凸优化研究中追求简单与统一

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

南京大学数学学院

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

江苏省运筹学会 2025 年 11 月 1 日

获颁江苏运筹《特殊贡献奖》的年会上的报告

四十年学术生涯给我的一些感悟

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

✠ 先贤名言，铭记在心。纵然可以没有本领，也不迷失价值标准 ✠

爱美之心，人皆有之。数学往往被人认为是枯燥的，计算更是被人看作是繁琐的。领悟了数学之美的数学工作者，他的职业生涯才可能是充满乐趣的。

✠ 世上三百六十行，热爱自己的职业，应是同一个道理 ✠

数学之美，不是纯数学的专利。为应用服务的最优化方法研究，同样可以追求简单与统一。简单，他人才会看懂使用；统一，自己才有美的享受。

✠ 发现了其中的数学之美，研究才变得满怀激情和欲罢不能 ✠

文革期间华罗庚先生普及双法对一位回乡知青的影响

- 华罗庚先生当年普及的双法— 统筹法和优选法。　　普及双法以优选法为主。
- 先生强调要“牢记把方法交给群众”。
  —华罗庚《数学工作者要大力为农业服务》
  人民日报 1960年10月30日
- 这成为从上世纪60年代开始的近20年间, 华罗庚从事数学普及工作的指导思想。
  — 王元《华罗庚》
- 随着全民族文化水平的提高, 群众有了新的定义. 提供工程师们容易掌握的方法, 可以作为部分优化学者的工作目标.

优选法求解单峰函数的极值点, 我们同样只致力于比较简单的凸优化问题的求解.

宣传优选法小分队和华先生本人的报告, 使我走上优化方法研究之路.

何旭初 (1921-1990)

南京大学何旭初教授

何旭初先生是国内优化界的前辈和当年领袖人物之一。

我大学毕业的时候, 想考研究生, 但我的英语水平太差, 何先生知道后, 说也可以考我尚能应付的俄语, 让我继续求学成为可能。

我考上了南大研究生后, 何先生又把学校分配到他名下的公派留学生名额给了我(当年家父曾问我们姓何是否也是其中一个原因)。

按要求必须先通过英语水平考试, 再速成学点德语, 去德国跟何先生为我选定的导师Stoer学习。

我从德国博士毕业回南京大学, 在何旭初先生直接领导下工作了几年

祝愿我的博士导师巴伐利亚科学院院士 Stoer 教授夫妇健康长寿

我的德国导师，是巴伐利亚科学院院士。今年 91 岁的他，身体健康，导师 80 岁时，他们夫妇和子女们的合影。

他根据我的实际情况，尊重我的个人兴趣，给我许多指导和帮助。

我获得博士学位以后，就回到母校南京大学工作。

我的中外两位导师都跟冯康先生有深厚的友谊.

何旭初先生跟冯康先生是中央大学同学. 我有幸聆听过冯先生受何先生邀请来南大数学系作的报告, 目睹大师讲课和书写黑板的风采. 成了我至今作报告希望有黑板的原因.

Stoer 教授非常尊敬冯康先生, 多次邀请冯先生顺访欧洲时讲学并请到自己家里做客. 他对冯康先生独立于西方创造了有限元方法有极高评价.

冯先生的算法价值标准

影响我一生的研究实践

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

打开我的主页，能看到我受教育的经历和工作经历，一些学业感悟和报告的PDF文件，我会维护和不断更新.

School of Mathematics,
Nanjing University,
Nanjing, 210093, China
E-mail: hebma@nju.edu.cn

Current Research Areas:
Mathematical Programming, Numerical Optimization,
Variational Inequalities, Projection and contraction methods for VI,
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：
2015-2020 Professor, Dept. of Math, Southern University of Science aand 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

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. 古稀回首 4. 说说我的主要研究兴趣(续) --- 我们在ADMM类方法的主要工作
5. 若只给我两页●那我就说这些

两页短文A-H的导读 A. 我的主要研究历程-从变分不等式的投影收缩算法到凸优化的分裂收缩算法
B. PC方法中的孪生方向和姊妹方法 C. 两页纸给出ADMM 收敛性证明
D. 凸优化问题相应的变分不等式以及为几类典型问题构造的邻近点算法
E. 统一框架下求解两个可分离块凸优化问题的 效率更高的ADMM算法
F. 统一框架下求解三个可分离块凸优化问题的算法-高斯回代和正则化方法
G. 古稀后的研究进展-I: 从好不容易凑出一个方法到并不费力构造一簇算法
H. 古稀后的研究进展-II：预测-校正的广义邻近点(Generalized PPA)算法

A-H八个单元扩展成较为详细并加些例子说明的20个两页短篇●试述自己一生主要学术历程与收获
中文概要的五个英文两页短篇 A-H八单元的配套习题 英文概要的五个英文两页短篇
《数学文化》2020 年/ 第 11 卷第 2 期刊登的我的自述: 四十年上下求索 一份珍贵的回忆材料

My Talks: 比较系统的知识建议阅读第3个报告. 也建议阅读最近的一些系列报告

For more systematic knowledge, it is recommended to read Talk 3, which is written in English.

23. 2024 年 春季在华师大_上海大学_复旦大学_南京应用数学中心_西工大_西交大_西电的报告 导读 报告
22. 2024 年2月在三亚“数学、图像科学与人工智能”研讨会的两个报告的内容摘要 报告I 报告II
21. 从PPA-ADMM到凸优化的广义PPA (2023年12 月6日应西电数学与统计学院邀请的90分钟的报告)
20. 2023 年11月在 曲阜师大两次（共计四小时）课程的PPT 两次课程的内容摘要 课程I 课程II
19. 2023 年10月在天元数学东北中心八次课程的汇总讲义 前言与目录 I II III IV V VI VII VIII

1 人生经历影响了我的价值取向和数学思维方式

说起函数连续和导数的关系, 一些结论理科大学生还是记得的.

例如, 二元函数在某一点有偏导数, 而函数在这一点不一定连续, 让学生随口举个例子就不一定会. 我自己学生时代想到这样一个例子:

$$f(x,y) = \begin{cases} -1, & xy = 0,\ \ (x,y)\text{ 在坐标轴上;} \\ 0, & xy \neq 0,\ \ (x,y)\text{ 不在坐标轴上.} \end{cases}$$

这样的 $f(x,y)$ 在原点的值为 -1, 在原点的任意小的邻域内都有值为 0 的点, 因此在原点这一点不连续.

但在 x 轴和 y 轴都是常数 -1, 因此在 $(0,0)$ 点偏导数存在, 均为 0.

跟几位教分析的老师交流, 都说这个例子不错, 学生容易懂, 就是教科书上没有看到过. 我不信是我第一个提出这样的例子. 当年就有辅导老师问我怎么想到这样的例子的, 我回答说那就得感谢我那 10 多年的种田经历了.

江南水乡是稻麦两季, 水稻要水, 麦子怕渍. 水稻成熟后, 要把田里的水排去, 让土地凉干. 待水稻收割后种麦子, 要在地里纵横开沟. 为多打一些粮食, 这沟开的还有些讲究.

沟要开的直, 沟底要平, 才便于灦水, 沟底可以看作是光滑的; 为了增加麦子生长的有效面积, 沟也要开的窄(不能如左图那么宽).

水田是平的, 那些开成的沟的纵横交叉处, 就可以看做有偏导数而不连续的点.

自己有了这样的例子, 书本上的例子我再也记不住.

这也并非好事, 说明我学习理解领会他人成果的能力有问题.

