我和交替方向法 (ADMM) 的 20 年

基于变分不等式和邻近点算法的分裂收缩算法

何 炳 生

南方科技大学数学系　　南京大学数学系

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

中学的数理基础　必要的社会实践

普通的大学数学　基本的优化常识

第十一届全国数学规划年会　桂林电子科技大学, 2017.05.13

✠ 经历影响人的价值取向 ✠

在我刚刚完成高中教育以后的 10 多年里, 接触到的不是以微积分为代表的高等数学, 而是华罗庚先生宣传的优选法。 那个年代, 我对数学的理解, 就像瞎子摸象。

✠ 上大学以后, 我才知道数学是年轻人的学问 ✠

华罗庚教授 1980 年在南京大学礼堂关于 “优选法” 和普及数学应用的报告 以及

南京大学何旭初先生在当年国内优化界的学术地位, 对我报考研究生时选择 “最优化” 作为专业方向产生了很大的影响。 所幸找了个适合自己能力的研究方向!

✠ 优选法的精髓是简单和有用 ✠

作为一个上大学前有 10 多年实践经验的专业优化工作者, 研究、宣传和开发科技工作者能用来解决问题的方法, 就成了我一生的追求。

我一生的主要工作: 变分不等式的投影收缩算法
和 ADMM 代表的凸优化分裂收缩算法.

两类算法有紧密联系, 所以我有时喜欢用的题目是:

从变分不等式的投影收缩算法到凸优化的 ADMM 类分裂收缩算法. 试图介绍这些算法研究的来龙去脉.

简单的方法, 得到了 期盼的应用 和 过望的引用

ADMM 为代表的分裂收缩算法 的热火

STATISTICAL DATA

DOI: 10.15978/j.cnki.1673-5668.201606004

中国热点论文榜

中国科学院文献情报中心科学计量中心

“热点论文”在科学界已经是耳熟能详的名词。顾名思义，热点论文即为众人所关注的论文。这种关注度在科学计量学领域可以用论文被引用的次数来量化和测度。我们以2011－2015年中国科学家的SCI论文为数据基础，分领域统计了自发表以来被引频次最高的论文，以展现颇具显示度的中国科技成果。本期发布的热点论文榜涉及数学、物理学、化学、生物学、医学、农学、地学、环境科学和工程技术等9个领域。

✠ 中国科学院 文献情报中心 科学计量中心 ✠

中国科学院文献情报中心科学计量中心公布了 2011 – 2015《中国热点论文榜》.《中国数学领域热点论文(2011 – 2015)》的 10 篇入选论文中, 3 篇是由我的学生完成的。

- 十篇中第 1 篇杨俊锋是第一作者, 杨是袁亚湘院士的硕士, Rice 大学张寅教授和我联合培养的博士。这篇文章用乘子交替方向法 (ADMM) 求解问题，成文时杨已经在南京大学工作。
- 第 4 篇陶敏是第一作者，香港浸会大学教授袁晓明第二作者。成文时陶是我的在读博士生，毕业后被南京大学引进。袁晓明在 2000 年前后跟我读过硕士。
- 第 10 篇是袁晓明跟我合作的文章。我耳顺以后，联名合作文章的投稿、通信、修改、校稿，都由学生完成。
- 杨、陶、袁都是我的学生。三篇中，两篇是袁晓明参与的。

《中国数学领域热点论文(2011 – 2015)》的 10 篇入选论文

表1　中国数学领域热点论文(2011－2015年)

序号	论文题目	作者机构	是否合作	被引频次
1	Yang Junfeng*等. Alternating Direction Algorithms for L(1)-Problems in Compressive Sensing. Siam Journal on Scientific Computing, 2011, 33(1), 250-278.	南京大学†	是	224
2	Zhou Yong*等. A Class of Fractional Evolution Equations and Optimal Controls. Nonlinear Analysis-Real World Applications, 2011, 12(1), 262-272.	湘潭大学†	-	165
3	Fan Engui等. Linear Superposition Principle Applying to Hirota Bilinear Equations. Computers & Mathematics with Applications, 2011, 61(4), 950-959.	复旦大学	是	137
4	Tao Min*等. Recovering Low-Rank and Sparse Components of Matrices from Incomplete and Noisy Observations. Siam Journal on Optimization, 2011, 21(1), 57-81.	南京大学† 南京邮电大学	-	111
5	Khan Yasir*等. Homotopy Perturbation Transform Method for Nonlinear Equations Using He's Polynomials. Computers & Mathematics with Applications, 2011, 61(8), 1963-1967.	浙江大学†	-	98
6	Small Michael等. Recurrence-Based Time Series Analysis by Means of Complex Network Methods. International Journal of Bifurcation and Chaos, 2011, 21(4), 1019-1046.	香港理工大学	是	97

