Very Simple Yet Powerful

2019 年 8 月，S. Becker 在 arXiv 上贴了一篇他于 2011 年写成的文章, 见 arXiv: 1908.036.33v1 [math.OC] 9 Aug 2019.

Becker 在这篇文章正文的第一句话就是 “Recent works such as [HY12] have proposed a very simple yet powerful technique for analyzing optimization methods” .

[HY12] 是这篇注记的参考文献 [10]. 应一些读者要求写下的这篇注记, 阐述了 [HY12] 的主要思想, 应用及新的进展.

Bingsheng HE

Department of Mathematics, SUSTech, and
Department of Mathematics, Nanjing University

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

Outline

Convex optimization problems concerned in this note

- $\min\text{-}\max$ problem $\quad \min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\{\Phi(x,y)=\theta_1(x)-y^TAx-\theta_2(y)\}$

- Linearly constrained COP $\quad \min\{\theta(x)|Ax=b\,(\text{or} \geq b),\; x\in\mathcal{X}\}$

  can be translated to the following min-max problem:

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\{L(x,y)=\theta_1(x)-y^TAx+b^Ty\},\quad \mathcal{Y}=\Re^m(\text{or }\ \Re^m_+).$$

- Convex Optimization with separable structure (ADMM)

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

变分不等式 (VI) 是瞎子爬山判定山顶的数学表达形式

邻近点算法 (PPA) 是步步为营 稳扎稳打的求解方法

1 Optimization problem and VI

1.1 Differential convex optimization in Form of 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 ?

$$x^* \in \Omega^* \quad \Leftrightarrow \quad x^* \in \Omega \text{ and any feasible direction is not descent direction.}$$

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.

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

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

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

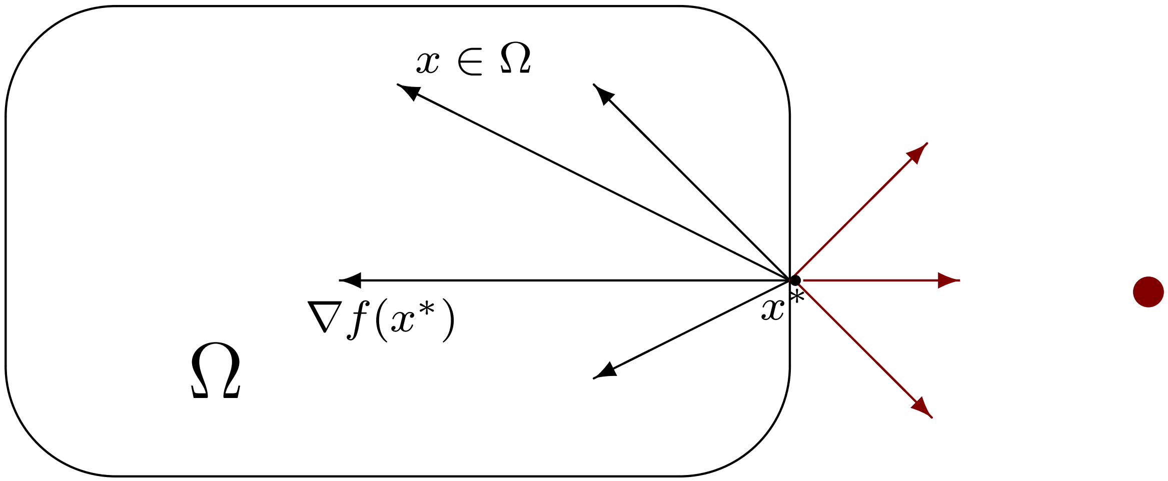


Fig. 1 Differentiable Convex Optimization and VI

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.

Let $\mathcal{X} \subset \Re^n$ be a closed convex set, $\theta(x)$ and $f(x)$ be convex functions and $f(x)$ is differentiable. Then, we have

$$x^* \in \arg\min_{x\in\mathcal{X}} \theta(x) \quad \Leftrightarrow \quad x^* \in \mathcal{X}, \;\; \theta(x) - \theta(x^*) \geq 0, \;\; \forall\, x \in \mathcal{X}.$$

$$x^* \in \arg\min_{x\in\mathcal{X}} f(x) \quad \Leftrightarrow \quad x^* \in \mathcal{X}, (x - x^*)^T \nabla f(x^*) \geq 0, \forall\, x \in \mathcal{X}.$$

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,*

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

if and only if

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

1.2 The Min-Max Problem

The min-max problem considered in this talk has the following mathematical form

$$\min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} \ \Phi(x,y) := \theta_1(x) - y^T A x - \theta_2(y), \tag{1.4}$$

where $A \in \Re^{m\times n}$, $\theta_1(x) : \Re^n \to \Re$ and $\theta_2(y) : \Re^m \to \Re$ are convex functions which are not necessarily differentiable.

Let (x^*, y^*) be the solution of (1.4), then we have

$$\Phi_{y\in\mathcal{Y}}(x^*, y) \le \Phi(x^*, y^*) \le \Phi_{x\in\mathcal{X}}(x, y^*).$$

These two inequalities can be written as

$$\begin{cases} x^* \in \mathcal{X}, \ \ \Phi(x, y^*) - \Phi(x^*, y^*) \ge 0, \ \ \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, \ \ \Phi(x^*, y^*) - \Phi(x^*, y) \ge 0, \ \ \forall\, y \in \mathcal{Y}. \end{cases}$$

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

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

We write it in a compact form of the variational inequality:

VI(Ω, F) $\quad u^* \in \Omega, \; \theta(u) - \theta(u^*) + (u - u^*)^T F(u^*) \geq 0, \; \forall\, u \in \Omega,$ (1.6a)

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad F(u) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}. \qquad (1.6b)$$

and $\Omega = \mathcal{X} \times \mathcal{Y}$. Notice that $(u - \tilde{u})^T(F(u) - F(\tilde{u})) \equiv 0$.

