变分不等式框架下结构型凸优化的分裂收缩算法

II 单块线性约束凸优化问题的PPA算法和均困的增广拉格朗日乘子法

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

何 炳 生　　南京大学数学系

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

天元数学东北中心　2023 年 10 月 17 – 27 日

1 Preliminaliers

定理 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,*

$$x^* \in \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{1.1a}$$

if and only if

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{1.1b}$$

引理 1 *Let the vectors* $a, b \in \Re^n$*,* $H \in \Re^{n\times n}$ *be a positive definite matrix. If* $b^T H(a - b) \geq 0$*, then we have*

$$\|b\|_H^2 \leq \|a\|_H^2 - \|a - b\|_H^2. \tag{1.2}$$

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

$\|x\| = (x^T x)^{\frac{1}{2}}$. H is positive definite, $\|x\|_H = (x^T H x)^{\frac{1}{2}}$

The optimal condition of the linearly constrained convex optimization

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

is characterized as a special mixed monotone variational inequality:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \qquad (1.3)$$

PPA with Relaxation for VI (1.3)

For given v^k and $H \succ 0$, find w^{k+1},

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) & \\ \geq (v - v^{k+1})^T H(v^k - v^{k+1}), \ \ \forall\, w \in \Omega. & \qquad (1.4) \end{aligned}$$

Relaxation: ($v = w$ or v is a sub-vector of w)

$$v^{k+1} := v^k - \alpha(v^k - v^{k+1}), \quad \alpha \in (0, 2). \qquad (1.5)$$

2 从原始-对偶混合梯度法到按需定制的邻近点算法

We consider the $\min - \max$ problem (*e. g.* 图像处理中的 ROF Model [3, 16])

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

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

$$\begin{cases} x^* \in \mathcal{X}, \;\; \Phi(x, y^*) - \Phi(x^*, y^*) \geq 0, \;\; \forall\, x \in \mathcal{X}, & \text{(2.2a)} \\ y^* \in \mathcal{Y}, \;\; \Phi(x^*, y^*) - \Phi(x^*, y) \geq 0, \;\; \forall\, y \in \mathcal{Y}. & \text{(2.2b)} \end{cases}$$

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

$$\begin{cases} x^* \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T y^*) \geq 0, \quad \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^*) \; + \; (y - y^*)^T(\; Ax^*) \;\; \geq 0, \quad \forall\, y \in \mathcal{Y}. \end{cases}$$

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

$$u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u - u^*)^T F(u^*) \geq 0, \; \forall\, u \in \Omega, \tag{2.3}$$

where

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

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

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

For the convex optimization problem $\min\{\theta(x) \mid Ax = b, x \in \mathcal{X}\}$,
whose Lagrangian function is $L(x, y) = \theta(x) - y^T (Ax - b)$, we can rewrite it as

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

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 $\theta_1(x) = \theta(x), \quad \theta_2(y) = -b^T y$ and $\mathcal{Y} = \Re^m$.

2.1 求解鞍点问题的 原始-对偶混合梯度法 PDHG [18]

For given (x^k, y^k), PDHG [18] 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

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

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

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

这里有人会说, 如果 (2.5a) 中的 $\theta_1(x)$ 是可微函数, 我们能得到 (2.6) 吗？ 能！

When $\theta_1(x)$ is differentiable, the optimal condition of (2.5a) is: $x^{k+1} \in \mathcal{X}$ and

$$(x - x^{k+1})^T \{\nabla\theta_1(x^{k+1}) - A^T y^k + r(x^{k+1} - x^k)\} \geq 0, \ \forall x \in \mathcal{X}.$$

We rewrite the above VI as $x^{k+1} \in \mathcal{X}$ and

$$\begin{aligned} &\nabla\theta_1(x^{k+1})^T(x - x^{k+1}) \\ &\quad + (x - x^{k+1})^T\{-A^T y^k + r(x^{k+1} - x^k)\} \geq 0, \quad \forall x \in \mathcal{X} \end{aligned} \tag{2.7}$$

Since $\theta_1(x)$ is convex function, we have

$$\theta_1(x) - \theta_1(x^{k+1}) \geq \nabla\theta_1(x^{k+1})^T(x - x^{k+1}).$$

Substituting it in (2.7), we get (2.6). $\square$

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

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{Ax^{k+1} + s(y^{k+1} - y^k)\} \geq 0, \ \forall y \in \mathcal{Y}. \tag{2.8}$$

Combining (2.6) and (2.8), we have $(x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$,

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

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

$$u^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T F(u^{k+1}) \ge (u - u^{k+1})^T Q(u^k - u^{k+1}), \ \ \forall\, u \in \Omega. \tag{2.9}$$

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 (1.4), 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 $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Note that its Lagrange function is

$$L(x, y) = c^T x - y^T(Ax - b) \tag{2.10}$$

which defined on $\Re^2_+ \times \Re$. $x^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $y^* = 1$. is the unique saddle point of the Lagrange function.

For solving the $\min$-$\max$ problem (2.10), by using (2.4), the iterative formula is

$$\begin{cases} x^{k+1} = \arg\min\{c^T x - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 | x \geq 0\} \\ \qquad\;\; = \arg\min\{\frac{r}{2}\|x - [x^k + \frac{1}{r}(A^T y^k - c)]\|^2 | x \geq 0\} \\ \qquad\;\; = P_{\Re^n_+}[x^k + \frac{1}{r}(A^T y^k - c)] \\ \qquad\;\; = \max\{[x^k + \frac{1}{r}(A^T y^k - c)], 0\}, \\ y^{k+1} = y^k - \frac{1}{s}(Ax^{k+1} - b). \end{cases}$$

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.