一生的代表性工作: 投影收缩算法和ADMM类方法

2 投影收缩算法中的简单与统一

2.1 由投影得到的三个基本不等式

设 $\Omega \subset \Re^n$ 是一个非空闭凸集, F 是 $\Re^n \to \Re^n$ 的一个映射. 变分不等式是求

$$\text{(VI)} \qquad u^* \in \Omega, \quad (u-u^*)^T F(u^*) \geq 0, \quad \forall\, u \in \Omega. \tag{2.1}$$

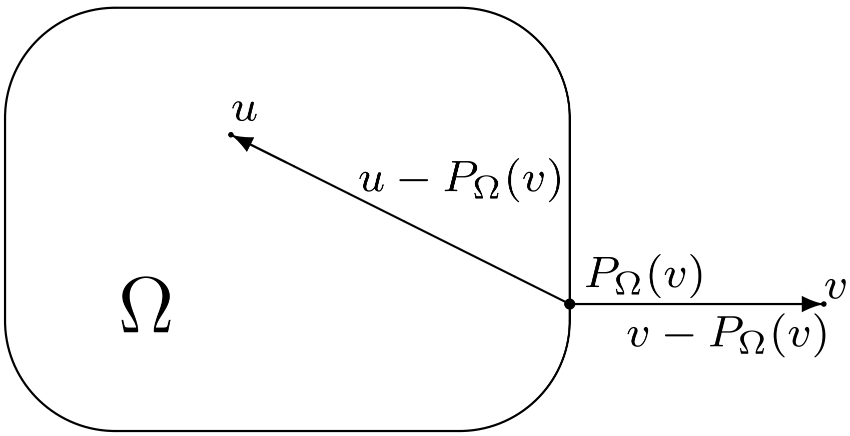
我们说一个变分不等式单调, 是指其中的算子 F 满足

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

在求解变分不等式(2.1)的投影收缩算法中, 对给定的当前点 u^k 和常数 $\beta > 0$, 我们利用投影

$$\tilde{u}^k = P_\Omega[u^k - \beta F(u^k)] \tag{2.2}$$

生成一个预测点 $\tilde{u}^k$.



假如 $\tilde{u}^k = u^k$, u^k 就是变分不等式 (2.1) 的解. $\|u^k - \tilde{u}^k\|^2$ 常用来度量误差.

此外, 凸集上投影有如右上图所示的重要性质: 对任意的 $v \in \Re^n$, 有

$$(u - P_\Omega(v))^T (v - P_\Omega(v)) \le 0, \quad \forall u \in \Omega. \tag{2.3}$$

基于变分不等式的定义 (2.1) 和投影的基本性质 (2.3), 我们有三个基本不等式:

由于 $\tilde{u}^k \in \Omega$, 根据变分不等式 (2.1) 的定义, 对任意的 $\beta > 0$, 都有

(FI-1) $$(\tilde{u}^k - u^*)^T \beta F(u^*) \ge 0, \quad \forall u^* \in \Omega^*. \tag{2.4}$$

在 (2.3) 中令 $v = u^k - \beta_k F(u^k)$, 那么由 (2.2) 得到 $\tilde{u}^k = P_\Omega(v)$, 并将那个属于 Ω 的 u 取成 u^*, 就得到

(FI-2) $$(\tilde{u}^k - u^*)^T \big(u^k - \beta F(u^k) - \tilde{u}^k\big) \ge 0, \quad \forall u^* \in \Omega^*. \tag{2.5}$$

根据单调性, 我们有

(FI-3) $$(\tilde{u}^k - u^*)^T \beta\big(F(\tilde{u}^k) - F(u^*)\big) \ge 0, \quad \forall u^* \in \Omega^*. \tag{2.6}$$

我们称它们为投影收缩算法中的三个基本不等式.

2.2 由 (FI-1)+(FI-2) 生成求解线性变分不等式的寻查方向

线性变分不等式中, 算子 $F(u)=Mu+q$, 其中 M 为满足 $M^T+M\succeq 0$ 的方阵, q 为相应的向量. (FI-1) 和 (FI-2) 就分别成了

$$(\tilde{u}^k-u^*)^T\beta(Mu^*+q)\ge 0,\quad \forall u^*\in\Omega^* \tag{2.7}$$

和

$$(\tilde{u}^k-u^*)^T\{(u^k-\tilde{u}^k)-\beta(Mu^k+q)\}\ge 0,\quad \forall u^*\in\Omega^*. \tag{2.8}$$

将这两式相加并将左边的 $(\tilde{u}^k-u^*)$ 拆解成 $\{(u^k-u^*)-(u^k-\tilde{u}^k)\}$, 就有

$$\{(u^k-u^*)-(u^k-\tilde{u}^k)\}^T\{(u^k-\tilde{u}^k)-\beta M(u^k-u^*)\}\ge 0,\quad \forall u^*\in\Omega^*.$$

利用单调变分不等式算子的单调性, $(u^k-u^*)^TM(u^k-u^*)\ge 0$, 从上式得到

$$\boxed{(u^k-u^*)^T\{(I+\beta M^T)(u^k-\tilde{u}^k)\}\ge\|u^k-\tilde{u}^k\|^2,\quad \forall u^*\in\Omega^*.} \tag{2.9}$$

不等式 (2.9) 表明由 (FI-1) + (FI-2) 导出的向量 $(I+\beta M^T)(u^k-\tilde{u}^k)$ 是距离函数 $\|u-u^*\|^2$ 在 u^k 处的上升方向. 我们可以选择适当的 $\alpha>0$, 通过

$$u^{k+1}(\alpha) = u^k - \alpha(I + \beta M^T)(u^k - \tilde{u}^k). \tag{2.10}$$

生成欧氏模下比 u^k 更靠近解集的, 依赖于步长 α 的新的迭代点 u^{k+1}.

2.3 由 (FI-1)+(FI-2)+(FI-3) 生成求解非线性变分不等式的寻查方向

对非线性单调变分不等式, 在生成预测点 $\tilde{u}^k$ 的投影 (2.2) 中, 要求参数 β_k 取得满足

$$\beta_k \|F(u^k) - F(\tilde{u}^k)\| \le \nu \|u^k - \tilde{u}^k\|, \quad \nu \in (0, 1). \tag{2.11}$$

对选定的 $\beta_k > 0$, 由 (FI-1)+(FI-2)+(FI-3), 我们得到

$$\{(u^k - u^*) - (u^k - \tilde{u}^k)\}^T \{(u^k - \tilde{u}^k) - \beta_k(F(u^k) - F(\tilde{u}^k))\} \ge 0, \quad \forall u^* \in \Omega^*.$$

利用记号

$$d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k(F(u^k) - F(\tilde{u}^k)), \tag{2.12}$$

我们得到

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k), \quad \forall u^* \in \Omega^*. \tag{2.13}$$

根据参数 β_k 满足的条件 (2.11) 和 Cauchy-Schwarz 不等式, (2.13) 右端的

$$\underline{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \geq (1-\nu)\|u^k - \tilde{u}^k\|^2},$$

上式跟 (2.13) 表明由 (FI-1) + (FI-2) + (FI-3) 生成的 $d(u^k, \tilde{u}^k)$ 是欧氏模下距离函数 $\|u - u^*\|^2$ 在 u^k 处的上升方向.

我们可以选择适当的 $\alpha > 0$, 通过

$$u^{k+1}(\alpha) = u^k - \alpha d(u^k, \tilde{u}^k) \tag{2.14}$$

生成欧氏模下比 u^k 更靠近解集的, 依赖于步长 α 的新的迭代点 u^{k+1}.

2.4 相关论文得到的重视

不同投影收缩算法的所有搜索方向都出自三个基本不等式的不同组合.

- B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization (1997), 35: 69-76.