We use the VI form (1.6), whether $\theta(u)$ is differentiable or not.

2 Proximal point algorithms and its Beyond

Lemma 2.1 *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*

$$\|b\|_H^2 \leq \|a\|_H^2 - \|a-b\|_H^2.$$

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

2.1 Proximal point algorithms for convex optimization

Convex Optimization

Now, let us consider the *simple* convex optimization

$$\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}, \tag{2.1}$$

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

For solving (2.1), the k-th iteration of the proximal point algorithm (abbreviated to

PPA) [13, 15] begins with a given x^k, offers the new iterate x^{k+1} via the recursion

$$x^{k+1} = \text{Argmin}\{\theta(x) + f(x) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \tag{2.2}$$

Since x^{k+1} is the solution of (2.2), it follows from Lemma 1.1 that $x^{k+1} \in \mathcal{X}$,

$$\theta(x)-\theta(x^{k+1})+(x - x^{k+1})^T\{\nabla f(x^{k+1}) + r(x^{k+1} - x^k)\} \geq 0, \ \forall x \in \mathcal{X}. \tag{2.3}$$

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

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

Since $(x^{k+1} - x^*)^T \nabla f(x^{k+1}) \geq (x^{k+1} - x^*)^T \nabla f(x^*) \geq 0$, we have

$$(x^{k+1} - x^*)^T(x^k - x^{k+1}) \geq 0. \tag{2.4}$$

Let $a = x^k - x^*$ and $b = x^{k+1} - x^*$ and using Lemma 2.1, we obtain

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \|x^k - x^{k+1}\|^2, \tag{2.5}$$

which is a nice convergence property of the Proximal Point Algorithm.

We write the problem (2.1) and its PPA (2.2) in VI form

Instead of the optimization problem form $x^* \in \arg\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}$, we use its equivalent VI statement

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \ge 0, \quad \forall\, x \in \mathcal{X}. \qquad (2.6)$$

For solving the optimization problem (2.1), the k-th iteration of the PPA (see (2.3)) is: $x^{k+1} = \arg\min\{\theta(x) + f(x) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}$, we prefer use its equivalent VI form:

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad & \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \nabla f(x^{k+1}) \\ & \ge (x - x^{k+1})^T r(x^k - x^{k+1}), \quad \forall\, x \in \mathcal{X}. \end{aligned} \qquad (2.7)$$

Using (2.6) and (2.7), we consider the PPA for the variational inequality (5.6)

2.2 Preliminaries of PPA for Variational Inequalities

The optimal condition of the min-max problem is characterized as a monotone variational inequality:

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

PPA for VI (2.8) in Euclidean-norm For given u^k and $r > 0$, find u^{k+1},

$$\begin{aligned} u^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (u - u^{k+1})^T F(u^{k+1}) \\ &\geq (u - u^{k+1})^T r(u^k - u^{k+1}), \ \ \forall\, u \in \Omega. \end{aligned} \tag{2.9}$$

u^{k+1} is called the proximal point of the k-th iteration for the problem (2.8).

✠ u^k is the solution of (2.8) if and only if $u^k = u^{k+1}$ ✠

Setting $u = u^*$ in (2.9), we obtain

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

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

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

and consequently (by using (2.8)) we obtain

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

Thus, we have

$$(u^{k+1} - u^*)^T (u^k - u^{k+1}) \geq 0. \tag{2.10}$$

By setting $a = u^k - u^*$ and $b = u^{k+1} - u^*$, the inequality (2.10) means that $b^T(a - b) \geq 0$. By using Lemma 2.1, we obtain

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \|u^k - u^{k+1}\|^2. \tag{2.11}$$

We get the nice convergence property of Proximal Point Algorithm.

For any positive definite matrix H, $\|u\|_H = (u^T H u)^{\frac{1}{2}}$ is a Norm.

PPA for monotone mixed VI in H-norm

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

$$\begin{aligned} u^{k+1} \in \Omega, \quad & \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T F(u^{k+1}) \\ & \geq (u - u^{k+1})^T H(u^k - u^{k+1}), \ \forall\, u \in \Omega, \end{aligned} \tag{2.12}$$

where H is a symmetric positive definite matrix.

✠ Again, u^k is the solution of (2.8) if and only if $u^k = u^{k+1}$ ✠

Convergence Property of Proximal Point Algorithm in H-norm

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

Any norms are equivalent ! $\|u - u^*\|_H \to 0 \quad \Leftrightarrow \quad \|u - u^*\| \to 0.$

3 PPA for VI arising from min-max problem

This section presents various applications of the proposed algorithms for the min-max problem, namely

$$\min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} \ \Phi(x,y) := \theta_1(x) - y^T Ax - \theta_2(y). \qquad (3.1)$$

The equivalent variational inequality of the $\min - \max$ problem (3.1) is

$$u^* \in, \Omega, \quad \theta(u) - \theta(u^*) + (u-u^*)^T F(u^*) \geq 0, \ \ \forall\, u \in \Omega, \qquad (3.2a)$$

where

$$u = \left[\begin{array}{c} x \\ y \end{array}\right], \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad F(u) = \left[\begin{array}{c} -A^T y \\ Ax \end{array}\right], \qquad (3.2b)$$

and $\Omega = \mathcal{X} \times \mathcal{Y}$.

