By-b\|^2, \quad (6.1)$$

where $\beta > 0$ is a penalty coefficient. Using the augmented Lagrange function, the augmented Lagrangian method (3.10)-(3.11) for solving the problem (3.6) can be written as

$$\left\{ \begin{array}{rcl} (x^{k+1}, y^{k+1}) & = & \arg\min\big\{\mathcal{L}_\beta^{[2]}(x,y,\lambda^k) \bigm| x \in \mathcal{X}, y \in \mathcal{Y}\big\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array} \right. \quad (6.2)$$

The recursion of the alternating direction method of multipliers (3.12) for the structured

convex optimization (3.6) can be written as

$$\left\{ \begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array} \right. \quad (6.3)$$

Thus, ADMM can be viewed as a relaxed Augmented Lagrangian Method. The main advantage of ADMM is that one can solve the x and y-subproblem separately. Note that the essential variable of ADMM (6.3) is $v = (y, \lambda)$.

Since 1997, we focus our attention to ADMM, see [23]. Later, in 2002, we have ADMM paper published in Mathematical Programming [18].

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

♣ B. S. He, L. Z. Liao, D. Han, and H. Yang, A new inexact alternating directions method for monontone variational inequalities, *Mathematical Programming* **92**(2002), 103–118.

6.1 Classical ADMM in the Unified Framework

This subsection shows that the ADMM scheme (6.3) is also a special case of the prototype algorithm (4.4) and the Convergence Condition is satisfied. Recall the model (3.6) can be explained as the VI (4.1) with the specification given in (3.8b).

In order to cast the ADMM scheme (6.3) into a special case of (4.4), let us first define the artificial vector $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ by

$$\tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1} \quad \text{and} \quad \tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b), \tag{6.4}$$

where (x^{k+1}, y^{k+1}) is generated by the ADMM (6.3).

According to the scheme (6.3), the defined artificial vector $\tilde{w}^k$ satisfies the following VI:

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, & \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T(-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k)) \geq 0, & \forall\, y \in \mathcal{Y}, \\ \qquad (A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. & \end{cases}$$

This can be written in form of (4.4a) as described in the following lemma.

Lemma 6.1 *For given v^k, let w^{k+1} be generated by* (6.3) *and $\tilde{w}^k$ be defined by* (6.4).

Then, we have

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

where

$$Q = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix}. \tag{6.5}$$

Recall the essential variable of the ADMM scheme (6.3) is (y, λ). Moreover, using the definition of $\tilde{w}^k$, the λ^{k+1} updated by (6.3) can be represented as

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b) \\ &= \lambda^k - \left[-\beta B(y^k - \tilde{y}^k) + \beta(A\tilde{x}^k + By^k - b)\right] \\ &= \lambda^k - \left[-\beta B(y^k - \tilde{y}^k) + (\lambda^k - \tilde{\lambda}^k)\right]. \end{aligned}$$

Therefore, the ADMM scheme (6.3) can be written as

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{6.6a}$$

which corresponds to the step (4.4b) with

$$M = \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \quad \text{and} \quad \alpha = 1. \tag{6.6b}$$

Now we check that the Convergence Condition is satisfied by the ADMM scheme (6.3). Indeed, for the matrix M in (6.6b), we have

$$M^{-1} = \begin{pmatrix} I & 0 \\ \beta B & I \end{pmatrix}.$$

Thus, by using (6.5) and (6.6b), we obtain

$$H = QM^{-1} = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ \beta B & I \end{pmatrix} = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix},$$

and consequently

$$\begin{aligned} G &= Q^T + Q - \alpha M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} \beta B^T B & -B^T \\ 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \\ &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \end{aligned} \tag{6.7}$$

Therefore, H is symmetric and positive definite under the assumption that B is full column rank; and G is positive semi-definite. The Convergence Condition is satisfied; and thus the convergence of the ADMM scheme (6.3) is guaranteed.

Note that Theorem 5.4 is true for $G \succeq 0$. Thus the classical ADMM (6.3) has $O(1/t)$ convergence rate in the ergodic sense.

Since $\alpha = 1$, according to (5.5) and the form of G in (6.7), we have

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \tfrac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2, \quad \forall v^* \in \mathcal{V}^*. \tag{6.8}$$

Lemma 6.2 *For given v^k, let w^{k+1} be generated by* (6.3) *and $\tilde{w}^k$ be defined by* (6.4)*. Then, we have*

$$\frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 \geq \|v^k - v^{k+1}\|_H^2. \tag{6.9}$$

Proof. According to (6.3) and (6.4), the optimal condition of the y-subproblem is

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

Because

$$\lambda^{k+1} = \tilde{\lambda}^k - \beta B(\tilde{y}^k - y^k) \quad \text{and} \quad \tilde{y}^k = y^{k+1},$$

it can be written as

$$y^{k+1} \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \geq 0, \; \forall\, y \in \mathcal{Y}. \tag{6.10}$$

The above inequality is hold also for the last iteration, *i. e.*, we have

$$y^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(y^k) + (y - y^k)^T\{-B^T\lambda^k\} \geq 0, \; \forall\, y \in \mathcal{Y}. \tag{6.11}$$

Setting $y = y^k$ in (6.10) and $y = y^{k+1}$ in (6.11), and then adding them, we get

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0. \tag{6.12}$$

Using $\lambda^k - \tilde{\lambda}^k = (\lambda^k - \lambda^{k+1}) + \beta B(y^k - y^{k+1})$ and the inequality (6.12), we obtain

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

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

Substituting (6.9) in (6.8), we get the following nice property of the classical ADMM.

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

which is the same as (3.27) in Section 3.2.

Notice that the sequence $\{\|v^k - v^{k+1}\|_H^2\}$ generated by the classical ADMM is monotone non-increasing [27]. In fact, in Theorem 5.5, we have proved that (see (5.19))

$$\|M(v^k - \tilde{v}^k)\|_H \leq \|M(v^{k-1} - \tilde{v}^{k-1})\|_H, \quad \forall\, k \geq 1. \tag{6.13}$$

Because (see the correction formula (6.6))

$$v^k - v^{k+1} = M(v^k - \tilde{v}^k),$$

it follows from (6.13) that

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

On the other hand, the inequality (3.27) tell us that

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

Thus, we have

$$\begin{aligned} \|v^t - v^{t+1}\|_H^2 &\le \frac{1}{t+1} \sum_{k=0}^{t} \|v^k - v^{k+1}\|_H^2 \\ &\le \frac{1}{t+1} \sum_{k=0}^{\infty} \|v^k - v^{k+1}\|_H^2 \;\le\; \frac{1}{t+1} \|v^0 - v^*\|_H^2. \end{aligned}$$

Therefore, ADMM (6.3) has $O(1/t)$ convergence rate in pointwise iteration-complexity.

6.2 ADMM in Sense of Customized PPA [3]

If we change the performance order of y and λ of the classical ADMM (6.3), it becomes

$$\left\{ \begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^{k+1}) \,|\, y \in \mathcal{Y}\}. \end{array} \right. \tag{6.14}$$

In this way we can get a positive semidefinite matrix Q in (4.4a). We define

$$\tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1}, \quad \tilde{\lambda}^k = \lambda^{k+1}, \tag{6.15}$$

where $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is the output of (9.3) and thus it can be rewritten as

$$\left\{ \begin{array}{l} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(\tilde{x}^k, y, \tilde{\lambda}^k) \,|\, y \in \mathcal{Y}\}. \end{array} \right. \tag{6.16}$$

Because $\tilde{\lambda}^k = \lambda^{k+1} = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, the optimal condition of the x-subproblem of (6.16) is

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \ge 0, \quad \forall\, x \in \mathcal{X}. \tag{6.17}$$

Notice that

$$\mathcal{L}_\beta^{[2]}(\tilde{x}^k, y, \tilde{\lambda}^k) = \theta_1(\tilde{x}^k) + \theta_2(y) - (\tilde{\lambda}^k)^T(A\tilde{x}^k + By - b) + \tfrac{\beta}{2}\|A\tilde{x}^k + By - b\|^2,$$

ignoring the constant term in the y optimization subproblem of (6.16), it turns to

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) - (\tilde{\lambda}^k)^T By + \tfrac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \mid y \in \mathcal{Y}\},$$

and consequently, the optimal condition is $\tilde{y}^k \in \mathcal{Y}$,

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\big[-B^T\tilde{\lambda}^k + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b)\big] \ge 0, \ \forall y \in \mathcal{Y}.$$

For the term $[\,\cdot\,]$ in the last inequality, using $\beta(A\tilde{x}^k + By^k - b) = -(\tilde{\lambda}^k - \lambda^k)$, we have

$$\begin{aligned} & -B^T\tilde{\lambda}^k + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b) \\ = \;& -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) + \beta B^T(A\tilde{x}^k + By^k - b) \\ = \;& -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k). \end{aligned}$$

Finally, the optimal condition of the y-subproblem can be written as $\tilde{y}^k \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\big[-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k)\big] \ge 0, \ \forall y \in \mathcal{Y}. \tag{6.18}$$

From the λ update form in (6.16) we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \tag{6.19}$$

Combining (6.17), (6.18) and (6.19), and using the notations of (3.8)-(3.8b), we get following lemma.

Lemma 6.3 *For given v^k, let $\tilde{w}^k$ be generated by* (6.16)*. Then, we have*

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

where

$$Q = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{6.20}$$

Because Q is symmetric and positive semidefinite, according to (4.4), we can take

$$M = I \qquad \alpha \in (0,2) \qquad \text{and thus} \qquad H = Q.$$

In this way, we get the new iterate by

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

The generated sequence $\{v^k\}$ has the convergence property

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

Ensure the matrix H to be positive definite If we add an additional proximal term $\frac{\delta\beta}{2}\|B(y-y^k)\|^2$ to the y-subproblem of (6.16) with any small $\delta > 0$, it becomes

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta^{(2)}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta^{(2)}(\tilde{x}^k, y, \tilde{\lambda}^k) + \frac{\delta\beta}{2}\|B(y-y^k)\|^2 \,|\, y \in \mathcal{Y}\}. \end{cases} \tag{6.21}$$

In the ADMM based customized PPA (6.16), the y-subproblem can be written as

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \frac{\beta}{2}\|By - p^k\|^2 \,|\, y \in \mathcal{Y}\}, \tag{6.22}$$

where

$$p^k = b + \frac{1}{\beta}\tilde{\lambda}^k - A\tilde{x}^k.$$

If we add an additional term $\frac{\delta\beta}{2}\|B(y-y^k)\|^2$ (with any small $\delta > 0$) to the objective function of the y-subproblem, we will get $\tilde{y}^k$ via

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \frac{\beta}{2}\|By - p^k\|^2 + \frac{\delta\beta}{2}\|B(y-y^k)\|^2 \,|\, y \in \mathcal{Y}\}.$$

By a manipulation, the solution point of the above subproblem is obtained via

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \frac{(1+\delta)\beta}{2}\|By - q^k\|^2 \,|\, y \in \mathcal{Y}\}, \tag{6.23}$$

where

$$q^k = \frac{1}{1+\delta}(p^k + \delta By^k).$$

In this way, the matrix Q in (6.20) will turn to

$$Q = \begin{pmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix}.$$

Take $H = Q$, for any $\delta > 0$, H is positive definite when B is a full rank matrix. In other words, instead of (6.22), using (6.23) to get $\tilde{y}^k$, it will ensure the positivity of H theoretically. However, in practical computation, it works still well by using $\delta = 0$.

ADMM in sense of customized PPA

1. Produce a predictor $\tilde{w}^k$ via (6.21) with given $v^k = (y^k, \lambda^k)$,
2. Update the new iterate by $v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k)$, $\alpha = 1.5 \in (0, 2)$.

Theorem 6.1 *The sequence* $\{v^k\}$*generated by the ADMM in Sense of PPA satisfies*

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

where

$$H = \begin{pmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}.$$

Since the correction formula is $v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k)$, the contraction inequality can be written as

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

Notice that the sequence $\{\|v^k - v^{k+1}\|_H^2\}$ generated by the ADMM in sense of PPA is also monotone non-increasing. Again, because (see (5.19))

$$\|M(v^k - \tilde{v}^k)\|_H \le \|M(v^{k-1} - \tilde{v}^{k-1})\|_H, \quad \forall\, k \ge 1. \tag{6.24}$$

it follows from (6.24) and the correction formula that

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

Thus, we have

$$\begin{aligned} \|v^t - v^{t+1}\|_H^2 &\le \frac{1}{t+1}\sum_{k=0}^{\infty}\|v^k - v^{k+1}\|_H^2 \\ &\le \frac{1}{t+1}\frac{\alpha}{2-\alpha}\|v^0 - v^*\|_H^2. \end{aligned}$$

Therefore, ADMM (in Sense of Customized PPA) has $O(1/t)$ convergence rate in pointwise iteration-complexity.

6.3 Symmetric ADMM [19]

In the problem (3.6) in Section 3.2, x and y are a pair of fair variables. It is nature to consider a symmetric method: Update the Lagrangian Multiplier after solving each x and y-subproblem. .

We take $\mu \in (0,1)$ (usually $\mu = 0.9$), the method is described as

$$\text{(S-ADMM)} \quad \left\{ \begin{array}{ll} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, & \text{(6.25a)} \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), & \text{(6.25b)} \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, & \text{(6.25c)} \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b). & \text{(6.25d)} \end{array} \right.$$

This method is called **Alternating direction method of multipliers with symmetric multipliers updating**, or **Symmetric Alternating Direction Method of Multipliers**.

The convergence of the proposed method is established via the unified framework.

For establishing the main result, we introduce an artificial vector $\tilde{w}^k$ by

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}, \tag{6.26}$$

where (x^{k+1}, y^{k+1}) is generated by the ADMM (6.25). First, by using (6.26), we interpret (6.25a)-(6.25c) as a prediction which only involves the variables w^k and $\tilde{w}^k$.

According to (6.26), by using $\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b)$, the optimal condition of the x-subproblem (6.25a) is

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{6.27}$$

Notice that the objective function of the y-subproblem (6.25c) is

$$\begin{array}{l} \mathcal{L}_\beta^{[2]}(\tilde{x}^k, y, \lambda^{k+\frac{1}{2}}) \\ \quad = \quad \theta_1(\tilde{x}^k) + \theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(A\tilde{x}^k + By - b) + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2. \end{array}$$

Ignoring the constant term in the y-subproblem, it turns to

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \,|\, y \in \mathcal{Y}\}.$$

Consequently, according to Lemma 1.1, we have

$$\begin{array}{l}\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) \\ \qquad +(y-\tilde{y}^k)^T\{-B^T\lambda^{k+\frac{1}{2}} + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b)\} \geq 0, \ \forall\, y \in \mathcal{Y}.\end{array}$$

Using $\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b)$, we get

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \mu(\lambda^k - \tilde{\lambda}^k) = \tilde{\lambda}^k + (1-\mu)(\lambda^k - \tilde{\lambda}^k),$$

and

$$\beta(A\tilde{x}^k + By^k - b) = (\tilde{\lambda}^k - \lambda^k).$$

Thus,

$$\begin{array}{l}-B^T\lambda^{k+\frac{1}{2}} + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b) \\ \begin{array}{rl} = & -B^T[\tilde{\lambda}^k + (1-\mu)(\lambda^k - \tilde{\lambda}^k)] + \beta B^T B(\tilde{y}^k - y^k) \\ & +\beta B^T(A\tilde{x}^k + By^k - b) \\ = & -B^T\tilde{\lambda}^k - (1-\mu)B^T(\lambda^k - \tilde{\lambda}^k) + \beta B^T B(\tilde{y}^k - y^k) \\ & +B^T(\lambda^k - \tilde{\lambda}^k) \\ = & -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - \mu B^T(\tilde{\lambda}^k - \lambda^k).\end{array}\end{array}$$

Finally, the optimal condition of the y-subproblem can be written as $\tilde{y}^k \in \mathcal{Y}$, and

$$\begin{array}{r}\theta_2(y) - \theta_2(\tilde{y}^k) \ + \ (y-\tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) \\ -\mu B^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \quad \forall y \in \mathcal{Y}.\end{array} \tag{6.28}$$

According to the definition of $\tilde{w}^k$ in (6.26), $\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b)$, we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \tag{6.29}$$

Combining (6.27), (6.28) and (6.29), and using the notations of (3.8), we get following lemma.