ISI® Citation Classic Award

Presented to

HE BING-SHENG

In recognition of influential research reflected in the publication of a highly cited paper from the period 1981-1998

25 September 2000
Beijing, China

ISI
THOMSON SCIENTIFIC™

Adam Klein, Senior Director
ISI Asia-Pacific Division

因为简单, 1997 年发表的关于三个基本不等式的文章, 入选 1981-1998 经典引文奖

投影收缩算法中的孪生方向, 姊妹方法一直是我多年来津津乐道的收获

2.5 其他领域的应用

- 被美国科学院和工程院院士, 2018 年世界数学家大会一小时被告人加州伯克利 **M. I. Jordan** 教授在内的学者引用. 一些相当细致的计算法则也被 **Jordan** 教授课题组借鉴, 相关的定理被写进他们论文的附录。
- 国内, 除了中科院岩土所的科技工作者将投影收缩算法成功用于许多岩土工程问题的求解以外, 这类方法也被成功应用到机器人的实时控制、电力系统以及航空航天领域的结构问题求解。
- 能被工程界应用的直接原因就是方法简单实用!

是不是所有的人都希望简单统一的? 当然不是! 道理不言自明.

3 Mathematical Preliminaries of Convex OP

3.1 Differential convex optimization in Form of VI

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

$$\min\{f(x) \mid x \in \Omega\}. \tag{3.1}$$

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.

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

推广优选法所说的瞎子爬山, 就是指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:

$$x^* \in \Omega, \quad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \Omega. \tag{3.2}$$

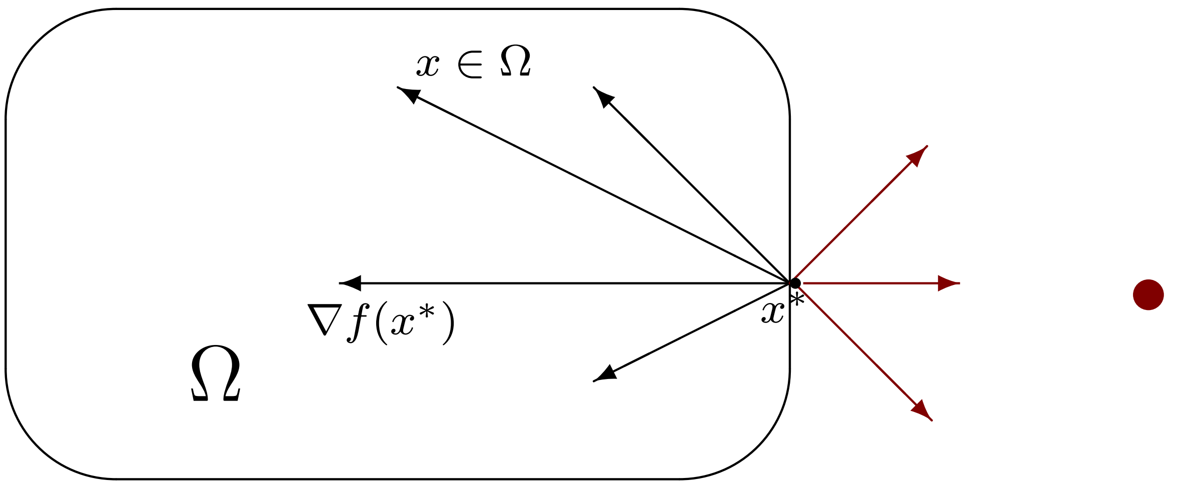


Fig. 1.1 Differential Convex Optimization and VI

The general form of variational inequality:

$$x^* \in \Omega, \quad (x - x^*)^T F(x^*) \geq 0, \quad \forall x \in \Omega. \tag{3.3}$$

When $(x-y)^T(F(x) - F(y)) \geq 0$, we say (3.3) is a monotone VI.

我们需要用到的普通的大学数学: 主要是基于微积分学的一个定理

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

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

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:

Theorem 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,*

$$x^*\in\arg\min\{\theta(x)+f(x)\,|\,x\in\mathcal{X}\} \tag{3.4a}$$

if and only if

我们称之为配对定理

$$x^*\in\mathcal{X},\ \ \theta(x)-\theta(x^*)+(x-x^*)^T\nabla f(x^*)\geq 0,\ \ \forall\, x\in\mathcal{X}. \tag{3.4b}$$

这样, 我们就把凸优化问题 (3.4a), 转换成了单调变分不等式 (3.4b).

3.2 Convex optimization and its related VI

我们需要用到的一般的优化原理：拉格朗日函数的鞍点及其性质

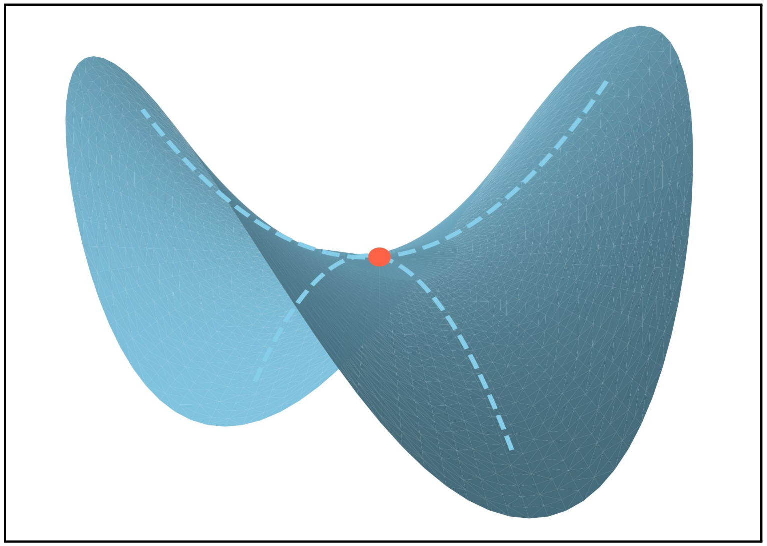
We consider the linearly constrained convex optimization problem

$$\min\{\theta(u) \mid \mathcal{A}u = b, \ u \in \mathcal{U}\}. \qquad (3.5)$$

The Lagrangian function of the problem (3.5) is

$$L(u, \lambda) = \theta(u) - \lambda^{\top}(\mathcal{A}u - b), \qquad (3.6)$$

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



A pair of $(u^*, \lambda^*) \in \mathcal{U} \times \Re^m$ is called a saddle point of the Lagrange function (3.6), 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

$$\begin{cases} u^* \in \mathcal{U}, \quad L(u, \lambda^*) - L(u^*, \lambda^*) \ge 0, \ \ \forall\, u \in \mathcal{U}, & (3.7a) \\ \lambda^* \in \Re^m, \ L(u^*, \lambda^*) - L(u^*, \lambda) \ge 0, \ \ \forall\, \lambda \in \Re^m. & (3.7b) \end{cases}$$

Using the notation of $L(u,\lambda)$, we get the following variational inequality:

$$\left\{\begin{array}{ll} u^* \in \mathcal{U}, \quad \theta(u)-\theta(u^*)+(u-u^*)^{\top}(-\mathcal{A}^{\top}\lambda^*) \geq 0, \ \ \forall\, u \in \mathcal{U}, & \text{(3.8a)} \\ \lambda^* \in \Re^m, \qquad\qquad\qquad\qquad (\lambda-\lambda^*)^{\top}(\mathcal{A}u^*-b) \geq 0, \ \ \forall\, \lambda \in \Re^m. & \text{(3.8b)} \end{array}\right.$$

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

$$w^* \in \Omega, \quad \theta(u)-\theta(u^*)+(w-w^*)^{\top}F(w^*) \geq 0, \quad \forall\, w \in \Omega, \qquad \text{(3.9a)}$$

where

$$w=\left(\begin{array}{c} u \\ \lambda \end{array}\right), \quad F(w)=\left(\begin{array}{c} -\mathcal{A}^{\top}\lambda \\ \mathcal{A}u-b \end{array}\right) \ \text{ and } \ \Omega=\mathcal{U}\times\Re^m. \qquad \text{(3.9b)}$$

Because F is a affine operator and

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

The matrix is skew-symmetric, thus we have $(w-\tilde{w})^{\top}(F(w)-F(\tilde{w})) \equiv 0$.