3.1 How to reach an implementable PPA

If we use the PPA form (2.9) to solve (3.2), start from a given u^k, the task is to find a u^{k+1}, such that

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

The concrete form is

$$(x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}, \quad \begin{bmatrix} \theta(x) - \theta(x^{k+1}) \\ \theta(y) - \theta(y^{k+1}) \end{bmatrix} + \begin{bmatrix} x - x^{k+1} \\ y - y^{k+1} \end{bmatrix}^T \left\{ \begin{bmatrix} -A^T y^{k+1} \\ A x^{k+1} \end{bmatrix} + \begin{bmatrix} r(x^{k+1} - x^k) & \\ & r(y^{k+1} - y^k) \end{bmatrix} \right\} \geq 0, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{3.3}$$

According to Lemma 1.1, the equivalent optimization problems of the VI (3.3) is

$$\begin{cases} x^{k+1} = \arg\min_{x\in\mathcal{X}}\{\theta_1(x) - x^T A^T y^{k+1} + \frac{r}{2}\|x - x^k\|^2\}, & \text{(3.4a)} \\ y^{k+1} = \arg\min_{y\in\mathcal{Y}}\{\theta_2(y) + y^T A x^{k+1} + \frac{r}{2}\|y - y^k\|^2\}. & \text{(3.4b)} \end{cases}$$

The problems (3.4a) and (3.4b) are coupled. Unfortunately, there are no appropriate methods for solving the problems (3.4a) and (3.4b) together.

Replaced y^{k+1} in (3.4a) with y^k, the optimization problems (3.4) are reduced to

$$\begin{cases} x^{k+1} = \arg\min_{x\in\mathcal{X}}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2\}, & \text{(3.5a)} \\ y^{k+1} = \arg\min_{y\in\mathcal{Y}}\{\theta_2(y) + y^T A x^{k+1} + \frac{r}{2}\|y - y^k\|^2\}. & \text{(3.5b)} \end{cases}$$

The problems (3.5) can be solved one by one, its equivalent VI form is

$$(x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}, \quad \begin{bmatrix} \theta(x) - \theta(x^{k+1}) \\ \theta(y) - \theta(y^{k+1}) \end{bmatrix} + \begin{bmatrix} x - x^{k+1} \\ y - y^{k+1} \end{bmatrix}^T \left\{ \begin{bmatrix} -A^T y^k \\ Ax^{k+1} \end{bmatrix} + \begin{bmatrix} r(x^{k+1} - x^k) & 0 \\ 0 & r(y^{k+1} - y^k) \end{bmatrix} \right\} \geq 0, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

Notice that $F(u^{k+1}) = \begin{bmatrix} -A^T y^{k+1} \\ Ax^{k+1} \end{bmatrix}$, we rewrite the above VI in the form

$$(x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}, \quad \begin{bmatrix} \theta(x) - \theta(x^{k+1}) \\ \theta(y) - \theta(y^{k+1}) \end{bmatrix} + \begin{bmatrix} x - x^{k+1} \\ y - y^{k+1} \end{bmatrix}^T \left\{ \begin{bmatrix} -A^T y^{k+1} \\ Ax^{k+1} \end{bmatrix} + \begin{bmatrix} r(x^{k+1} - x^k) + A^T(y^{k+1} - y^k) \\ 0 \quad r(y^{k+1} - y^k) \end{bmatrix} \right\} \geq 0, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{3.6}$$

The compact form of (3.6) is

$$u^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \left\{ F(u^{k+1}) + Q(u^{k+1} - u^k) \right\} \ge 0, \quad \forall u \in \Omega. \tag{3.7}$$

where

$$Q = \begin{bmatrix} rI_n & A^T \\ 0 & rI_m \end{bmatrix}$$ is not symmetric.

If we change the block upper-triangular matrix

$$Q = \begin{bmatrix} rI_n & A^T \\ 0 & rI_m \end{bmatrix}$$ to a symmetric matrix $$H = \begin{bmatrix} rI_n & A^T \\ A & sI_m \end{bmatrix},$$

the variational inequality (3.7) becomes

$$u^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \left\{ F(u^{k+1}) + H(u^{k+1} - u^k) \right\} \ge 0, \quad \forall u \in \Omega. \tag{3.8}$$

Notice that the concrete form of (3.8) is

$$(x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}, \quad \begin{bmatrix} \theta(x) - \theta(x^{k+1}) \\ \theta(y) - \theta(y^{k+1}) \end{bmatrix} + \begin{bmatrix} x - x^{k+1} \\ y - y^{k+1} \end{bmatrix}^T \left\{ \begin{bmatrix} -A^T y^{k+1} \\ A x^{k+1} \end{bmatrix} + \begin{bmatrix} r(x^{k+1} - x^k) + A^T (y^{k+1} - y^k) \\ A(x^{k+1} - x^k) + s(y^{k+1} - y^k) \end{bmatrix} \right\} \geq 0, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{3.9}$$

According to Lemma 1.1, the variational inequality (3.9) is implemented by

$$\begin{cases} x^{k+1} = \arg\min_{x \in \mathcal{X}} \{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2\}, & \text{(3.10a)} \\ y^{k+1} = \arg\min_{y \in \mathcal{Y}} \{\theta_2(y) + y^T A(2x^{k+1} - x^k) + \frac{s}{2}\|y - y^k\|^2\}. & \text{(3.10b)} \end{cases}$$

Ignoring the constant term in the objective function,

For given (x^k, y^k), we get x^{k+1} via solving

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

With the getting x^{k+1}, we obtain y^{k+1} by solving the following problem:

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

Using the notation of $\Phi(x, y)$, the iterative scheme (3.10) can be written as

$$\begin{cases} x^{k+1} = \arg\min_{x\in\mathcal{X}}\Big\{\Phi(x, y^k) + \frac{r}{2}\big\|x - x^k\big\|^2\Big\}, & \text{(3.11a)} \\ y^{k+1} = \arg\max_{y\in\mathcal{Y}}\Big\{\Phi\big([2x^{k+1} - x^k], y\big) - \frac{s}{2}\big\|y - y^k\big\|^2\Big\}. & \text{(3.11b)} \end{cases}$$

Assumption:

1. The sub-problems

$$\min_{x\in\mathcal{X}}\{\theta_1(x)+\frac{r}{2}\|x-p\|^2\} \quad \text{and} \quad \min_{y\in\mathcal{Y}}\{\theta_2(y)+\frac{s}{2}\|y-q\|^2\}$$

have closed solution. Thus, solving the sub-problems in (3.11) is simple.