Lemma 6.4 *For given v^k, let w^{k+1} be generated by* (6.25) *and $\tilde{w}^k$ be defined by* (6.26). *Then, we have*

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

where

$$Q = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{6.30}$$

We have finished to interpret (6.25a)-(6.25c) as a prediction. By a manipulation, the new iterate λ^{k+1} in (6.25d) can be represented as

$$\begin{aligned}\lambda^{k+1} &= [\lambda^k - \mu(\lambda^k - \tilde{\lambda}^k)] - \mu\big[-\beta B(y^k - \tilde{y}^k) + \beta(Ax^{k+1} + By^k - b)\big] \\ &= \lambda^k - \big[-\mu\beta B(y^k - \tilde{y}^k) + 2\mu(\lambda^k - \tilde{\lambda}^k)\big]. \end{aligned} \quad (6.31)$$

Thus, together with $y^{k+1} = \tilde{y}^k$, the correction step can be represented as

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

This can be rewritten into a compact form:

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \quad (6.32a)$$

with

$$M = \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix}. \quad (6.32b)$$

These relationships greatly simplify our analysis and presentation.

In order to use the unified framework, we only need to verify the positiveness of H and G. For the matrix M given by (6.32b), we have

$$M^{-1} = \begin{pmatrix} I & 0 \\ \frac{1}{2}\beta B & \frac{1}{2\mu} I_m \end{pmatrix}.$$

For $H = QM^{-1}$, it follows that

$$H = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} I & 0 \\ \frac{1}{2}\beta B & \frac{1}{2\mu} I_m \end{pmatrix} = \begin{pmatrix} (1 - \frac{1}{2}\mu)\beta B^T B & -\frac{1}{2} B^T \\ -\frac{1}{2} B & \frac{1}{2\mu\beta} I_m \end{pmatrix}.$$

Thus

$$H = \frac{1}{2} \begin{pmatrix} \sqrt{\beta} B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix} \begin{pmatrix} (2-\mu) I & -I \\ -I & \frac{1}{\mu} I \end{pmatrix} \begin{pmatrix} \sqrt{\beta} B & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix}.$$

Notice that

$$\begin{pmatrix} (2-\mu) & -1 \\ -1 & \frac{1}{\mu} \end{pmatrix} = \begin{cases} \succ 0, & \mu \in (0,1); \\ \succeq 0, & \mu = 1. \end{cases}$$

Therefore, H is positive definite for any $\mu \in (0,1)$ when B is a full column rank matrix.

It remains to check the positiveness of $G = Q^T + Q - M^T HM$. Note that

$$\begin{aligned} M^T HM &= M^T Q = \begin{pmatrix} I & -\mu\beta B^T \\ 0 & 2\mu I_m \end{pmatrix} \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} (1+\mu)\beta B^T B & -2\mu B^T \\ -2\mu B & \frac{2\mu}{\beta} I_m \end{pmatrix}. \end{aligned}$$

Using (6.30) and the above equation, we have

$$G = (Q^T + Q) - M^T HM = (1-\mu) \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix}.$$

Thus

$$G = (1-\mu) \begin{pmatrix} \sqrt{\beta} B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix} \begin{pmatrix} I & -I \\ -I & 2I \end{pmatrix} \begin{pmatrix} \sqrt{\beta} B & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix}.$$

Because the matrix

$$\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

is positive definite, for any $\mu \in (0,1)$, G is essentially positive definite (positive definite when B is a full column rank matrix). The convergence conditions (4.5) are satisfied.

Take $\mu = 0.9$, it will accelerate the convergence much. For the numerical experiments of this method, it is refereed to consult [19].

The symmetric ADMM is a special version of the unified framework (4.4) - (4.5) whose $\alpha = 1$,

$$H = \begin{pmatrix} (1-\frac{1}{2}\mu)\beta B^T B & -\frac{1}{2} B^T \\ -\frac{1}{2} B & \frac{1}{2\mu\beta} I_m \end{pmatrix} \quad \text{and} \quad G = (1-\mu) \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix}.$$

Both the matrices H and G are positive definite for $\mu \in (0,1)$. According to Theorem 5.2, we have

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

7 Splitting Methods for m-block Problems

We consider the linearly constrained convex optimization with m separable operators

$$\min\Big\{\sum_{i=1}^m \theta_i(x_i) \,\big|\, \sum_{i=1}^m A_i x_i = b,\ x_i \in \mathcal{X}_i\Big\}. \tag{7.1}$$

Its Lagrange function is

$$L(x_1,\dots,x_m,\lambda) = \sum_{i=1}^m \theta_i(x_i) - \lambda^T\big(\sum_{i=1}^m A_i x_i - b\big), \tag{7.2}$$

which defined on $\Omega := \prod_{i=1}^m \mathcal{X}_i \times \Re^m$. The related VI$(\Omega, F, \theta)$ has the form

$$\text{VI}(\Omega, F, \theta) \quad w^* \in \Omega,\ \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0,\ \forall\, w \in \Omega, \tag{7.3a}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_m \\ \lambda \end{pmatrix}, \quad \theta(x) = \sum_{i=1}^m \theta_i(x_i), \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_m^T\lambda \\ \sum_{i=1}^m A_i x_i - b \end{pmatrix}. \tag{7.3b}$$

$$\mathcal{L}_\beta(x_1,\dots,x_m,\lambda) = L(x_1,\dots,x_m,\lambda) + \frac{\beta}{2}\|\textstyle\sum_{i=1}^m A_i x_i - b\|^2 \tag{7.4}$$

is the augmented Lagrangian function.

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

$$\left\{\begin{array}{l} x_1^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k,\dots,x_m^k,\lambda^k) \mid x_1 \in \mathcal{X}_1\}, \\ x_2^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1^{k+1}, x_2, x_3^k,\dots,x_m^k,\lambda^k) \mid x_2 \in \mathcal{X}_2\}, \\ \qquad\vdots \\ x_i^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1^{k+1},\dots,x_{i-1}^{k+1}, x_i, x_{i+1}^k,\dots,x_m^k,\lambda^k) \mid x_i \in \mathcal{X}_i\}, \\ \qquad\vdots \\ x_m^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1^{k+1},\dots,x_{m-1}^{k+1}, x_m,\lambda^k) \mid x_m \in \mathcal{X}_m\}, \\ \lambda^{k+1} = \lambda^k - \beta\big(\sum_{i=1}^m A_i x_i^{k+1} - b\big). \end{array}\right. \tag{7.5}$$

There is counter example [6], it is not necessary convergent for the problem with $m \ge 3$.

7.1 ADMM with Gaussian Back Substitution [20]

Let $(x_1^{k+1}, x_2^{k+1}, \ldots, x_m^{k+1}, \lambda^{k+1})$ be the output of (7.5). By denoting

$$\tilde{x}_i^k = x_i^{k+1}, \quad i = 1, \ldots, m \tag{7.6}$$

the x_i-subproblems of (7.5) can be written as

$$\left\{\begin{array}{l} \tilde{x}_1^k = \arg\min\big\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \ldots, x_m^k, \lambda^k) \mid x_1 \in \mathcal{X}_1\big\}; \\ \tilde{x}_2^k = \arg\min\big\{\mathcal{L}_\beta(\tilde{x}_1^k, x_2, x_3^k, \ldots, x_m^k, \lambda^k) \mid x_2 \in \mathcal{X}_2\big\}; \\ \qquad\quad \vdots \\ \tilde{x}_i^k = \arg\min\big\{\mathcal{L}_\beta(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k) \mid x_i \in \mathcal{X}_i\big\}; \\ \qquad\quad \vdots \\ \tilde{x}_m^k = \arg\min\big\{\mathcal{L}_\beta(\tilde{x}_1^k, \ldots, \tilde{x}_{m-1}^k, x_m, \lambda^k) \mid x_m \in \mathcal{X}_m\big\}. \end{array}\right. \tag{7.7}$$

Additionally, we define

$$\tilde{\lambda}^k = \lambda^k - \beta\big(A_1\tilde{x}_1^k + \sum_{j=2}^m A_j x_j^k - b\big). \tag{7.8}$$

Using the notation of the augmented Lagrangian function (see (7.4)), the optimal condition of the x_1-subproblem of (7.7) can be written as

$$\tilde{x}_1^k \in \mathcal{X}_1,\ \theta_1(x_1) - \theta_1(\tilde{x}_1^k) + (x_1 - \tilde{x}_1^k)^T\{-A_1^T\lambda^k + \beta A_1^T(A_1\tilde{x}_1^k + \sum_{j=2}^m A_j x_j^k - b)\} \ge 0.$$

According to the definition of $\tilde{\lambda}^k$ (see (7.8)), it follows from the last inequality

$$\tilde{x}_1^k \in \mathcal{X}_1,\ \theta_1(x_1) - \theta_1(\tilde{x}_1^k) + (x_1 - \tilde{x}_1^k)^T\{-A_1^T\tilde{\lambda}^k\} \ge 0,\ \forall\, x_1 \in \mathcal{X}_1. \tag{7.9}$$

For $i = 2, \ldots, m$, the optimal condition of the x_i-subproblem of (7.7) is

$$\begin{array}{l} \tilde{x}_i^k \in \mathcal{X}_i, \quad \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{-A_i^T\lambda^k \\ \qquad\quad + \beta A_i^T\big[A_1\tilde{x}_1^k + \sum_{j=2}^i A_j\tilde{x}_j^k + \sum_{j=i+1}^m A_j x_j^k - b\big]\} \ge 0,\ \forall x_i \in \mathcal{X}_i. \end{array}$$

Consequently, by using the definition of $\tilde{\lambda}^k$, we have $\tilde{x}_i^k \in \mathcal{X}_i$ and

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{-A_i^T\tilde{\lambda}^k + \beta A_i^T\big[\sum_{j=2}^i A_j(\tilde{x}_j^k - x_j^k)\big]\} \ge 0, \tag{7.10}$$

for all $x_i \in \mathcal{X}_i$. In addition, (7.8) can be written as

$$(\sum_{j=1}^m A_j\tilde{x}_j^k - b) - \sum_{j=2}^m A_j(\tilde{x}_j^k - x_j^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \tag{7.11}$$

Combining (7.9), (7.10) and (7.11) together and using the notations in (7.3), we obtain

Lemma 7.1 *Let* $\tilde{w}^k$ *be generated by* (7.7)-(7.8) *from the given vector* v^k*. Then, we have* $\tilde{w}^k \in \Omega$ *and*

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall\, w \in \Omega, \quad (7.12)$$

where

$$Q = \begin{pmatrix} \beta A_2^T A_2 & 0 & \cdots & \cdots & 0 \\ \beta A_3^T A_2 & \beta A_3^T A_3 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \beta A_m^T A_2 & \beta A_m^T A_3 & \cdots & \beta A_m^T A_m & 0 \\ -A_2 & -A_3 & \cdots & -A_m & \frac{1}{\beta} I_m \end{pmatrix}. \quad (7.13)$$

After having (7.12), we have finished the prediction step (4.4a). The rest task is to complete the correction step (4.4b), finding a matrix M and a constant $\alpha > 0$, which satisfy the convergence conditions (4.5). In the following we give some examples.

The first choice of Matrix M In the first choice, we take

$$M = Q^{-T} D, \quad (7.14)$$

where

$$D = \text{diag}(\beta A_2^T A_2, \beta A_3^T A_3, \ldots, \beta A_m^T A_m, \frac{1}{\beta} I).$$

By using the notation of D, we have

$$Q^T + Q = D + P^T P$$

where

$$P = (\sqrt{\beta} A_2, \sqrt{\beta} A_3, \ldots, \sqrt{\beta} A_m, \sqrt{\frac{1}{\beta}} I). \quad (7.15)$$

For the matrix H, according to the definition (4.5a), we have

$$H = Q M^{-1} = Q D^{-1} Q^T$$

and thus H is symmetric and positive definite. Because

$$M^T HM = Q^T M = D,$$

it follows that

$$G = Q^T + Q - \alpha M^T HM = (1-\alpha)D + P^T P.$$

For any $\alpha \in (0,1)$, G is positive definite.

How to implement the correction step ?

Because $M = Q^{-T}D$ and the correction is $v^{k+1} = v^k - \alpha Q^{-T}D(v^k - \tilde{v}^k)$, we have

$$Q^T(v^{k+1} - v^k) = \alpha D(\tilde{v}^k - v^k). \tag{7.16}$$

According to the matrix P in (7.15), we define

$$\text{diag}(P) = \begin{pmatrix} \sqrt{\beta}A_2 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta}A_3 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta}A_m & 0 \\ 0 & \cdots & \cdots & 0 & \frac{1}{\sqrt{\beta}}I_m \end{pmatrix}.$$

In addition, we denote

$$L = \begin{pmatrix} I_m & 0 & \cdots & \cdots & 0 \\ I_m & I_m & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & I_m & 0 \\ -I_m & \cdots & \cdots & -I_m & I_m \end{pmatrix}. \tag{7.17}$$

Using these definitions, we have

$$Q = \big[\text{diag}(P)\big]^T L \big[\text{diag}(P)\big], \qquad D = \big[\text{diag}(P)\big]^T \big[\text{diag}(P)\big].$$

According to (7.16), we need only to solve

$$L^T \big[\text{diag}(P)\big](v^{k+1} - v^k) = \alpha \big[\text{diag}(P)\big](\tilde{v}^k - v^k). \tag{7.18}$$

In order to start the next iteration, we only need $\big[\text{diag}(P)\big]v^{k+1}$, which is easy to be obtained by a Gaussian substitution form (7.18). This kind of method is proposed in [20].

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

Using the uniform framework, the convergence proof is much simple !