凸优化问题 (3.5), 转换成的变分不等式 (3.9) 跟变分不等式 (2.1) 没有本质的区别

Two block separable convex optimization

We consider the following structured separable convex optimization

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{3.10}$$

This is a special problem of (3.5) 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 (3.10) is

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

The same analysis tells us that the saddle point is a solution of the following VI:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^\top F(w^*) \ge 0, \quad \forall\, w \in \Omega. \tag{3.11a}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \tag{3.11b}$$

$$F(w) = \begin{pmatrix} -A^\top \lambda \\ -B^\top \lambda \\ Ax + By - b \end{pmatrix}, \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{3.11c}$$

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

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

Again, due to the skew-symmetry, we have $(w - \tilde{w})^\top (F(w) - F(\tilde{w})) \equiv 0$.

ADMM 主攻的可分离线性约束凸优化问题 (3.10), 转换成了变分不等式 (3.11)

Convex optimization problem with three separable functions

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

is a special problem of (3.5) with three blocks. The Lagrangian function is

$$L^{(3)}(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^{\top}(Ax+By+Cz-b).$$

The same analysis tells us that the saddle point is a solution of the following VI:

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

where $\theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z)$,

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

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

线性约束的凸优化问题, 都统一转换成一个 **VI**. 问题归结为求 **VI** 的解点.

4 PPA for monotone Variational Inequalities

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

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

4.1 PPA for the mixed Variational Inequality (4.1)

PPA for monotone mixed VI in Euclidean-norm

For given w^k and $r > 0$, find w^{k+1}, which satisfies

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) & \\ \geq (w - w^{k+1})^T r(w^k - w^{k+1}), \ \forall\, w \in \Omega. & \end{aligned} \tag{4.2}$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (4.1).

✠ **w^k is the solution of (4.1) if and only if $w^k = w^{k+1}$** ✠

Setting $w = w^*$ in (4.2), we obtain

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

Note that (see the structure of $F(w)$ in (3.9b))

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

and consequently (by using (4.1)) we obtain

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

Thus, we have

$$(w^{k+1} - w^*)^T (w^k - w^{k+1}) \geq 0. \tag{4.3}$$

By setting $a = w^k - w^*$ and $b = w^{k+1} - w^*$, the inequality (4.3) means that $b^T(a - b) \geq 0$. It leads $\|b\|^2 \leq \|a\|^2 - \|a - b\|^2$. Thus, we have

$$\|\boldsymbol{w^{k+1} - w^*}\|^2 \leq \|\boldsymbol{w^k - w^*}\|^2 - \|\boldsymbol{w^k - w^{k+1}}\|^2. \tag{4.4}$$

This is a 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

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(x) - \theta(x^{k+1}) &+ (w - w^{k+1})^T F(w^{k+1}) \\ &\geq (w - w^{k+1})^T H(w^k - w^{k+1}), \ \forall\, w \in \Omega, \end{aligned} \tag{4.5}$$

where H is a symmetric positive definite matrix.

✠ Again, w^k is the solution of (4.1) if and only if $w^k = w^{k+1}$ ✠

Convergence Property of Proximal Point Algorithm in H-norm

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \tag{4.6}$$

By using the block-matrix technique, we can get a proper positive definite matrix H. And the solutions of the subproblems (4.5) have the closed forms.

这些观点在判定算法是否收敛、修正和设计算法方面有巨大功效

4.2 Customized PPA [12]

根据预设正定矩阵 构造PPA算法. 方法可以在[19]中查到.

Customized PPA for the variational inequality (3.9) :

$$\theta(u)-\theta(u^{k+1})+(w-w^{k+1})^{\top}F(\tilde{w}^k)\geq(w-\tilde{w}^k)^{\top}H(w^k-\tilde{w}^k), \tag{4.7a}$$

for all $w\in\Omega$, where

$$H=\begin{pmatrix}\beta A^{\top}A+\delta I & A^{\top}\\ A & \frac{1}{\beta}I\end{pmatrix} \tag{4.7b}$$

The concrete formula of (4.7) is

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

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

$$\begin{cases}\theta(u)-\theta(u^{k+1})+(u-u^{k+1})^{\top}\\ \quad\{\underline{-A^{\top}\lambda^{k+1}}+\boldsymbol{(\beta A^{\top}A+\delta I)}(u^{k+1}-u^k)+\boldsymbol{A}^{\top}(\lambda^{k+1}-\lambda^k)\}\geq 0,\\ \quad(\underline{Au^{k+1}-b})\quad+\quad\boldsymbol{A}(u^{k+1}-u^k)\quad+\quad\boldsymbol{(1/\beta)}(\lambda^{k+1}-\lambda^k)=0.\end{cases} \tag{4.8}$$

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

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (4.8),

we need only use the following procedure: (Primal-Dual) 根据“配对”基本定理.

$$\begin{cases} u^{k+1} = \text{Argmin}\left\{ \begin{array}{c} \theta(u) - u^\top A^\top \lambda^k + \\ \frac{1}{2}(u - u^k)^\top(\beta A^\top A + \delta I)(u - u^k) \end{array} \middle| u \in \mathcal{U} \right\} & \text{(4.9a)} \\ \lambda^{k+1} = \lambda^k - \beta\big(A[2u^{k+1} - u^k] - b\big). & \text{(4.9b)} \end{cases}$$

Then, we use the form

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

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

Besides $\theta(u)$, the objective function of the u-subproblem (4.9a) contains a non-trivial quadratic function, and thus it usually does not has closed form solution.

4.3 Balanced ALM [14]

Balanced PPA for the variational inequality (3.9) : Find $\tilde{w}^k \in \Omega$, such that

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (w - \tilde{w}^k)^\top H(w^k - \tilde{w}^k), \quad \text{(4.10a)}$$

for all $w \in \Omega$, where

$$H = \begin{pmatrix} rI_n & A^\top \\ A & \frac{1}{r}AA^\top + \delta I_m \end{pmatrix} \quad \text{is positive definite.} \quad \text{(4.10b)}$$

Then, we use the form

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

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

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

to update the new iterate w^{k+1}. The concrete form of (4.10) is

$$\begin{cases} \theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^\top \{\underline{-A^\top \tilde{\lambda}^k} + \boldsymbol{rI}(\tilde{u}^k - u^k) + \boldsymbol{A}^\top(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ \qquad (\underline{A\tilde{u}^k - b}) + \boldsymbol{A}(\tilde{u}^k - u^k) + (\tfrac{\boldsymbol{1}}{\boldsymbol{r}}\boldsymbol{A}\boldsymbol{A}^\top + \boldsymbol{\delta I})(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

It can written as

$$\begin{cases} \tilde{u}^k \in \mathcal{U}, \;\; \theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^\top \{-A^\top \lambda^k + r(\tilde{u}^k - u^k)\} \geq 0, \;\; \forall x \in \mathcal{X}, \\ \qquad A[(2\tilde{u}^k - u^k) - b] + (\frac{1}{r} AA^\top + \delta I_m)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases} \tag{4.11}$$

Thus, the predictor w^{k+1} in (4.10) is implemented by 根据“配对”基本定理.

$$\begin{cases} \tilde{u}^k = \arg\min\{\theta(u) - u^\top A^\top \lambda^k + \frac{1}{2} r \|u - u^k\|^2 \mid u \in \mathcal{U}\}, & \text{(4.12a)} \\ \tilde{\lambda}^k = \arg\min\Big\{\lambda^\top \big(A[2\tilde{u}^k - u^k] - b\big) + \frac{1}{2} \big\|\lambda - \lambda^k\big\|^2_{(\frac{1}{r} AA^\top + \delta I_m)}\Big\}. & \text{(4.12b)} \end{cases}$$

Remark. $\tilde{\lambda}^k$ in (4.12b) is the solution of the following system of linear equations:

$$H_0(\lambda - \lambda^k) + \big(A[2\tilde{u}^k - u^k] - b\big) = 0, \tag{4.13}$$

where

$$H_0 = \frac{1}{r} AA^\top + \delta I_m. \tag{4.14}$$

The u-subproblem (4.12a) is simpler than (4.9a). H_0 is positive definite, there are efficient algorithms in literature for solving such a systems of linear equations.

