瞎子爬山 和 步步为营

凸优化算法中的变分不等式和邻近点策略

中学的数理基础	必要的社会实践
普通的大学数学	一般的优化常识

何 炳 生

南京大学数学系

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

南京大学数学系本科生论坛 2021 年 3 月 17 日

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

Department of Mathematics, SUSTech
(Southern University of Science and Technology),
Shenzhen, 518055, China
Phone: +86-755-88018721   Email: hebs@sustech.edu.cn
&
Department of Mathematics,
Nanjing University, Nanjing, 210093, China
E-mail: hebma@nju.edu.cn

Current Research Areas:
  Mathematical Programming, Numerical Optimization, Variational Inequalities,
  ADMM-like splitting contraction methods for convex optimization

Education:
  PhD: Applied Mathematics, The University of Wuerzburg, Germany, 1986
    Thesis Advisor: Professor Dr. Josef Stoer
  BSc: Computational Mathematics, Nanjing University, 1981

Work:
  2015---   Professor, Dept. of Math, Southern University of Science and Technology (SUSTech)
  2013-2015 Professor, School of Management Science and Engineering, Nanjing University
  1997~2013 Professor, Department of Mathematics, Nanjing University
  1992~1997 Associate Professor, Department of Mathematics,Nanjing University

我的几类主要研究工作的分类论文和简要介绍（附阅读建议）

1. 变分不等式的投影收缩算法(Projection-Contraction Methods)
2. 两个可分离函数的乘子交替方向法(ADMM)
3. 多个可分离函数的交替方向类算法(ADMM-Like Methods)
4. 变分不等式框架下的邻近点算法(VI & PPA)

My Thinkings:

1. 关门感想
2. 说说我的主要研究兴趣 — 兼谈华罗庚推广优选法对我的影响
3. 说说我的主要研究兴趣(续) --- 我们在ADMM类方法的主要工作
4. 古稀回首
5. 两页纸简述我职业生涯中的主要研究工作

My Talks:

1. 从变分不等式的投影收缩算法到凸规划的分裂收缩算法 — 我研究生涯的来龙去脉
2. 生活理念对设计优化分裂算法的帮助 — 以改造 ADMM 求解三个可分离算子问题为例
3. 凸优化的分裂收缩算法 — 变分不等式为工具的统一框架 (适合打印的 综合文本)
4. 从商业谈判的角度看一些优化方法的设计 — 从 min-max 问题的求解谈起
5. 我和乘子交替方向法(ADMM) 的20年 — 2017 年5月全国数学规划会议报告 综述版本
6. 图像处理中的凸优化问题及其相应的分裂收缩算法 — ISICDM会议报告I 报告II 报告III
7. 介绍：构造求解凸优化的分裂收缩算法—用好变分不等式和邻近点算法两大法宝
8. 线性化ALM-ADMM等方法中的"替代"参数严重影响收敛速度—提升空间有多少？
9. 被 S. Becker 誉为 Very Simple yet Powerful Technique 的主要思想-应用及新的进展

我的报告的 PDF 文件, 都可以在我的主页上查到.

凸优化: 约束集合为闭凸集, 目标函数为凸函数.

连续性凸优化的一些数学模型

1. 简单约束凸优化问题 $\min\{f(x) \mid x \in \mathcal{X}\}$ 其中 $\mathcal{X}$ 是一个凸集.
2. 线性约束的凸优化问题 $\min\{\theta(x) | Ax = b \,(\text{or} \geq b),\, x \in \mathcal{X}\}$
3. 结构型凸优化 $\min\{\theta_1(x) + \theta_2(y) | Ax + By = b,\, x \in \mathcal{X},\, y \in \mathcal{Y}\}$
4. 变分不等式 $w^* \in \Omega,\ \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0,\ \forall w \in \Omega$

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.

报告介绍这些理念在鉴定、修正和设计算法方面的巨大功效

凸函数的定义和基本性质

A function $f(x)$ is convex iff

$$f((1-\mu)x+\mu y) \le (1-\mu)f(x)+\mu f(y)$$

$$\forall \mu \in [0,1].$$

Properties of convex function

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

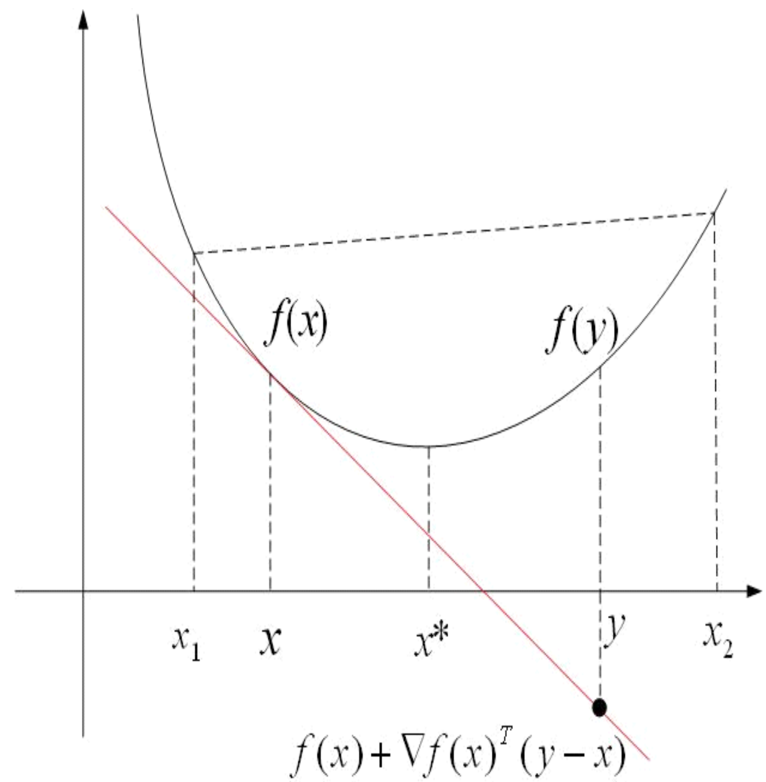
  $$f(y) - f(x) \ge \nabla f(x)^T (y-x).$$

  Thus, we have also

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

- Adding above two inequalities, we get

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



Convex function

- $f \in \mathcal{C}^1$, ∇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.

1 Optimization problem and VI （瞎子爬山）

1.1 Differential convex optimization in Form of VI

Let $\Omega \subset \Re^n$ be a closed convex set, we consider the convex minimization problem

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

What is the first-order optimal condition ?