The second choice of Matrix M For the second choice, we decompose Q in form

$$Q = \begin{pmatrix} \beta Q_0 & 0 \\ -\mathcal{A} & \frac{1}{\beta} I \end{pmatrix}.$$

Thus, in comparison with the matrix Q in (7.13), we have

$$Q_0 = \begin{pmatrix} A_2^T A_2 & 0 & \cdots & 0 \\ A_3^T A_2 & A_3^T A_3 & \ddots & \vdots \\ \vdots & & \ddots & \vdots \\ A_m^T A_2 & A_m^T A_3 & \cdots & A_m^T A_m \end{pmatrix}$$

and

$$\mathcal{A} = (A_2, A_3, \ldots, A_m).$$

In addition, we denote

$$D_0 = \text{diag}(A_2^T A_2, A_3^T A_3, \ldots, A_m^T A_m).$$

Thus, by using the notation of D_0, we have

$$Q_0^T + Q_0 = D_0 + \mathcal{A}^T \mathcal{A}$$

and

$$Q^T + Q = \begin{pmatrix} \beta(D_0 + \mathcal{A}^T\mathcal{A}) & -\mathcal{A}^T \\ -\mathcal{A} & \frac{2}{\beta}I \end{pmatrix}.$$

We take the matrix M in the correction step (4.4b) by a ν-dependent matrix

$$M_\nu = \begin{pmatrix} \nu Q_0^{-T} D_0 & 0 \\ -\beta\mathcal{A} & I \end{pmatrix} \ \Big(\text{thus } \ M_\nu^{-1} = \begin{pmatrix} \frac{1}{\nu} D_0^{-1} Q_0^T & 0 \\ \frac{1}{\nu}\beta\mathcal{A}D_0^{-1}Q_0^T & I \end{pmatrix}\Big) \tag{7.19}$$

and set $\alpha = 1$. In other words, the new iterate v^{k+1} is given by

$$v^{k+1} = v^k - M_\nu(v^k - \tilde{v}^k). \tag{7.20}$$

Now, we check if the convergence conditions (4.5) are satisfied. First,

$$\begin{aligned} H &= QM_\nu^{-1} = \begin{pmatrix} \beta Q_0 & 0 \\ -\mathcal{A} & \frac{1}{\beta}I \end{pmatrix} \begin{pmatrix} \frac{1}{\nu} D_0^{-1} Q_0^T & 0 \\ \frac{1}{\nu}\beta\mathcal{A}D_0^{-1}Q_0^T & I \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\nu}\beta Q_0 D_0^{-1} Q_0^T & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix}, \end{aligned}$$

is symmetric and positive definite and the condition (4.5a) is satisfied. Because

$$\begin{aligned} M_\nu^T H M_\nu &= Q^T M_\nu = \begin{pmatrix} \beta Q_0^T & -\mathcal{A}^T \\ 0 & \frac{1}{\beta}I \end{pmatrix} \begin{pmatrix} \nu Q_0^{-T} D_0 & 0 \\ -\beta\mathcal{A} & I \end{pmatrix} \\ &= \begin{pmatrix} \beta(\nu D_0 + \mathcal{A}^T\mathcal{A}) & -\mathcal{A}^T \\ -\mathcal{A} & \frac{1}{\beta}I \end{pmatrix}, \end{aligned}$$

it follows that

$$\begin{aligned} G &= Q^T + Q - M_\nu^T H M_\nu \\ &= \begin{pmatrix} \beta(D_0 + \mathcal{A}^T\mathcal{A}) & -\mathcal{A}^T \\ -\mathcal{A} & \frac{2}{\beta}I \end{pmatrix} - \begin{pmatrix} \nu\beta D_0 + \beta\mathcal{A}^T\mathcal{A} & -\mathcal{A}^T \\ -\mathcal{A} & \frac{1}{\beta}I \end{pmatrix} \\ &= \begin{pmatrix} (1-\nu)\beta D_0 & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix}. \end{aligned} \tag{7.21}$$

For any $\nu \in (0, 1)$ (resp. $\nu = 1$), G is positive definite (resp. positive semi-definite).

We call this method Alternating direction method with Gaussian back substitution, because

- The predictor $\tilde{w}^k$ is obtained via (7.7)-(7.8), in an alternating direction manner;
- In the correction step (7.20),

$$v^{k+1} = v^k - M_\nu(v^k - \tilde{v}^k).$$

Since (see (7.19))

$$M_\nu = \begin{pmatrix} \nu Q_0^{-T} D_0 & 0 \\ -\beta \mathcal{A} & I \end{pmatrix} \quad \text{and} \quad \mathcal{A} = (A_2, A_3, \ldots, A_m),$$

it follows from (7.8) that

$$\lambda^{k+1} = \lambda^k - \beta\big(\textstyle\sum_{j=1}^m A_j \tilde{x}_j^k - b\big).$$

The x-part of the new iterate v^{k+1} is obtained by

$$Q_0^T \begin{pmatrix} x_2^{k+1} - x_2^k \\ x_3^{k+1} - x_3^k \\ \vdots \\ x_m^{k+1} - x_m^k \end{pmatrix} = \nu D_0 \begin{pmatrix} \tilde{x}_2^k - x_2^k \\ \tilde{x}_3^k - x_3^k \\ \vdots \\ \tilde{x}_m^k - x_m^k \end{pmatrix}. \tag{7.22}$$

- Q_0^T is an upper-triangular matrix, it can be viewed as a Gaussian back substitution.

In practice, to begin the k-th iteration, we need to have $(A_2 x_2^k, A_3 x_3^k, \ldots, A_m x_m^k, \lambda^k)$ (see (7.4) and (7.7)). Thus, in order to begin the next iteration, we need only to get $(A_2 x_2^{k+1}, A_3 x_3^{k+1}, \ldots, A_m x_m^{k+1}, \lambda^{k+1})$ from (7.22). Because

$$Q_0^T \begin{pmatrix} x_2^{k+1} - x_2^k \\ x_3^{k+1} - x_3^k \\ \vdots \\ x_m^{k+1} - x_m^k \end{pmatrix} = \begin{pmatrix} A_2^T & 0 & \cdots & 0 \\ 0 & A_3^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_m^T \end{pmatrix} \begin{pmatrix} I & I & \cdots & I \\ 0 & I & \cdots & I \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & I \end{pmatrix} \begin{pmatrix} A_2(x_2^{k+1} - x_2^k) \\ A_3(x_3^{k+1} - x_3^k) \\ \vdots \\ A_m(x_m^{k+1} - x_m^k) \end{pmatrix}$$

and

$$D_0 \begin{pmatrix} \tilde{x}_2^k - x_2^k \\ \tilde{x}_3^k - x_3^k \\ \vdots \\ \tilde{x}_m^k - x_m^k \end{pmatrix} = \begin{pmatrix} A_2^T & 0 & \cdots & 0 \\ 0 & A_3^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_m^T \end{pmatrix} \begin{pmatrix} A_2(\tilde{x}_2^k - x_2^k) \\ A_3(\tilde{x}_3^k - x_3^k) \\ \vdots \\ A_m(\tilde{x}_m^k - x_m^k) \end{pmatrix},$$

we can get $(A_2 x_2^{k+1}, A_3 x_3^{k+1}, \ldots, A_m x_m^{k+1})$ which satisfies (7.22) via solving the following system equations:

$$\begin{pmatrix} I & I & \cdots & I \\ 0 & I & \cdots & I \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & I \end{pmatrix} \begin{pmatrix} A_2(x_2^{k+1} - x_2^k) \\ A_3(x_3^{k+1} - x_3^k) \\ \vdots \\ A_m(x_m^{k+1} - x_m^k) \end{pmatrix} = \nu \begin{pmatrix} A_2(\tilde{x}_2^k - x_2^k) \\ A_3(\tilde{x}_3^k - x_3^k) \\ \vdots \\ A_m(\tilde{x}_m^k - x_m^k) \end{pmatrix}. \quad (7.23)$$

Indeed, $(A_2x_2^{k+1}, A_3x_3^{k+1}, \ldots, A_mx_m^{k+1})$ from (7.23) satisfies the systems of equations (7.22). The solution of (7.23) can be obtained via the following update form:

$$\begin{pmatrix} A_2x_2^{k+1} \\ A_3x_3^{k+1} \\ \vdots \\ A_mx_m^{k+1} \end{pmatrix} = \begin{pmatrix} A_2x_2^k \\ A_3x_3^k \\ \vdots \\ A_mx_m^k \end{pmatrix} - \nu \begin{pmatrix} I & -I & & \\ & \ddots & \ddots & \\ & & \ddots & -I \\ & & & I \end{pmatrix} \begin{pmatrix} A_2(x_2^k - \tilde{x}_2^k) \\ A_3(x_3^k - \tilde{x}_3^k) \\ \vdots \\ A_m(x_m^k - \tilde{x}_m^k) \end{pmatrix}.$$

ADMM with Gaussian back substitution

1. Produce a predictor $\tilde{w}^k$ via (7.7)-(7.8) with given $v^k = (x_2^k, \ldots, x_m^k, \lambda^k)$,
2. Update the new iterate by $v^{k+1} = v^k - M_\nu(v^k - \tilde{v}^k)$ (see (7.20))

Theorem 7.1 *The sequence $\{v^k\}$generated by the ADMM (with Gaussian back substitution) satisfies*

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

where

$$H = \begin{pmatrix} \frac{1}{\nu}\beta Q_0 D_0^{-1} Q_0^T & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} (1-\nu)\beta D_0 & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}.$$

Implementation of the above method for three block problems

Let us see how to implement the methods for the problem with three separable operators

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) | Ax + By + Cz = b, \; x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}. \quad (7.24)$$

Notice that its Lagrange function is

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

which defined on $\Omega := \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m$. The variational inequality $\mathsf{VI}(\Omega, F, \theta)$ is:

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

where

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

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

The related augmented Lagrangian function is defined by

$$\mathcal{L}_\beta^{[3]}(x, y, z, \lambda) = L^{[3]}(x, y, z, \lambda) + \frac{\beta}{2}\|Ax + By + Cz - b\|^2. \tag{7.25}$$

Note that the essential variable is $v = (y, z, \lambda)$, and the prediction (7.7)-(7.8) becomes

$$\left\{ \begin{array}{ll} \tilde{x}^k = \arg\min\big\{\mathcal{L}_\beta^{[3]}(x, y^k, z^k, \lambda^k) \;\big|\; x \in \mathcal{X}\big\}, & \text{(7.26a)} \\ \tilde{y}^k = \arg\min\big\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, y, z^k, \lambda^k) \;\big|\; y \in \mathcal{Y}\big\}, & \text{(7.26b)} \\ \tilde{z}^k = \arg\min\big\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, \tilde{y}^k, z, \lambda^k) \;\big|\; z \in \mathcal{Z}\big\}, & \text{(7.26c)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b). & \text{(7.26d)} \end{array} \right.$$

For this special case, the matrix Q in (7.13) has the form

$$Q = \begin{pmatrix} \beta B^T B & 0 & 0 \\ \beta C^T B & \beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I_m \end{pmatrix}. \tag{7.27}$$

We take the second choice of the matrix M (see (7.19)) in the correction step, namely,

$$M_\nu = \begin{pmatrix} \nu Q_0^{-T} D_0 & 0 \\ -\beta\mathcal{A} & I \end{pmatrix}.$$

Because

$$Q_0 = \begin{pmatrix} B^TB & 0 \\ C^TB & C^TC \end{pmatrix}, \quad \text{and} \quad D_0 = \text{diag}(B^TB, C^TC),$$

we obtain

$$Q_0^{-T}D_0 = \begin{pmatrix} I & -(B^TB)^{-1}B^TC \\ 0 & I \end{pmatrix},$$

and thus

$$M_\nu = \begin{pmatrix} \nu I & -\nu(B^TB)^{-1}B^TC & 0 \\ 0 & \nu I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{7.28}$$

The correction is updated by

$$v^{k+1} = v^k - M_\nu(v^k - \tilde{v}^k). \tag{7.29}$$

For the dual variable

$$\lambda^{k+1} = \lambda^k - [-\beta B(y^k - \tilde{y}^k) - \beta C(z^k - \tilde{z}^k) + (\lambda^k - \tilde{\lambda}^k)].$$

Using the definition of $\tilde{\lambda}^k$, we have

$$-\beta B(y^k - \tilde{y}^k) - \beta C(z^k - \tilde{z}^k) + (\lambda^k - \tilde{\lambda}^k) = \beta(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b),$$

the correction step (7.29) can be written as

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \\ \lambda^{k+1} \end{pmatrix} := \begin{pmatrix} By^k \\ Cz^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \nu I & -\nu I & 0 \\ 0 & \nu I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} B(y^k - y^{k+1}) \\ C(z^k - z^{k+1}) \\ \lambda^k - \lambda^{k+1} \end{pmatrix}, \tag{7.30}$$

where the $(y^{k+1}, z^{k+1}, \lambda^{k+1})$ in the right hand side is the output of the direct extension of ADMM (7.5) for the problem with three separate operators (7.24). The details of (7.30) is

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \\ \lambda^{k+1} \end{pmatrix} := \begin{pmatrix} (1-\nu)By^k + \nu By^{k+1} + \nu C(z^k - z^{k+1}) \\ (1-\nu)Cz^k + \nu Cz^{k+1} \\ \lambda^{k+1} \end{pmatrix}. \tag{7.31}$$

Recall, for $\nu = 1$, the matrix G in (7.21) is positive semi-definite and the related method has $O(1/t)$ convergence rate in an ergodic sense.

7.2 ADMM + Prox-Parallel Splitting ALM

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

$$\left\{\begin{array}{l} x_1^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \dots, x_m^k, \lambda^k) \mid x_1 \in \mathcal{X}_1\big\}, \\ \text{for} \quad i = 2, \dots, m, \quad \text{do :} \\ x_i^{k+1} = \arg\min\limits_{x_i \in \mathcal{X}_i} \left\{ \begin{array}{c} \mathcal{L}_\beta(x_1^{k+1}, x_2^k \dots, x_{i-1}^k, x_i, x_{i+1}^k, \dots, x_m^k, \lambda^k) \\ + \frac{\tau\beta}{2}\|A_i(x_i - x_i^k)\|^2 \end{array} \right\}, \\ \lambda^{k+1} = \lambda^k - \beta\big(\sum_{i=1}^m A_i x_i^{k+1} - b\big). \end{array}\right. \tag{7.32}$$

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

An equivalent recursion of (7.32) $\mu = \tau + 1$ and τ is given in (7.32).

$$\left\{\begin{array}{l} x_1^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \dots, x_m^k, \lambda^k) \mid x_1 \in \mathcal{X}_1\big\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta\big(A_1 x_1^{k+1} + \sum_{i=2}^m A_i x_i^k - b\big), \\ \text{for} \quad i = 2, \dots, m, \quad \text{do :} \\ x_i^{k+1} = \arg\min \left\{ \begin{array}{c} \theta_i(x_i) - (\lambda^{k+\frac{1}{2}})^T A_i x_i \\ + \frac{\mu\beta}{2}\|A_i(x_i - x_i^k)\|^2 \end{array} \middle| \; x_i \in \mathcal{X}_i \right\}, \\ \lambda^{k+1} = \lambda^k - \beta\big(\sum_{i=1}^m A_i x_i^{k+1} - b\big). \end{array}\right. \tag{7.33}$$

The method (7.33) is proposed in IMA Numerical Analysis [21]:

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

Equivalence of (7.32) and (7.33)

It needs only to check the optimization conditions of their x_i-subproblems for $i = 2, \ldots, m$. Note that the optimal condition of the x_i-subproblem of (7.32) is

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

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

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \beta\big(A_1 x_1^{k+1} + \textstyle\sum_{j=2}^m A_j x_j^k - b\big); \tag{7.34}$$

it can be written as

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

and consequently

$$x_i^{k+1} \in \mathcal{X}_i, \quad \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T \big\{ -A_i^T \lambda^{k+\frac{1}{2}} \\ +(1+\tau)\beta A_i^T A_i (x_i^{k+1} - x_i^k) \big\} \geq 0, \ \forall \, x_i \in \mathcal{X}_i. \tag{7.35}$$

Setting $\mu = 1 + \tau$, (7.35) just is the optimal condition of the x_i-subproblem of (7.33). Notice that the

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

For $i = 2, \ldots, m$,

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

We use (7.33) to analyze the convergence conditions. By denoting

$$\tilde{x}_i^k = x_i^{k+1}, \quad i = 1, \ldots, m \qquad \text{and} \qquad \tilde{\lambda}^k = \lambda^{k+\frac{1}{2}}, \tag{7.36}$$

the optimal condition of the x_i-subproblems of (7.33) can be written as

$$\begin{cases} \theta_1(x_1) - \theta_1(\tilde{x}_1^k) + (x_1 - \tilde{x}_1^k)^T(-A_1^T\tilde{\lambda}^k) \ge 0, \ \forall x_1 \in \mathcal{X}_1; \\ \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\big(-A_i^T\tilde{\lambda}^k + \mu\beta A_i^T A_i(\tilde{x}_i^k - x_i^k)\big) \ge 0, \\ \qquad \forall x_i \in \mathcal{X}_i; i = 2, \ldots, m. \end{cases} \tag{7.37}$$

Since $\tilde{\lambda}^k = \lambda^{k+\frac{1}{2}}$, we have

$$\tilde{\lambda}^k = \lambda^k - \beta\big(A_1\tilde{x}_1^k + \sum_{j=2}^m A_j x_j^k - b\big)$$

and thus

$$(\sum_{i=1}^m A_i\tilde{x}_i^k - b) - \sum_{j=2}^m A_j(\tilde{x}_j^k - x_j^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \tag{7.38}$$

Combining (7.37) and (7.38) together and using the notations in (7.3), we obtain

Lemma 7.2 *Let w^{k+1} be generated by* (7.33) *from the given vector v^k and $\tilde{w}^k$ be defined by* (7.36)*. Then, we have $\tilde{w}^k \in \Omega$ and*

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall\, w \in \Omega, \tag{7.39}$$

where

$$Q = \begin{pmatrix} \mu\beta A_2^T A_2 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & \mu\beta A_m^T A_m & 0 \\ -A_2 & \cdots & -A_{m-1} & -A_m & \frac{1}{\beta}I_m \end{pmatrix}. \tag{7.40}$$

This is a prediction as described in (4.4a). Here, the matrix Q is not symmetric.