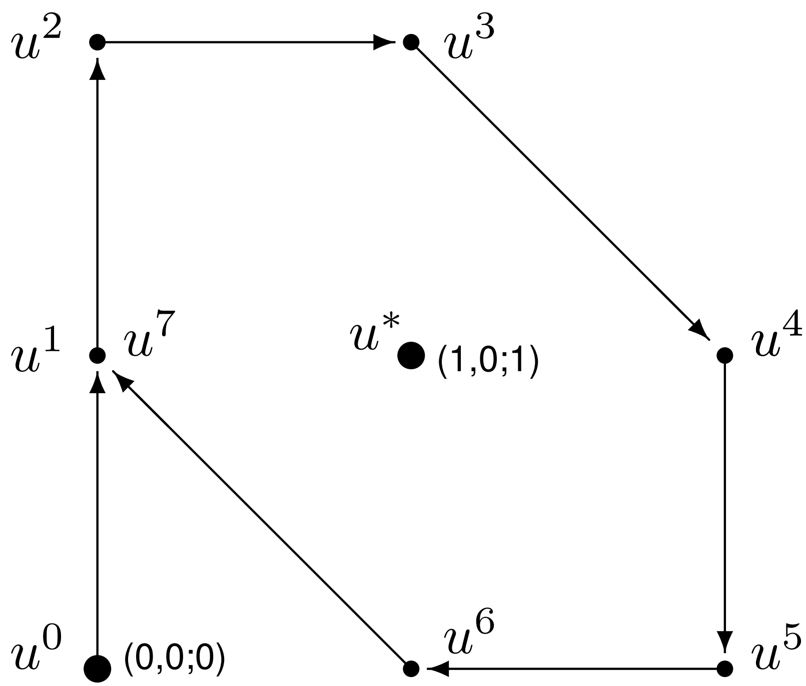
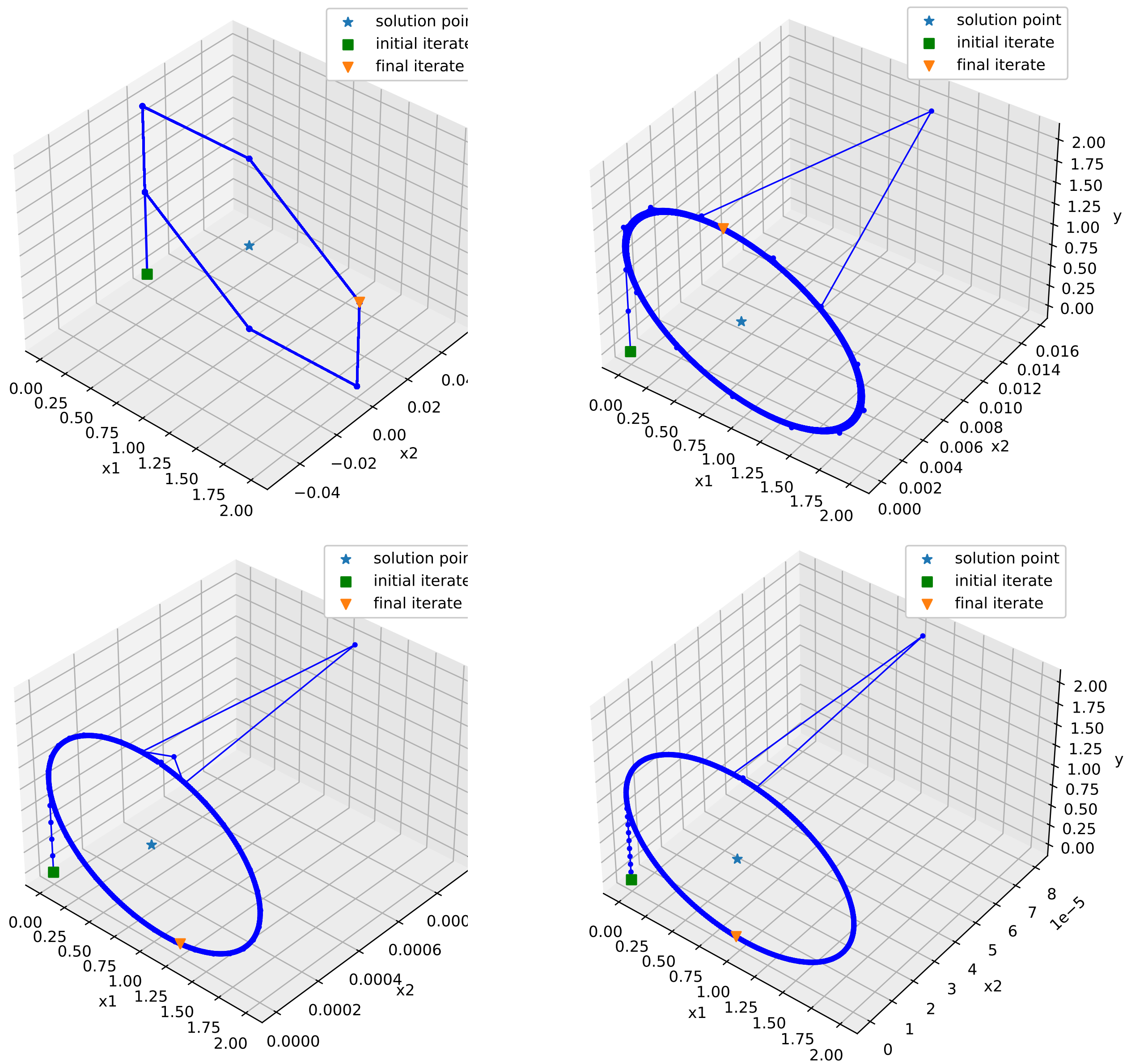


Fig. 2.1 The sequence generated by PDHG Method with $r = s = 1$

$$u^0 = (0, 0; 0)$$
$$u^1 = (0, 0; 1)$$
$$u^2 = (0, 0; 2)$$
$$u^3 = (1, 0; 2)$$
$$u^4 = (2, 0; 1)$$
$$u^5 = (2, 0; 0)$$
$$u^6 = (1, 0; 0)$$
$$u^7 = (0, 0; 1)$$

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



对 $r = s = 1, 2, 5, 10$, PDHG 方法都不收敛

2.2 Customized Proximal Point Algorithm-Classical Version

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 (2.9) 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 (2.8) 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

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

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{A[2x^{k+1} - x^k] + s(y^{k+1} - y^k)\} \geq 0. \quad (2.11)$$

Thus, for given (x^k, y^k), producing a proximal point (x^{k+1}, y^{k+1}) via (2.4a) and (2.11) can be summarized as:

$$x^{k+1} = \text{argmin}\Big\{\Phi(x, y^k) + \frac{r}{2}\big\|x - x^k\big\|^2 \,\Big|\, x \in \mathcal{X}\Big\}. \tag{2.12a}$$

$$y^{k+1} = \text{argmax}\Big\{\Phi\big([2x^{k+1} - x^k], y\big) - \frac{s}{2}\big\|y - y^k\big\|^2\Big\} \tag{2.12b}$$

By ignoring the constant term in the objective function, getting x^{k+1} from (2.12a) is equivalent to obtaining x^{k+1} from

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

The solution of (2.12b) is given by

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

According to the assumption, there is no difficulty to solve (2.12a)-(2.12b).