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

Optimal condition in variational inequality form

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

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

瞎子爬山判定山顶的准则是: 所有可行方向都不再是上升方向.

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

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

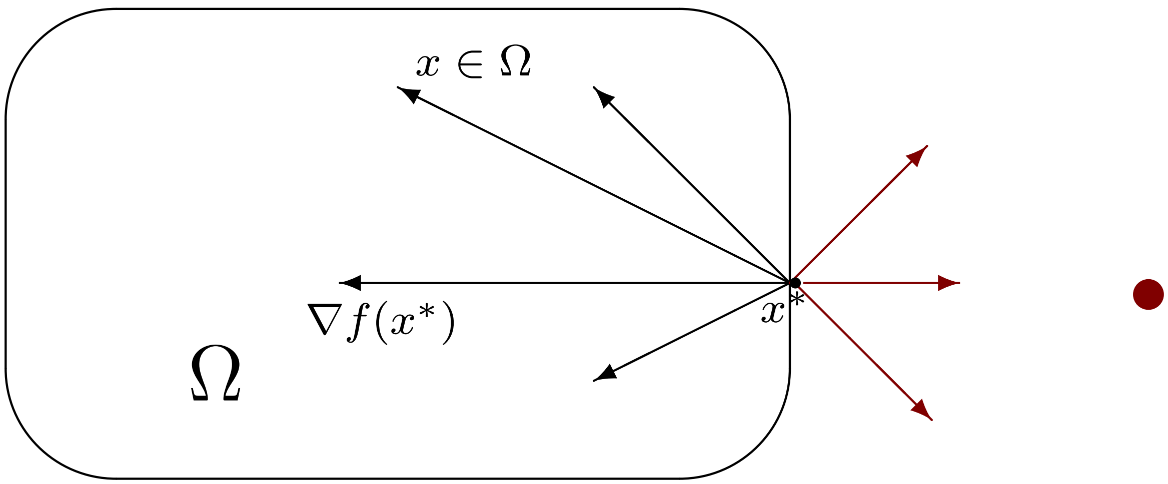


Fig. 1.1 Differential Convex Optimization and VI

The general form of variational inequality:

$$x^* \in \Omega, \quad (x - x^*)^T F(x^*) \geq 0, \quad \forall x \in \Omega. \tag{1.3}$$

When $(x - y)^T (F(x) - F(y)) \geq 0$, we say (1.3) is a monotone VI.

通篇我们需要用到的大学数学　主要是基于微积分学的一个引理

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

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

上面的凸优化最优性条件是最基本的, 合在一起就是下面的引理:

Lemma 1 *Let $\mathcal{X}\subset\Re^n$ be a closed convex set, $\theta(x)$ and $f(x)$ be convex functions and $f(x)$ is differentiable. Assume that the solution set of the minimization problem $\min\{\theta(x)+f(x)\,|\,x\in\mathcal{X}\}$ is nonempty. Then,*

$$x^*\in\arg\min\{\theta(x)+f(x)\,|\,x\in\mathcal{X}\} \tag{1.4a}$$

if and only if

$$x^*\in\mathcal{X},\quad \theta(x)-\theta(x^*)+(x-x^*)^T\nabla f(x^*)\geq 0,\quad \forall x\in\mathcal{X}. \tag{1.4b}$$

这样, 我们就把凸优化问题 (1.4a), 转换成了单调变分不等式 (1.4b).

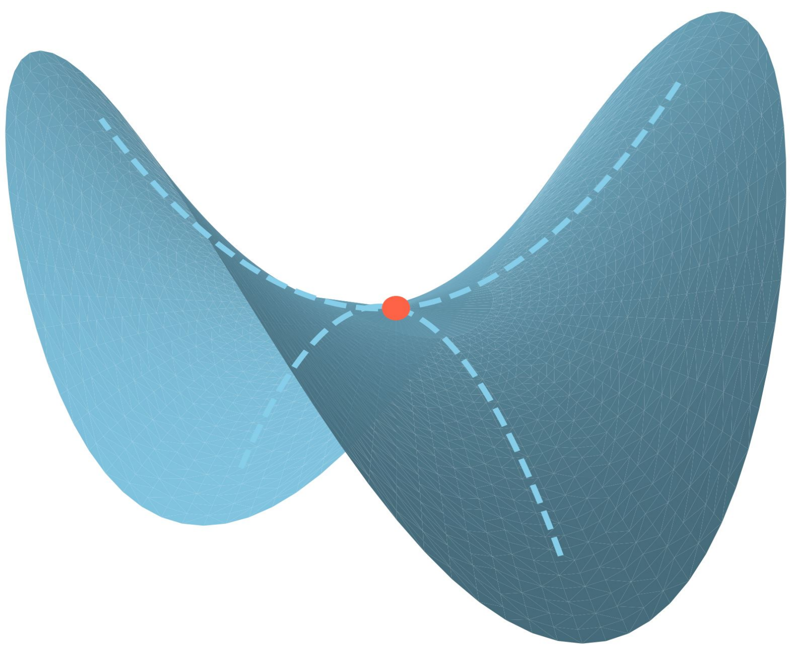
1.2 Linearly constrained Optimization in form of VI

We consider the linearly constrained convex optimization problem

$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{1.5}$$

The Lagrange function of (1.5) is

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \qquad (u,\lambda) \in \mathcal{U} \times \Re^m. \tag{1.6}$$



Saddle Point of the Lagrange Function

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

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

The above inequalities can be written as

$$\left\{\begin{array}{ll} u^* \in \mathcal{U}, \;\; L(u, \lambda^*) - L(u^*, \lambda^*) \ge 0, \;\; \forall\, u \in \mathcal{U}, & \text{(1.7a)} \\ \lambda^* \in \Re^m, \;\; L(u^*, \lambda^*) - L(u^*, \lambda) \ge 0, \;\; \forall\, \lambda \in \Re^m. & \text{(1.7b)} \end{array}\right.$$

In other words,

$$\left\{\begin{array}{ll} u^* \in \arg\min\{L(u, \lambda^*) \,|\, u \in \mathcal{U}\}, & \text{(1.8a)} \\ \lambda^* \in \arg\max\{L(u^*, \lambda) \,|\, \lambda \in \Re^m\}. & \text{(1.8b)} \end{array}\right.$$