Since $\tilde{\lambda}^k = \lambda^k - \beta\big(A_1\tilde{x}_1^k + \sum_{j=2}^m A_j x_j^k - b\big)$ and $\tilde{x}_i^k = x_i^{k+1}$, we have

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - \beta\big(\textstyle\sum_{j=1}^m A_j x_j^{k+1} - b\big) \\ &= \lambda^k - \big[-\beta\textstyle\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k) + (\lambda^k - \tilde{\lambda}^k)\big] \end{aligned} \qquad (7.41)$$

Thus, letting

$$M = \begin{pmatrix} I & 0 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & I & 0 \\ -\beta A_2 & \cdots & -\beta A_{m-1} & -\beta A_m & I_m \end{pmatrix}, \qquad (7.42)$$

the v^{k+1} obtained by (7.33) can be written as

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

Now, we check if the convergence conditions (4.5) is satisfied.

For analysis convenience, we denote

$$D_0 = \text{diag}(A_2^T A_2, A_3^T A_3, \ldots, A_m^T A_m) \qquad (7.43a)$$

and

$$\mathcal{A} = (A_2, A_3, \ldots, A_m). \qquad (7.43b)$$

Thus the matrix Q in (7.40) and M in (7.42) can be written in compact form,

$$Q = \begin{pmatrix} \mu\beta D_0 & 0 \\ -\mathcal{A} & \frac{1}{\beta} I \end{pmatrix},$$

and

$$M = \begin{pmatrix} I & 0 \\ -\beta\mathcal{A} & I \end{pmatrix},$$

respectively. By a simple manipulation, it shows that

$$H = QM^{-1} = \begin{pmatrix} \mu\beta D_0 & 0 \\ -\mathcal{A} & \frac{1}{\beta}I \end{pmatrix} \begin{pmatrix} I & 0 \\ \beta\mathcal{A} & I \end{pmatrix} = \begin{pmatrix} \mu\beta D_0 & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix}$$

is positive definite.

For the matrix Q defined in (7.40) and M defined in (7.42), we have

$$\begin{aligned} Q^T M &= \begin{pmatrix} \mu\beta D_0 & -\mathcal{A}^T \\ 0 & \frac{1}{\beta}I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta\mathcal{A} & I \end{pmatrix} \\ &= \begin{pmatrix} \mu\beta D_0 + \beta\mathcal{A}^T\mathcal{A} & -\mathcal{A}^T \\ -\mathcal{A} & \frac{1}{\beta}I \end{pmatrix}. \end{aligned}$$

Consequently, we have

$$\begin{aligned} G &= Q^T + Q - M^T H M \;=\; Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2\mu\beta D_0 & -\mathcal{A}^T \\ -\mathcal{A} & \frac{2}{\beta}I \end{pmatrix} - \begin{pmatrix} \mu D_0 + \beta\mathcal{A}^T\mathcal{A} & -\mathcal{A}^T \\ -\mathcal{A} & \frac{1}{\beta}I \end{pmatrix} \\ &= \begin{pmatrix} \mu\beta D_0 - \beta\mathcal{A}^T\mathcal{A} & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix} := \begin{pmatrix} G_0 & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix}. \end{aligned}$$

Notice that

$$G \succ 0 \quad \Longleftrightarrow \quad G_0 = \mu\beta D_0 - \beta\mathcal{A}^T\mathcal{A} \succ 0.$$

Thus, we need only to check the positivity of G_0. Since

$$D_0 = \mathrm{diag}(A_2^T A_2, A_3^T A_3, \ldots, A_m^T A_m),$$

and

$$\mathcal{A} = (A_2, A_3, \ldots, A_m),$$

by a manipulation, we obtain

$$
\begin{aligned}
G_0 &= \mu\beta D_0 - \beta\mathcal{A}^T\mathcal{A} \\
&= \beta \begin{pmatrix} (\mu-1)A_2^TA_2 & -A_2^TA_3 & \cdots & -A_2^TA_m \\ -A_3^TA_2 & (\mu-1)A_3^TA_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & -A_{m-1}^TA_m \\ -A_m^TA_2 & \cdots & -A_m^TA_{m-1} & (\mu-1)A_m^TA_m \end{pmatrix} \\
&= \begin{pmatrix} A_2^T & & & \\ & A_3^T & & \\ & & \ddots & \\ & & & A_m^T \end{pmatrix} \beta G_0(\mu) \begin{pmatrix} A_2 & & & \\ & A_3 & & \\ & & \ddots & \\ & & & A_m \end{pmatrix},
\end{aligned}
$$

where $G_0(\mu)$ is an $(m-1)\times(m-1)$ blocks matrix

$$
\begin{aligned}
G_0(\mu) &= \begin{pmatrix} (\mu-1)I & -I & \cdots & -I \\ -I & (\mu-1)I & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I \\ -I & \cdots & -I & (\mu-1)I \end{pmatrix} \\
&= \mu \begin{pmatrix} I & & & \\ & \ddots & & \\ & & \ddots & \\ & & & I \end{pmatrix} - \begin{pmatrix} I & I & \cdots & I \\ I & I & \cdots & I \\ \vdots & \vdots & & \vdots \\ I & I & \cdots & I \end{pmatrix}_{\substack{(m-1)\times(m-1) \\ \text{blocks}}}
\end{aligned}
$$

It is clear that G_μ is positive definite if and only if $\mu > m-1$.

Since $\tau = \mu - 1$, the method (7.32) is convergent when $\tau > m-2$.

ADMM + Prox-Parallel Splitting ALM

- Produce a predictor w^{k+1} via (7.33) with given $v^k = (x_2^k, \ldots, x_m^k, \lambda^k)$, where $\mu > m-1$.

Theorem 7.2 *The sequence $\{v^k\}$generated by the ADMM (+Prox-Parallel Splitting ALM) satisfies*

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

where

$$H = \begin{pmatrix} \mu\beta D_0 & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}, \qquad G = \begin{pmatrix} \mu\beta D_0 - \beta \mathcal{A}^T\mathcal{A} & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix},$$

$$D_0 = \textit{diag}(A_2^T A_2, A_3^T A_3, \ldots, A_m^T A_m), \quad \mathcal{A} = (A_2, A_3, \ldots, A_m).$$

Implementation of the method for three block problems

For the problem with three separable operators

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

we have

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

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

$$\begin{cases} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, \\ z^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{cases} \tag{7.44}$$

An equivalent recursion of (7.44) is

$$\begin{cases} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\ y^{k+1} = \text{Argmin}\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \text{Argmin}\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T Cz + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{cases} \tag{7.45}$$

where $\mu = \tau + 1 \geq 2$. Implementation of (7.45) is via

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

This method is accepted by Osher's research group

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

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 21, NO. 7, JULY 2012 3239

A Convex Model for Nonnegative Matrix Factorization and Dimensionality Reduction on Physical Space

Ernie Esser, Michael Möller, Stanley Osher, Guillermo Sapiro, *Senior Member, IEEE*, and Jack Xin

$$\min_{T \geq 0, V_j \in D_j, e \in E} \zeta \sum_i \max_j (T_{i,j}) + \langle R_w \sigma C_w, T \rangle$$
$$\text{such that} \quad YT - X_s = V - X_s \text{diag}(e). \quad (15)$$

Since the convex functional for the extended model (15) is slightly more complicated, it is convenient to use a variant of ADMM that allows the functional to be split into more than two parts. The method proposed by He *et al.* in [34] is appropriate for this application. Again, introduce a new variable Z

Using the ADMM-like method in [34], a saddle point of the augmented Lagrangian can be found by iteratively solving the subproblems with parameters $\delta > 0$ and $\mu > 2$, shown in the

tion refinement step. Due to the different algorithm used to solve the extended model, there is an additional numerical parameter μ, which for this application must be greater than two according to [34]. We set μ equal to 2.01. There are also model parame-

[33] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," 2009 [Online]. Available: http://arxiv.org/PS cache/arxiv/pdf/0912/0912.3599v1.pdf

[34] B. He, M. Tao, and X. Yuan, "A splitting method for separate convex programming with linking linear constraints," Tech. Rep., 2011 [Online]. Available: http://www.optimization-online.org/DB_FILE/2010/06/ 2665.pdf

What is the optimal regular factor –Main result in OO6235

Recent Advance in : Bingsheng He, Xiaoming Yuan: On the Optimal Proximal Parameter of an ADMM-like Splitting Method for Separable Convex Programming http://www.optimization-online.org/DB_HTML/2017/ 10/6235.html

Our new assertion: Solving the problem (7.24). The parameter τ in (7.44)

- if $\tau > 0.5$, the method is still convergent;
- if $\tau < 0.5$, there is divergent example.

Equivalently the parameter μ in (7.45) :

- if $\mu > 1.5$, the method is still convergent;
- if $\mu < 1.5$, there is divergent example.

For convex optimization problem (7.24) with three separable objective functions, the parameters in the equivalent methods (7.44) and (7.45) :

- **0.5** is the threshold factor of the parameter τ in (7.44) !
- **1.5** is the threshold factor of the parameter μ in (7.45) !

8 Self-adaptive gradient descent method for convex optimization

This section is relatively independent of other sections. Let $f : \Re^n \to \Re$ be a differentiable function, $\Omega \subset \Re^n$ be a closed convex set ($\Omega = \Re^n$ is the possible simplest case). Assume that the projection on Ω is easy to be carried out. For example, $\Omega = \Re^n$, $\Re^n_+$, or Ω is a "box".

We study the gradient descent method for convex optimization problem

$$\min \{f(x) \mid x \in \Omega\}. \tag{8.1}$$

The solution set of (8.1) is denoted by Ω^* and assumed to be non-empty.

According to the analysis in Sect. 1.1, the problem (8.1) is equivalent to finding a $x^* \in \Omega$, such that

$$\text{VI}(\Omega, \nabla f) \qquad x^* \in \Omega, \quad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \Omega. \tag{8.2}$$

For the discussion in this section, we need some basic concepts of projection. Let x^* be a solution of VI$(\Omega, \nabla f)$, for any given $\beta > 0$, we have

$$x^* = P_\Omega[x^* - \beta \nabla f(x^*)].$$

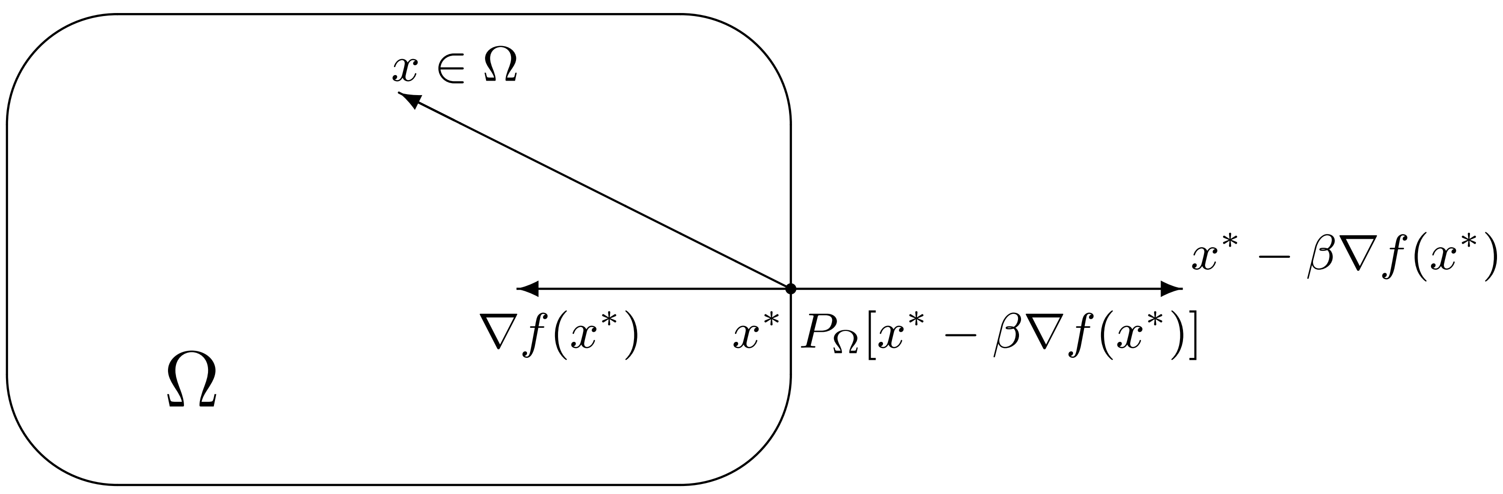


Fig. 2.1 x^* is a solution of VI$(\Omega, \nabla f)$ (8.1) $\quad \Leftrightarrow \quad x^* = P_\Omega[x^* - \beta \nabla f(x^*)]$

For given x^k and $\beta > 0$, we denote

$$\tilde{x}^k \quad = \quad P_\Omega[x^k - \beta \nabla f(x^k)], \tag{8.3}$$

which is projection of a given vector $[x^k - \beta\nabla f(x^k)]$ on Ω. In other words,

$$\tilde{x}^k = \arg\min\{\frac{1}{2}\|x - [x^k - \beta_k \nabla f(x^k)]\|^2 \,|\, x \in \Omega\}.$$

One of the important property of the projection mapping is

$$(x - P_\Omega(z))^T(z - P_\Omega(z)) \le 0, \quad \forall z \in R^n, \forall x \in \Omega. \qquad (8.4)$$

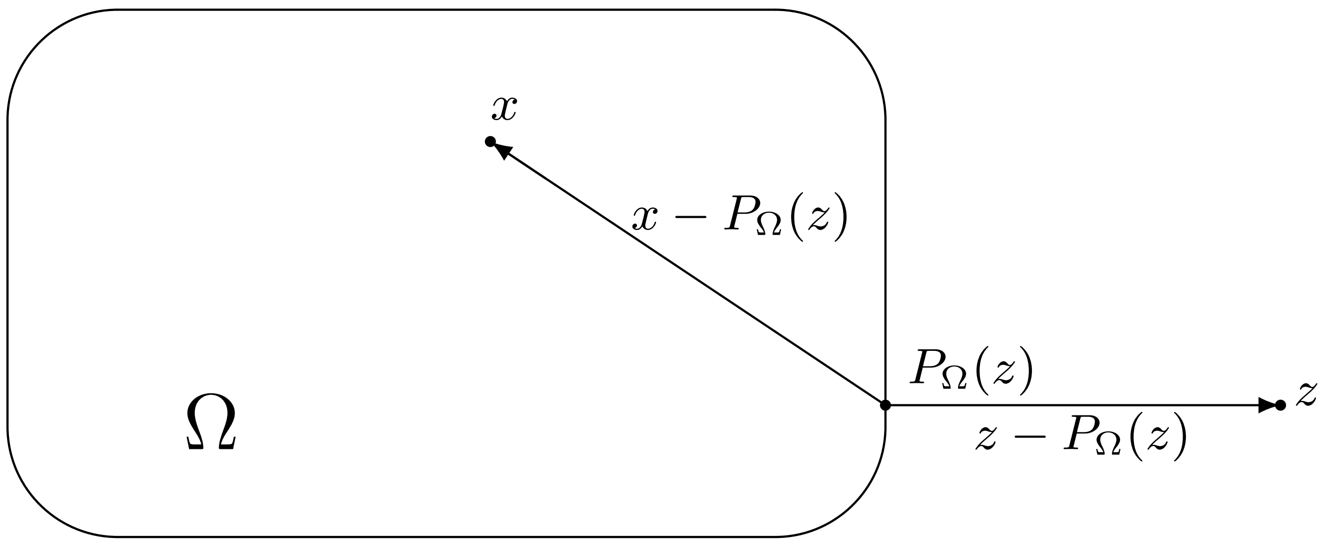


Fig. 2.2 Geometric interpretation of the inequality (8.4)

Since $\tilde{x}^k \in \Omega$, according to the definition of the variational inequality formulation (see (8.2)), for any $\beta > 0$, we have

$$\text{(FI1)} \qquad (\tilde{x}^k - x^*)^T \beta \nabla f(x^*) \ge 0. \qquad (8.5)$$

We call (8.5) the first fundamental inequality.