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

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

定理 2 *The sequence* $\{u^k = (x^k, y^k)\}$ *generated by the customized PPA* (2.12) *satisfies*

$$\|u^{k+1} - u^*\|_H^2 \le \|u^k - u^*\|_H^2 - \|u^k - u^{k+1}\|_H^2. \tag{2.13}$$

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

the iterative scheme is

$$x^{k+1} = \text{argmin}\Big\{\theta(x) + \frac{r}{2}\Big\|x - \big[x^k + \frac{1}{r}A^Ty^k\big]\Big\|^2 \,\Big|\, x \in \mathcal{X}\Big\}. \tag{2.14a}$$

$$y^{k+1} = y^k - \frac{1}{s}\big[A(2x^{k+1} - x^k) - b\big]. \tag{2.14b}$$

For solving the $\min$-$\max$ problem (2.10), by using (2.12), the iterative formula is

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

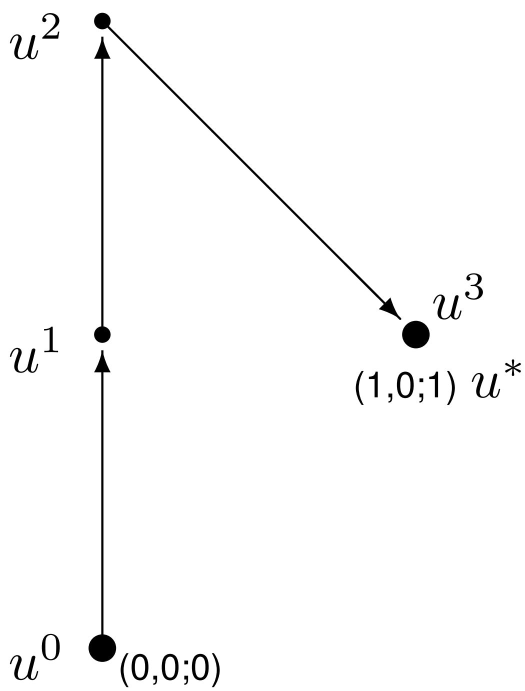
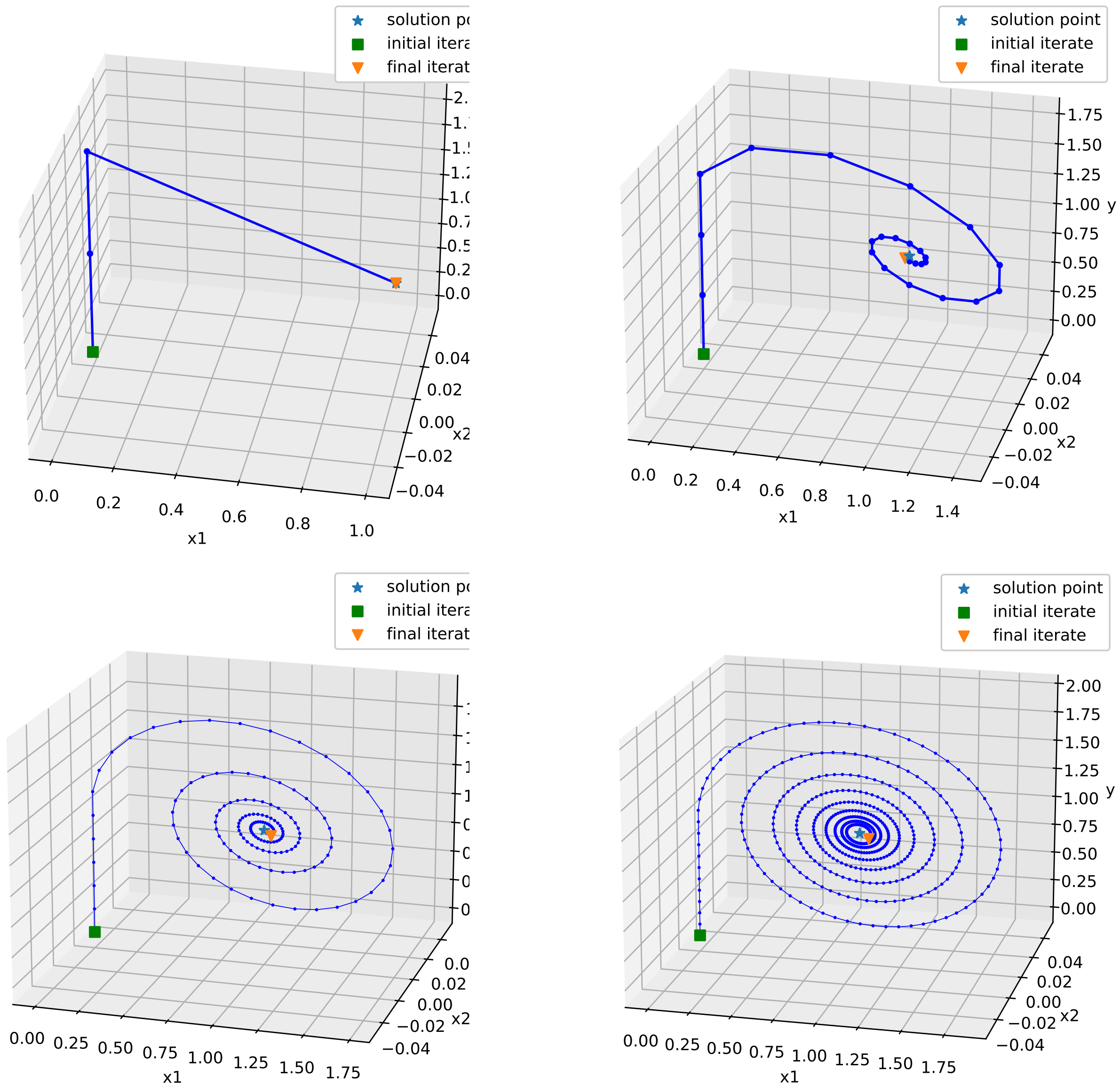