According to the definition of $L(u, \lambda)$ (see(1.6)),

$$\begin{aligned} & L(u, \lambda^*) - L(u^*, \lambda^*) \\ = \;\; & [\theta(u) - (\lambda^*)^T(\mathcal{A}u - b)] - [\theta(u^*) - (\lambda^*)^T(\mathcal{A}u^* - b)] \\ = \;\; & \theta(u) - \theta(u^*) + (u - u^*)^T(-\mathcal{A}^T\lambda^*) \end{aligned}$$

it follows from (1.8a) that

$$u^* \in \mathcal{U}, \quad \theta(u) - \theta(u^*) + (u - u^*)^T(-\mathcal{A}^T\lambda^*) \geq 0, \quad \forall\, u \in \mathcal{U}. \qquad (1.9)$$

Similarly, for (1.8b), since

$$\begin{aligned} & L(u^*, \lambda^*) - L(u^*, \lambda) \\ & \quad = \quad [\theta(u^*) - (\lambda^*)^T(\mathcal{A}u^* - b)] - [\theta(u^*) - (\lambda)^T(\mathcal{A}u^* - b)] \\ & \quad = \quad (\lambda - \lambda^*)^T(\mathcal{A}u^* - b), \end{aligned}$$

we have

$$\lambda^* \in \Re^m, \quad (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \geq 0, \quad \forall\, \lambda \in \Re^m. \qquad (1.10)$$

Note that the expression (1.10) is equivalent to $\mathcal{A}u^* = b$.

Writing (1.9) and (1.10) together, we get the following variational inequality:

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

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega. \qquad \text{(1.11a)}$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \ \text{and} \ \ \Omega = \mathcal{U} \times \Re^m. \qquad \text{(1.11b)}$$

Because F is a affine operator and

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

The matrix is skew-symmetric, we have

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

线性约束凸优化问题 (1.5), 转换成了单调变分不等式 (1.11).

Convex optimization problem with two separable functions

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{1.12}$$

This is a special problem of (1.5) 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 (1.12) 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 is a solution of the following VI:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{1.13}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \tag{1.14a}$$

$$F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{1.14b}$$

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

线性约束凸优化问题 (1.12), 转换成了单调变分不等式 (1.13)–(1.14).

2 Proximal Point Algorithms （步步为营）

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

$$\|b\|_H^2 \le \|a\|_H^2 - \|a-b\|_H^2. \tag{2.1}$$

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

2.1 Proximal point algorithms for convex optimization

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

$$\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}, \tag{2.2}$$

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

For solving (2.2), the k-th iteration of the proximal point algorithm (abbreviated to

PPA) [16, 19] begins with a given x^k, offers the new iterate x^{k+1} via the recursion

$$x^{k+1} = \text{Argmin}\{\theta(x) + f(x) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \tag{2.3}$$

Since x^{k+1} is the optimal solution of (2.3), it follows from Lemma 1 that

$$\theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \{\nabla f(x^{k+1}) + r(x^{k+1} - x^k)\} \geq 0, \ \forall\, x \in \mathcal{X}. \tag{2.4}$$

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

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

It remains true by changing the last $\nabla f(x^{k+1})$ to $\nabla f(x^*)$. Thus, we have

$$(x^{k+1} - x^*)^T (x^k - x^{k+1}) \geq 0. \tag{2.5}$$

Let $a = x^k - x^*$ and $b = x^{k+1} - x^*$ and using Lemma 2, we obtain

$$\|\boldsymbol{x^{k+1}} - \boldsymbol{x^*}\|^{\mathbf{2}} \leq \|\boldsymbol{x^k} - \boldsymbol{x^*}\|^{\mathbf{2}} - \|\boldsymbol{x^k} - \boldsymbol{x^{k+1}}\|^{\mathbf{2}}, \tag{2.6}$$

which is the nice convergence property of Proximal Point Algorithm.

We write the problem (2.2) **and its PPA** (2.3) **in VI form**

For the optimization problem (2.2) , namely,

$$\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\},$$

the equivalent variational inequality form is

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.7}$$

For solving the problem (2.2), the variational inequality form of the k-th iteration of the PPA (see (2.4)) is:

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \nabla f(x^{k+1}) & \\ \geq (x - x^{k+1})^T r(x^k - x^{k+1}), \quad \forall\, x \in \mathcal{X}. & \end{aligned} \tag{2.8}$$

PPA 通过求解一系列的 (2.3), 求得 (2.2) 的解, 采用的是步步为营的策略.

The solution of (2.8) is Proximal Point, it has the contraction property (2.6).

2.2 Preliminaries of PPA for Variational Inequalities

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{2.9}$$

PPA for VI (2.9) in Euclidean-norm For given w^k and $r > 0$, find w^{k+1},

which satisfies

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) & \\ \geq (w - w^{k+1})^T r(w^k - w^{k+1}), \quad \forall\, w \in \Omega. & \end{aligned} \tag{2.10}$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (2.9).

✠ **w^k is the solution of (2.9) if and only if $w^k = w^{k+1}$** ✠

Setting $w = w^*$ in (2.10), we obtain

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

Note that (see the structure of $F(w)$ in (1.11b))

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

and consequently (by using (2.9)) we obtain

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

Thus, we have

$$(w^{k+1} - w^*)^T (w^k - w^{k+1}) \geq 0. \qquad (2.11)$$

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

the inequality (2.11) means that $\boldsymbol{b^T(a - b) \geq 0}$.

By using Lemma 2, we obtain

$$\|\boldsymbol{w^{k+1} - w^*}\|^2 \leq \|\boldsymbol{w^k - w^*}\|^2 - \|\boldsymbol{w^k - w^{k+1}}\|^2. \qquad (2.12)$$

We get the nice convergence property of Proximal Point Algorithm.

PPA for monotone mixed VI in H-norm

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

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq (w - w^{k+1})^T H(w^k - w^{k+1}), \ \forall\, w \in \Omega, \qquad (2.13)$$

where H is a symmetric positive definite matrix.

✠ Again, w^k is the solution of (2.9) if and only if $w^k = w^{k+1}$ ✠

Convergence Property of Proximal Point Algorithm in H-norm

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \qquad (2.14)$$

By using the block-matrix technique, we can get a proper positive definite matrix H. And the solutions of the subproblems (2.13) have the closed forms.