Notice that $\tilde{x}^k$ is the projection of $x^k - \beta\nabla f(x^k)$ on Ω and $x^* \in \Omega$.

Set $v = x^k - \beta\nabla f(x^k)$ and $u = x^*$ in the projection inequality (8.4), since $P_\Omega(v) = \tilde{x}^k$, we have

$$\text{(FI2)} \qquad (\tilde{x}^k - x^*)^T \{[x^k - \beta\nabla f(x^k)] - \tilde{x}^k\} \ge 0. \qquad (8.6)$$

We call (8.6) the second fundamental inequality.

8.1 Motivation from Projection and Contraction Method

The projection and contraction (P-C) method is an iterative predict-correct (P-C) method. We say a method is a contractive, if the distance of the iterates $\{x^k\}$ to the solution set is strictly monotone decreasing.

For given x^k, the projection and contraction method offers its predictor $\tilde{x}^k$ by

$$\tilde{x}^k = P_\Omega[x^k - \beta \nabla f(x^k)].$$

Let $H \in \Re^{n\times n}$ be a symmetric semi-definite matrix. Although the initial purpose of constructing projection and contraction methods [13, 14, 15] are not for solving convex quadratic programming

$$\min\{\frac{1}{2}x^T Hx + c^T x \,|\, x \in \Omega\}, \tag{8.7}$$

we still illustrate our idea with problem (8.7). For the problem (8.7), the

corresponding linear variational inequality is

$$x^* \in \Omega, \quad (x - x^*)^T \beta(Hx^* + c) \geq 0, \quad \forall\, x \in \Omega. \tag{8.8}$$

For given x^k, the predictor $\tilde{x}^k$ is given by

$$\tilde{x}^k = P_\Omega[x^k - \beta(Hx^k + c)]. \tag{8.9}$$

8.1.1 Projection and contraction for convex QP

For the linear variational inequality (8.8), the fundamental inequalities (FI1) (8.5) and (FI2) (8.6) are reduced to

$$\begin{cases} (\tilde{x}^k - x^*)^T \beta(Hx^* + c) \geq 0, & \text{(FI1)} \\ \qquad \text{and} & \\ (\tilde{x}^k - x^*)^T \big([x^k - \beta(Hx^k + c)] - \tilde{x}^k\big) \geq 0, & \text{(FI2)} \end{cases}$$

respectively.

Adding (FI1) and (FI2), we get

$$\{(x^k - x^*) - (x^k - \tilde{x}^k)\}^T \{(x^k - \tilde{x}^k) - \beta H(x^k - x^*)\} \geq 0.$$

Since H is positive semi-definite, from the above inequality, we obtain

$$(x^k - x^*)^T (I + \beta H)(x^k - \tilde{x}^k) \geq \|x^k - \tilde{x}^k\|^2, \quad \forall x^* \in \Omega.$$

The last inequality can be interpreted as

$$\Big\langle \nabla \big(\frac{1}{2}\|x - x^*\|^2_{(I+\beta H)}\big)\big|_{x=x^k}, (x^k - \tilde{x}^k) \Big\rangle \geq \|x^k - \tilde{x}^k\|^2, \quad \forall x^* \in \Omega.$$

In other words, $-(x^k - \tilde{x}^k)$ is a descent direction of

the unknown distance function $\frac{1}{2}\|x - x^*\|^2_{(I+\beta H)}$ at x^k.

By letting

$$G = I + \beta H, \tag{8.10}$$

we get

$$(x^k - x^*)^T G(x^k - \tilde{x}^k) \geq \|x^k - \tilde{x}^k\|^2, \quad \forall\, x^* \in \Omega^*. \tag{8.11}$$

The projection and contraction methods requires the sequence

$\{\|x^k - x^*\|^2_G\}$ to be strictly monotone decreasing. We let

$$x(\alpha) = x^k - \alpha(x^k - \tilde{x}^k), \tag{8.12}$$

be the new iterate depends the step-size α, and consider the function

$$\vartheta(\alpha) = \|x^k - x^*\|^2_G - \|x(\alpha) - x^*\|^2_G. \tag{8.13}$$

Using (8.11), it follows that

$$\begin{aligned} \vartheta(\alpha) &= \|x^k - x^*\|^2_G - \|x^k - x^* - \alpha(x^k - \tilde{x}^k)\|^2_G \\ &\geq 2\alpha\|x^k - \tilde{x}^k\|^2 - \alpha^2\|x^k - \tilde{x}^k\|^2_G. \end{aligned} \tag{8.14}$$

In other words, we get a quadratic function

$$q(\alpha) = 2\alpha\|x^k - \tilde{x}^k\|^2 - \alpha^2\|x^k - \tilde{x}^k\|_G^2. \tag{8.15}$$

which is a low bound of $\vartheta(\alpha)$. The function $q(\alpha)$ reaches its maximum at

$$\alpha_k^* = \frac{\|x^k - \tilde{x}^k\|^2}{\|x^k - \tilde{x}^k\|_G^2}. \tag{8.16}$$

In practical computation, we use

$$x^{k+1} = x^k - \gamma\alpha_k^*(x^k - \tilde{x}^k), \quad \gamma \in (0, 2) \tag{8.17}$$

to produce the new iterate x^{k+1} (corrector). The sequence $\{x^k\}$ satisfies

$$\begin{aligned} \|x^{k+1} - x^*\|_G^2 &\leq \|x^k - x^*\|_G^2 - q(\gamma\alpha^*) \\ &= \|x^k - x^*\|_G^2 - \gamma(2-\gamma)\alpha_k^*\|x^k - \tilde{x}^k\|^2, \end{aligned} \tag{8.18}$$

where $G = I + \beta H$.

Note that $G = (I + \beta H)$ and the "optimal step-size" (see (8.16)) is

$$\alpha_k^* = \frac{\|x^k - \tilde{x}^k\|^2}{(x^k - \tilde{x}^k)^T(I + \beta H)(x^k - \tilde{x}^k)}. \tag{8.19}$$

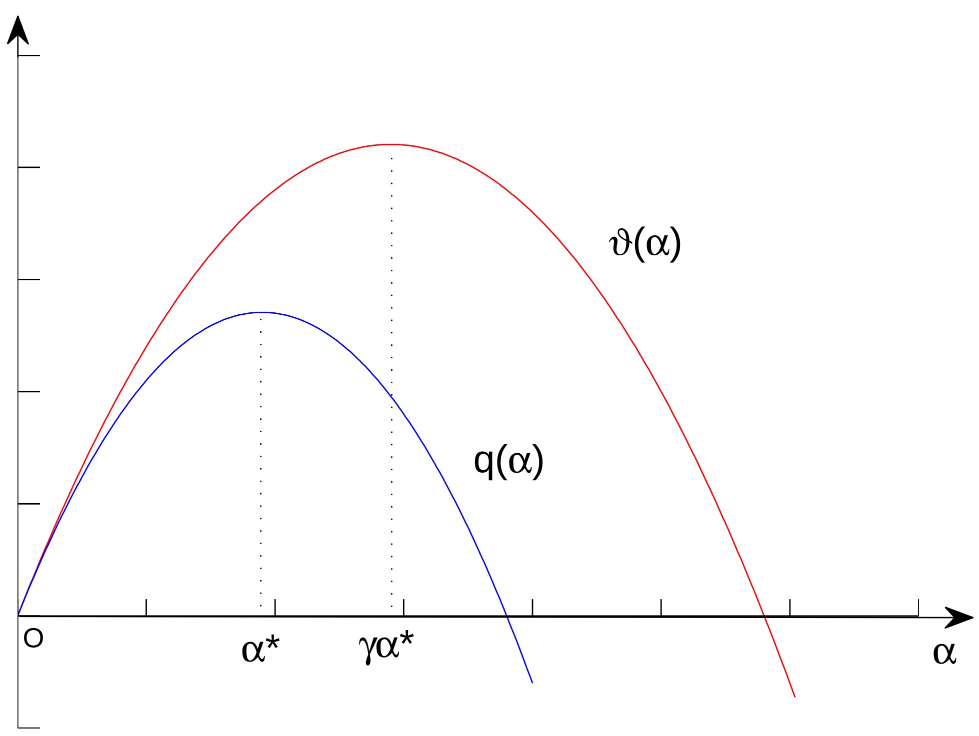


Fig. 2.3 The meaning of the relaxed factor $\gamma \in [1, 2)$

The convergence speed is dependent on the parameter β !

Self adaptive gradient descent method for convex QP (8.7).

For a given x^k, for the chosen parameter β, the predictor is given by

$$\tilde{x}^k = P_\Omega[x^k - \beta(Hx^k + c)].$$

Additionally, if the condition

$$(x^k - \tilde{x}^k)^T(\beta H)(x^k - \tilde{x}^k) \le \nu\|x^k - \tilde{x}^k\|^2, \quad \nu \in (0,1) \tag{8.20}$$

is satisfied, then according to (8.19), we have

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

Self adaptive gradient descent method for convex QP (8.7).

Then, we can in (8.17) dynamically choose

$$\gamma_k = 1/\alpha_k^*, \qquad \text{thus} \qquad 1 < \gamma_k \le 1 + \nu < 2. \tag{8.21}$$

In this case, $\gamma_k\alpha_k^* = 1$, the corrector formula (8.17), namely,

$$x^{k+1} = x^k - \gamma_k\alpha_k^*(x^k - \tilde{x}^k)$$

becomes

$$x^{k+1} = \tilde{x}^k = P_\Omega[x^k - \beta(Hx^k + c)]. \tag{8.22}$$

The contraction inequality (8.18)

$$\begin{aligned}\|x^{k+1} - x^*\|^2_{(I+\beta_k H)} &\le \|x^k - x^*\|^2_{(I+\beta_k H)} - (2-\gamma_k)\|x^k - x^{k+1}\|^2 \\ &\le \|x^k - x^*\|^2_{(I+\beta_k H)} - (1-\nu)\|x^k - x^{k+1}\|^2.\end{aligned}$$

The last inequality follows from $1 - \nu \le 2 - \gamma_k$ (see (8.21), $\gamma_k \le 1 + \nu$).

We get a simple projected gradient method, the only condition is

$$(x^k - \tilde{x}^k)^T(\beta H)(x^k - \tilde{x}^k) \le \nu \|x^k - \tilde{x}^k\|^2, \quad \nu \in (0,1). \qquad (8.23)$$

8.1.2 Comparison with the Steepest descent method

How good is the self adaptive gradient descent method discussed in §8.1.1 ?

When $\Omega = \Re^n$, the problem (8.7) becomes a unconstrained convex quadratic programming

$$\min\{\tfrac{1}{2}x^T H x + c^T x\}. \qquad (8.24)$$

If we use the steepest descent method to solve (8.24), in k-th step, the iterative formula is

$$x^{k+1} = x^k - \alpha_k^{SD}(Hx^k + c),$$

where the step-size $\quad \alpha_k^{SD} = \dfrac{\|Hx^k + c\|^2}{(Hx^k + c)^T H (Hx^k + c)}.$

If we use the self adaptive gradient descent method discussed in §8.1.1 to solve (8.24), in k-th step, the iterative formula is

$$x^{k+1} = x^k - \beta_k(Hx^k + c)$$

where

$$\beta_k \le \nu \cdot \frac{\|Hx^k + c\|^2}{(Hx^k + c)^T H (Hx^k + c)} = \nu \cdot \alpha_k^{SD}.$$

In comparison with the steepest descent method, we have

- ✠ the same search direction,
- ✠ reduced step-size.

What is the different numerical behaviour ?

Preliminary numerical tests for the problem (8.24)

The Hessian Matrix In the test example, the Hessian matrix is the Hilbert matrix.

$$H = \{h_{ij}\}, \quad h_{ij} = \frac{1}{i+j-1}, \quad i = 1, \cdots, n; \quad j = 1, \cdots, n.$$

$$n \text{ from } 100 \text{ to } 500.$$

We set

$$x^* = (1, 1, \ldots, 1)^T \in R^n \quad \text{and} \quad c = -Hx^*.$$

Different start points:

$$x^0 = 0, \qquad x^0 = c, \quad \text{or} \quad x^0 = -c.$$

Stop criteri0n:

$$\|Hx^k + c\|/\|Hx^0 + c\| \le 10^{-7}.$$

The reduced step size:

$$\beta = r\alpha_k^{SD}.$$

Table 1. Iteration number with different r ($r = 1$ is the SD method) Start point $x^0 = 0$

n=	0.1	0.3	0.5	0.7	0.8	**0.9**	0.95	0.99	**1.00**	1.20
100	2863	1346	853	627	582	**437**	565	1201	**13169**	22695
200	3283	1398	923	804	541	**669**	898	1178	**14655**	21083
300	3497	1323	856	739	720	**568**	619	1545	**17467**	24027
500	3642	1351	1023	773	667	**578**	836	2024	**17757**	22750

Start with $x^0 = 0$. Stop with x^k. In average: $\|x^k - x^*\|/\|x^0 - x^*\| = 3.0\text{e}-3$.

Table 2. Iteration number with different r ($r = 1$ is the SD method) Start point $x^0 = c$

n=	0.1	0.3	0.5	0.7	0.8	**0.9**	0.95	0.99	**1.00**	1.2
100	2129	1034	544	424	302	**438**	568	919	**5527**	9667
200	1880	808	568	482	372	**339**	446	713	**6625**	11023
300	1852	1002	741	531	610	**452**	450	917	**6631**	10235
500	2059	939	568	573	379	**547**	558	874	**7739**	11269

Start with $x^0 = c$. Stop with x^k. In average: $\|x^k - x^*\|/\|x^0 - x^*\| = 1.8\text{e}-3$.

Table 3. Iteration number with different r ($r = 1$ is the SD method) Start point $x^0 = -c$

n=	0.1	0.3	0.5	0.7	0.8	**0.9**	0.95	0.99	**1.00**	1.2
100	2545	1221	666	591	498	**482**	638	1581	**14442**	20380
200	2826	990	874	470	526	**455**	578	841	**15222**	18892
300	2891	1299	918	738	549	**571**	608	2552	**18762**	21208
500	3158	1769	909	678	506	**512**	678	1240	**17512**	19790

Start with $x^0 = -c$. Stop with x^k. In average: $\|x^k - x^*\|/\|x^0 - x^*\| = 3.8\text{e}-3$.