6	Small Michael等. Recurrence-Based Time Series Analysis by Means of Complex Network Methods. International Journal of Bifurcation and Chaos, 2011, 21(4), 1019-1046.	香港理工大学	是	97
7	Liu Weidong等. A Constrained L(1) Minimization Approach to Sparse Precision Matrix Estimation. Journal of the American Statistical Association, 2011, 106(494), 594-607.	上海交通大学	是	94
8	Zhang Xiaoqun*等. A Unified Primal-Dual Algorithm Framework Based on Bregman Iteration. Journal of Scientific Computing, 2011, 46(1), 20-46.	上海交通大学†	是	92
9	Wang Guotao*等. Impulsive Anti-Periodic Boundary Value Problem for Nonlinear Differential Equations of Fractional Order. Nonlinear Analysis-Theory Methods & Applications, 2011, 74(3), 792-804.	山西师范大学†	是	90
10	Yuan Xiaoming*等. On the O(1/N) Convergence Rate of the Douglas-Rachford Alternating Direction Method. Siam Journal on Numerical Analysis, 2012, 50(2), 700-709.	香港浸会大学†	-	90

可遇而不可求. 我们为优化在数学界争了一次光.

“热点论文”是根据论文被引用次数来量化的“为众人所关注的论文”，时有良莠不齐，为人诟病。最初见到这个“榜上有名”的消息时，我也并不那么在意。

然而，有海外著名华人学者在向我祝贺时提到：我们这三篇文章，都发在 SIAM 系列期刊上！

三篇文章都与优化领域的乘子交替方向法有关。或许可以这样说, 优化方法提供给“数据科学”的工具中, 受人关注的首推 ADMM。

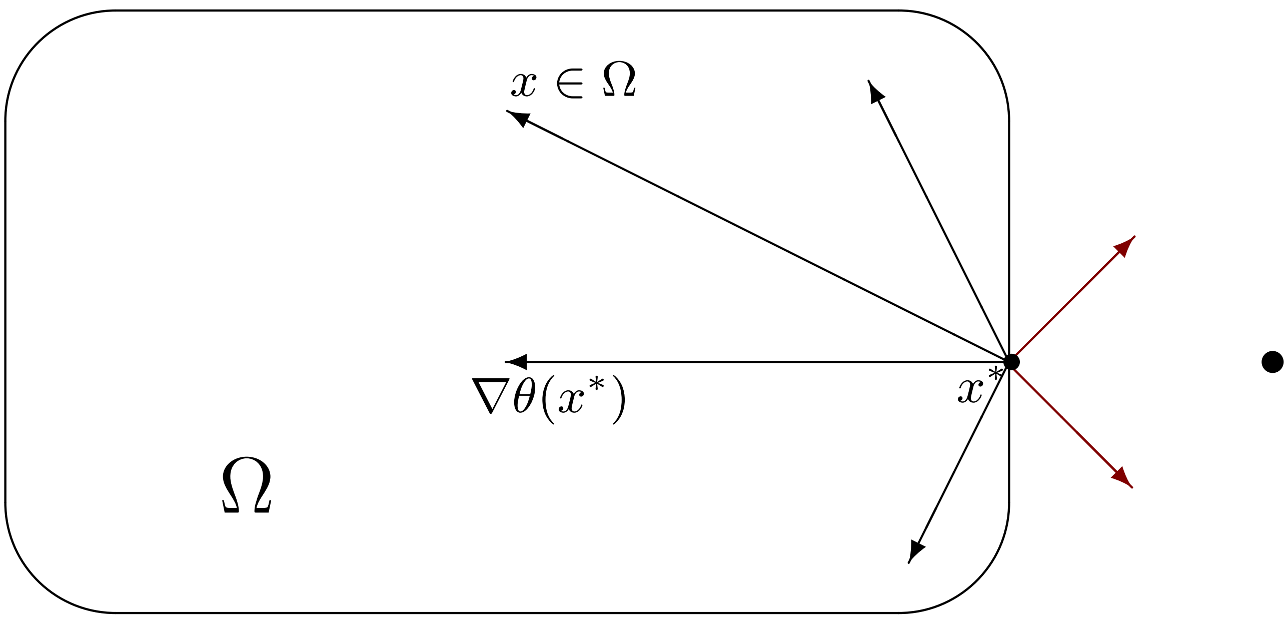
我 40 岁开始带研究生，一直战战兢兢，生怕引错方向，耽误学生。现在看来，我从 1997 年就关注并研究 ADMM，这一步跨的多么幸运！