Fig. 2.2 The sequence generated by C-PPA Method with $r = s = 1$

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



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

Besides (2.12), (x^{k+1}, y^{k+1}) can be produced by using the dual-primal order:

$$y^{k+1} = \text{argmax}\Big\{\Phi\big(x^k, y\big) - \frac{s}{2}\big\|y - y^k\big\|^2\Big\} \tag{2.15a}$$

$$x^{k+1} = \text{argmin}\Big\{\Phi(x, (2y^{k+1} - y^k)) + \frac{r}{2}\big\|x - x^k\big\|^2 \,\big|\, x \in \mathcal{X}\Big\}. \tag{2.15b}$$

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

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

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, $rs > \|A^T A\|$, 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 $Ax \ge b$, namely,

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

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 (2.12b), it needs to change the update dual update form

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

In the dual-primal order (2.15a), it needs to change the update dual update form

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

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

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

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].

作者 C-P 说到我们的 PPA 解释极大地简化了收敛性分析.

我们依然认为, 只有当左边 (6) 式的矩阵 M 对称正定, 才是收敛的 PPA 方法.

否则, 就像我们前面给出的例子, 方法是不一定收敛的.

由 CP 方法演译得来的矩阵M, 当 $\theta = 0$, 方法不能保证收敛.

对 $\theta \in (0, 1)$, 收敛性没有证明, 至今还是一个 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 [9].

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

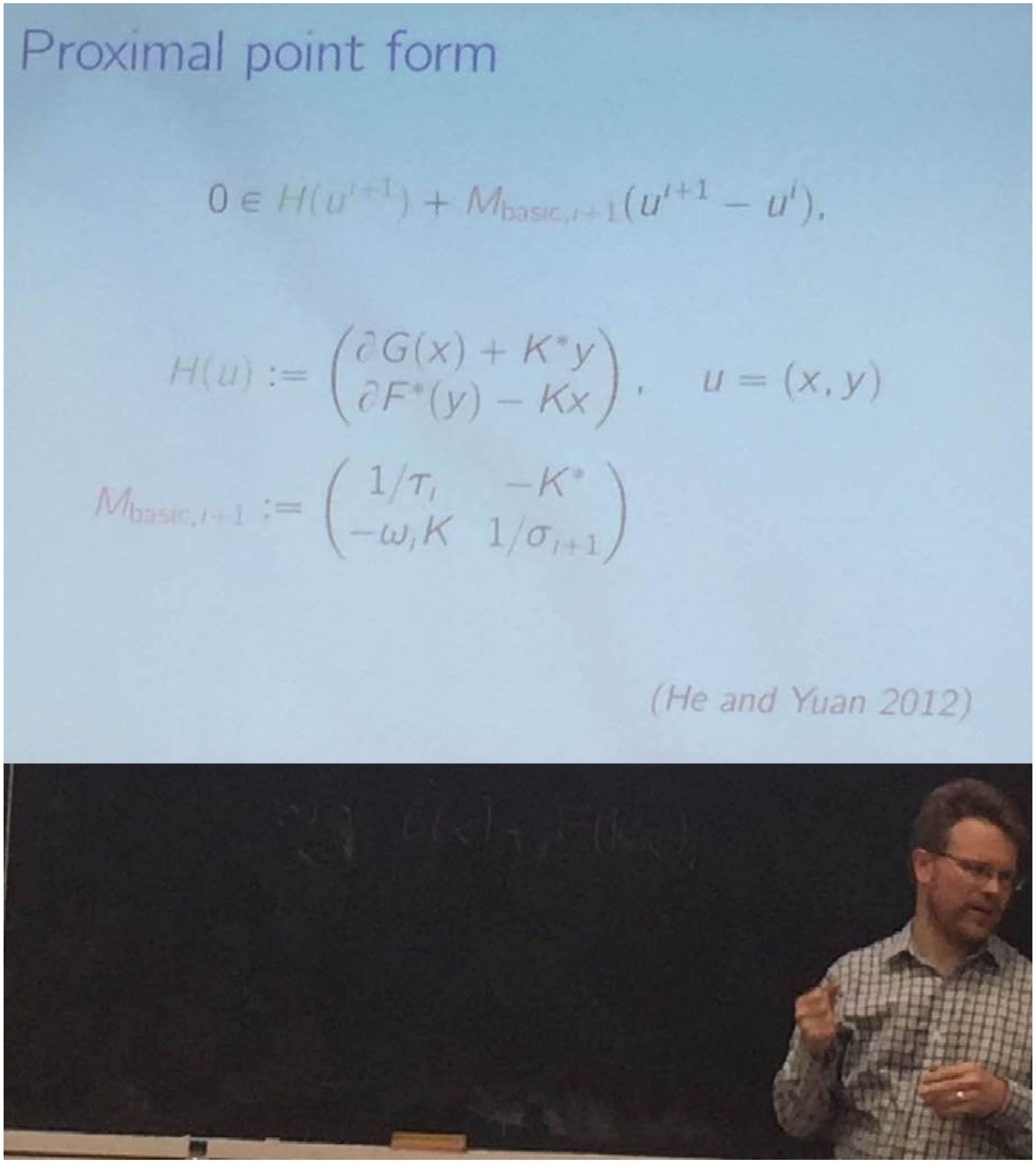
- Input: Operator norm $L := \|K\|$, Lipschitz constant L_f of ∇f, and Bregman distance functions D_x and D_y.
- Initialization: Choose $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$, $\tau, \sigma > 0$
- Iterations: For each $n \geq 0$ let

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

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

♣ 该文的文献 [16] 是我们发表在 SIAM J. Imaging Science 上的文章.

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.



2017 年 7 月, 南方科技大学数学系的一位副主任去英国访问. 在他参加的一个学术会议上, 首位报告人讲: 用 He and Yuan 提出的邻近点形式 (PPF), 处理图像问题。

见到一幅幻灯片介绍我们的工作, 我的同事抢拍了一张照片发给我。

这也说明, 只有简单的思想才容易得到传播, 被人接受。

University of Colorado **Boulder** Technical Report, Department of Applied Mathematics

The Chen-Teboulle algorithm is the proximal point algorithm

Stephen Becker *

November 22, 2011; posted August 13, 2019

Abstract

We revisit the ... on the step-size p...