The different positive definite matrix H in (4.7) and (4.10) leads the different PPA.

4.4 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

- 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.

[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 [12].

几年后, Chambolle and Pock 在 Math. Program. 上发表了相关文章

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.

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.

S. Becker 是 E. Candès 在加州理工的学生, 经过将近八年才挂出这篇短文.

5 Algorithmic unified framework for monotone VI

We focus on how to solve the variational inequality (3.9).

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

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \end{aligned} \tag{5.1a}$$

where the prediction matrix Q is not necessarily symmetric, but the kernel of $Q^T + Q$ is positive definite.

并不要求 Q 对称的预测非常容易实现.

v is called the essential variable in the iteration which can be equal to w, or a part of the whole vector of w.

[Correction Step]. Find the correction matrix M which satisfied (5.2). The new iteration v^{k+1} is given by

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{5.1b}$$

[Convergence Conditions]. For the prediction matrix Q in (5.1a) and the correction matrix M in (5.1b), there is a positive definite matrix H, such that

$$HM = Q \quad \text{and} \quad G = Q^T + Q - M^T HM \succ 0. \tag{5.2}$$

5.1 Convergence Analysis

Theorem 2 *For solving the VI* (3.9)*, let* $\{v^k\}$, $\{\tilde{w}^k\}$ *be the sequences generated by* (5.1)*. If the conditions* (5.2) *are satisfied, then we have* $\tilde{w}^k \in \Omega$,

$$\begin{aligned} &\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ &\quad \geq \frac{1}{2}\left(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\right) + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \ \forall w \in \Omega, \end{aligned} \tag{5.3}$$

and

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.4}$$

Proof. Treating the term $Q(v^k - \tilde{v}^k)$ in the RHS of (5.1a) by using $Q = HM$ (see (5.2)) and the correction formula (5.1b), we obtain

$$Q(v^k - \tilde{v}^k) = HM(v^k - \tilde{v}^k) = H(v^k - v^{k+1}).$$

Thus, from (5.1) we get

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad & \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ & \geq \quad (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \quad \forall w \in \Omega. \end{aligned} \tag{5.5}$$

Applying the identity

$\boxed{\text{自觉运用这里的恒等式得益于中学的数理基础}}$

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

to the RHS of (5.5) with $a = v$, $b = \tilde{v}^k$, $c = v^k$ and $d = v^{k+1}$, we obtain

$\boxed{\text{上面的积化和差的恒等式起了很重要的作用}}$

$$\begin{aligned} & (v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ & = \frac{1}{2}\left(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\right) + \frac{1}{2}\left(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2\right). \end{aligned} \tag{5.6}$$

To the second part of the RHS of (5.6), by using

$HM = Q$ and; $2v^TQv = v^T(Q^T + Q)v$, it follows that

$$
\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&\overset{(5.1b)}{=} \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
&= (v^k - \tilde{v}^k)^T(Q^T + Q - M^THM)(v^k - \tilde{v}^k) \\
&\overset{(5.2)}{=} \|v^k - \tilde{v}^k\|_G^2.
\end{aligned}
\tag{5.7}
$$

这里用的是一些简单矩阵向量运算.

Substituting (5.7) in (5.6), and then in (5.5), we get the assertion (5.3) directly.

To prove (5.4), setting the $w \in \Omega$ in (5.3) by any fixed w^*, and then using

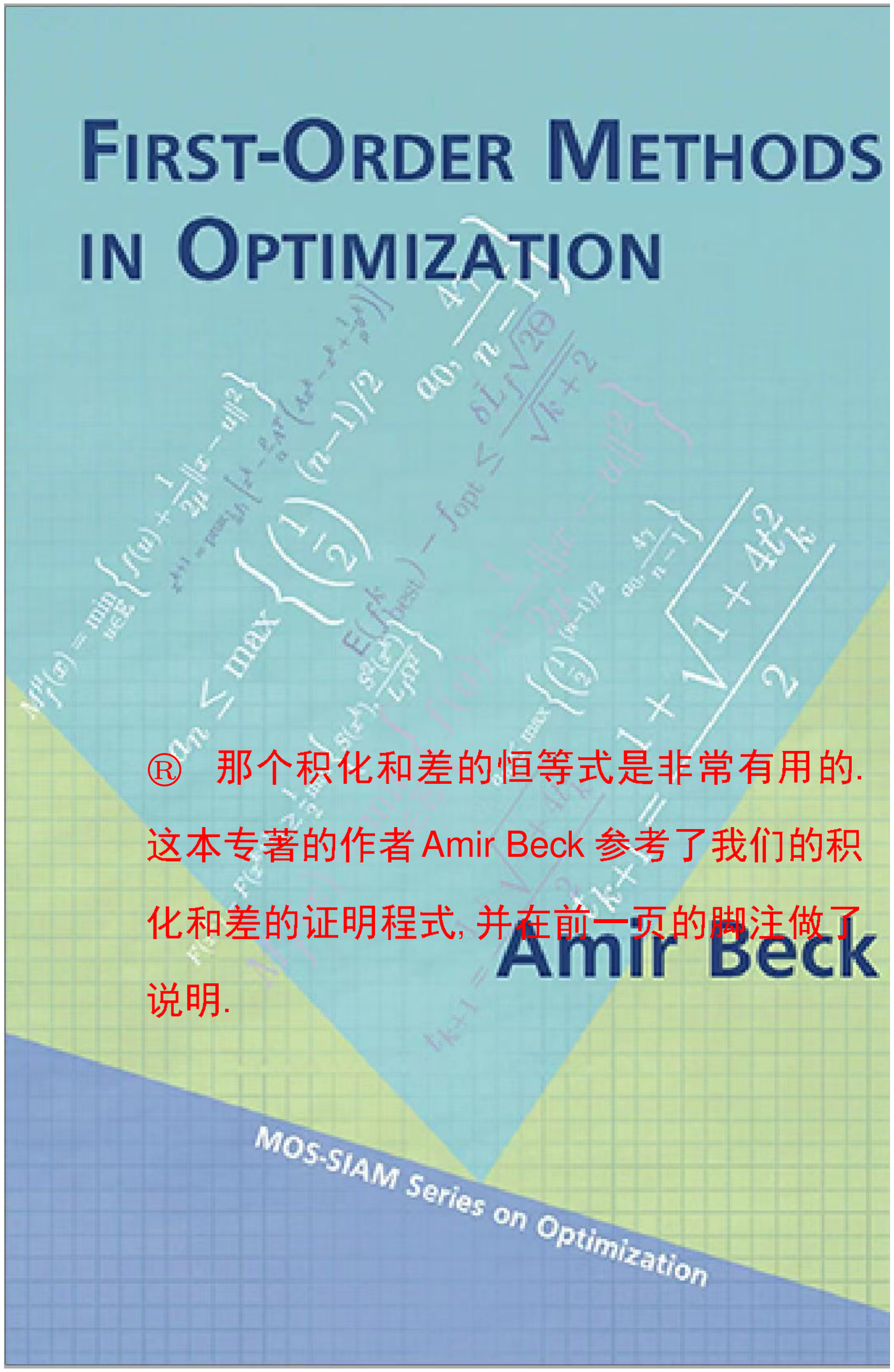
$$
\begin{aligned}
&\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \\
&\quad = \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \geq 0.
\end{aligned}
$$

Thus, from (5.3) we get

定理的证明纯粹是用些简单的加加减减.