这个报告并非要说 ADMM 是永远的热点, 而是回顾在机会面前我们做了些什么。如果真有道理, 说不定可以供年轻同行参考借鉴。

1 变分不等式框架下的最优性条件

设 Ω 是 $\Re^n$ 中的非空闭凸集, 我们先来讨论可微凸优化问题

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



可微凸优化的最优性条件与变分不等式

最优性条件: 所有的可行方向都不再是下降方向

可以写成 $x^* \in \Omega, \quad (x-x^*)^T \nabla\theta(x^*) \geq 0, \quad \forall x \in \Omega$

将 $\nabla\theta(x)$ 写成 $F(x)$, 就是

$$x^* \in \Omega, \quad (x-x^*)^T F(x^*) \geq 0, \quad \forall x \in \Omega$$

上式就是变分不等式的标准形式.

需要说明的是, 如果 $\nabla\theta(x)$ 可微, $\nabla^2\theta(x)$ 一定对称. 而对一般变分不等式, 向量函数 $F(x)$ 即使可微, $F(x)$ 的 Jacobian 矩阵不一定对称.

变分不等式 — 瞎子爬山的数学表达形式.

普通的大学数学 — 基于微积分学的一个引理

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

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

上面的凸优化最优性条件是最最基本的, 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,*

$$x^*=\arg\min\{\theta(x)+f(x)\,|\,x\in\mathcal{X}\} \tag{1.2a}$$

if and only if

$$x^*\in\mathcal{X},\quad \theta(x)-\theta(x^*)+(x-x^*)^T\nabla f(x^*)\ge 0,\quad \forall x\in\mathcal{X}. \tag{1.2b}$$

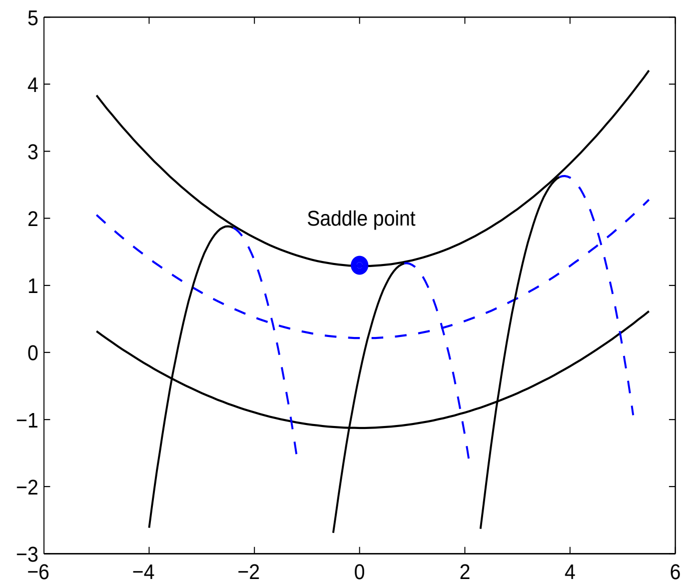
基本的优化常识 — 线性约束凸优化的鞍点原理

考虑线性约束的凸优化问题

$$\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}, \quad (1.3)$$

其中 $\theta(x)$ 是凸函数, $A \in \Re^{m\times n}$, $b \in \Re^m$, $\mathcal{X} \subset \Re^n$ 为凸集. 其 Lagrange 函数是定义在 $\mathcal{X} \times \Re^m$ 上的

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



如果一对 (x^*, λ^*) 满足

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

就称为 Lagrange 函数在 $\mathcal{X} \times \Re^m$ 上的鞍点。它的等价意义就是

$$\begin{cases} x^* &= \arg\min\{L(x, \lambda^*) \,|\, x \in \mathcal{X}\}, \\ \lambda^* &= \arg\max\{L(x^*, \lambda) \,|\, \lambda \in \Lambda\}. \end{cases}$$

利用前面的引理, 就有

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

它可以被写成一种紧致的形式:

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

其中

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

注意到其中的算子 F 是单调的. 所谓单调是指

$(u-v)^T(F(u)-F(v)) \geq 0.$ 这里恰有 $(u-v)^T(F(u)-F(v)) = 0.$

两个可分离算子线性约束凸优化问题 与 等价的变分不等式

对两个算子的可分离线性结构型约束凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\}, \tag{1.4}$$

通过类似的分析, 最优性条件同样可以表示成一个变分不等式:

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

其中

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

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

三个可分离算子线性约束凸优化问题 与 等价的变分不等式

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

通过类似的分析, 最优性条件同样可以表示成一个变分不等式:

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

其中

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

和

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

2 前一个十年的主要工作和影响(1997-2007)

2.1 为什么国内我们起步最早, 开始我们做了些什么?

我们关心 ADMM, 与香港科技大学做交通的杨海知道我研究变分不等式, 20 年前邀请我去访问有关. 交通网络分析中不少问题都用变分不等式去描述. 我的 ADMM 研究, 是从 1997 年 4 月开始的.

设 $f(x), g(y)$ 是单调算子. 结构型单调变分不等式是指

$$u^* \in \Omega, \quad (u-u^*)^T F(u^*) \ge 0, \quad \forall \;\; u \in \Omega,$$

其中

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad F(u) = \begin{pmatrix} f(x) \\ g(y) \end{pmatrix},$$

$$\Omega = \{(x,y) | x \in \mathcal{X}, y \in \mathcal{Y}, Ax + By = b\}.$$

那段时间，我刚把投影收缩算法的机理搞清楚。接下来究竟做什么，我花两个星期的时间浸在香港科技大学图书馆查资料，觉得可以用投影收缩算法有关思想去研究交替方向法。

ELSEVIER

我们 ADMM 的第一篇文章

97 年投稿　　98 年发表

Operations Research Letters 23 (1998) 151–161

operations research letters

Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities

Bingsheng He[a], Hai Yang[b,*]

[a] *Department of Mathematics, Nanjing University, Nanjing, 210093, People's Republic of China*

[b] *Department of Civil Engineering, The Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong, People's Republic of China*

Received 1 June 1997; received in revised form 1 July 1998

求解可分离结构变分不等式的 ADMM 方法

Step 1. 从给定的 (y^k, λ^k) 出发, 求得 $x^{k+1} \in \mathcal{X}$, 使得

$$(x' - x^{k+1})^T \Big(f(x^{k+1}) - A^T[\lambda^k - \beta(Ax^{k+1} + By^k - b)] \Big) \ge 0, \ \forall\, x' \in \mathcal{X}.$$

Step 2. 由已有的 (x^{k+1}, λ^k), 求得 $y^{k+1} \in \mathcal{Y}$, 使得

$$(y' - y^{k+1})^T \Big(g(y^{k+1}) - B^T[\lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)] \Big) \ge 0, \ \forall\, y' \in \mathcal{Y}.$$

Step 3. 校正 Lagrange 乘子 λ

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

严重影响收敛速度的是参数 β 的选取

2000 年发表在 JOTA 的文章

JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 106, No. 2, pp. 337–356, AUGUST 2000

合作者汪圣利是首批博士生, 现在是某单位首席专家

该文提出的调比法则是被 Boyd 等实际采用的

Alternating Direction Method with Self-Adaptive Penalty Parameters for Monotone Variational Inequalities[1]

B. S. He,[2] H. Yang,[3] and S. L. Wang[4]

Communicated by R. Glowinski

✠ 调 比 准 则 ✠

Strategy S3. $\{\beta_k\}$ is a self-adaptive variable,

$$\beta_{k+1} = \begin{cases} \beta_k(1+\tau_k), & \text{if } \|x^k - P_{\mathcal{X}}[x^k - (f(x^k) - A^T\lambda^k)]\| < \mu\|Ax^k + By^k - b\|, \\ \beta_k/(1+\tau_k), & \text{if } \mu\|x^k - P_{\mathcal{X}}[x^k - (f(x^k) - A^T\lambda^k)]\| > \|Ax^k + By^k - b\|, \\ \beta_k, & \text{otherwise.} \end{cases}$$

In any cases, if $\beta_{k+1} \neq \beta_k$, then

$$\beta_{k+1} = (1+\tau_k)\beta_k \quad \text{or} \quad \beta_{k+1} = \beta_k/(1+\tau_k).$$

- 交替方向法的收敛速度，严重依赖 β 的选择.
- 我们不用固定的罚参数 β, 而是采用罚参数序列 $\{\beta_k\}$.
- 这里的罚参数序列 $\{\beta_k\}$ 的自调比准则,
  应该算是我们早期对 ADMM 算法最重要的实质性贡献。
- 问题还是只有得到部分解决！ 原因在哪里？ 后面再提.