Recent works such as [HY12] have proposed a very simple yet powerful technique for analyzing optimization methods.

1 Background

Recent works such as [HY12] have proposed a very simple yet powerful technique for analyzing optimization methods. The idea consists simply of working with a different norm in the *product* Hilbert space. We fix an inner product $\langle x, y \rangle$ on $\mathcal{H} \times \mathcal{H}^*$. Instead of defining the norm to be the induced norm, we define the primal norm as follows (and this induces the dual norm)

$$\|x\|_V = \sqrt{\langle Vx, x \rangle} = \sqrt{\langle x, x \rangle_V}, \quad \|y\|_V^* = \|y\|_{V^{-1}} = \sqrt{\langle y, V^{-1}y \rangle} = \sqrt{\langle y, y \rangle_{V^{-1}}}$$

for any Hermitian positive definite $V \in \mathcal{B}(\mathcal{H}, \mathcal{H})$; we write this condition as $V \succ 0$. For finite dimensional spaces $\mathcal{H}$, this means that V is a positive definite matrix.

2.4 Relationship to Chambolle-Pock Method

Chambolle and Pock [3] have proposed a method for solving the convex-concave $\min - \max$ problem, in short, C-P method. Applied C-P method to the problem (2.1), it is also required $rs > \|A^T A\|$.

CP method. For given (x^k, y^k), C-P method obtains x^{k+1} via

$$x^{k+1} = \arg\min\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \tag{2.16a}$$

Then, y^{k+1} is given by

$$y^{k+1} = \arg\max\{\Phi([x^{k+1} + \tau(x^{k+1} - x^k)], y) - \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\} \tag{2.16b}$$

where $\tau \in [0, 1]$.

当 $\tau = 1$ 并且 $rs > \|A^T A\|$, PPA 算法收敛. 当 $\tau = 0$, 方法不能保证收敛.

对 $\tau \in (0, 1)$, 收敛性没有证明, 至今还是一个 Open Problem.

- 原始-对偶混合梯度法(PDHG) (2.4) 和按需定制的邻近点算法(C-PPA) (2.12) 都是 Chambolle-Pock 方法 [3] 分别取 $\tau = 0$ 和 $\tau = 1$ 的特例.
- 对 $\tau = 0$ 的 PDHG 方法 (2.4), $\S$2.1 中已经说明不能保证收敛. 对 $\tau = 1$ 的 CPPA 方法 (2.12), 其收敛性在 $\S$2.2 中有了结论.
- 根据我们的知识, 对于 $\tau \in (0, 1)$ 的 CP 方法 (2.16), 收敛性还没有定论.

CP 方法十年记　　2020 年9 月

- Chambolle 和 Pock 在 2010 年提出的求解 $\min - \max$ 问题的原始-对偶方法, 在图像处理领域有着广泛的应用和很大的影响, 被称为CP 方法。
- Chambolle 和 Pock 方法的第一个版本公布于2010 年6 月. 他们的方法中有个 $[0, 1]$ 之间的参数, 但在文章中, 只对参数为 1 的方法给了证明. 读了他们的这篇文章以后, 我们对这类方法的收敛性进行了研究.
- 由于我们多年研究单调变分不等式的求解方法, 很快发现, 参数为 1 的 CP 方法, 可以解释为变分不等式 H-模(H 为对称正定矩阵) 的邻近点算法 (PPA), 因此收敛性证明特别简单. 五个月不到的 2010 年 11 月 4 日, 我

们把相关证明的第一稿, OO-2790, 公布在 Optimization Online 上. 同时, 对参数为 0 的 CP 方法, 我们找到了不收敛的例子.

- 参数在 $(0,1)$ 间的 CP 方法, 能不能保证收敛, 这个问题至今没有解决.
- Chambolle 和 Pock 很快发现了我们的工作, 一个多月后的 2010 年 12 月 21 日, 他们的文章在 J. MIV online 正式发表. 我们高兴地看到, Chambolle 和 Pock 已经引用了我们的文章, 也提到了我们的证明. 我们的文章正式发表以后, CP 后来就不再提参数在 $[0,1)$ 间的方法了.
- 特别感谢CP 方法的原创者认可我们给出的简单证明. 他们在2011年的IEEE ICCV 会议论文中, 称赞我们的工作极大地简化了收敛性分析 (which greatly simplifies the convergence analysis).
- 后来CP 方法的作者又有多篇相关的文章发表(后面的文章他们都只讨论参数为 1 的方法). 他们于2016 年在Math. Progr. 发表的文章中, 继续利用我们的 PPA 解释, 文章的引言中就开诚布公 (In particular, exploiting a proximal-point interpretation due to [16], we are able to give a very elementary proof). 这里的$[16]$ 是我们 2010 年的预印本 OO-2790, 2012 年春发表在 SIAM Imaging Science.

3 From ALM to Balanced ALM

We consider the generic convex minimization model with linear constraints

$$\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}, \tag{3.1}$$

where $\theta : \Re^n \to \Re$ is a closed proper convex but not necessarily smooth function; $\mathcal{X} \subseteq \Re^n$ is a closed convex set; $A \in \Re^{m\times n}$ and $b \in \Re^m$.

The Lagrangian function of (3.1) is

$$L(x,\lambda) = \theta(x) - \lambda^T(Ax - b), \tag{3.2}$$

which is defined on $\Omega = \mathcal{X} \times \Re^m$. A pair of (x^*, λ^*) defined on $\mathcal{X} \times \Lambda$ is called a saddle point of the Lagrangian function (3.2) if it satisfies the inequalities

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

Alternatively, we can rewrite these inequalities as the variational inequalities:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{3.3a}$$

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Re^m. \tag{3.3b}$$

Note that for the operator F defined in (3.3b) is affine with a skew-symmetric matrix. Thus we have

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \equiv 0. \tag{3.4}$$

We denote by Ω^* the solution set of the variational inequality (3.3).