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

We obtain (5.4) and the theorem is completely proved. $\Box$



® 那个积化和差的恒等式是非常有用的. 这本专著的作者 Amir Beck 参考了我们的积化和差的证明程式, 并在前一页的脚注做了说明.

We will use the following notation:

$$\tilde{\mathbf{x}}^k = \mathbf{x}^{k+1},$$
$$\tilde{\mathbf{z}}^k = \mathbf{z}^{k+1},$$
$$\tilde{\mathbf{y}}^k = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^k - \mathbf{c}).$$

Using (15.15), (15.16), the subgradient inequality, and the above notation, we obtain that for any $\mathbf{x} \in \mathrm{dom}(h_1)$ and $\mathbf{z} \in \mathrm{dom}(h_2)$,

$$h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \left\langle \rho\mathbf{A}^T\left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right) + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \right\rangle \geq 0,$$
$$h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \left\langle \rho\mathbf{B}^T\left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right) + \mathbf{Q}(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \right\rangle \geq 0.$$

Using the definition of $\tilde{\mathbf{y}}^k$, the above two inequalities can be rewritten as

$$h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \langle \mathbf{A}^T\tilde{\mathbf{y}}^k + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \rangle \geq 0,$$
$$h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \langle \mathbf{B}^T\tilde{\mathbf{y}}^k + (\rho\mathbf{B}^T\mathbf{B} + \mathbf{Q})(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \rangle \geq 0.$$

Adding the above two inequalities and using the identity

$$\mathbf{y}^{k+1} - \mathbf{y}^k = \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c}),$$

we can conclude that for any $\mathbf{x} \in \mathrm{dom}(h_1)$, $\mathbf{z} \in \mathrm{dom}(h_2)$, and $\mathbf{y} \in \mathbb{R}^m$,

$$H(\mathbf{x},\mathbf{z}) - H(\tilde{\mathbf{x}}^k, \tilde{\mathbf{z}}^k) + \left\langle \begin{pmatrix} \mathbf{x} - \tilde{\mathbf{x}}^k \\ \mathbf{z} - \tilde{\mathbf{z}}^k \\ \mathbf{y} - \tilde{\mathbf{y}}^k \end{pmatrix}, \begin{pmatrix} \mathbf{A}^T\tilde{\mathbf{y}}^k \\ \mathbf{B}^T\tilde{\mathbf{y}}^k \\ -\mathbf{A}\tilde{\mathbf{x}}^k - \mathbf{B}\tilde{\mathbf{z}}^k + \mathbf{c} \end{pmatrix} - \begin{pmatrix} \mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) \\ \mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) \\ \frac{1}{\rho}(\mathbf{y}^k - \mathbf{y}^{k+1}) \end{pmatrix} \right\rangle \geq 0, \tag{15.17}$$

where $\mathbf{C} = \rho\mathbf{B}^T\mathbf{B} + \mathbf{Q}$. We will use the following identity that holds for any positive semidefinite matrix $\mathbf{P}$:

$$(\mathbf{a} - \mathbf{b})^T\mathbf{P}(\mathbf{c} - \mathbf{d}) = \frac{1}{2}\left(\|\mathbf{a} - \mathbf{d}\|_{\mathbf{P}}^2 - \|\mathbf{a} - \mathbf{c}\|_{\mathbf{P}}^2 + \|\mathbf{b} - \mathbf{c}\|_{\mathbf{P}}^2 - \|\mathbf{b} - \mathbf{d}\|_{\mathbf{P}}^2\right).$$

Using the above identity, we can conclude that

$$\begin{aligned}(\mathbf{x} - \tilde{\mathbf{x}}^k)^T\mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) &= \frac{1}{2}\left(\|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2 + \|\tilde{\mathbf{x}}^k - \mathbf{x}^k\|_{\mathbf{G}}^2\right) \\ &\geq \frac{1}{2}\|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2,\end{aligned} \tag{15.18}$$

as well as

$$(\mathbf{z} - \tilde{\mathbf{z}}^k)^T\mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) = \frac{1}{2}\|\mathbf{z} - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 - \frac{1}{2}\|\mathbf{z} - \mathbf{z}^k\|_{\mathbf{C}}^2 + \frac{1}{2}\|\mathbf{z}^k - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 \tag{15.19}$$

and

$$\begin{aligned}&2(\mathbf{y} - \tilde{\mathbf{y}}^k)^T(\mathbf{y}^k - \mathbf{y}^{k+1}) \\ &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \|\tilde{\mathbf{y}}^k - \mathbf{y}^k\|^2 - \|\tilde{\mathbf{y}}^k - \mathbf{y}^{k+1}\|^2 \\ &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2 \\ &-\|\mathbf{y}^k + \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}) - \mathbf{y}^k - \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c})\|^2 \\ &= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2 - \rho^2\|\mathbf{B}(\mathbf{z}^k - \tilde{\mathbf{z}}^k)\|^2.\end{aligned}$$

This theorem is proved under weak conditions:

$$Q^T + Q \succeq 0, \;\; H \succeq 0, \;\; HM = Q, \;\; G = Q^T + Q - M^T HM \succeq 0$$

Assertion (5.3) is useful for the convergence rate proof of ADMM, see H and Yuan, SIAM Numer. Anal. 2012, 50:700-709; Numer. Math. 2015, 130: 567-577

6 The equivalent convergence conditions and the generalized PPA

Under the condition that the prediction matrix Q is nonsingular, it is easy to construct the correction matrix M which satisfies the convergence conditions (5.2). In fact, because $Q^T + Q \succ 0$, we can take

$$D \succ 0, \qquad G \succ 0 \qquad \text{and} \qquad D + G = Q^T + Q. \tag{6.1}$$

Afterwards, we let

$$HM = Q \qquad \text{and} \qquad M^T HM = D. \tag{6.2}$$

Because

$$\begin{cases} HM=Q, \\ M^THM=D \end{cases} \Leftrightarrow \begin{cases} Q^TM=D, \\ HM=Q \end{cases} \Leftrightarrow \begin{cases} M=Q^{-T}D, \\ H=QD^{-1}Q^T \end{cases},$$

we get the matrices M, H and G, which satisfy the conditions (5.2).

Infinite combinations of D and G which satisfy conditions (6.1).

符合条件(6.1)的各种不同的D有无穷, 由此可以构造一簇算法.

Choosing matrix D that satisfies condition (6.1), we get

$$M = Q^{-T}D \quad \text{and} \quad H = QD^{-1}Q^T.$$

Because $M = Q^{-T}D$, the correction $v^{k+1} = v^k - M(v^k - \tilde{v}^k)$ can be achieved by solving the equation

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

Generalized PPA by choosing a special D.

We can take a special pair of D and G in (6.1) by

$$D = G = \tfrac{1}{2}(Q^T + Q). \tag{6.3}$$

In this case, $M = \frac{1}{2}Q^{-T}(Q^T + Q)$. Because $D = G$, the contractive inequality (5.4) becomes

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_D^2, \quad \forall v^* \in \mathcal{V}^*. \tag{6.4}$$

Moreover, since $D = M^T H M$ (see (6.2)) and $M(v^k - \tilde{v}^k) = v^k - v^{k+1}$ (see (5.1b)), it follows that

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{6.5}$$

The inequality (6.5) is just the main convergence result of the classical PPA. Because the each iteration consists of a prediction and a correction, we call the related method as a generalized PPA.

In practice, we suggest to take $D = \alpha(Q^T + Q)$ and $\alpha \in [0.5, 0.8)$.