我们将见识这些理念在鉴定、修正和设计算法方面的巨大功效

3 From PDHG to Customized-PPA

We consider the convex optimization

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

The Lagrange function is

$$L(x,\lambda) = \theta(x) - \lambda^T(Ax-b), \qquad (x,\lambda) \in \mathcal{X} \times \Re^m. \tag{3.2}$$

The corresponding variational inequality is

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + \begin{pmatrix} x - x^* \\ \lambda - \lambda^* \end{pmatrix}^T \begin{pmatrix} -A^T\lambda^* \\ Ax^* - b \end{pmatrix} \ge 0, \quad \forall\, w \in \Omega.$$

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

In this section, we assume that the subproblem

$$\min\{\theta(x) + \frac{r}{2}\|x-a\|^2 \mid x \in \mathcal{X}\} \quad \text{is simple}.$$

3.1 Original primal-dual hybrid gradient algorithm [20]

For given (x^k, λ^k), **PDHG** [20] produces a pair of (x^{k+1}, λ^{k+1}). First,

$$x^{k+1} = \text{Argmin}\{L(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \tag{3.3a}$$

and then we obtain λ^{k+1} via

$$\lambda^{k+1} = \text{Argmax}\{L(x^{k+1}, \lambda) - \frac{s}{2}\|\lambda - \lambda^k\|^2 \mid \lambda \in \Re^m\}. \tag{3.3b}$$

Applying Lemma 1 to the subproblem (3.3a), we get

$$\theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T\{-A^T\lambda^k + r(x^{k+1} - x^k)\} \geq 0, \ \forall\, x \in \mathcal{X}. \tag{3.4}$$

The problem (3.3b) is an unconstrained optimization, thus we have

$$(Ax^{k+1} - b) + s(\lambda^{k+1} - \lambda^k) = 0, \tag{3.5}$$

and it can be written as

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

Combining (3.4) and (3.5), we get

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

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

where

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

The compact form is

$$\theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T \{F(w^{k+1}) + Q(w^{k+1} - w^k)\} \geq 0, \ \forall w \in \Omega, \tag{3.6}$$

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 (2.13), and we can not expect its convergence.

The following example of linear programming indicates the original PDHG (3.3) is not necessary convergent.

Consider a pair of the primal-dual linear programming:

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

We take the following example

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

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

$$L(x, y) = c^T x - y^T (Ax - b) \tag{3.7}$$

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

For solving the $\min$-$\max$ problem (3.7), by using (3.3), 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.

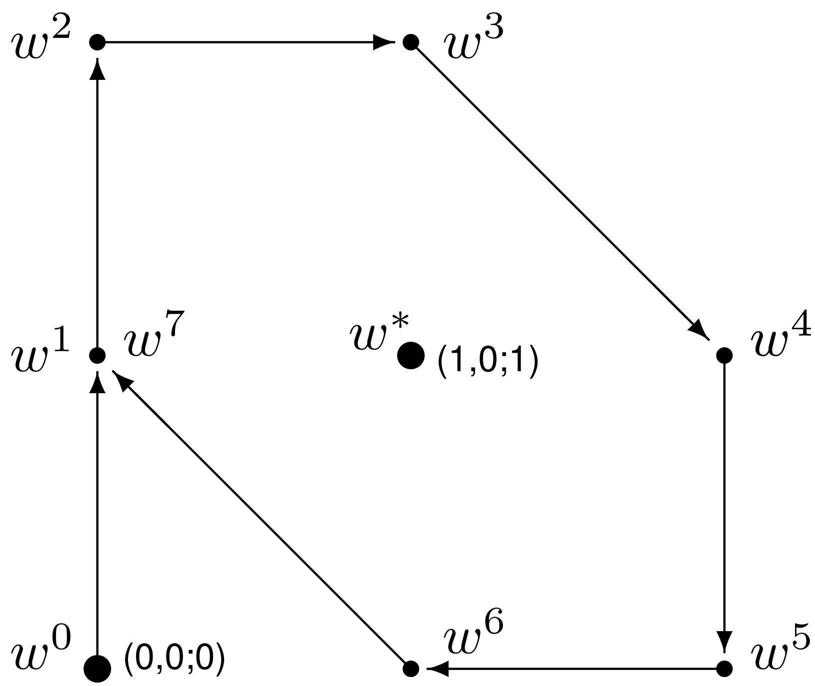
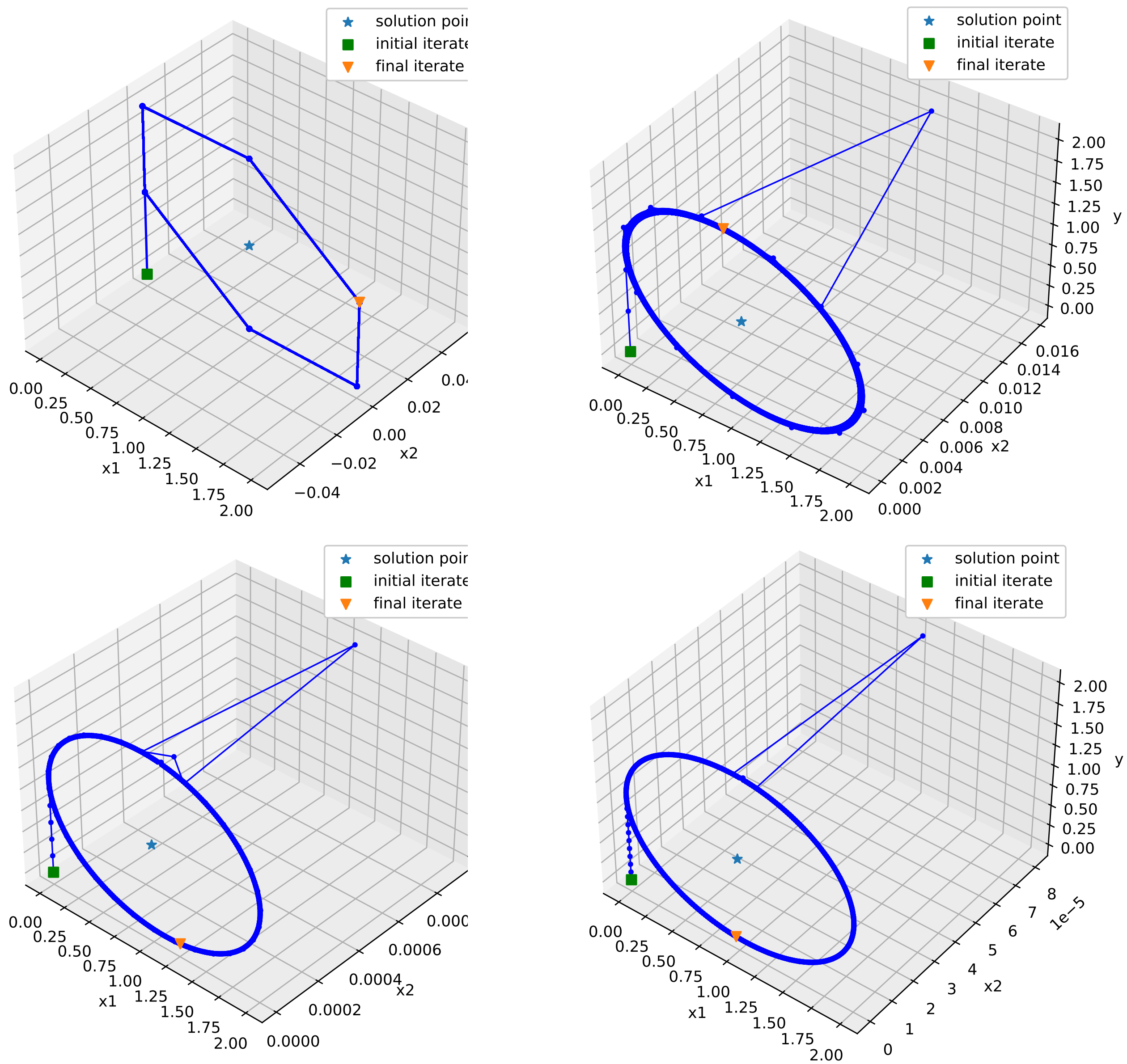