✠ 我们 2002 年发表在 MP 的文章 ✠

合作者中的韩德仁, 也是我的第一批博士, 当时在读

Math. Program., Ser. A 92: 103–118 (2002)

Digital Object Identifier (DOI) 10.1007/s101070100280

Bingsheng He · Li-Zhi Liao · Deren Han · Hai Yang

A new inexact alternating directions method for monotone variational inequalities

Received: January 4, 2000 / Accepted: October 2001
Published online February 14, 2002 –

Abstract. The alternating directions method (ADM) is an effective method for solving a class of variational inequalities (VI) when the proximal and penalty parameters in sub-VI problems are properly selected. In this paper, we propose a new ADM method which needs to solve two strongly monotone sub-VI problems in each iteration approximately and allows the parameters to vary from iteration to iteration. The convergence of the proposed ADM method is proved under quite mild assumptions and flexible parameter conditions.

Key words. variational inequality – alternating directions method – inexact method

A variational inequality problem is to find a vector $u^* \in \Omega$ such that

$$(u - u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega, \tag{1.1}$$

where Ω is a nonempty closed convex subset of $\mathcal{R}^n$, and F is a mapping from $\mathcal{R}^n$ into itself. In this paper, we consider the VI problem with the following structure:

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(u) = \begin{pmatrix} f(x) \\ g(y) \end{pmatrix}, \tag{1.2}$$

$$\Omega = \{(x, y) | x \in \mathcal{X}, \ \ y \in \mathcal{Y}, \ \ Ax + By = b\}, \tag{1.3}$$

- 从调比 $\frac{\beta}{2}\|Ax + By - b\|^2$ 中 β 到调比 $\frac{1}{2}\|Ax + By - b\|_H^2$ 中 H.
- 加邻近点项, 利用一个不精确准则, 使得子问题允许不精确求解.
- 不精确准则很难操作的, 需要借助一个关于不等式的引理, 利用上述引理, 得到误差的一个上界估计式。
- 我要做的是线性化和预测校正. 无奈必须按审稿意见改, 最后我的意图还是在发表在 JCM 2006 的文章中实现的。见论文 [14].

ADMM 求解两个可分离算子的线性约束凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)\mid Ax+By=b,\ x\in\mathcal{X},y\in\mathcal{Y}\} \tag{2.1}$$

问题 (2.1) 的增广 Lagrange 函数是

$$\mathcal{L}_\beta(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)+\frac{\beta}{2}\|Ax+By-b\|^2. \tag{2.2}$$

其中 $\beta>0$ 是常数. **The convergence time seriously depend on β !**

交替方向法 Alternating Direction Method of Multipliers (ADMM)

交替方向法 ADMM 的 k 次迭代是(从给定的 (y^k,λ^k) 开始

$$\left\{\begin{array}{l} x^{k+1}=\text{Argmin}\{\mathcal{L}_\beta(x,y^k,\lambda^k)\mid x\in\mathcal{X}\},\\ y^{k+1}=\text{Argmin}\{\mathcal{L}_\beta(x^{k+1},y,\lambda^k)\mid y\in\mathcal{Y}\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b).\end{array}\right. \tag{2.3}$$

写开来就是

$$\begin{cases} x^{k+1} = \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, \\ y^{k+1} = \text{Argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{cases}$$

利用 x 和 y-子问题的最优性条件, 相当于变分不等式

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

2.2 最受海外学者重视的—罚参数选择的自调比准则

交替方向法的工作: Stanford 大学 Boyd 课题组的引用

我们的自调比准则，被 Boyd 教授称为一个简单而有效的公式.