2. The matrix $H=\begin{bmatrix} rI_n & A^T \\ A & sI_m \end{bmatrix}$ is positive definite.

$$\boldsymbol{rs > \|A^TA\| \iff H=\begin{bmatrix} rI_n & A^T \\ A & sI_m \end{bmatrix}}$$ **is positive definite.**

Theorem 3.1 *The method* (3.10) *is a PPA for VI* (3.2)*. The generated sequence* $\{u^k=(x^k,y^k)\}$ *satisfies*

$$\|u^{k+1}-u^*\|_H^2 \le \|u^k-u^*\|_H^2 - \|u^k-u^{k+1}\|_H^2, \quad \forall u^*\in\Omega^*.$$

3.2 Chambolle-Pock method

The Chambolle-Pock algorithm [3] is a well known approach for solving the min-max problems arising from imaging processing. Following is their iterative scheme:

For given (x^k, y^k), produce a pair of (x^{k+1}, y^{k+1}). First,

$$x^{k+1} = \arg\min_{x\in\mathcal{X}}\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2\}. \tag{3.12a}$$

Then, set

$$\bar{x}^k = x^{k+1} + \tau(x^{k+1} - x^k), \ \tau \in [0, 1] \tag{3.12b}$$

Finally, obtain y^{k+1} via

$$y^{k+1} = \text{Argmax}\{\Phi(\bar{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}, \tag{3.12c}$$

Using Lemma 1.1, we interpreted the output of the Chambolle-Pock algorithm as the solution of the solution of the following variational inequality:

$$(x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}, \quad \begin{bmatrix} \theta(x) - \theta(x^{k+1}) \\ \theta(y) - \theta(y^{k+1}) \end{bmatrix} + \begin{bmatrix} x - x^{k+1} \\ y - y^{k+1} \end{bmatrix}^T \left\{ \begin{bmatrix} -A^T y^{k+1} \\ A x^{k+1} \end{bmatrix} \right.$$

$$\left. + \begin{bmatrix} r(x^{k+1} - x^k) + A^T(y^{k+1} - y^k) \\ \tau A(x^{k+1} - x^k) + r(y^{k+1} - y^k) \end{bmatrix} \right\} \geq 0, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \quad (3.13)$$

The compact form of (3.13) is

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

where

$$Q = \begin{bmatrix} rI_n & A^T \\ \tau A & sI_m \end{bmatrix}$$

is not symmetric unless $\tau = 1$.

Set $\tau = 1$ and let $rs > \|A^T A\|$, (3.13) is the PPA form (2.12). Our re-normed PPA interpretation greatly simplifies the convergence analysis.

The method (3.12) is first proposed by Chambolle and Pock [3] and is called C-P method. Thanks to the authors for mentioning our proof in a footnote of this paper.

3.3 Simplicity recognition

VI-PPA Form is recognized by Researchers in Image Science

In the first paper about C-P method

- 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.

the authors mentioned our proof (interpretation) in the footnote of page 121.

- T. Pock and A. Chambolle, IEEE ICCV, 1762-1769, 2011.

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

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

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

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].

[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.