Fig. 4.1 The sequence generated by PDHG Method with $r = s = 1$

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

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



对 $r = s = 1, 2, 5, 10$, PDHG 方法都不收敛

3.2 Proximal Point Algorithm in H-norm

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

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

then the variational inequality (3.6) will become the following desirable form:

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

For this purpose, we need only to change (3.5) in PDHG, namely,

$$(Ax^{k+1} - b) + s(\lambda^{k+1} - \lambda^k) = 0,$$

to

$$(Ax^{k+1} - b) + A(x^{k+1} - x^k) + s(\lambda^{k+1} - \lambda^k) = 0. \tag{3.8}$$

Because x^{k+1} is known, with the given x^k and λ^k, λ^{k+1} in (3.8) is given by

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

Thus, for given (x^k, λ^k), producing a proximal point (x^{k+1}, λ^{k+1}) via (3.3a) and (3.8) can be summarized as:

$$\text{(PPA)} \quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\big\{L(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 \;\big|\; x \in \mathcal{X}\big\}, & \text{(3.9a)} \\ \lambda^{k+1} = \arg\max\big\{L\big([2x^{k+1} - x^k], \lambda\big) - \frac{s}{2}\|\lambda - \lambda^k\|^2\big\} & \text{(3.9b)} \end{array} \right.$$

By ignoring the constant term in the objective function, getting x^{k+1} from (3.9a) is equivalent to obtaining x^{k+1} from

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

The solution of (3.9b) is given by

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

Assumption: $\min\{\theta(x) + \frac{r}{2}\|x - a\|^2 \mid x \in \mathcal{X}\}$ **is simple**

Indeed, under the assumption, the sub-problem (3.9a) is simple.

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

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

Theorem 1 *The sequence $\{w^k = (x^k, \lambda^k)\}$ generated by the customized PPA satisfies*

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \tag{3.10}$$

For solving the $\min$-$\max$ problem (3.7), by using (3.9), the iterative formula is

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

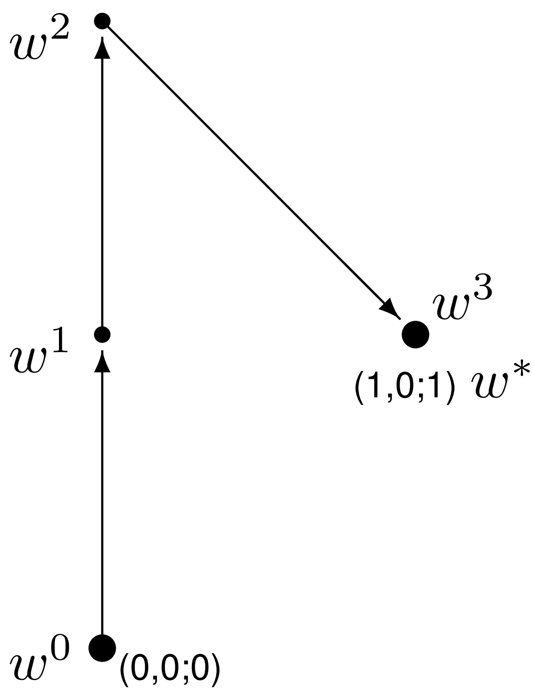
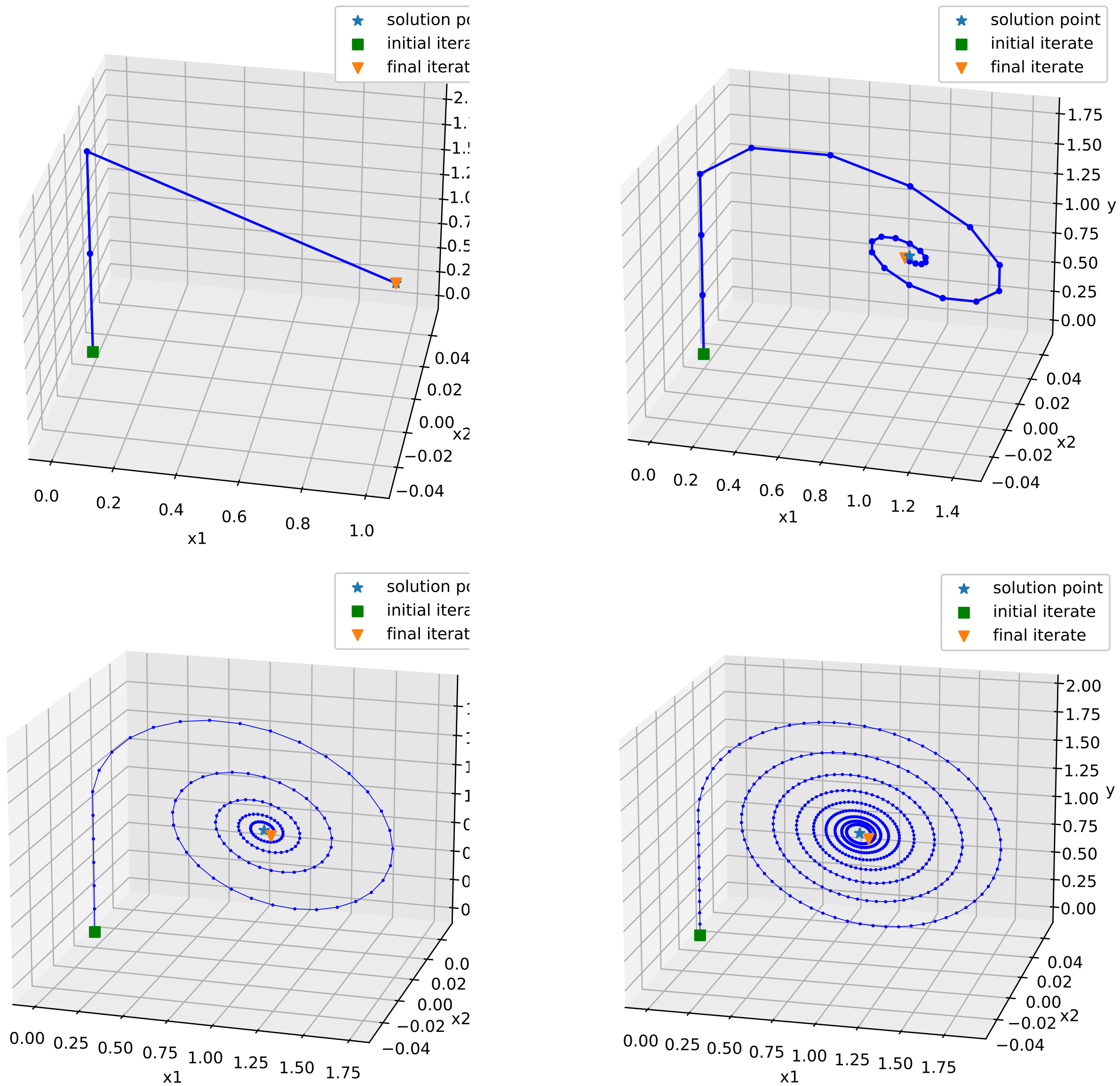