定理 3 **[PPA for VI (3.3)]** *The sequence*

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (v - v^{k+1})^T H(v^k - v^{k+1}), \quad \forall w \in \Omega. \end{aligned} \tag{3.5}$$

Then we have

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall w^* \in \Omega^*. \tag{3.6}$$

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

3.1 Augmented Lagrangian Method

The augmented Lagrangian method originally proposed in [12, 13, 14] for (3.1) reads as

$$
\text{(ALM)}\quad \left\{ \begin{array}{ll} x^{k+1} \in \arg\min\big\{L(x,\lambda^k) + \frac{r}{2}\|Ax-b\|^2 \;\big|\; x\in\mathcal{X}\big\} & \text{(3.7a)} \\ \lambda^{k+1} = \arg\max\big\{L(x^{k+1},\lambda) - \frac{1}{2r}\|\lambda-\lambda^k\|^2\big\}. & \text{(3.7b)} \end{array} \right.
$$

The method is implemented by

$$
\left\{ \begin{array}{ll} x^{k+1} \in \arg\min\big\{\theta(x) - x^T A^T\lambda^k + \frac{r}{2}\|Ax-b\|^2 \;\big|\; x\in\mathcal{X}\big\}, & \text{(3.8a)} \\ \lambda^{k+1} = \lambda^k - r\big(Ax^{k+1}-b\big). & \text{(3.8b)} \end{array} \right.
$$

$$
(x^{k+1},\lambda^{k+1}) \in \mathcal{X}\times\Re^m,
$$

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

引理 2 *For given λ^k, let w^{k+1} be generated by* (3.7)*, then we have*

$$w^{k+1} \in \Omega, \;\; \theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \geq (\lambda - \lambda^{k+1})^T \frac{1}{r}(\lambda^k - \lambda^{k+1}), \;\; \forall\, w \in \Omega. \tag{3.9}$$

It is a form of (3.3) with $v = \lambda$, $H = \frac{1}{r} I_m$.

According to Theorem 3, the sequence $\{\lambda^k\}$ generated by ALM (3.7) satisfied

$$\|\lambda^{k+1} - \lambda^*\|^2 \leq \|\lambda^k - \lambda^*\|^2 - \|\lambda^k - \lambda^{k+1}\|^2, \quad \forall \lambda^* \in \Lambda^*. \tag{3.10}$$

Disadvantages: The x-subproblem of of the k-th iteration of ALM has the mathematical form

$$\min\Big\{\theta(x) + \frac{r}{2}\|Ax - p^k\|^2 \;\Big|\; x \in \mathcal{X}\Big\}. \tag{3.11}$$

Because of the quadratic term $\frac{r}{2}\|Ax - p^k\|^2$, sometimes it is difficult to get a solution of (3.8a).

3.2 CP-PPA method [9]

The scheme of CP-PPA method [3, 4, 9] is appropriate for (3.1). It reads as

$$\text{(CP-PPA)} \quad \begin{cases} x^{k+1} = \arg\min\big\{L(x,\lambda^k) + \frac{r}{2}\|x - x^k\|^2 \;\big|\; x \in \mathcal{X}\big\}, & \text{(3.12a)} \\ \lambda^{k+1} = \arg\max\big\{L([2x^{k+1} - x^k], \lambda) - \frac{s}{2}\|\lambda - \lambda^k\|^2\big\}. & \text{(3.12b)} \end{cases}$$

The method is implemented by

$$\begin{cases} x^{k+1} = \arg\min\big\{\theta(x) + \frac{r}{2}\|x - (x^k + \frac{1}{r}A^T\lambda^k)\|^2 \;\big|\; x \in \mathcal{X}\big\}, & \text{(3.13a)} \\ \lambda^{k+1} = \lambda^k - \frac{1}{s}\big(A[2x^{k+1} - x^k] - b\big). & \text{(3.13b)} \end{cases}$$

引理 3 *For given* w^k*, let* w^{k+1} *be generated by* (3.12)*, then we have*

$$\begin{aligned} w^{k+1} \in \Omega, \quad & \theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ & \ge (w - w^{k+1})^T H(w^k - w^{k+1}), \quad \forall\, w \in \Omega, \end{aligned} \qquad \text{(3.14a)}$$

where

$$H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}. \tag{3.14b}$$

According to Theorem 3, the sequence $\{w^k\}$ generated by CP-PPA (3.12) satisfied (3.6) where H is defined in (3.14b).

Disadvantages. In order to guarantee the convergence, the parameters r and s should satisfy

$$rs > \|A^T A\|. \tag{3.15}$$

Unless the matrix $A^T A$ is well-conditioned, the condition (3.15) will lead slow convergence.

- CP-PPA 算法的 x-子问题(3.12a) 中, 用 $\frac{1}{2}r\|x - x^k\|^2$ 去替代ALM 算法 x-子问题(3.7a) 中的 $\frac{1}{2}r\|Ax - b\|^2$. 方法是简单了, 但为了使矩阵 H 正定, 我们必须取 $rs > \|A^T A\|$. rs 要大于 $A^T A$ 的谱半径.
- 从迭代公式(3.12)可以看出, r 和 s 大, 会迫使新的迭代点 $w^{k+1} = (x^{k+1}, \lambda^{k+1})$ 靠近原来的点 $w^k = (x^k, \lambda^k)$ 太近. 在很多时候, 这会影响收敛速度.

3.3 Balanced ALM [10]

Our balanced ALM [10, 17] is to share the difficulty equally in the primal-dual steps.

$$\begin{cases} x^{k+1} = \arg\min\{L(x,\lambda^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & \text{(3.16a)} \\ \lambda^{k+1} = \arg\max\Big\{L([2x^{k+1} - x^k],\lambda) - \frac{1}{2}\|\lambda - \lambda^k\|^2_{(\frac{1}{r}AA^T + \delta I_m)}\Big\}. & \text{(3.16b)} \end{cases}$$

Replaced

$$\lambda^{k+1} = \arg\max\{L([2x^{k+1} - x^k],\lambda) - \frac{s}{2}\|\lambda - \lambda^k\|^2\},$$

in (3.12b) by

$$\lambda^{k+1} = \arg\max\Big\{L([2x^{k+1} - x^k],\lambda) - \frac{1}{2}\|\lambda - \lambda^k\|^2_{(\frac{1}{r}AA^T + \delta I_m)}\Big\}.$$

The balanced ALM (3.16) is implemented by

$$\begin{cases} x^{k+1} = \arg\min\{\theta(x) - x^T A^T \lambda^k + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, & \text{(3.17a)} \\ \lambda^{k+1} = \arg\min\Big\{\lambda^T\big(A[2x^{k+1} - x^k] - b\big) + \frac{1}{2}\|\lambda - \lambda^k\|^2_{(\frac{1}{r}AA^T + \delta I_m)}\Big\}. & \text{(3.17b)} \end{cases}$$

Remark. λ^{k+1} in (3.17b) is the solution of the following system of linear equations:

$$H_0(\lambda - \lambda^k) + \big(A[2x^{k+1} - x^k] - b\big) = 0, \tag{3.18}$$

where

$$H_0 = \frac{1}{r}AA^T + \delta I_m. \tag{3.19}$$

Because the matrix H_0 is positive definite, there are efficient algorithms in literature for solving such a systems of linear equations.