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

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

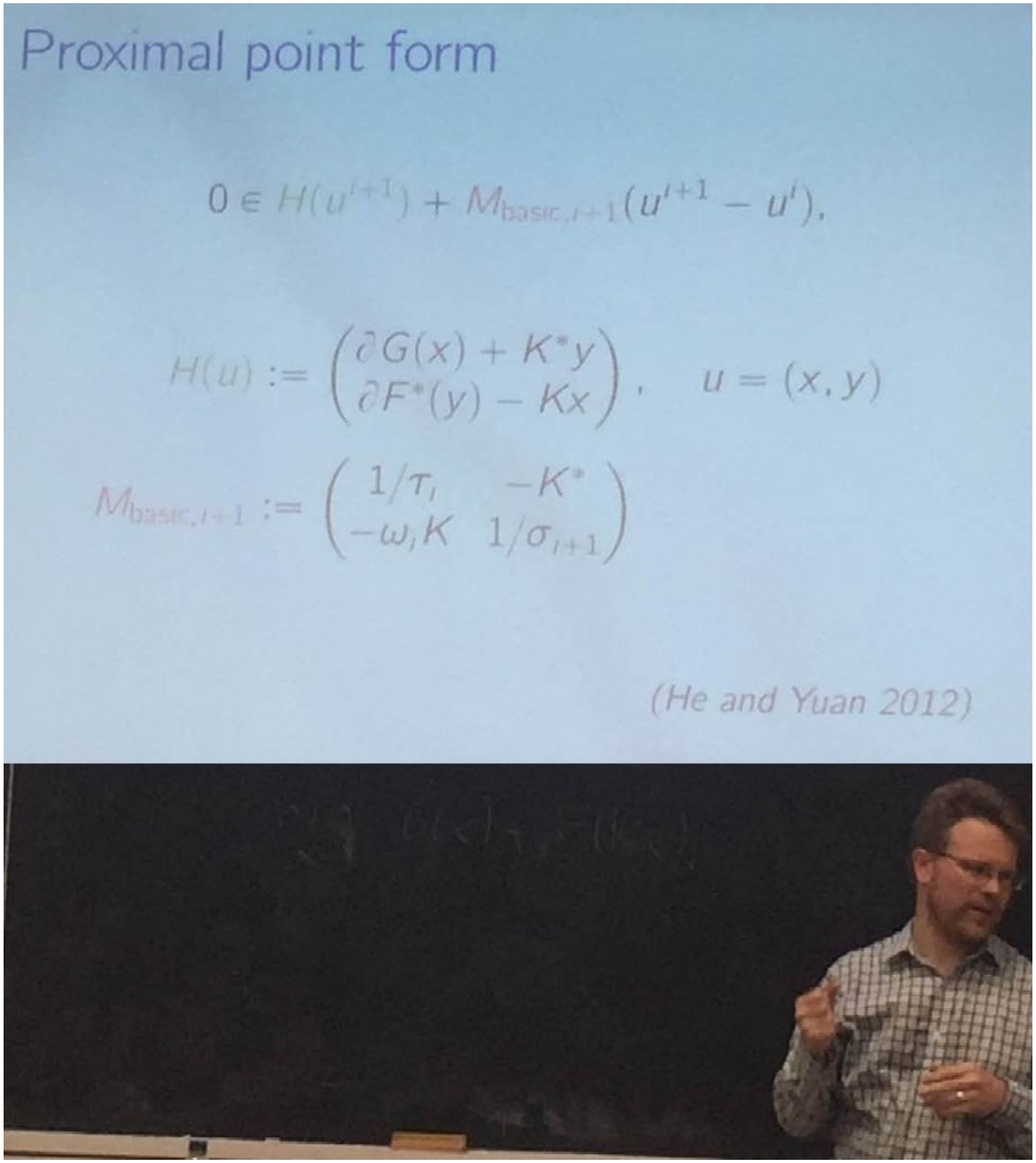
- Input: Operator norm $L := \|K\|$, Lipschitz constant L_f of ∇f, and Bregman distance functions D_x and D_y.
- Initialization: Choose $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$, $\tau, \sigma > 0$
- Iterations: For each $n \geq 0$ let

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

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

♣ **The cited paper [16] published in SIAM J. Imaging Science, 2012**

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, 首位报告人讲到, 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。

4 Extended PPA for the Variational Inequality

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 Chen-Teboulle algorithm using recent insights and show that this allows a better bound on the step-size parameter.

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.

Recent insights allows a better bound on the step-size parameter.

S. Becker: Recent works such as [HY12] have proposed a very simple yet powerful technique for analysing optimization methods.

For given $u^k = (x^k, y^k)$, set the solution of (3.10) as a predictor. Namely,

$$\text{(CPPA)} \quad \begin{cases} \tilde{x}^k = \arg\min_{x\in\mathcal{X}}\{\Phi(x,y^k) + \frac{r}{2}\|x - x^k\|^2\}, & \text{(4.1a)} \\ \tilde{y}^k = \arg\max_{y\in\mathcal{Y}}\{\Phi\big([\boldsymbol{2\tilde{x}^k - x^k}],\lambda\big) - \frac{s}{2}\|y - y^k\|^2\} & \text{(4.1b)} \end{cases}$$

where $\Phi(x,y) = \theta_1(x) - y^T Ax - \theta_2(y)$.

For given $u^k = (x^k, y^k)$, set the solution of (3.10) as a predictor. Namely,

$$\begin{cases} \tilde{x}^k = \arg\min\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, & \text{(4.2a)} \\ \tilde{y}^k = \arg\min\Big\{\theta_2(y) + y^T A[2\tilde{x}^k - x^k] + \frac{s}{2}\|y - y^k\|^2 \,\Big|\, y \in \mathcal{Y}\Big\}. & \text{(4.2b)} \end{cases}$$

The output $\tilde{u}^k \in \Omega$ of the iteration (4.1) is a predictor which satisfies

$$\theta(u)-\theta(\tilde{u}^k)+(u-\tilde{u}^k)^T F(\tilde{u}^k) \ge (u-\tilde{u}^k)^T H(u^k-\tilde{u}^k), \ \forall u \in \Omega. \quad \text{(4.3a)}$$

where

$$H = \begin{bmatrix} rI & A^T \\ A & sI \end{bmatrix} \quad \text{is positive definite.} \quad \text{(4.3b)}$$

Correction-Extension

The new iterate is given by

$$u^{k+1} = u^k - \alpha(u^k - \tilde{u}^k), \quad \alpha \overset{\text{say}}{=} 1.5 \in [1,2). \quad \text{(4.4)}$$

- ◇ B.S. He and X.M. Yuan, Convergence analysis of primal-dual algorithms for a saddle -point problem: From contraction perspective, SIAM J. Imag. Sci., 5, 119-149, 2012.
- ◇ B.S. He, X.M. Yuan and W.X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, Comput. Optim. Appl., 56: 559-572, 2013.
- ◇ 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.

Setting $u = u^*$ in (4.3a), and using $(\tilde{u}^k - u^*)F(\tilde{u}^k) = (\tilde{u}^k - u^*)F(u^*)$, we get

$$(\tilde{u}^k - u^*)^T H(u^k - \tilde{u}^k) \geq 0.$$

Lemma 4.1 *For given* u^k*, let the predictor* $\tilde{u}^k$ *be generated by* (4.3a)*, then we have*

$$(u^k - u^*)^T H(u^k - \tilde{u}^k) \geq \|u^k - \tilde{u}^k\|_H^2, \tag{4.5}$$

where H *is a positive definite matrix given by* (4.3b)*.*

For the given positive definite matrix H, (4.5) means that

$$\Big(\nabla\big(\frac{1}{2}\|u - u^*\|_H^2\big)\Big|_{u=u^k}\Big)^T \big(u^k - \tilde{u}^k\big) \geq \|u^k - \tilde{u}^k\|_H^2.$$

The above inequality tells us that $-(u^k - \tilde{u}^k)$ is a decent direction of the unknown distance function $\frac{1}{2}\|u - u^*\|_H^2$ at the current point u^k.