Fig. 4.2 The sequence generated by
C-PPA Method with $r = s = 1$

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



对 $r = s = 1, 2, 5, 10$, 按需定制的 PPA 方法都收敛

Remark Let the linear constraints become to a system of inequalities.

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

In this case, the Lagrange multiplier λ should be nonnegative. $\Omega = \mathcal{X} \times \Re^m_+$.

We need only to make a slight change in the prediction procedure:

In the primal-dual order:

$$\lambda^{k+1} = \lambda^k - \tfrac{1}{s}\big(A(2x^{k+1} - x^k) - b\big) \qquad \Rightarrow$$

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

3.3 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

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

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].

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 [13].

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

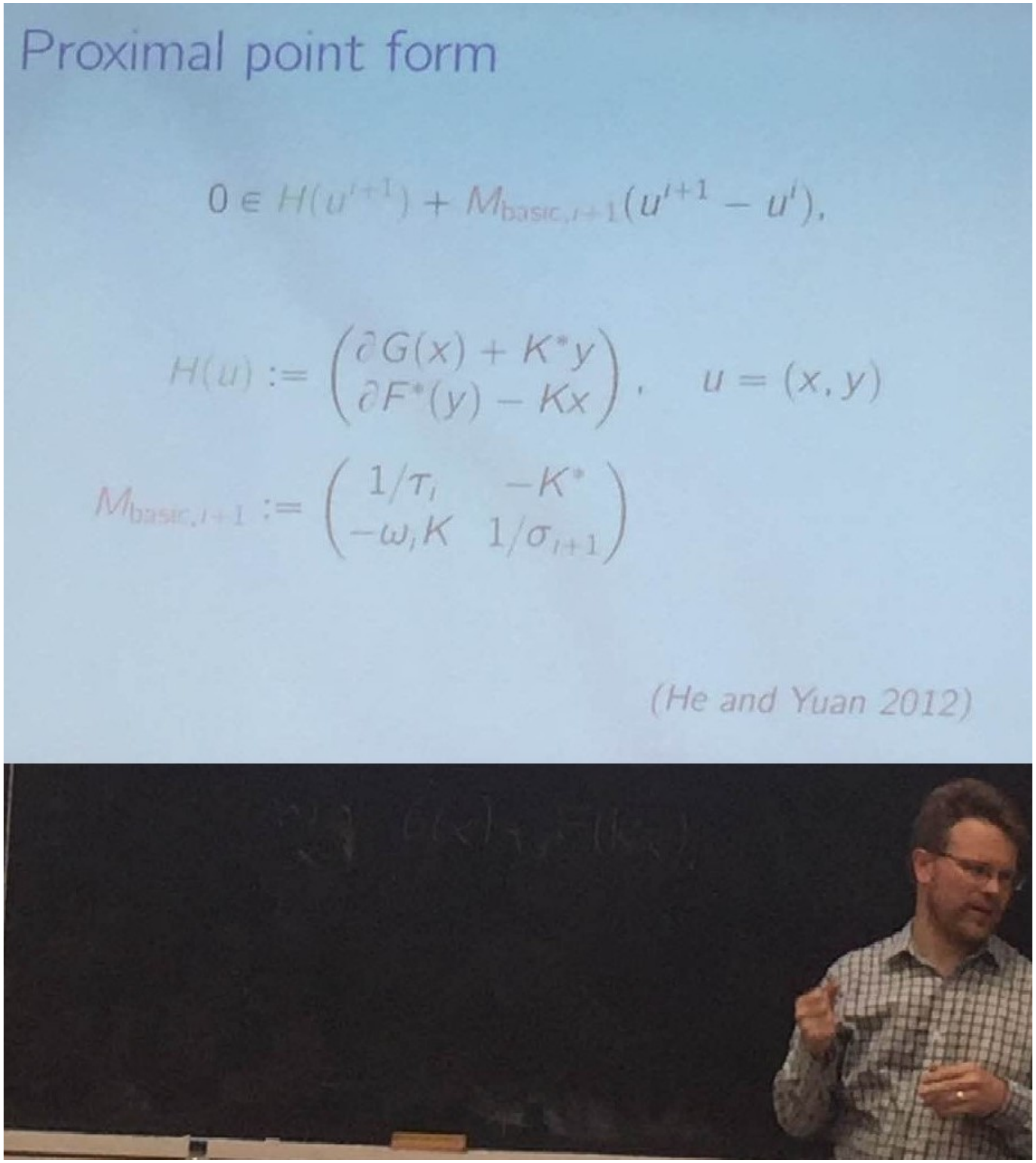
- Input: Operator norm $L := \|K\|$, Lipschitz constant L_f of ∇f, and Bregman distance functions D_x and D_y.
- Initialization: Choose $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$, $\tau, \sigma > 0$
- Iterations: For each $n \geq 0$ let

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

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。

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.