Foundations and Trends® in
Machine Learning
Vol. 3, No. 1 (2010) 1–122

DOI: 10.1561/2200000016

now
the essence of knowledge

Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers

By Stephen Boyd, Neal Parikh, Eric Chu,
Borja Peleato and Jonathan Eckstein

Boyd 等在这篇论文的第 20 页中说到:

A simple scheme that often works well is (see, *e.g.*, [96, 169]):

$$\rho^{k+1} := \begin{cases} \tau^{\text{incr}}\rho^k & \text{if } \|r^k\|_2 > \mu\|s^k\|_2 \\ \rho^k/\tau^{\text{decr}} & \text{if } \|s^k\|_2 > \mu\|r^k\|_2 \\ \rho^k & \text{otherwise,} \end{cases} \tag{3.13}$$

where $\mu > 1$, $\tau^{\text{incr}} > 1$, and $\tau^{\text{decr}} > 1$ are parameters. Typical choices might be $\mu = 10$ and $\tau^{\text{incr}} = \tau^{\text{decr}} = 2$. The idea behind this penalty parameter update is to try to keep the primal and dual residual norms within a factor of μ of one another as they both converge to zero.

Boyd 等这篇论文的部分参考文献:

[96] B. S. He, H. Yang, and S. L. Wang, "Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities," *Journal of Optimization Theory and Applications*, vol. 106, no. 2, no. 2, pp. 337–356, 2000.

[169] S. L. Wang and L. Z. Liao, "Decomposition method with a variable parameter for a class of monotone variational inequality problems," *Journal of Optimization Theory and Applications*, vol. 109, no. 2, no. 2, pp. 415–429, 2001.

✠ 斯坦福大学 Boyd 等利用 ADMM 开发了一个凸优化的求解器. 介绍这个求解器的文章, 作为 2017 年 《Journal of Machine Learning Research》的第一篇文章。

Journal of Machine Learning Research 18 (2017) 1-5 Submitted 9/15; Revised 10/16; Published 2/17

SnapVX: A Network-Based Convex Optimization Solver

David Hallac* HALLAC@STANFORD.EDU
Christopher Wong† CRWONG@CS.STANFORD.EDU
Steven Diamond† DIAMOND@CS.STANFORD.EDU
Abhijit Sharang† ABHISG@CS.STANFORD.EDU
Rok Sosič† ROK@CS.STANFORD.EDU
Stephen Boyd* BOYD@STANFORD.EDU
Jure Leskovec† JURE@CS.STANFORD.EDU

- 文章的 《Abstract》中提到, 基于 ADMM, 能够有效地求解各种不同应用领域的大型凸优化问题.

- 文章中说, ADMM 的收敛时间依赖于罚参数 ρ 的选取（我们的文章里一般用 β 表示）. 校正 ρ 可以采用我们 2000 年文章中的做法.

> - **ADMM ρ-update** - The convergence time of ADMM depends on the value of the penalty parameter ρ (Nishihara et al., 2015), as it affects the tradeoff between primal and dual convergence, both of which need to be obtained for the overall problem to be solved. SnapVX users are not only able to select the value of ρ (it defaults to $\rho = 1$), but can also define a function to update ρ after each iteration based on the primal and dual residual values (He et al., 2000; Fougner and Boyd, 2015).

- 文章列出的部分参考文献.

B. He, H. Yang, and S. Wang. Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. *Journal of Optimization Theory and Applications*, 106(2):337–356, 2000.

P. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

交替方向法的收敛性质, 建议参阅 `maths.nju.edu.cn/~hebma` 上系列讲义《凸优化和单调变分不等式的收缩算法》的第十一讲.

效果还是不如意，可能的问题在哪里？

凸优化问题

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

和目标函数乘上一个大于 0 的常数 τ 以后生成的问题

$$\min\{\boldsymbol{\tau}(\theta_1(x)+\theta_2(y)) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\} \tag{2.4}$$

等价. 上面的问题的增广 Lagrange 函数是

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

The convergence time depend on $\boldsymbol{\tau}$ and $\boldsymbol{\beta}$!

参数 $\boldsymbol{\tau}$ 的选择, 相当于把问题中目标函数和约束条件的量纲协调好.

3 后一个十年研究的主要进展 (2007-2017)

- 2007-2008 年, 我们做些有管理背景的逆变分不等式. 研究 2-3 个算子的平行分裂的增广 Lagrange 乘子法 [11], 用预测-校正的套路.
- 2009 年春, 杨俊锋从 Rice 联合培养回国, 介绍了用交替极小化算法 (Alternating Minimization Algorithm) 求解图像处理问题。
- 两个可分离算子问题罚函数方法松弛后就是交替极小化算法。
- 增广 Lagrange 乘子法优于罚函数方法(Nocedal & Wright [27]).
- 两个可分离算子问题的增广 Lagrange 乘子法松弛就是 ADMM。
- 我知道, 也告知学生 ADMM 会有新的用武之地了。
- 赶快继续这方面研究！（我本人只做几乎没有任何附加条件的）

$$\min\{\theta_1(x)+\theta_2(y)\,|\,Ax+By=b,\ x\in\mathcal{X},\ y\in\mathcal{Y}\} \tag{3.1}$$