Then, we can define an α- dependent new iterate u_{α}^{k+1} given by

$$u^{k+1}(\alpha) = u^k - \alpha(u^k - \tilde{u}^k), \quad \text{where} \quad \alpha \in (0, 2). \tag{4.6}$$

and consider to maximize the profit function

$$\vartheta_k(\alpha) = \|u^k - u^*\|_H^2 - \|u^{k+1}(\alpha) - u^*\|_H^2. \tag{4.7}$$

Thus, it follows from (4.6) that

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

By using (4.5), we get

$$\begin{aligned}\vartheta_k(\alpha) &\geq 2\alpha\|u^k - \tilde{u}^k\|_H^2 - \alpha^2\|u^k - \tilde{u}^k\|_H^2 \\ &= \alpha(2-\alpha)\|u^k - \tilde{u}^k\|_H^2 \; = \; q_k(\alpha). \qquad \Box\end{aligned}$$

$q_k(\alpha)$ reaches its maximum at α_k^* which is given by $\alpha_k^* = 1$

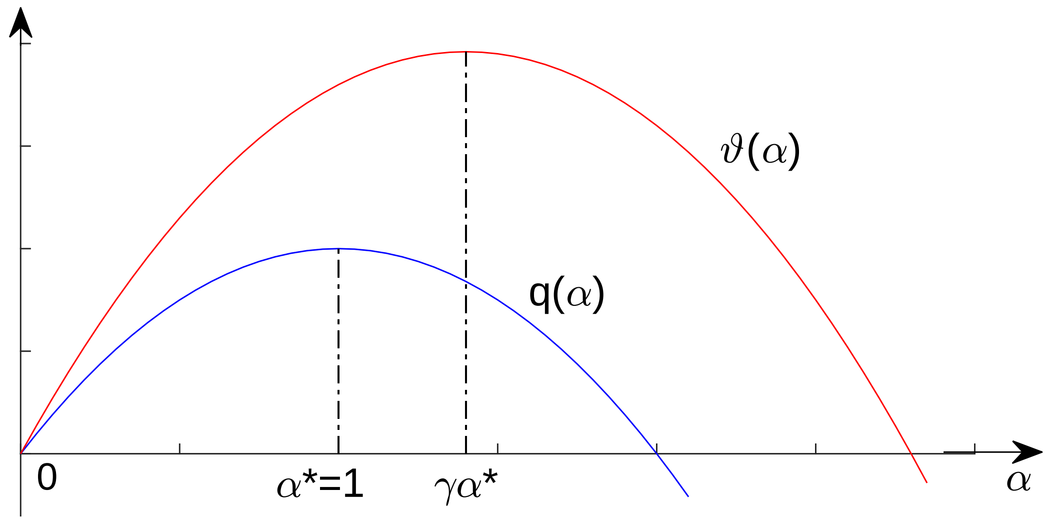


Fig 2. The reason for taking $\alpha = \gamma\alpha^*$, $\gamma \in [1, 2)$

Theorem 4.1 *For given* u^k, *let* $\tilde{u}^k$ *and* u^{k+1} *be generated by* (4.3) - (4.4), *then we have*

$$\|u^{k+1} - u^*\|_H^2 \le \|u^k - u^*\|_H^2 - \alpha(2-\alpha)\|u^k - \tilde{u}^k\|_H^2, \ \ \forall u^* \in \Omega^*. \quad (4.8)$$

5 Linearly constrained Optimization in form of VI

We consider the linearly constrained convex optimization problem

$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{5.1}$$

The Lagrange function of (5.1) is

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \qquad (u,\lambda) \in \mathcal{U} \times \Re^m. \tag{5.2}$$