3.4 Relationship to Chambolle-Pock Method

Chambolle and Pock [2] have proposed a method for solving the convex-concave $\min - \max$ problem, in short, C-P method. Applied C-P method to the problem (3.1), it is also required $rs > \|A^T A\|$.

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

$$x^{k+1} = \arg\min\{L(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \tag{3.11a}$$

Then, λ^{k+1} is given by

$$\lambda^{k+1} = \arg\max\{L([x^{k+1} + \tau(x^{k+1} - x^k)], \lambda) - \frac{s}{2}\|\lambda - \lambda^k\|^2 \mid \lambda \in \Lambda\} \tag{3.11b}$$

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

- 原始-对偶混合梯度法(PDHG) (3.3) 和按需定制的邻近点算法(C-PPA) (3.9) 都是 Chambolle-Pock 方法 [2] 分别取 $\tau = 0$ 和 $\tau = 1$ 的特例.
- 对 $\tau = 0$ 的 PDHG 方法(3.3), §3.1 中已经说明不能保证收敛. 对 $\tau = 1$ 的 CPPA 方法(3.9), 其收敛性在 §3.2 中有了结论.
- 根据我们的知识, 对于 $\tau \in (0, 1)$ 的 CP 方法 (3.11), 收敛性还没有定论.

关于 CP 方法十年的一些小故事 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. 五个月后的 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 这么快就注意到并引用了我们的文章, 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.

4 Splitting Methods in a Unified Framework

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

4.1 Algorithms in a unified framework

Algorithmic Framework for VI (4.1)

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

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

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

[Correction Step.] Determine a nonsingular matrix M and the step size $\alpha > 0$, update the new iterate by

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \qquad (4.2b)$$

If $v^k = \tilde{v}^k$, $\tilde{w}^k$ is a solution of (4.1).

Convergence Conditions

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

$$HM = Q. \tag{4.3a}$$

For the given H, M and Q in (4.3a), and the step size α, the matrix

$$G = Q^T + Q - \alpha M^T HM \succ 0. \tag{4.3b}$$

4.2 Methods for Linearly Constrained Problems

We consider the convex optimization, namely

$$\min\{\theta(u) \mid Au = b, \; u \in \mathcal{U}\}. \tag{4.4}$$

The related Lagrange function is

$$L(u, \lambda) = \theta(u) - \lambda(Au - b),$$

and the corresponding variational inequality is

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

where

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

采用前一节的 PPA 方法

$$\text{(PPA)} \quad \left\{ \begin{array}{ll} u^{k+1} = \arg\min\big\{L(u,\lambda^k) + \frac{r}{2}\|u - u^k\|^2 \;\big|\; u \in \mathcal{U}\big\}, & \text{(4.5a)} \\ \lambda^{k+1} = \arg\max\big\{L\big([2u^{k+1} - u^k], \lambda\big) - \frac{s}{2}\|\lambda - \lambda^k\|^2\big\} & \text{(4.5b)} \end{array} \right.$$

Customized PPA

For given $v^k = w^k = (u^k, \lambda^k)$, the predictor is given by

$$\text{(C-PPA)} \quad \begin{cases} \tilde{u}^k = \arg\min\big\{L(u,\lambda^k) + \frac{r}{2}\|u - u^k\|^2 \;\big|\; u \in \mathcal{U}\big\}, & \text{(4.6a)} \\ \tilde{\lambda}^k = \arg\max\big\{L\big([2\tilde{u}^k - u^k], \lambda\big) - \frac{s}{2}\|\lambda - \lambda^k\|^2\big\} & \text{(4.6b)} \end{cases}$$

The output $\tilde{w}^k \in \Omega$ of the iteration (4.6) satisfies

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

It is a form of (4.2a) where

$$Q = \begin{pmatrix} rI & A^T \\ A & sI \end{pmatrix} \quad \text{is symmetric}$$

The new iterate is updated by

$$w^{k+1} = w^k - \alpha(w^k - \tilde{w}^k), \quad \alpha \in (0,2). \tag{4.7}$$

The subproblem (4.6a) is a problem of mathematical form

$$\min\{\theta(u) + \frac{r}{2}\|u - a^k\|^2 \,|\, u \in \mathcal{U}\} \tag{4.8}$$

where $r > 0$ is a given scalar and $a^k = u^k + \frac{1}{r}A^T\lambda^k$

According to the correction (4.2b)

the matrix M in the updated form (4.7) is the identity matrix I .

Then, we have and

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

The convergence conditions (4.3) are satisfied. More about C-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.

4.3 Convergence proof in the unified framework

Theorem 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*

$$\alpha\big\{\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\big\} \\ \geq \ \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{\alpha}{2}\|v^k - \tilde{v}^k\|_G^2, \ \forall w \in \Omega. \quad (4.9)$$

Proof. Using $Q = HM$ (see (4.3a)) and the relation (4.2b), the right hand side of (4.3a) can be written as $(v - \tilde{v}^k)^T \frac{1}{\alpha} H(v^k - v^{k+1})$ and hence

$$\alpha\{\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\} \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \ \forall w \in \Omega. \quad (4.10)$$

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 (4.10) with

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

we thus obtain

$$\begin{aligned} & 2(v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ & \quad = \big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + (\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \end{aligned} \tag{4.11}$$

For the last term of (4.11), we have

$$\begin{aligned} & \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ & \quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ & \quad \overset{\text{(4.3a)}}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - \alpha M(v^k - \tilde{v}^k)\|_H^2 \\ & \quad = \quad (v^k - \tilde{v}^k)^T \big[\alpha(Q^T + Q) - \alpha^2 M^T H M\big](v^k - \tilde{v}^k) \\ & \quad \overset{\text{(4.3b)}}{=} \quad \alpha\|v^k - \tilde{v}^k\|_G^2. \end{aligned} \tag{4.12}$$

Substituting (4.11), (4.12) in (4.10), the assertion of this theorem is proved. $\square$

Theorem 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 ($G \succ 0$), then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{4.13}$$

Proof. Set $w = w^*$ in (4.9), the assertion follows directly from

$$\begin{aligned} & \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \\ = \quad & \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \ge 0. \qquad \Box \end{aligned}$$

用这个框架, 容易证明ADMM 算法遍历意义和点列意义下的收敛速率.