求解问题 (3.1) 的罚函数方法

$$(x^{k+1}, y^{k+1}) = \text{Argmin}\{\theta_1(x) + \theta_2(y) + \tfrac{\beta}{2}\|Ax + By - b\|^2 \,|\, x \in \mathcal{X},\ y \in \mathcal{Y}\}$$

求解问题 (3.1) 的增广 Lagrange 乘子法 从给定的 λ^k 开始

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

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

子问题难能度完全一样, 增广 Lagrange 乘子法优于罚函数方法。

共同的缺点 没有利用 x 和 y 的可分离结构！ 求解会无从着手。

求解问题 (3.1) 的松弛的罚函数方法 — 交替极小化方法(AMA)

从给定的 y^k 开始

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

$$y^{k+1} = \text{Argmin}\{\theta_2(y) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 | y \in \mathcal{Y}\}.$$

求解问题 (3.1) 的松弛的增广 Lagrange 乘子法 — ADMM

从给定的 (y^k, λ^k) 开始

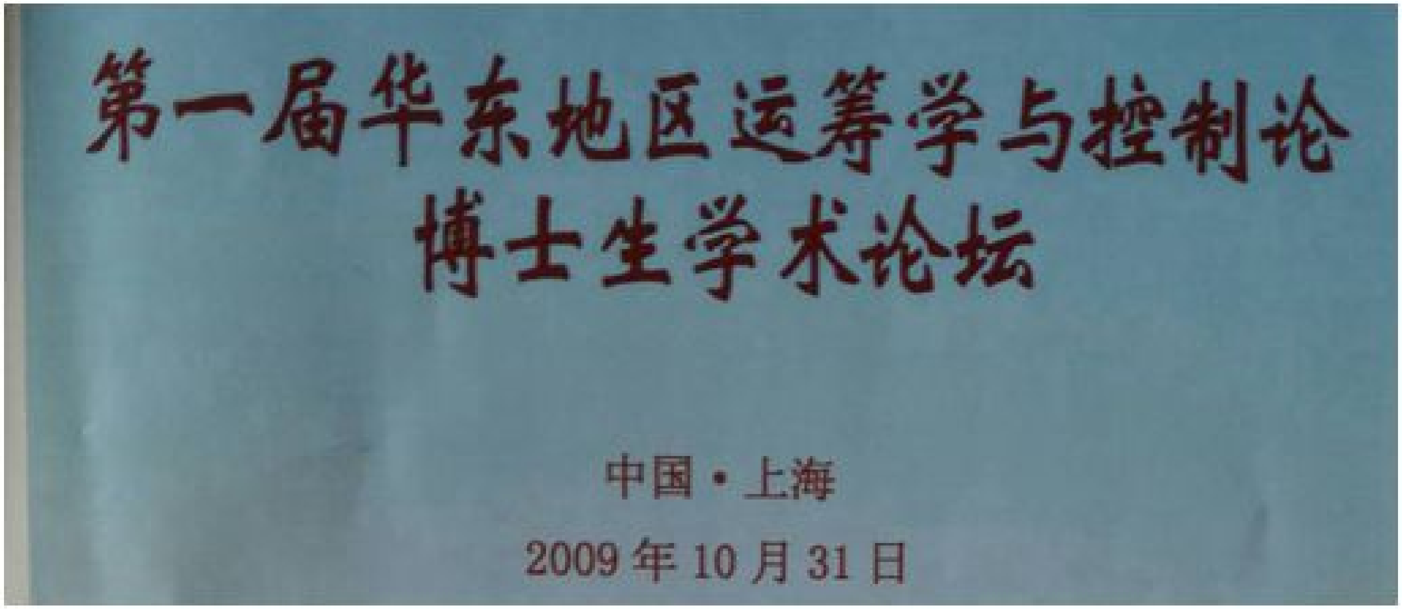
$$x^{k+1} = \text{Argmin}\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + By^k - b\|^2 | x \in \mathcal{X}\},$$

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

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

都松弛, 乘子交替方向法 (ADMM) 应该优于交替极小化方法 (AMA)