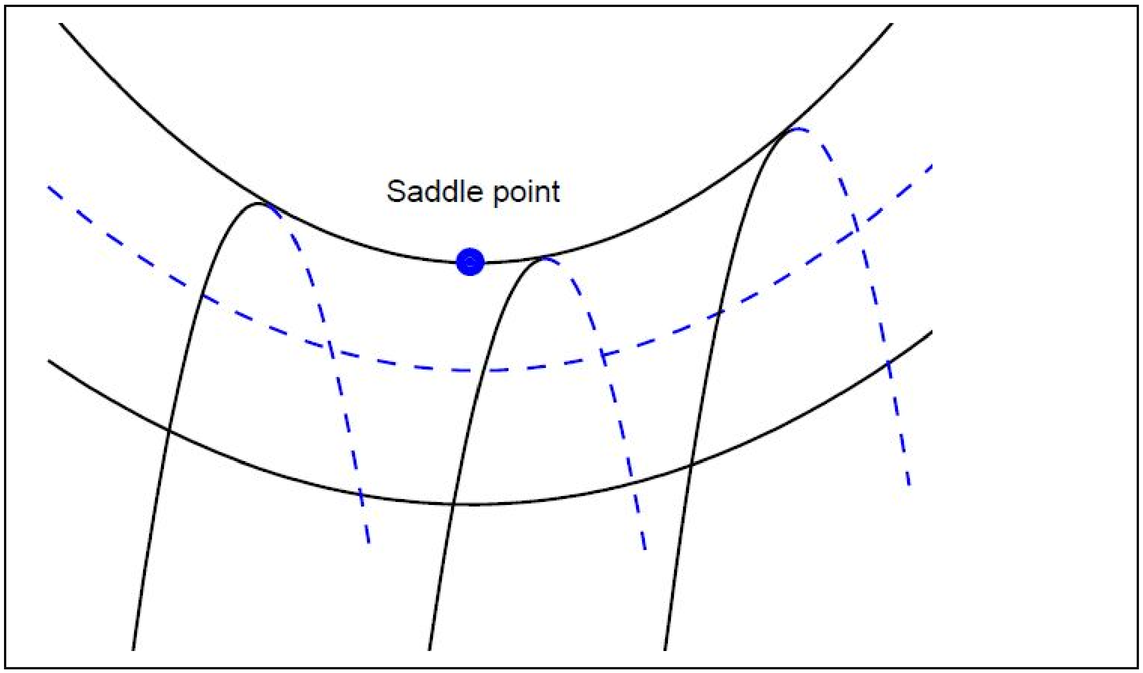


Fig 2. The saddle point of the Lagrangian function

5.1 Saddle point and the equivalent variational inequality

A pair of (u^*, λ^*) is called a saddle point of the Lagrange function (5.2), if

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

The above inequalities mean that

$$\begin{cases} u^* \in \mathcal{U}, \;\; L(u, \lambda^*) - L(u^*, \lambda^*) \ge 0, \;\; \forall\, u \in \mathcal{U}, & \text{(5.3a)} \\ \lambda^* \in \Lambda, \;\; L(u^*, \lambda^*) - L(u^*, \lambda) \ge 0, \;\; \forall\, \lambda \in \Lambda. & \text{(5.3b)} \end{cases}$$

The inequality (5.3a) represents that

$$u^* \in \mathcal{U}, \;\; \theta(u) - \theta(u^*) + (u - u^*)^T(-\mathcal{A}^T\lambda^*) \ge 0, \;\; \forall\, u \in \mathcal{U}. \qquad (5.4)$$

Similarly, for (5.3b), we have

$$\lambda^* \in \Re^m, \;\; (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \ge 0, \;\; \forall\, \lambda \in \Re^m. \qquad (5.5)$$

Notice that the above expression is equivalent to

$$\mathcal{A}u^* = b.$$

Writing (5.4) and (5.5) together, we get the following variational inequality:

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

The saddle-point can be characterized as the solution of the following VI:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \qquad (5.6)$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \text{ and } \quad \Omega = \mathcal{U} \times \Re^m. \qquad (5.7)$$

Notice that F is a affine operator with a skew-symmetric matrix, namely,

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

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

Convex optimization problem with two separable functions

We consider the convex optimization problem which has the following form:

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{5.8}$$

This is a special problem of (5.1) with

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \mathcal{U} = \mathcal{X} \times \mathcal{Y}, \quad \mathcal{A} = (A, B).$$

The Lagrangian function of the problem (5.8) is

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

The same analysis tells us that the saddle point of the Lagrange function $L^2(x, y, \lambda)$ is a solution of the following variational inequality:

$$w^* \in \Omega, \;\; \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \;\; \forall\, w \in \Omega, \tag{5.9a}$$

where

$$u = \begin{bmatrix} x \\ y \end{bmatrix}, \qquad \theta(u) = \theta_1(x) + \theta_2(y), \tag{5.9b}$$

$$w = \begin{bmatrix} x \\ y \\ \lambda \end{bmatrix}, \qquad F(w) = \begin{bmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{bmatrix}. \tag{5.9c}$$

and $\Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m$.

The affine operator $F(w)$ has the form

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

Again, we have

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

The augmented Lagrangian Function of the problem (5.8) is

$$\begin{aligned}\mathcal{L}_\beta(x,y,\lambda) \quad = \quad & \theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b) \\ & +\frac{\beta}{2}\|Ax+By-b\|^2. \end{aligned} \qquad (5.10)$$

Alternating direction method of multipliers (ADMM)

Solving the problem (5.8) by using ADMM [4, 5], the k-th iteration begins with a given $v^k=(y^k,\lambda^k)$, it offers the new iterate $v^{k+1}=(y^{k+1},\lambda^{k+1})$ via

$$\text{(ADMM)} \quad \left\{ \begin{array}{ll} x^{k+1}=\arg\min\{\mathcal{L}_\beta(x,y^k,\lambda^k) \mid x\in\mathcal{X}\}, & \text{(5.11a)} \\ y^{k+1}=\arg\min\{\mathcal{L}_\beta(x^{k+1},y,\lambda^k) \mid y\in\mathcal{Y}\}, & \text{(5.11b)} \\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). & \text{(5.11c)} \end{array} \right.$$

Since x^{k+1} is a computational result dependent on the given $v^k=(y^k,\lambda^k)$, we call it the intermediate variable. The variables $v = (y, \lambda)$ are called essential variables in ADMM.

We denote the solution set of (5.9) by Ω^*. The sequence $\{v^k\}$ generated by ADMM(5.11) satisfies

$$\|v^{k+1} - v^*\|_G^2 \le \|v^k - v^*\|_G^2 - \|v^k - v^{k+1}\|_G^2, \quad \forall\, v \in \mathcal{V}^*, \tag{5.12}$$

where

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

and

$$\mathcal{V}^* = \{(y^*, \lambda^*) \,|\, (x^*, y^*, \lambda^*) \in \Omega^*\}.$$

For a short proof, the reader may refer to our paper [11]. Besides the contractive property (5.12), it was proved that the residue sequence $\{\|v^k - v^{k+1}\|_G^2\}$ generated by ADMM(5.11) is monotonically no-increasing, namely,

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

5.2 Extended PPA for Variational Inequalities (5.9)

The optimal condition of the problem (5.8) is characterized as the variational inequality (5.9), namely

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

Guided by (4.3) - (4.4), we consider the following extended PPA for the above VI.