- 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.
- 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.

5 Special prediction-correction methods

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{5.1}$$

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

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \; \forall w \in \Omega, \tag{5.2a}$$

where the matrix H is symmetric and positive definite.

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

$$v^{k+1} = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha \in (0, 2) \tag{5.2b}$$

H is a symmetric positive definite matrix.

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

The sequence $\{v^k\}$ generated by the prediction-correction method (5.2) satisfies

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

The above inequality is the Key for convergence analysis !

Set $\alpha = 1$ in (4.2b), the prediction (5.2a) becomes: $w^{k+1} \in \Omega$ such that

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

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

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

上式是跟 (2.14) 类似的不等式, 是关于核心变量 v 的 PPA 方法.

5.2 Applications for separable problems

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

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\}. \tag{5.1}$$

Its VI-form is

$$w^*\in\Omega, \quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0, \quad \forall\, w\in\Omega. \tag{5.2}$$

where

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix}, \quad u=\begin{pmatrix} x\\ \\ y\end{pmatrix}, \quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}, \tag{5.3a}$$

and

$$\theta(u)=\theta_1(x)+\theta_2(y), \qquad\qquad \Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m. \tag{5.3b}$$

The augmented Lagrangian Function of the problem (5.1) is

$$\mathcal{L}_\beta(x,y,\lambda) = \theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)+\frac{\beta}{2}\|Ax+By-b\|^2. \tag{5.4}$$

Solving the problem (5.1) by using ADMM, the k-th iteration begins with given (y^k,λ^k), it offers the new iterate (y^{k+1},λ^{k+1}) via

$$\text{(ADMM)}\quad \left\{ \begin{array}{ll} x^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x,y^k,\lambda^k)\;\big|\;x\in\mathcal{X}\big\}, & \text{(5.5a)} \\ y^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x^{k+1},y,\lambda^k)\;\big|\;y\in\mathcal{Y}\big\}, & \text{(5.5b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1}+By^{k+1}-b). & \text{(5.5c)} \end{array} \right.$$

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

The main convergence result is

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

where

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

Ignoring some constant term in the objective function, ADMM (5.5) **is implemented by**

(ADMM)

$$\begin{cases} x^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T p^k \\ +\frac{\beta}{2}\|A(x - x^k)\|^2. \end{array} \middle| \, x \in \mathcal{X} \right\}, & \text{(5.6a)} \\ y^{k+1} = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T q^k \\ +\frac{\beta}{2}\|B(y - y^k)\|^2. \end{array} \middle| \, y \in \mathcal{Y} \right\}, & \text{(5.6b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(5.6c)} \end{cases}$$

where

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

根据给定的 Special 框架 (5.2)，设计预测公式.

5.3 ADMM in PPA-sense

In order to solve the separable convex optimization problem (5.1), we construct a method whose prediction-step is

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{5.7a}$$

where

$$H = \left[\begin{array}{cc} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{array} \right], \quad \text{(a small } \delta > 0\text{, say } \delta = 0.05\text{).} \tag{5.7b}$$

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

The concrete form of (5.7) is

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

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

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k)\,|\,x\in\mathcal{X}\}, & \text{(5.8a)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), & \text{(5.8b)} \\ \tilde{y}^k = \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T [\boldsymbol{2\tilde{\lambda}^k - \lambda^k}] \\ +\frac{1+\delta}{2}\beta\|B(y-y^k)\|^2 \end{array} \middle|\, y\in\mathcal{Y} \right\}. & \text{(5.8c)} \end{cases}$$

这个预测与ADMM (5.6)中子问题难度一样, 采用(5.2b) 校正, 会加快收敛.

5.4 Linearized ADMM-Like Method

By using the linearized version of (5.8), the prediction step becomes

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (v-\tilde{v}^k)^T H(v^k-\tilde{v}^k), \ \forall w \in \Omega, \quad (5.9)$$

where

$$H = \begin{bmatrix} sI & -B^T \\ -B & \frac{1}{\beta} I_m \end{bmatrix}, \text{ 代替 (5.7) 中的 } \begin{bmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{bmatrix}. \quad (5.10)$$

The concrete formula of (5.9) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T \tilde{\lambda}^k}\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T \tilde{\lambda}^k} + \ \boldsymbol{s}(\tilde{y}^k - y^k) \boldsymbol{- B^T} (\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b}) \boldsymbol{- B}(\tilde{y}^k - y^k) + \boldsymbol{(1/\beta)}(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases} \quad (5.11)$$

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

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

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (5.11),

we need only use the following procedure:

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

用 $\frac{s}{2}\|y-y^k\|^2$ 代替 $\frac{1+\delta}{2}\beta\|B(y-y^k)\|^2$, 为保证收敛, 需要 $s > \beta\|B^TB\|$.

对给定的 $\beta > 0$，要求 $s > \beta\|B^TB\|$，太大的 s 会影响收敛速度

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}.

Conclusions

- 变分不等式(VI) 和邻近点策略(PPA) 是凸优化算法的两大法宝.
- 用变分不等式框架照一照, 很快就能知道一个“想当然的方法”收敛性到底可靠不可靠. 修改策略也随之而来了.
- 我们提供的算法框架, 收敛性证明用到的知识非常简单, 证明篇幅也很短. 这个框架, 可以帮助思考如何去设计效率更高的方法.
- 根据“邻近点策略”设计的算法, 有“步步为营”的稳健, 也有一阶算法收敛慢的困扰.
- 我的观点, 正确的你就接受; 胡说八道的, 你就付之一笑！

References

[1] 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.

[2] 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.

[3] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, Mathematical Programming, 155 (2016) 57-79.

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

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

[6] 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] D. Hallac, Ch. Wong, S. Diamond, A. Sharang, R. Sosi$\breve{c}$, S. Boyd and J. Leskovec, SnapVX: A Network-Based Convex Optimization Solver, Journal of Machine Learning Research 18 (2017) 1-5.

[8] 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] 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.

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

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

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

[17] J. Nocedal and S. J. Wright, Numerical Optimization, Springer Verlag, New York, 1999.

[18] 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.

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

[20] 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.

Thank you very much for your attention！！