引理 4 *For given w^k, let w^{k+1} be generated by* (3.16)*, then we have*

$$\begin{aligned} w^{k+1} \in \Omega, \quad & \theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ & \geq (w - w^{k+1})^T H(w^k - w^{k+1}), \quad \forall\, w \in \Omega, \end{aligned} \tag{3.20a}$$

where

$$H = \begin{pmatrix} rI_n & A^T \\ A & \frac{1}{r}AA^T + \delta I_m \end{pmatrix} \quad \textit{is positive definite.} \tag{3.20b}$$

Proof. According to Lemma 1, x^{k+1} offered by (3.17a) is characterized as $x^{k+1} \in \mathcal{X}$ and

$$\theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \big\{ -A^T\lambda^k + r(x^{k+1} - x^k) \big\} \ge 0, \ \forall\, x \in \mathcal{X}.$$

Then, for any unknown λ^{k+1}, we have

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \ \ \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T(-A^T\lambda^{k+1}) \\ \ge (x - x^{k+1})^T \big\{ r(x^k - x^{k+1}) + A^T(\lambda^k - \lambda^{k+1}) \big\}, \ \forall x \in \mathcal{X}. \end{aligned} \quad (3.21)$$

Similarly, according to Lemma 1, λ^{k+1} offered by (3.17b) is characterized by the variational inequality $\lambda^{k+1} \in \Re^m$,

$$(\lambda - \lambda^{k+1})^T \Big\{ \Big(A[2x^{k+1} - x^k] - b \Big) + \Big(\frac{1}{r} AA^T + \delta I_m \Big) (\lambda^{k+1} - \lambda^k) \Big\} \ge 0, \quad \forall\, \lambda \in \Re^m.$$

It can be rewritten as $\lambda^{k+1} \in \Lambda$ as

$$\begin{aligned} (\lambda - \lambda^{k+1})^T(Ax^{k+1} - b) \\ \ge (\lambda - \lambda^{k+1})^T \big\{ (A(x^k - x^{k+1}) + \Big(\frac{1}{r} AA^T + \delta I_m \Big) (\lambda^k - \lambda^{k+1}) \big\}, \\ \forall\, \lambda \in \Lambda. \end{aligned} \quad (3.22)$$

Combining (3.21) and (3.22), and using the notation in (3.3), we get the assertion of this lemma. $\square$

Notice that the matrix H in

$$H = \begin{pmatrix} \sqrt{r}I_n \\ \sqrt{\frac{1}{r}}A \end{pmatrix} \left(\sqrt{r}I_n, \sqrt{\frac{1}{r}}A^T\right) + \begin{pmatrix} 0 & 0 \\ 0 & \delta I_m \end{pmatrix},$$

for any $w = (x, \lambda) \neq 0$. Thus, we have

$$w^T H w = \left\|\sqrt{r}x + \sqrt{\tfrac{1}{r}}A^T\lambda\right\|^2 + \delta\|\lambda\|^2 > 0,$$

and therefore the matrix H is positive definite. $\square$

均困的增广拉格朗日乘子法, x-子问题 (3.16a) 和 CP-PPA 中的 x-子问题 (3.12a) 完全一样. λ-子问题 (3.17b) 要求解一个系数矩阵正定的线性方程组. 我们用这个替换了严重影响收敛速度的 $rs > \|A^T A\|$ (see (3.15)). 注意到, 在整个迭代过程中, 我们只要对矩阵 H_0 (see (3.19)) 做一次 Cholesky 分解.

4 ALM in PPA-sense

The methods introduced in this section are recently published in [19].

根据预设正定矩阵 构造PPA算法. 方法可以在[19]中查到.

The convex optimization problem,

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

is translated to the equivalent variational inequality :

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, u \in \Omega, \qquad \text{(4.1a)}$$

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Re^m. \qquad \text{(4.1b)}$$

4.1 Relaxed PPA in Primal-Dual Order

Relaxed PPA for the variational inequality (4.1) :

$$\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (w-\tilde{w}^k)^T H(w^k-\tilde{w}^k), \ \forall w\in\Omega, \quad \text{(4.2a)}$$

where

$$H=\begin{pmatrix} \beta A^TA+\delta I_n & A^T \\ A & \frac{1}{\beta}I_m \end{pmatrix} \quad \text{(4.2b)}$$

The concrete formula of (4.2) is

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

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

$$\begin{cases} \theta(x)-\theta(\tilde{x}^k)+(x-\tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k}+\ \boldsymbol{(\beta A^TA+\delta I_n)}(\tilde{x}^k-x^k)+\boldsymbol{A^T}(\tilde{\lambda}^k-\lambda^k)\}\ge 0, \\ \qquad (\underline{A\tilde{x}^k-b}) \quad +\boldsymbol{A}(\tilde{x}^k-x^k) \quad + \quad \boldsymbol{(1/\beta)}(\tilde{\lambda}^k-\lambda^k)=0. \end{cases}$$

(4.3)

$$\begin{cases} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\lambda^k + \ (\beta A^T A + \delta I_n)(\tilde{x}^k - x^k)\} \geq 0, \\ \qquad\qquad \big(A[2\tilde{x}^k - x^k] - b\big) \quad + \quad (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (4.3),

we need only use the following procedure: (Primal-Dual)

$$\begin{cases} \tilde{x}^k = \text{Argmin}\left\{ \begin{array}{c} \theta(x) - x^T A^T \lambda^k \\ +\frac{1}{2}(x - x^k)^T(\beta A^T A + \delta I_n)(x - x^k) \end{array} \middle| x \in \mathcal{X} \right\}, \\ \tilde{\lambda}^k = \lambda^k - \beta\big(A[2\tilde{x}^k - x^k] - b\big). \end{cases}$$