Let H be a proper positive definite matrix.

[Prediction]. Start with a given v^k, find a predictor $\tilde{w}^k$ which satisfies

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) &+ (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ &\geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \ \ \forall\, w \in \Omega. \end{aligned} \tag{5.13}$$

[Correction]. Update the new iterate v^{k+1} by

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \overset{\text{say}}{=} 1.5 \in [1, 2). \tag{5.14}$$

✠ $\tilde{w}^k$ is the solution of (5.9) if and only if $v^k = \tilde{v}^k$ ✠

Similarly as in Section 4, setting $w = w^*$ in (5.13), we obtain

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

By using $(\tilde{w}^k - w^*)^T F(\tilde{w}^k) = (\tilde{w}^k - w^*)^T F(w^*)$ (see $F(w)$ in (5.9c)) and the optimality, we obtain

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

and consequently,

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_H^2. \tag{5.15}$$

Finally, we have the following results which is key-inequality of convergence for the prediction- correction method (5.13) - (5.14).

Theorem 5.1 *For given* v^k*, let* $\tilde{w}^k$ *and* v^{k+1} *be generated by the prediction-correction method* (5.13) - (5.14)*. Then we have*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \alpha(2-\alpha)\|v^k - \tilde{v}^k\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.16}$$

6 Design the extended PPA for solving VI (5.9)

Design the extended PPA for VI (5.9) guided by (5.13) - (5.14).

6.1 ADMM in PPA-sense

In order to solve the separable convex optimization problem (5.8), we construct a method whose prediction-step is

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

where

$$H = \begin{pmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad \text{(a small } \delta > 0\text{, say } \delta = 0.05\text{).} \tag{6.1b}$$

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

The concrete form of (6.1) is

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

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

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

In fact, the prediction can be arranged by

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x,y^k,\lambda^k)\,|\,x\in\mathcal{X}\}, & \text{(6.2a)} \\ \tilde{\lambda}^k = \lambda^k-\beta(A\tilde{x}^k+By^k-b), & \text{(6.2b)} \\ \tilde{y}^k = \text{Argmin}\left\{ \begin{array}{c} \theta_2(y)-y^TB^T\boldsymbol{[2\tilde{\lambda}^k-\lambda^k]} \\ +\frac{1+\delta}{2}\beta\|B(y-y^k)\|^2 \end{array} \Bigg| \, y\in\mathcal{Y}\right\}. & \text{(6.2c)} \end{cases}$$

The computational load of the prediction (6.2) equals the one of the ADMM (5.11).
The correction $v^{k+1}=v^k-\alpha(v^k-\tilde{v}^k)$ will accelerate the convergence.

6.2 Linearized ADMM-Like Method

Simplify the subproblem (6.2c). Replace $\frac{1+\delta}{2}\beta\|B(y-y^k)\|^2$ with $\frac{s}{2}\|y-y^k\|^2$.

By using the linearized version of (6.2), the prediction step becomes

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T H(v^k-\tilde{v}^k), \ \forall w \in \Omega, \quad (6.3)$$

where

$$H = \begin{bmatrix} sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}, \text{ 代替 (6.1) 中的 } \begin{bmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}. \quad (6.4)$$

The concrete formula of (6.3) is

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

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

$$\begin{cases} \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + \boldsymbol{s}(\tilde{y}^k-y^k) \ \ \boldsymbol{-B^T} \ (\tilde{\lambda}^k-\lambda^k)\} \ge 0, \\ (\underline{A\tilde{x}^k+B\tilde{y}^k-b})\boldsymbol{-B}(\tilde{y}^k-y^k)+\boldsymbol{(1/\beta)}(\tilde{\lambda}^k-\lambda^k)=0. \end{cases} \quad (6.5)$$

Then, we use the form

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

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

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (6.5),

we need only use the following procedure:

$$\left\{ \begin{array}{ll} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, & \text{(6.6a)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), & \text{(6.6b)} \\ \tilde{y}^k = \text{Argmin}\{\theta_2(y) - y^T B^T [\mathbf{2\tilde{\lambda}^k - \lambda^k}] + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\} & \text{(6.6c)} \end{array} \right.$$

The term $\frac{1+\delta}{2}\beta\|B(y - y^k)\|^2$ in (6.2c) is replaced by $\frac{s}{2}\|y - y^k\|^2$. In order to ensure the positivity of the matrix H in (6.4), $s > \beta\|B^T B\|$ is necessary.

Solving the problem (6.6c) is somewhat easy than solving the problem (6.2c), however, sometimes the large scalar s will lead a slow convergence.

总结：对求解凸优化问题

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

我们在 §6 中提出两种预测-校正方法

- 如果子问题中求解过程中, 二次项不带来任何困难的时候, 建议采用 §6.1 中的方法.
- 如果子问题中求解过程中, 对一个子问题中的二次项线性化后才比较容易求解, 建议采用 §6.2 中的方法.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Foundations and Trends in Machine Learning Vol. 3, No. 1 (2010) 1 – 122.

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

[3] 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] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appli., 2 (1976), pp. 17-40.

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

[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, pp. 135-161, 2014.

[7] 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.

[8] B.S. He, M. Tao and X.M. Yuan, A splitting method for separable convex programming, *IMA J. Numerical Analysis* **31**(2015), 394-426.

[9] 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.

[10] 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.

[11] 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.

[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] 何炳生, 我和乘子交替方向法 20 年,《运筹学学报》22 卷第1期, pp. 1-31, 2018.

VI 如“瞎子爬山”问是否最优， PPA 以“步步为营”向目标逼近.

Thank you very much for your attention ！