With the same direction and the reduced step-size, the method is 10∼30 times faster than the Steepest descent method !

$r \in (0.4, 0.95)$ is the suitable reduced factor !

What is the findings ? By setting

$$f(x) = \frac{1}{2}x^T Hx + c^T x$$

in (8.7), the iterative formula of the self adaptive gradient descent method (8.22) becomes

$$x^{k+1} = P_\Omega[x^k - \beta_k \nabla f(x^k)],$$

and the strategy for choosing β_k (8.20) is ($\nu \in [0.4, 0.95]$)

$$(x^k - x^{k+1})^T \beta_k [\nabla f(x^k) - \nabla f(x^{k+1})] \leq \nu \cdot \|x^k - x^{k+1}\|^2.$$

8.2 Projected Gradient Descent (PDG) method for nonlinear convex optimization

The findings on projection and contraction method for solving the quadratic programming also contribute to solving the following differentiable convex optimization problem.

Let Ω be a convex closed set in R^n. The problem concerted in this subsection is to find $x^* \in \Omega$, such that

$$(x - x^*)^T g(x^*) \geq 0, \quad \forall\, x \in \Omega, \tag{8.25}$$

where $g(x)$ is a mapping from R^n into itself. We assume that $g(x)$ is the gradient of a certain convex function, say $f(x)$, however, $f(x)$ is not provided. Only for given x, $g(x)$ is observable (sometimes with costly expenses).

In other words, (8.25) is equivalent to the following convex optimization problem

$$\min \{f(x) \mid x \in \Omega\}. \tag{8.26}$$

We call (8.26) an **oracle** convex optimization problem, because only the gradient information $g(x)$ can be used for solving (8.26). For $x^* \in \Omega^*$, we assume $f(x^*) > -\infty$.

In addition, we also assume that $g(x)$ is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall\, x,\, y \in \Re^n. \tag{8.27}$$

We require that $g(x)$ is Lipschitz continuous while it does not need to know the value of L in (8.27).

The methods presented in this section do not involve the value of $f(x)$, but they can guarantee that $f(x^k)$ is strict monotonically decreasing, hence they belong to the descending methods.

8.2.1 Steepest descent method for convex programming

Single step projected gradient method.

Step k. ($k \geq 0$) Set

$$x^{k+1} = P_\Omega[x^k - \beta_k g(x^k)], \tag{8.28a}$$

where the step size β_k satisfies the following condition:

$$(x^k - x^{k+1})^T (g(x^k) - g(x^{k+1})) \leq \frac{\nu}{\beta_k}\|x^k - x^{k+1}\|^2. \tag{8.28b}$$

Note that the condition (8.28b) automatically holds when $\beta_k \leq \nu/L$, where L is the Lipschitz modulus of $g(x)$. The reason is

$$\begin{aligned} &(x^k - x^{k+1})^T \beta_k (g(x^k) - g(x^{k+1})) \\ &\quad \le \;\; \|x^k - x^{k+1}\| \cdot \beta_k L \|x^k - x^{k+1}\| \;\le\; \nu \|x^k - x^{k+1}\|^2. \end{aligned}$$

8.2.2 Global convergence of the proposed method

In the following, we show an important lemma by using the basic properties of the projection and convex function.

Lemma 8.1 *For given x^k, let x^{k+1} be generated by* (8.28a). *If the step-size β_k satisfies* (8.28b), *then we have*

$$(x - x^{k+1})^T g(x^k) \ge \frac{1}{\beta_k}(x - x^{k+1})^T (x^k - x^{k+1}), \quad \forall x \in \Omega, \qquad (8.29)$$

and

$$\begin{aligned} &\beta_k (f(x) - f(x^{k+1})) \\ &\quad \ge \; (x - x^{k+1})^T (x^k - x^{k+1}) - \nu \|x^k - x^{k+1}\|^2, \; \forall\, x \in \Omega. \end{aligned} \qquad (8.30)$$

Proof. Note that x^{k+1} is the projection of $[x^k - \beta_k g(x^k)]$ on Ω (see (8.28a)), according to the projection's property (8.4), we have

$$(x - x^{k+1})^T \{[x^k - \beta_k g(x^k)] - x^{k+1}\} \le 0, \quad \forall\, x \in \Omega.$$

It follows that

$$(x - x^{k+1})^T \beta_k g(x^k) \ge (x - x^{k+1})^T (x^k - x^{k+1}), \quad \forall\, x \in \Omega, \qquad (8.31)$$

and the first assertion (8.29) is proved. Using the convexity of f , we have

$$f(x) \ge f(x^k) + (x - x^k)^T g(x^k), \qquad (8.32)$$

and

$$\begin{aligned} f(x^k) \quad &\ge \quad f(x^{k+1}) + (x^k - x^{k+1})^T g(x^{k+1}) \\ &= \quad f(x^{k+1}) + (x^k - x^{k+1})^T g(x^k) \\ &\qquad\qquad - (x^k - x^{k+1})^T \big(g(x^k) - g(x^{k+1})\big) \\ &\ge \quad f(x^{k+1}) + (x^k - x^{k+1})^T g(x^k) - \frac{\nu}{\beta_k} \|x^k - x^{k+1}\|^2. \qquad (8.33) \end{aligned}$$

The last " $\geq$ " is due to (8.28b). From (8.32) and (8.33), we get

$$\begin{aligned} & f(x) - f(x^{k+1}) \\ & \quad \geq \quad f(x^k) + (x - x^k)^T g(x^k) \\ & \qquad\quad -\{f(x^k) + (x^{k+1} - x^k)^T g(x^k) + \frac{\nu}{\beta_k}\|x^k - x^{k+1}\|^2\} \\ & \quad = \quad (x - x^{k+1})^T g(x^k) - \frac{\nu}{\beta_k}\|x^k - x^{k+1}\|^2. \end{aligned} \tag{8.34}$$

Substituting (8.29) in (8.34), we obtain

$$f(x) - f(x^{k+1}) \geq \frac{1}{\beta_k}(x - x^{k+1})^T (x^k - x^{k+1}) - \frac{\nu}{\beta_k}\|x^k - x^{k+1}\|^2,$$

and the second assertion of this lemma is proved. $\square$

The following theorem shows that the projected gradient method (8.28) is a descent method whose objective function value $\{f(x^k)\}$ is monotonically decreasing.

Theorem 8.1 *Let $\{x^k\}$ be the sequence generated by the single step projected gradient method* (8.28). *Then, we have*

$$f(x^{k+1}) \leq f(x^k) - \frac{1-\nu}{\beta_k}\|x^k - x^{k+1}\|^2, \tag{8.35}$$

and

$$\begin{aligned} \|x^{k+1} - x^*\|^2 \quad \leq \quad & \|x^k - x^*\|^2 - (1-2\nu)\|x^k - x^{k+1}\|^2 \\ & -2\beta_k(f(x^{k+1}) - f(x^*)). \end{aligned} \tag{8.36}$$

Proof. Setting $x = x^k$ in (8.30) in Lemma 8.1, we obtain the assertion (8.35) immediately. Next, setting $x = x^*$ in (8.30), we have

$$\begin{aligned} & \beta_k(f(x^*) - f(x^{k+1})) \\ & \quad \geq \quad (x^* - x^{k+1})^T (x^k - x^{k+1}) - \nu\|x^k - x^{k+1}\|^2, \end{aligned}$$

and thus

$$
\begin{aligned}
&(x^k - x^*)^T(x^k - x^{k+1}) \\
&\quad \geq \quad (1-\nu)\|x^k - x^{k+1}\|^2 + \beta_k(f(x^{k+1}) - f(x^*)).
\end{aligned}
$$

Using the above inequality, we get

$$
\begin{aligned}
&\|x^{k+1} - x^*\|^2 \\
&\quad = \quad \|(x^k - x^*) - (x^k - x^{k+1})\|^2 \\
&\quad = \quad \|x^k - x^*\|^2 - 2(x^k - x^*)^T(x^k - x^{k+1}) + \|x^k - x^{k+1}\|^2 \\
&\quad \leq \quad \|x^k - x^*\|^2 - 2(1-\nu)\|x^k - x^{k+1}\|^2 \\
&\qquad\qquad -2\beta_k(f(x^{k+1}) - f(x^*)) + \|x^k - x^{k+1}\|^2 \\
&\quad = \quad \|x^k - x^*\|^2 - (1-2\nu)\|x^k - x^{k+1}\|^2 \\
&\qquad\qquad -2\beta_k(f(x^{k+1}) - f(x^*)).
\end{aligned}
$$

This completes the proof of the assertion (8.36). $\square$

Directly from (8.36), it follows the following corollary:

Corollary 8.1 *Let $\{x^k\}$ be the sequence generated by the single step projected gradient method* (8.28)*. If we set $\nu \leq \frac{1}{2}$, then $\|x^{k+1} - x^*\|^2 < \|x^k - x^*\|^2$, for any $x^* \in \Omega^*$. The generated sequence $\{x^k\}$ is in a compact set.*

8.2.3 Convergence rate of the proposed method

Below we show that the iteration-complexity of the single projected gradient method is $O(1/k)$. For the convenience, we assume $\beta_k \equiv \beta$.

Theorem 8.2 *Let $\{x^k\}$ be generated by the single step projected gradient method* (8.28)*. Then, we have*

$$
\begin{aligned}
&2k\beta(f(x^k) - f(x^*)) \\
&\quad \leq \|x^0 - x^*\|^2 - \sum_{l=0}^{k-1}\Big((1-2\nu) + 2l(1-\nu)\Big)\|x^l - x^{l+1}\|^2. \qquad (8.37)
\end{aligned}
$$

Proof. First, it follows from (8.36) that, for any $x^* \in \Omega^*$ and all $l \geq 0$, we have

$$2\beta(f(x^*)-f(x^{l+1})) \geq \|x^{l+1}-x^*\|^2-\|x^l-x^*\|^2+(1-2\nu)\|x^l-x^{l+1}\|^2.$$

Summing the above inequality over $l = 0, \ldots, k-1$, we obtain

$$\begin{aligned} & 2\beta\Big(kf(x^*) - \sum_{l=0}^{k-1} f(x^{l+1})\Big) \\ \geq \quad & \|x^k - x^*\|^2 - \|x^0 - x^*\|^2 + \sum_{l=0}^{k-1}(1-2\nu)\|x^l - x^{l+1}\|^2. \end{aligned} \tag{8.38}$$

It follows from (8.35) that

$$2\beta l\big(f(x^l) - f(x^{l+1})\big) \geq 2l(1-\nu)\|x^l - x^{l+1}\|^2,$$

which can be rewritten as

$$2\beta\big(lf(x^l) - (l+1)f(x^{l+1}) + f(x^{l+1})\big) \geq 2l(1-\nu)\|x^l - x^{l+1}\|^2.$$

Summing the above inequality over $l = 0, \ldots, k-1$, it follows that

$$2\beta \sum_{l=0}^{k-1}\Big(lf(x^l)-(l+1)f(x^{l+1})+f(x^{l+1})\Big) \geq \sum_{l=0}^{k-1} 2l(1-\nu)\|x^l - x^{l+1}\|^2,$$

which simplifies to

$$2\beta\Big(-kf(x^k) + \sum_{l=0}^{k-1} f(x^{l+1})\Big) \geq \sum_{l=0}^{k-1} 2l(1-\nu)\|x^l - x^{l+1}\|^2. \tag{8.39}$$

Adding (8.38) and (8.39), we get

$$\begin{aligned} & 2k\beta\big(f(x^*) - f(x^k)\big) \\ \geq \quad & -\|x^0 - x^*\|^2 + \sum_{l=0}^{k-1}\Big((1-2\nu) + 2l(1-\nu)\Big)\|x^l - x^{l+1}\|^2, \end{aligned}$$

which implies (8.37) and the theorem is proved. $\square$

From (8.37) follows directly the following theorem.

Theorem 8.3 *Let $\{x^k\}$ be generated by the single step projected gradient method. If $\nu \le \frac{1}{2}$, then we have*

$$f(x^k) - f(x^*) \le \frac{\|x^0 - x^*\|^2}{2k\beta}, \tag{8.40}$$

and thus the iteration-complexity of this method is $O(1/k)$.

What is about for any $\nu \in (0.5, 1)$? For such ν, we define

$$p(\nu) = \arg\min\{l \,|\, l \ge 0 \text{ is a integer}, \ (1-2\nu) + 2l(1-\nu) \ge 0\}. \tag{8.41}$$

For any $\nu \in (0.5, 1)$, $p(\nu)$ is finite number. For example,we have

$\nu =$	0.9	0.8	0.7	$(0.5, 0.7)$
$p(\nu) =$	4	2	1	1

.

Since the term $\sum_{p(\nu)}^{k-1}\Big((1-2\nu) + 2l(1-\nu)\Big)\|x^l - x^{l+1}\|^2$ is positive, it

follows from Theorem 8.2 (see(8.37)) that

$$2k\beta(f(x^k) - f(x^*)) \le \|x^0 - x^*\|^2 - \sum_{l=0}^{p(\nu)-1}\Big((1-2\nu)+2l(1-\nu)\Big)\|x^l - x^{l+1}\|^2.$$

The last inequality implies that $\lim_{k\to\infty}(f(x^k) - f(x^*)) = 0$ and the iteration-complexity of this method is $O(1/k)$ for any $\nu \in (0, 1)$.

Theorem 8.4 *Let $\{x^k\}$ be generated by the single step projected gradient method, then we have*

$$f(x^k) - f(x^*) \le \frac{\|x^0 - x^*\|^2 + D}{2k\beta}, \tag{8.42}$$

where

$$D = -\sum_{l=0}^{p(\nu)-1}\Big((1-2\nu) + 2l(1-\nu)\Big)\|x^l - x^{l+1}\|^2.$$

and $p(\nu)$ is a finite integer defined in (8.41).

Self-adaptive projected gradient descent method

Self-adaptive projected gradient descent method.

Set $\beta_0 = 1$, $\mu = 0.5$, $\nu = 0.9$, $x^0 \in \Omega$ and $k = 0$. Provide $g(x^0)$.

For $k = 0, 1, \ldots$, if the stopping criterium is not satisfied, do

Step 1. $\tilde{x}^k = P_\Omega[x^k - \beta_k g(x^k)]$,

$r_k = \beta_k \|g(x^k) - g(\tilde{x}^k)\| / \|x^k - \tilde{x}^k\|$.

while $r_k > \nu$

$\boxed{\beta_k := \beta_k * 0.8/r_k,}$

$\tilde{x}^k = P_\Omega[x^k - \beta_k g(x^k)]$,

$r_k = \beta_k \|g(x^k) - g(\tilde{x}^k)\| / \|x^k - \tilde{x}^k\|$.

end(while)

$x^{k+1} = \tilde{x}^k$,

$g(x^{k+1}) = g(\tilde{x}^k)$.

$\boxed{\textbf{If} \quad r_k \le \mu \quad \text{then} \quad \beta_k := \beta_k * 1.5, \quad \textbf{end(if)}}$

Step 2. $\beta_{k+1} = \beta_k$ and $k = k + 1$, go to Step 1.