Applications to multi-blocks problem (应用举例)

We consider the multi-blocks convex optimization problems

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i)\;\Big|\;\sum_{i=1}^{p}A_ix_i=b\;(\text{or}\geq b),\quad x_i\in\mathcal{X}_i\Big\}. \tag{6.6}$$

The saddle point of the Lagrange function of the convex optimization problem (6.6) is equivalent to the solution of the following variational inequality:

$$w^*\in\Omega,\quad \theta(x)-\theta(x^*)+(w-w^*)^\top F(w^*)\geq 0,\quad \forall\, w\in\Omega, \tag{6.7a}$$

where

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

and

$$\theta(x)=\sum_{i=1}^{p}\theta_i(x_i),\qquad \Omega=\prod_{i=1}^{p}\mathcal{X}_i\times\Lambda.$$

Prediction for the multi-block problem (6.6)

From $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ to $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_p^k, \tilde{\lambda}^k)$:

Prediction Step. With given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, find $\tilde{w}^k \in \Omega$:

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^\top A_1^\top \lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\
\tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^\top A_2^\top \lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\
\qquad\qquad \vdots \\
\tilde{x}_i^k \in \arg\min_{x_i \in \mathcal{X}_i}\{\theta_i(x_i) - x_i^\top A_i^\top \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\
\qquad\qquad \vdots \\
\tilde{x}_p^k \in \arg\min_{x_p \in \mathcal{X}_p}\{\theta_p(x_p) - x_p^\top A_p^\top \lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}; \\
\tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^{p} A_j\tilde{x}_j^k - b\big)\big].
\end{array}\right. \tag{6.8}$$

Solve the separable sub-problems one by one in sequence

预测的基本程式就像求解线性方程组的消元, 得到一个预测矩阵 Q.

According to Theorem 1, the matrix Q in (5.1a) by the prediction (6.8) is

$$Q = \begin{pmatrix} \beta A_1^\top A_1 & 0 & \cdots & 0 & A_1^\top \\ \beta A_2^\top A_1 & \beta A_2^\top A_2 & \ddots & \vdots & A_2^\top \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^\top A_1 & \beta A_p^\top A_2 & \cdots & \beta A_p^\top A_p & A_p^\top \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{6.9}$$

The kernel matrix of Q and Q^{-T} is

$$\mathcal{Q} = \begin{pmatrix} I & 0 & \cdots & 0 & I \\ I & I & \ddots & \vdots & I \\ \vdots & & \ddots & 0 & \vdots \\ I & I & \cdots & I & I \\ 0 & 0 & \cdots & 0 & I \end{pmatrix} \quad \text{and} \quad \mathcal{Q}^{-T} = \begin{pmatrix} I & -I & 0 & \cdots & 0 \\ 0 & I & \ddots & \ddots & \vdots \\ \vdots & & \ddots & -I & 0 \\ 0 & \cdots & 0 & I & 0 \\ -I & 0 & \cdots & 0 & I \end{pmatrix},$$

respectively. Notice that $\mathcal{Q}$ has a special structure.

It is very easy to calculate its inverse. By using Q and Q^{-T}, the correction formula for producing $(A_1x_1^{k+1},\cdots,A_px_p^{k+1},\lambda^{k+1})$ of the generalized PPA is

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \frac{\alpha}{2}\begin{pmatrix} I & -I & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & I & -I & 0 \\ I & \cdots & I & 2I & \frac{1}{\beta}I \\ -\beta I & 0 & \cdots & 0 & I \end{pmatrix}\begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

Having $(A_1x_1^{k+1},\cdots,A_px_p^{k+1},\lambda^{k+1})$, we can start the prediction (6.8) of a new iteration.

In the above correction formula of the generalized PPA, the parameter of the step-length $\alpha = 1$, in practical computation, $\alpha \in [0.6, 0.8] \subset (0,2)$ is suggested.

校正的程式就像Gauss消去法求解线性方程组消元结束后的回代.

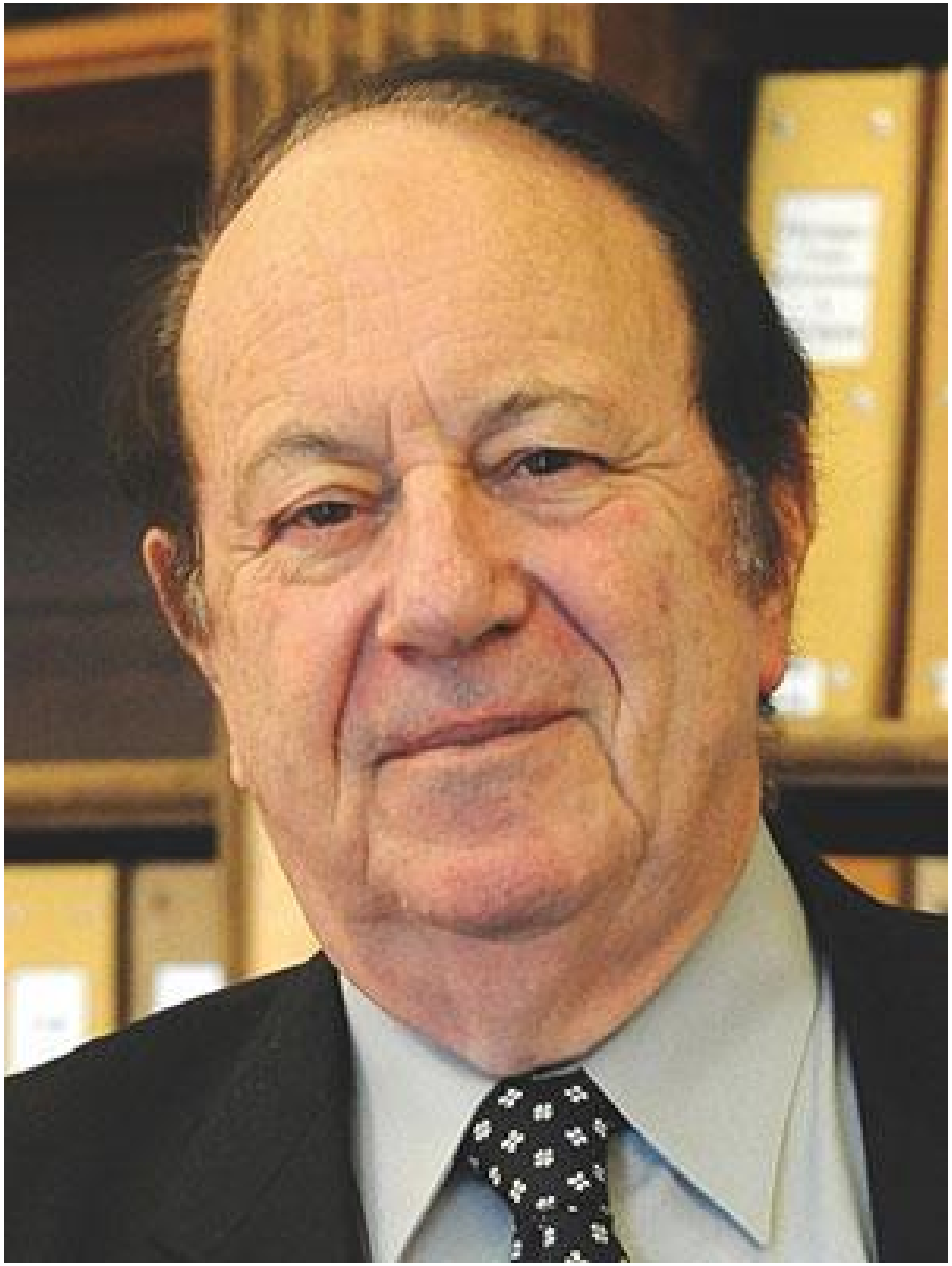
对自己的研究经历和收获的一点认识

- 我们提出的算法统一框架, 不仅涵盖了邻近点算法 (PPA), 增广拉格朗日乘子法 (ALM) 和交替方向法 (ADMM) 这些耳熟能详的经典算法, 还为多块可分离凸优化问题的求解提供了一簇算法, 关键技术是在原有的逐步推进的基础上加了个简单的校正。我们称这类方法为预测-校正的 ADMM 类分裂收缩算法，一是因为它保留了 ADMM 的收敛性质, 二是又可以把已有的 ADMM 方法中的每步迭代故意拆解成预测和校正去看待. 这样做，还让我们进一步发现 ADMM 的一些容易证明的漂亮性质。
- 预测-校正, 是计算数学中的常用思想方法. 在优化算法领域, 我比较自觉地用校正策略。说起来, 那还得感谢文化大革命期间的那段务农经历! 当年的生产队, 除了水稻和麦子, 还有些副业。栽桑, 要剪枝; 种瓜, 要掐蔓, 这样的活, 我干了五, 六年。那剪枝掐蔓, 不就是校正!? 学过和干过的, 即便在其他领域, 需要的时候, 想法也会自然冒出来。