Then, we use the form

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

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

4.2 Relaxed PPA in Dual-Primal Order

Relaxed PPA for the variational inequality (4.1) :

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \tag{4.4a}$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_n & -A^T \\ -A & \frac{1}{\beta} I_m \end{pmatrix}, \quad \text{(a small } \delta > 0\text{, say } \delta = 0.05\text{)}. \tag{4.4b}$$

Then, we use the form

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

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

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

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

The concrete form of (4.4) is

$$\left\{ \begin{array}{l} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_{n_2})}(\tilde{x}^k - x^k) \boldsymbol{- A^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ \qquad (\underline{A\tilde{x}^k - b}) \quad \boldsymbol{-A}(\tilde{x}^k - x^k) \quad + \quad \boldsymbol{(1/\beta)}\,(\tilde{\lambda}^k - \lambda^k) = 0. \end{array} \right.$$

$$\left\{ \begin{array}{l} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{-A^T(2\tilde{\lambda}^k - \lambda^k) + (\beta A^T A + \delta I_{n_2})(\tilde{x}^k - x^k)\} \geq 0, \\ \qquad (Ax^k - b) \quad + \quad (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{array} \right.$$

Implementation of (4.4) is (Dual-Primal)

$$\left\{ \begin{array}{ll} \tilde{\lambda}^k = \lambda^k - \beta(Ax^k - b), & \text{(4.5a)} \\ \tilde{x}^k = \text{Argmin}\left\{ \begin{array}{c} \theta(x) - x^T A^T[2\tilde{\lambda}^k - \lambda^k] + \\ \frac{1}{2}(x - x^k)^T(\underline{\beta A^T A + \delta I_n})(x - x^k) \end{array} \,\middle|\, x \in \mathcal{X} \right\}. & \text{(4.5b)} \end{array} \right.$$

4.3 PPA in Primal-Dual Order

Relaxed PPA for the variational inequality (4.1) :

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \tag{4.6a}$$

where

$$H = \begin{pmatrix} \delta I_n & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{4.6b}$$

Then, we use the form

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

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

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

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

The concrete form of (4.6) is

$$\begin{cases} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{\delta I_n}(\tilde{x}^k - x^k)\} \geq 0, \\ (\underline{A\tilde{x}^k - b}) + \boldsymbol{(1/\beta)}\,(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

Using $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k - b) = [\lambda^k - \beta(Ax^k - b)] - \beta A(\tilde{x}^k - x^k)$

$$\begin{cases} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \left\{ \begin{array}{c} -A^T[\lambda^k - \beta(Ax^k - b)] \\ +(\delta I_n + A^TA)(\tilde{x}^k - x^k) \end{array} \right\} \geq 0, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k - b). \end{cases}$$

Implementation

$$\begin{cases} \tilde{x}^k = \text{Argmin}\left\{ \begin{array}{l} \theta(x) - x^TA^T[\lambda^k - \beta(Ax^k - b)] + \\ \frac{1}{2}(x - x^k)^T(\underline{\beta A^TA + \delta I_n})(x - x^k) \end{array} \middle| x \in \mathcal{X} \right\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k - b). \end{cases}$$

5 Different positive definite matrices H in PPA

$$H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}, \qquad H = \begin{pmatrix} rI_n & -A^T \\ -A & sI_m \end{pmatrix}, \quad rs > \|A^TA\|.$$

$$H = \begin{pmatrix} rI_n & A^T \\ A & \frac{1}{r}AA^T + \delta I_m \end{pmatrix}, \qquad H = \begin{pmatrix} rI_n & -A^T \\ -A & \frac{1}{r}AA^T + \delta I_m \end{pmatrix}$$

$$H = \begin{pmatrix} \beta A^TA + \delta I_n & A^T \\ A & \frac{1}{\beta}I_m \end{pmatrix}, \qquad H = \begin{pmatrix} \beta A^TA + \delta I_n & -A^T \\ -A & \frac{1}{\beta}I_m \end{pmatrix}$$

$$H = \begin{pmatrix} \delta I_n & 0 \\ 0 & \dfrac{1}{\beta}I_m \end{pmatrix}, \qquad H = \begin{pmatrix} I_n & 0 \\ 0 & I_m \end{pmatrix}$$

可以根据问题的实际需要, 选择不同的正定矩阵 H

References

[1] S. Becker, The Chen-Teboulle algorithm is the proximal point algorithm, manuscript, 2011, arXiv: 1908.03633[math.OC].

[2] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A proximal point algorithms revisit on the alternating direction method of multipliers, Science China Mathematics, 56 (2013), 2179-2186.

[3] 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.

[4] A. Chambolle and T Pock, On the ergodic convergence rates of a first-order primal – dual algorithm， Math. Program., A 159 (2016) 253-287.

[5] G. Golub and C. F. Van Loan, Matrix Computations, The Johns Hopkins University Press, The Fourth Edition, 2013.

[6] G.Y. Gu, B.S. He and X.M. Yuan, Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach, Comput. Optim. Appl., 59(2014), 135-161.

[7] B.S. He and Y. Shen, On the convergence rate of customized proximal point algorithm for convex optimization and saddle-point problem (in Chinese). Sci Sin Math, 2012, 42(5): 515 – 525, doi: 10.1360/012011-1049

[8] 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.

[9] 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.

[10] B.S. He and X.M. Yuan, Balanced Augmented Lagrangian Method for Convex Optimization. manusript, 2021. arXiv:2108.08554 [math.OC]

[11] B. S. He, S. J. Xu, and X. M. Yuan, On Convergence of the Arrow-Hurwicz Method for Saddle Point Problems, J. Math. Imaging Vis., (2022) https://doi.org/10.1007/s10851-022-01089-9

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

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

[14] 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.

[15] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Cont. Optim.*, **14**, 877-898, 1976.

[16] L. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D, 60 (1992), pp. 227 − 238

[17] Shengjie Xu, A dual-primal balanced augmented Lagrangian method for linearly constrained convex programmin, manusript, 2021. arXiv:2109.02106 [math.OC]

[18] 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.

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