Remark 8.1 *Instead of the condition* (8.28b)*, here we have*

$$\beta_k \|g(x^k) - g(x^{k+1})\| \le \nu \|x^k - x^{k+1}\|.$$

Remark 8.2 *If $r_k \le \nu$, we direct take $x^{k+1} = \tilde{x}^k$, and $g(x^{k+1}) = g(\tilde{x}^k)$ for the next iteration. We call the method Self-adaptive single step projected gradient method because it needs only once evaluation of the gradient $g(x^k)$ in each iteration when adjusting the parameter β_k is not necessary.*

Remark 8.3 *If $r_k > \nu$, we adjust the parameter β_k by $\beta_k := \beta_k * 0.8/r_k$. According to our limited numerical experiments, using the reduced β_k, the condition $r_k \le \nu$ is satisfied.*

Remark 8.4 *Too small step size β_k will leads to slow convergence. If $r_k \le \mu$, we will enlarge the trial step size β for the by $\beta_k := \beta_k * 1.5$.*

9 Conclusions and Remarks

9.1 ADMM vs AMA

The mathematical form of the linearly constrained convex optimization problem

$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{9.1}$$

The penalty function method (PFM) for solving the problem (9.1)

$$u^{k+1} = \text{Argmin}\big\{\theta(u) + \tfrac{\beta_k}{2}\|\mathcal{A}u - b\|^2 \big| u \in \mathcal{U}\big\}$$

Augmented Lagrangian Method (ALM) for solving the problem (9.1)

λ^k is given

$$u^{k+1} = \text{Argmin}\big\{\theta(u) - (\lambda^k)^T(\mathcal{A}u - b) + \tfrac{\beta}{2}\|\mathcal{A}u - b\|^2 \big| x \in \mathcal{X}\big\},$$

$$\lambda^{k+1} = \lambda^k - \beta(\mathcal{A}u^{k+1} - b).$$

Our objective is to solve the following separable convex optimization problem

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X},\ y \in \mathcal{Y}\} \tag{9.2}$$

Applied the penalty function method for solving Problem (9.2)

$$(x^{k+1}, y^{k+1}) = \text{Argmin}\big\{\theta_1(x) + \theta_2(y) + \frac{\beta_k}{2}\|Ax + By - b\|^2 \big| x \in \mathcal{X},\ y \in \mathcal{Y}\big\}$$

Applied ALM to solve the problem (9.2)

The k-th iteration begins with λ^k

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

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

Two kinds of different methods. The difficulties of their subproblems are equal.

It is well known: ALM is much efficient than the penalty function method.

See J. Nocedal and S. J. Wright: Numerical Optimization

The common disadvantage of both methods when they applied to solve (9.2):

✠ The methods do not use the separable structure of the problem (9.2).

Either ALM or the penalty function methods, their subproblem involves the both variables x and y. Sometimes we have no way to solve such subproblems.

Relaxed versions: Relax one of the variables as the known vector

- Relaxed penalty function method (PFM) for the problem (9.2)

  — **Alternating Minimization Algorithm (AMA)**.

- Relaxed augmented Lagrangian method (ALM) for the problem (9.2)

  — **Alternating Direction Method of Multipliers (ADMM)**.

Applying Alternating Minimization Algorithm (AMA) for the problem (9.2)

The k-th iteration begins with given y^k,

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

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

Applying Alternating Direction Method of Multipliers (ADMM) for (9.2)

The k-th iteration begins with given (y^k, λ^k),

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

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

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

ALM is better than PFM $\Rightarrow$ Their relaxed versions ADMM is better than AMA

9.2 Developments of ADMM for two-block problems

1. ADMM in sense of PPA The k-th iteration begins with given (y^k, λ^k)

$$
\left\{
\begin{array}{l}
\left\{
\begin{array}{lcl}
x^{k+1} & = & \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\
\lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^k - b), \\
y^{k+1} & = & \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+1}) \,|\, y \in \mathcal{Y}\},
\end{array}
\right. \qquad (9.3a) \\
\left\{
\begin{array}{lcl}
y^{k+1} & := & y^k - \gamma(y^k - y^{k+1}), \\
\lambda^{k+1} & := & \lambda^k - \gamma(\lambda^k - \lambda^{k+1}).
\end{array}
\right. \quad \text{(Extended)} \qquad (9.3b)
\end{array}
\right.
$$

The formula (9.3a) is obtained by changing the order of y and λ in the classical ADMM. The notation " $:=$ " in (9.3b) means that the (y^{k+1}, λ^{k+1}) in the right hand side of (9.3b) is the output of (9.3a). $\gamma \in [1, 2)$ is the extended factor.

- 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 Math, 56 (2013), 2179-2186.

2. Symmetric ADMM

The primal variables x and y are essentially equal, so it is recommended to adopt the symmetrical alternating direction method of multipliers.

Symmetric Alternating Direction Method of Multipliers

The k-th iteration begins with given (y^k, λ^k),

$$
\left\{
\begin{array}{l}
x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\[2mm]
\lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), \\[2mm]
y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\[2mm]
\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b),
\end{array}
\right. \qquad (9.4)
$$

wehre $\mu \in (0, 1)$ (usually $\mu = 0.9$).

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

9.3 Multi-block separable convex optimization

We take the three-block separable convex optimization problem as an example

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz=b,\ x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}\}. \quad (9.5)$$

Its augmented Lagrangian function is

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

$$\left\{\begin{array}{lcl} x^{k+1} & = & \arg\min\big\{\mathcal{L}_\beta^3(x,y^k,z^k,\lambda^k)\ \big|\ x\in\mathcal{X}\big\},\\ y^{k+1} & = & \arg\min\big\{\mathcal{L}_\beta^3(x^{k+1},y,z^k,\lambda^k)\ \big|\ y\in\mathcal{Y}\big\},\\ z^{k+1} & = & \arg\min\big\{\mathcal{L}_\beta^3(x^{k+1},y^{k+1},z,\lambda^k)\ \big|\ z\in\mathcal{Z}\big\},\\ \lambda^{k+1} & = & \lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b). \end{array}\right. \quad (9.6)$$

The above formula is the direct extension of ADMM for the three-block separable convex optimization problem (9.5). Unfortunately, it is not necessarily convergent [6].

- 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, 155 (2016) 57-79.

Direct extension of ADMM. The main example in our Math. Prog. Paper [6]:

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

where $\theta_1(x)=\theta_2(y)=\theta_3(z)=0, \mathcal{X}=\mathcal{Y}=\mathcal{Z}=\Re, \quad b=0\in\Re^3.$

$$[A,B,C]=\begin{bmatrix}1&1&1\\1&1&2\\1&2&2\end{bmatrix}.$$

Applied the direct extension of ADMM (9.6) to this example, the method is not convergent. However, this exmple have only theoretical meaning.

It is worth to consider a class of three-block problem whose constrained matrix

$$[A,B,C]=[A,B,I]$$ one of the submatrix is identity.

It is convergent when the direct extension of ADMM is applied to solve such more practical problems ? It is a challenging open problem !

Neither convergence nor counterexamples have been provided ! !

It is valuable to study a class of the following problems :

- Applying ADMM to the problem

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

- Change the equality to inequality, the considered problem becomes

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

- Reconvert it to a equality constrained special three-block problem:

$$\min\{\theta_1(x) + \theta_2(y) + 0 \,|\, Ax + By + z = b,\ x \in \mathcal{X}, y \in \mathcal{Y}, z \ge 0\}$$

- Peoples have tried to solve the above problems with direct extended ADMM, but so far neither convergence nor counterexamples have been proved.

Based on the above recognition, we propose some modified algorithms for the three operator problem. Our method does not add any conditions to the problem! No restrictions on Operate only on the method itself !

ADMM Like-Method I : Partial parallel Splitting ALM with reduced step-size

Begin with (y^k, z^k, λ^k), after solving the x-subproblem, we solve the y and z-subproblems in parallel.

$$\left\{\begin{array}{l} x^{k+\frac{1}{2}} = \arg\min\big\{\mathcal{L}^3_\beta(x, y^k, z^k, \lambda^k) \;\big|\; x \in \mathcal{X}\big\}, \\ y^{k+\frac{1}{2}} = \arg\min\big\{\mathcal{L}^3_\beta(x^{k+\frac{1}{2}}, y, z^k, \lambda^k) \big| y \in \mathcal{Y}\big\}, \\ z^{k+\frac{1}{2}} = \arg\min\big\{\mathcal{L}^3_\beta(x^{k+\frac{1}{2}}, y^k, z, \lambda^k) \big| z \in \mathcal{Z}\big\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+\frac{1}{2}} + By^{k+\frac{1}{2}} + Cz^{k+\frac{1}{2}} - b). \end{array}\right. \tag{9.7}$$

The output $(y^{k+\frac{1}{2}}, z^{k+\frac{1}{2}}, \lambda^{k+\frac{1}{2}})$ is a predictor, the new iterate is given by

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ z^k \\ \lambda^k \end{pmatrix} - \alpha \begin{pmatrix} y^k - y^{k+\frac{1}{2}} \\ z^k - z^{k+\frac{1}{2}} \\ \lambda^k - \lambda^{k+\frac{1}{2}} \end{pmatrix}, \quad \alpha \in (0, 2 - \sqrt{2}).$$

- B. S. He, Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities, *COA* **42**(2009), 195–212.

It is too free to deal with problems, thus, reducing the step length is necessary !

ADMM Like-Method II : ADMM with Gaussian back substitution

Taking the output of (9.6) as the predict point. We need only to correct the (y,z)-parts by

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} := \begin{pmatrix} y^{k} \\ z^{k} \end{pmatrix} - \alpha \begin{pmatrix} I & -(B^TB)^{-1}B^TC \\ 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}.$$

where $\alpha \in (0,1)$. Because we just need to provide $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$ for the next iteration,

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \end{pmatrix} := \begin{pmatrix} By^{k} \\ Cz^{k} \end{pmatrix} - \alpha \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} B(y^k - y^{k+1}) \\ C(z^k - z^{k+1}) \end{pmatrix}. \tag{9.8}$$

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

There is priority or unfairness in (9.6) for the essential primal variables y and z (resp. By and Cz). Thus, it is necessary to make up some adjustment !

ADMM Like-Method III : ADMM+Prox-Parallel Splitting ALM

After solving the x-subproblem, we solve the y and z-subproblems in parallel. Since we don't want to do post-processing (correction), adding a regular term to both of the y and z-subproblems in advance is necessary

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}^3_\beta(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} = \arg\min\{\mathcal{L}^3_\beta(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\beta\|B(y-y^k)\|^2 | y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}^3_\beta(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\beta\|C(z-z^k)\|^2 | z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \qquad (\tau \geq 1) \end{cases}$$

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

If you are too free and don't correct, adding the regular terms is necessary !
This can be explained as: people should not forget what they promised yesterday!

求解线性约束凸优化，增广拉格朗日乘子法 (ALM) 优于罚函数方法，有点优化基础知识的人都知道。

对两个可分离算子的线性约束凸优化问题，增广拉格朗日乘子法 (ALM) 和罚函数方法, 松弛后分别成了乘子交替方向法 (ADMM) 和交替极小化方法 (AMA)。

人们因此有理由对 ADMM 格外关心。

ADMM 不是我们提出来的。有了 10 年投影收缩算法的研究的基础, 使得我对 ADMM 类方法格外感兴趣。带领学生对ADMM 方法做一些有价值的改进和证明一些重要的理论结果，便顺理成章。

方法上，交换了原始变量 y 和对偶变量 λ 次序，进而得到因需定制的 PPA 意义下的 ADMM (Science in China, Mathematics, 2013)；

平等对待原始变量 x 和 y，两次校正对偶变量 λ，就得到对称型的 ADMM (SIAM Optimization, 2014)。

这些方法，道理上能站住脚，计算表现也不俗。

理论上，我们证明了 ADMM 在遍历意义下 (SIAM Numerical Analysis， 2012) 和点列意义下(Numer. Mathematik, 2015) 的 $O(1/t)$ 的收敛速率. 证明都不复杂.

ADMM 的广泛应用，人们自然想到向三个算子的问题推广。

我们在不能证明“直接推广的方法”收敛的时候，提出了一些处理多个算子问题的ADMM 类方法

(Computational Optimization and Applications, 2009).

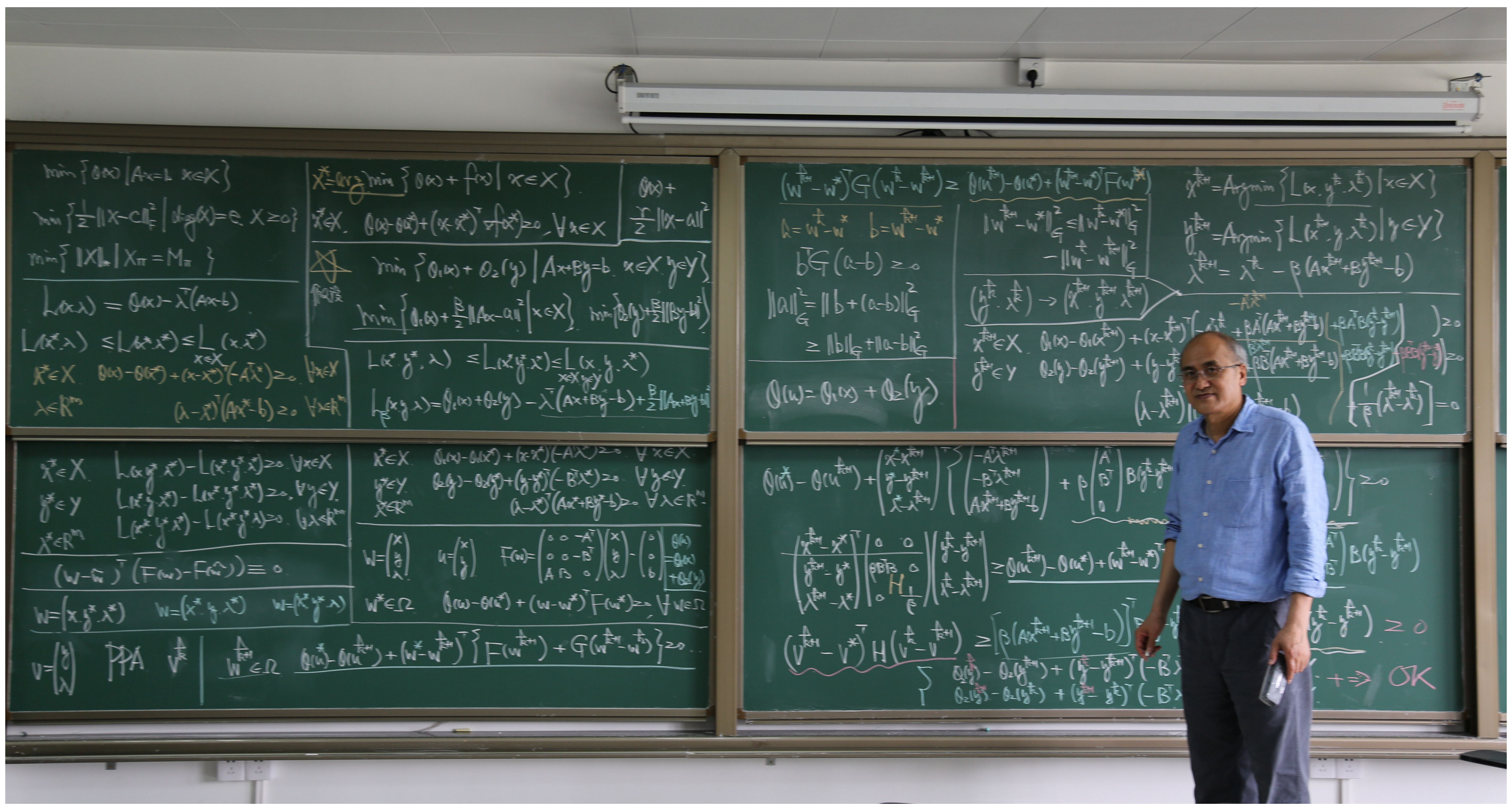
(SIAM Optimization, 2012; IMA Numerical Analysis, 2015).