- 我们提出的这类方法的每步迭代, 包含预测和校正两部分。用它处理多块的线性约束凸优化问题, 有点像高斯消去法求解线性方程组。预测, 相当于高斯消去法中自上而下的按序消元; 而校正, 则更像高斯消去法中自下往上的回代。
- 麻省理工学院教授 G. Strang 有本名为《Linear Algebra and its Applications》的线性代数教科书, 配合这本教材, 他还有个讲课视频。记得 Strang 在他的视频里有段话的大意是：高斯消去法这么重要, 但高斯想出这个方法大概只花了他十分钟时间。
- 我们在优化方法中做出这一套思想体系上跟高斯消去法有联系的工作, 前前后后大概花了 10 年。想起自己在美国数学评论的《数学谱系》中是 Gauss 的后代, 工作中还用到了高斯消去法的思想, 自然尤感欣慰！

$\boxed{\text{2010 年以后主要研究 ADMM 为代表的凸优化的分裂收缩算法}}$

机会: 香港回归前在港科大图书馆从 Glowinski 的书上学到 ADMM



我们从 1997 年 4 月开始在变分不等式投影收缩算法的基础上开展 ADMM 的研究。

Glowinski 从我们的稿件得知 ADMM 可以用来求解优化问题, 非常高兴. 但那时我们没有非用ADMM 不可的算例.

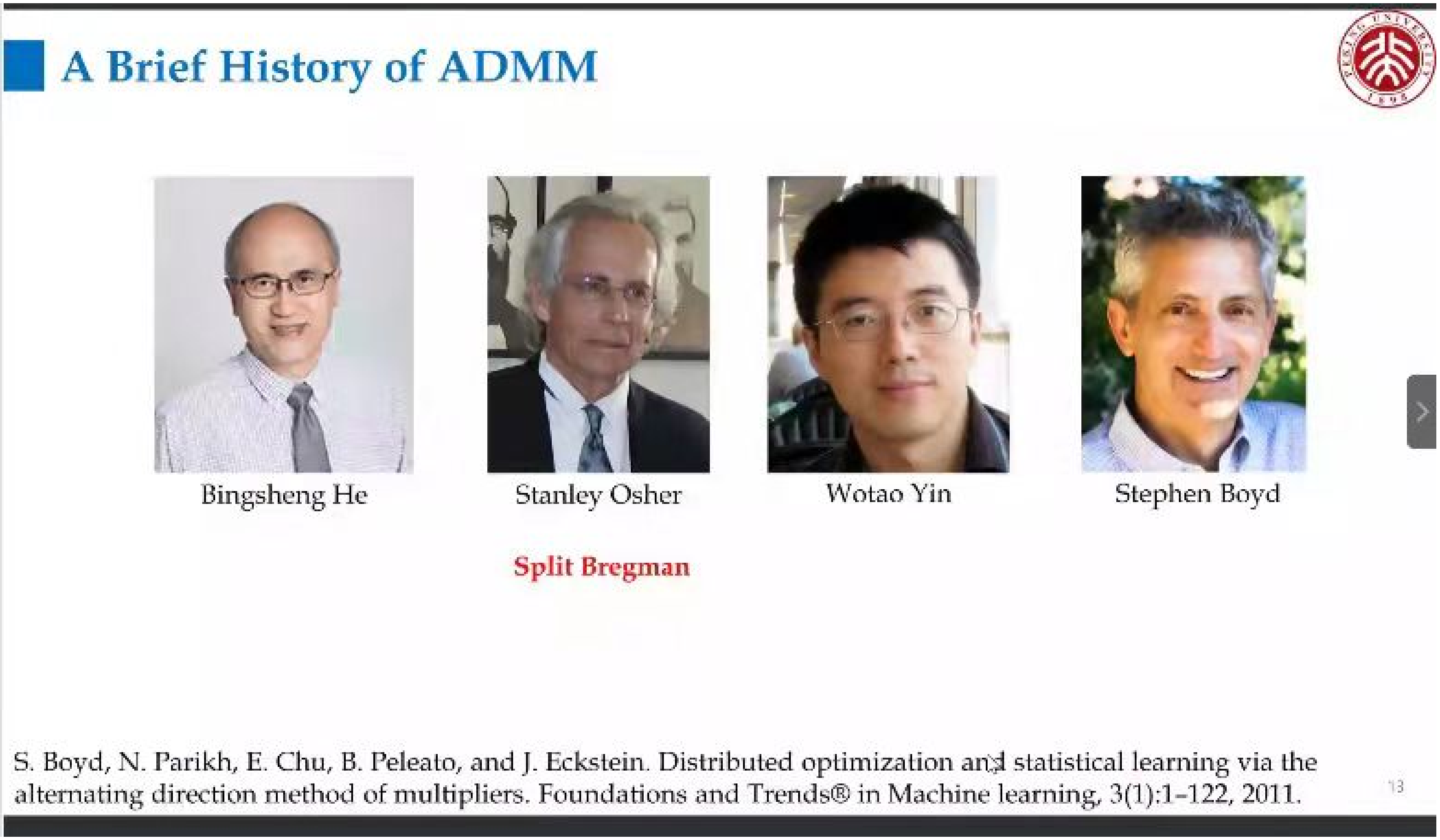
直到 2009 年, 发现了一大批 ADMM 的最新应用, 适逢第一届华东地区运筹学与控制论博士生学术论坛在上海大学召开, 我在大会报告中以“信息技术中的凸优化问题及交替方向法求解”为题呼请年轻人注意 ADMM。

2010 年初, Boyd 等发表关于 ADMM 的综述文章。

在国内，我们的研究工作也开始得到越来越多的认可。

这是北京大学林宙辰教授做报告。

我教过的学生从会议上拍了传给我的照片。



- 这里的History, 说明我做ADMM比后面三位学者早了好些年. 这里我也只是代表, 我的一些合作者和学生参与ADMM也比较早。
- 两位美国院士在引用我们文章的时候都给出了实质性的赞语。
- 科研运气很重要, 我们冒了泡是依靠独立思考和动手比人家早。

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.
- 从好不容易凑成一个方法到并不费劲构造一簇方法, 我把这些结果写在祖国大地上 [19, 20].
- 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.

希望各位以质疑的态度审视我的观点, 对的就相信, 不对的请批评指正.

Thank you very much for your attention ！

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

[5] 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.

[6] 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.

[7] 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.

[8] B. S. He, M. Tao and X.M. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, *SIAM Journal on Optimization* **22**(2012), 313-340.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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

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

[16] 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.

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

[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] 何炳生，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.

[20] 何炳生，凸优化分裂收缩算法统一框架的新进展–从好不容易凑出一个方法到并不费劲构造一族算法, 高等学校计算数学学报, 2024, 46: 1-24.