这些方法的共同特点是不需要对问题加任何条件！对 β 不加限制, 只对方法动手术！

后来我们又给出“直接推广的ADMM 方法处理三个算子问题不保证收敛”的例子(Math. Progr., 2016),说明:

以前提出的一些策略，手段上是必须的，机制上也是合理的。

- The analysis is guided by variational inequality.
- The most methods mentioned fall in a unified prediction-correction framework, in which the convergence analysis is quite simple.
- All the discussed methods are closed related to Proximal Point Algorithms.
- All the discussed ADMM-like splitting methods are rooted from Augmented Lagrangian Method.
- A through reading will acquaint you with the ADMM, while a more carefully reading may make you familiar with the tricks on constructing splitting methods according to the problem you met.
- The discussed first order splitting contraction methods are only appropriate for some structured convex optimization in some practical applications.



✠ VI is a powerful tool to analyze the splitting contraction algorithm of convex optimization It is very simple to prove the convergence in the framework of variational inequality.

✠ This picture shows that we can use a small space on the blackboard The origin and development of the alternating direction method of multipliers and the proof of convergence are all clear

感想：数学之美, 不是纯数学的专利。为应用服务的最优化方法研究, 同样可以追求简单与统一。简单, 他人才会看懂使用; 统一, 自己才有美的享受。

✠ 追求简单与统一, 是我研究工作欲罢不能的原因 ✠

授人以鱼不如授人以渔。一个好的优化方法, 应该是容易被工程师们掌握, 让人用来自己解决问题的方法。

✠ 为相关学科所用, 恰是我们从事优化方法研究的本源追求 ✠

✠ 感谢导师的培养, 朋友的支持, 还有经常提出问题一起讨论的学生 ✠

✠ 希望各位能以怀疑的态度审视我的观点, 对的就相信, 不对的就批评 ✠

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

[3] 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.

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

[5] C. H. Chen, B. S. He and X. M. Yuan, Matrix completion via alternating direction method, *IMA Journal of Numerical Analysis* **32**(2012), 227-245.

[6] 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*. 155 (2016) 57-79.

[7] J. Douglas and H. H. Rachford, On the numerical solution of the heat conduction problem in 2 and 3 space variables, *Transactions of the American Mathematical Society* **82** (1956), 421–439.

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

[9] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity problems, Volume I*, Springer Series in Operations Research, Springer-Verlag, 2003.

[10] 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.

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

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

[13] B.S. He, *A new method for a class of linear variational inequalities*, Math. Progr., **66**, pp. 137-144, 1994.

[14] B.S. He, *Solving a class of linear projection equations*, Numer. Math., **68**, pp. 71-80, 1994.

[15] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, 69-76, 1997.

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

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

[18] B. S. He, L. Z. Liao, D. Han, and H. Yang, A new inexact alternating directions method for monontone variational inequalities, *Mathematical Programming* **92**(2002), 103–118.

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

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

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

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

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

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

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

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

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

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

[29] Z. Lin, R. Liu and H. Li, *Linearized alternating direction method with parallel splitting and adaptive*

penalty for separable convex programs in machine learning, Machine Learning, 99 (2) (2015), pp. 287-325.

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

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

[32] 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.

[33] X. F. Wang and X. M. Yuan, *The linearized alternating direction method of multipliers for Dantzig selector*, SIAM J. Sci. Comput., 34 (2012), pp. A2792–A2811.

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

[35] P. Tseng, On accelerated proximal gradient methods for convex-concave optimization, manuscript, 2008.

[36] J. F. Yang and X. M. Yuan, *Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization*, Math. Comp., 82 (2013), pp. 301-329.

[37] X. Q. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2010), pp. 20-46.

[38] M. Zhu and T. F. Chan, An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration, CAM Report 08-34, UCLA, Los Angeles, CA, 2008.

Appendix: From PC Methods for VI to SC Methods for Convex Optimization

A1. Projection and contraction method for monotone variational inequality

Let $\Omega \subset \Re^n$ be a nonempty closed convex set, $F: \Re^n \to \Re^n$ be a mapping. Consider the following variational inequality:

$$u^* \in \Omega, \quad (u - u^*)^\top F(u^*) \geq 0, \quad \forall\, u \in \Omega. \tag{A1.1}$$

We say a variational inequality is monotone, when

$$(u - v)^\top (F(u) - F(v)) \geq 0.$$

For solving the problem (A1.1), the k-th iteration of the projection and contraction begins with a given u^k, produces a predictor $\tilde{u}^k$ via the projection

$$\begin{aligned} \tilde{u}^k &= P_\Omega[u^k - \beta_k F(u^k)] \\ &= \arg\min\{\tfrac{1}{2}\|u - [u^k - \beta_k F(u^k)]\|^2 \mid u \in \Omega\}. \end{aligned} \tag{A1.2}$$

In the projection (A1.2), the chosen parameter β_k satisfies

$$\beta_k \|F(u^k) - F(\tilde{u}^k)\| \leq \nu \|u^k - \tilde{u}^k\|, \quad \nu \in (0,1). \tag{A1.3}$$

Because $\tilde{u}^k = \arg\min\{\frac{1}{2}\|u - [u^k - \beta_k F(u^k)]\|^2 \mid u \in \Omega\}$, according to Lemma 1.1, we have

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^\top \{\tilde{u}^k - [u^k - \beta_k F(u^k)]\} \geq 0, \quad \forall u \in \Omega. \tag{A1.4}$$

Let $\quad d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k[F(u^k) - F(\tilde{u}^k)]. \quad$ (A1.5)

Adding the term $(u - \tilde{u}^k)^\top d(u^k, \tilde{u}^k)$ to the both sides of (A1.4), we get the $\tilde{u}^k$ based prediction formula

Prediction:

$$\tilde{u}^k \in \Omega, \; (u - \tilde{u}^k)^\top \beta_k F(\tilde{u}^k) \geq (u - \tilde{u}^k)^\top d(u^k, \tilde{u}^k), \; \forall u \in \Omega. \tag{A1.6}$$

Setting $u = u^*$ in (A1.6), we get

$$(\tilde{u}^k - u^*)^\top d(u^k, \tilde{u}^k) \geq \beta_k (\tilde{u}^k - u^*)^\top F(\tilde{u}^k). \tag{A1.7}$$

Due to the monotonicity of F, $(\tilde{u}^k - u^*)^\top F(\tilde{u}^k) \geq (\tilde{u}^k - u^*)^\top F(u^*)$. Since $\tilde{u}^k \in \Omega$, according to (A1.1), $(\tilde{u}^k - u^*)^\top F(u^*) \geq 0$. Thus, the right hand side of (A1.7) is nonnegative. Consequently, we have

$$(u^k - u^*)^\top d(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k). \qquad \text{(A1.8)}$$

According to the expression of $d(u^k, \tilde{u}^k)$ (A1.5) and the assumption (A1.3), by using the Cauchy-Schwarz inequality, we get

$$(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) \geq (1 - \nu)\|u^k - \tilde{u}^k\|^2. \qquad \text{(A1.9)}$$

Therefore, the right hand side of the inequality (A1.8) is positive. This means, for any positive definite matrix $H \in \Re^{n\times n}$, $H^{-1}d(u^k, \tilde{u}^k)$ is an ascent direction of the unknown distance function $\frac{1}{2}\|u - u^*\|_H^2$ at u^k. By using

Coirrection $$u_\alpha^{k+1} = u^k - \alpha H^{-1} d(u^k, \tilde{u}^k), \qquad \text{(A1.10)}$$

we get a new iterate which is more closed to the solution set in H-norm, where $d(u^k, \tilde{u}^k)$ is given by (A1.5). Consider the α-dependent profit

$$\vartheta_k(\alpha) := \|u^k - u^*\|_H^2 - \|u_\alpha^{k+1} - u^*\|_H^2. \qquad \text{(A1.11)}$$

According to (A1.10) and (A1.8), we get

$$\begin{aligned}
\vartheta_k(\alpha) &= \|u^k - u^*\|_H^2 - \|u^k - u^* - \alpha H^{-1} d(u^k, \tilde{u}^k)\|_H^2 \\
&= 2\alpha(u^k - u^*)^\top d(u^k, \tilde{u}^k) - \alpha^2 \|H^{-1} d(u^k, \tilde{u}^k)\|_H^2 \\
&\overset{\text{(A1.8)}}{\geq} 2\alpha(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) - \alpha^2 \|H^{-1} d(u^k, \tilde{u}^k)\|_H^2 \\
&=: q_k(\alpha).
\end{aligned} \qquad \text{(A1.12)}$$

The last inequality tells us that $q_k(\alpha)$ is a low bound of $\vartheta_k(\alpha)$.

Usually, we consider the projection and contraction method in the Euclidean-norm. In this case, $H = I$ and $q(\alpha)$ reaches its maximum at

$$\alpha_k^* = \text{argmax}\{q_k(\alpha)\} = \frac{(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}. \qquad \text{(A1.13)}$$

From the assumption (A1.3), we get $2(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2 > 0$ and thus $\alpha_k^* > \frac{1}{2}$.

$$\begin{aligned}
&\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2 \geq q(\alpha_k^*) \\
&= \alpha_k^*(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) \overset{\text{(A1.9)}}{\geq} \tfrac{1}{2}(1 - \nu)\|u^k - \tilde{u}^k\|^2.
\end{aligned}$$

In this way, we get the following key-inequality for convergence proof of the PC method:

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \tfrac{1}{2}(1-\nu)\|u^k - \tilde{u}^k\|^2. \qquad \text{(A1.14)}$$

✠ In PC methods, the directions in the left and right hand sides of (A1.6)

$\beta_k F(\tilde{u}^k)$ and $d(u^k, \tilde{u}^k)$ are called a pair of twin directions.

It is very interesting that two different correction methods with twin directions and same step length have the common contraction inequality for convergence.

In practical computation, we use correction formula

$$\text{(PC-Method I)} \quad u_I^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k) \qquad \text{(A1.15)}$$

or

$$\text{(PC Method II)} \quad u_{II}^{k+1} = P_\Omega[u^k - \gamma\alpha_k^*\beta_k F(\tilde{u}^k)] \qquad \text{(A1.16)}$$

to offer the new iterate u^{k+1}, where $\gamma \in (1.5, 1.8) \subset (0, 2)$, α_k^* is given by(A1.13).

✠ The detailed proof can be found in the series of Lecture 3 on my Homepage ✠

A2. Splitting-contraction (SC) methods for linearly constrained convex optimization

The linearly constrained separable convex optimization problems, as illustrated in Section 1.2, can be translated to the following variational inequality

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^\top F(w^*) \ge 0, \quad \forall\, w \in \Omega. \qquad \text{(A2.1)}$$

In (A2.1), the function $\theta(u)$ is convex and the mapping $F(w)$ is also monotone, specially,

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

For solving such variational inequality (A2.1), we have a unified framework of the prediction-correction methods.

Prediction. For given v^k (v involves some (or total) elements of the vector w), produce a $\tilde{w}^k \in \Omega$ which satisfies

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \ge (v - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \quad \text{(A2.2)}$$

where Q is not necessarily symmetric, but $Q^\top + Q$ is essentially positive definite.

The inequality (A2.2) is similar as (A1.6). Set $w = w^*$ in (A2.2), we get

$$(\tilde{v}^k - v^*)^\top Q(v^k - \tilde{u}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^\top F(\tilde{w}^k). \quad \text{(A2.3)}$$

Because $(\tilde{w}^k - w^*)^\top \{F(\tilde{w}^k) - F(w^*)\} = 0$, we have

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^\top F(\tilde{w}^k) = \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^\top F(w^*).$$

Since $\tilde{w}^k \in \Omega$, it follows from (A2.1) that

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

Thus, the right hand side of (A2.3) is nonnegative. Consequently, we get

$$\boxed{(v^k - v^*)^\top Q(v^k - \tilde{v}^k) \geq (v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k). \quad \text{(A2.4)}}$$

The inequality (A2.4) is similar as (A1.8). Since $Q^\top + Q$ is essential positive definite, the right hand side of (A2.4) is positive (otherwise, $\tilde{w}^k$ is a solution).

Thus, for any positive definite matrix H, $H^{-1}Q(v^k - \tilde{v}^k)$ is an ascent direction of the unknown distance function $\frac{1}{2}\|v - v^*\|_H^2$ at v^k. By using

$$v^{k+1} = v^k - \alpha H^{-1}Q(v^k - \tilde{v}^k), \quad \text{(A2.5)}$$

we get a new iterate which is more closed to $\mathcal{V}^*$ in H-norm, where $Q(v^k - \tilde{u}^k)$ is given in the right hand side of (A2.2). Let

$$M = H^{-1}Q, \quad \text{(A2.6)}$$

the correction formula becomes

$$\boxed{\textbf{Correction} \qquad v_\alpha^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \quad \text{(A2.7)}}$$

Consider the α-dependent profit

$$\vartheta_k(\alpha) := \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2. \quad \text{(A2.8)}$$

Thus, we have

$$\begin{array}{rcl} \vartheta_k(\alpha) & = & \|v^k - v^*\|_H^2 - \|v^k - v^* - \alpha M(v^k - \tilde{v}^k)\|_H^2 \\ & = & 2\alpha(v^k - v^*)^\top Q(v^k - \tilde{v}^k) - \alpha^2\|M(v^k - \tilde{v}^k)\|_H^2 \\ & \overset{\text{(A2.4)}}{\geq} & 2\alpha(v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k) - \alpha^2\|M(v^k - \tilde{v}^k)\|_H^2 \\ & =: & q_k(\alpha) \end{array} \quad \text{(A2.9)}$$

Now, $q_k(\alpha)$ is a low bound of $\vartheta_k(\alpha)$ and it reaches its maximum at

$$\alpha_k^* = \text{argmax}\{q_k(\alpha)\} = \frac{(v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k)}{\|M(v^k - \tilde{v}^k)\|_H^2}. \tag{A2.10}$$

Moreover, if the matrices satisfy

$$G = Q^\top + Q - M^\top HM \succ 0, \tag{A2.11}$$

it is easy to see that

$$\begin{aligned} & 2(v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k) \\ & \quad = \quad (v^k - \tilde{v}^k)^\top (Q^\top + Q)(v^k - \tilde{v}^k) \\ & \quad > \quad \|M(v^k - \tilde{v}^k)\|_H^2. \end{aligned} \tag{A2.12}$$

Together with (A2.10), we have $\alpha_k^* > \frac{1}{2}$. Consequently, it follows that

$$\begin{aligned} & \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ & \quad \geq \quad q(\alpha_k^*) \;=\; \alpha_k^*(v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k) \\ & \quad > \quad \frac{1}{4}\|v^k - \tilde{v}^k\|_{(Q^T+Q)}^2. \end{aligned} \tag{A2.13}$$

In other words, if the assumption (A2.11) is satisfied, we can take the unit step size $\alpha = 1$ in (4.4b). In other words, the correction formula becomes

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

It follows from (A2.9) that $q(1) = \|v^k - \tilde{v}^k\|_G^2$. Thus, the generated sequence $\{v^k\}$ has the contraction property

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

which is the key-inequality for convergence analysis of the proposed methods.

From the projection and contraction method for monotone variational inequality, to the splitting-contraction methods for linearly constrained convex optimization, it obeys

✠ one main line, a common model. ✠

所谓数学家, 就是能够发现定理之间相似之处的人。

好的数学家能够看到证明之间的相似。

最优秀的数学家则能够看到理论之间的相似。

可以想象, 终极的数学家应该能够看到相似之间的相似。

— Stefan Banach, 1892–1945

Thank you very much for reading ！