我们新的一轮工作是从 2009 年开始的



2009 年 10 月华东地区运筹学与控制论博士论坛

与周青教授一起应邀在博士论坛做大会报告

大会报告 1

题　目：TBA

报告人：周　青 教授

华东师范大学数学系

Email: qzhou@math.ecnu.edu.cn

简　介：1957 年 3 月出生，获北京大学学士学位、美国加利福尼亚大学洛杉矶分校硕士和博士学位，教授、博士生指导教师。主要研究兴趣为低维拓扑、大范围分析和随机复杂网络。曾主持国家自然科学基金面上项目和重点项目，获

大会报告 2

题　目：信息技术中的凸优化问题及交替方向法求解

报告人：何炳生 教授

南京大学数学系

Email: hebma@nju.edu.cn

简　介：1981 年于南京大学数学系获理学学士，1986 年获博士学位，

题目：信息技术中的凸优化问题和交替方向法求解

摘　要：根据不完整的（或者受污染）信息求得全部（正确的）信息，在信息技术中有着广泛的应用背景。这些一度被认为是没法求解的数学问题，经过数学工作者（包括菲尔兹奖获得者和世界数学家大会一小时报告人）近几年的努力，已经证明，在一些实际问题能满足的假设条件下，通过求解相应的松弛问题，可以求得原问题的真解。这些松弛了的（光滑或非光滑）凸优化问题一般具有特殊结构、形式简单明了，问题条件也不是太坏。来源于实际的问题一般是大型的，有的问题变量是阶数很高的矩阵，需要极小化的目标函数是矩阵的核模——矩阵奇异值之和，问题规模超大。求解这些问题给数学规划工作者提供了新的用武之地的同时也提出了新的挑战。我们介绍这些问题的基本模型，说明这些问题都可以用交替方向法求解。在用数值计算结果说明交替方向法是求解这些问题的一个简单有效方法的同时提出提高算法效率方面一些值得继续研究的问题。

2009 年, 我的大会报告提请青年学者注意交替方向法

3.1 收敛速率方面的理论结果

证明了交替方向法在遍历意义下和非遍历意义下的 $O(1/t)$ 收敛性.

- B. S. He and X. M. Yuan, On the $O(1/n)$ convergence rate of the Douglas - Rachford alternating direction method, *SIAM J. Numerical Analysis* **50**(2012), 700-709. 遍历意义下的收敛速率.

$$\tilde{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \left\{ \theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \right\} \le \epsilon.$$

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

- 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. 点列意义下的收敛速率.

SIAM Numerical Analysis 2012 文章的影响

siam Society for Industrial and Applied Mathematics

Sign in | Help | View Cart

Home | Journals | E-books | Proceedings | For Authors | Subscriptions | Interactive Features

SIAM Journal on Numerical Analysis

This list is updated daily based on the volume of full text article downloads over the past 12 calendar months.

2012 年发表以来, 五年来常在该刊物的热点论文中排名前三.

Add to Favorites | Track Citations | E-mail to a friend | View Abstracts | Send to Citation Manager

Select/unselect all items

The Local Discontinuous Galerkin Method for Time-Dependent Convection-Diffusion Systems
Bernardo Cockburn and **Chi-Wang Shu**
SIAM Journal on Numerical Analysis 1998, Vol. 35, No. 6, pp. 2440-2463: 2440-2463.
Abstract | References | PDF (461 KB) | Add to Favorites

+ Show Abstract

On the $O(1/n)$ Convergence Rate of the Douglas–Rachford Alternating Direction Method
Bingsheng He and **Xiaoming Yuan**
SIAM Journal on Numerical Analysis 2012, Vol. 50, No. 2, pp. 700-709: 700-709.
Abstract | References | PDF (165 KB) | Add to Favorites

+ Show Abstract

Boyd 的 ADMM 网页，提到我们的有关理论工作:

Stephen P. Boyd

- Home
- Teaching
- Biography

Research

- Books
- Papers
- Software
- Students

Classes

- EE103

ADMM

The *alternating direction method of multipliers* (ADMM) is an algorithm that solves convex optimization problems by breaking them into smaller pieces, each of which are then easier to handle. It has recently found wide application in a number of areas. On this page, we provide a few links to to interesting applications and implementations of the method, along with a few primary references.

ADMM is used in a large number of papers at this point, so it is impossible to be comprehensive here. We only intend to highlight a few representative examples in different areas. To keep the listing light, we have only listed more detailed bibliographic information for papers that are not easy to find online; in any case, the information given should be more than enough to track down the papers.

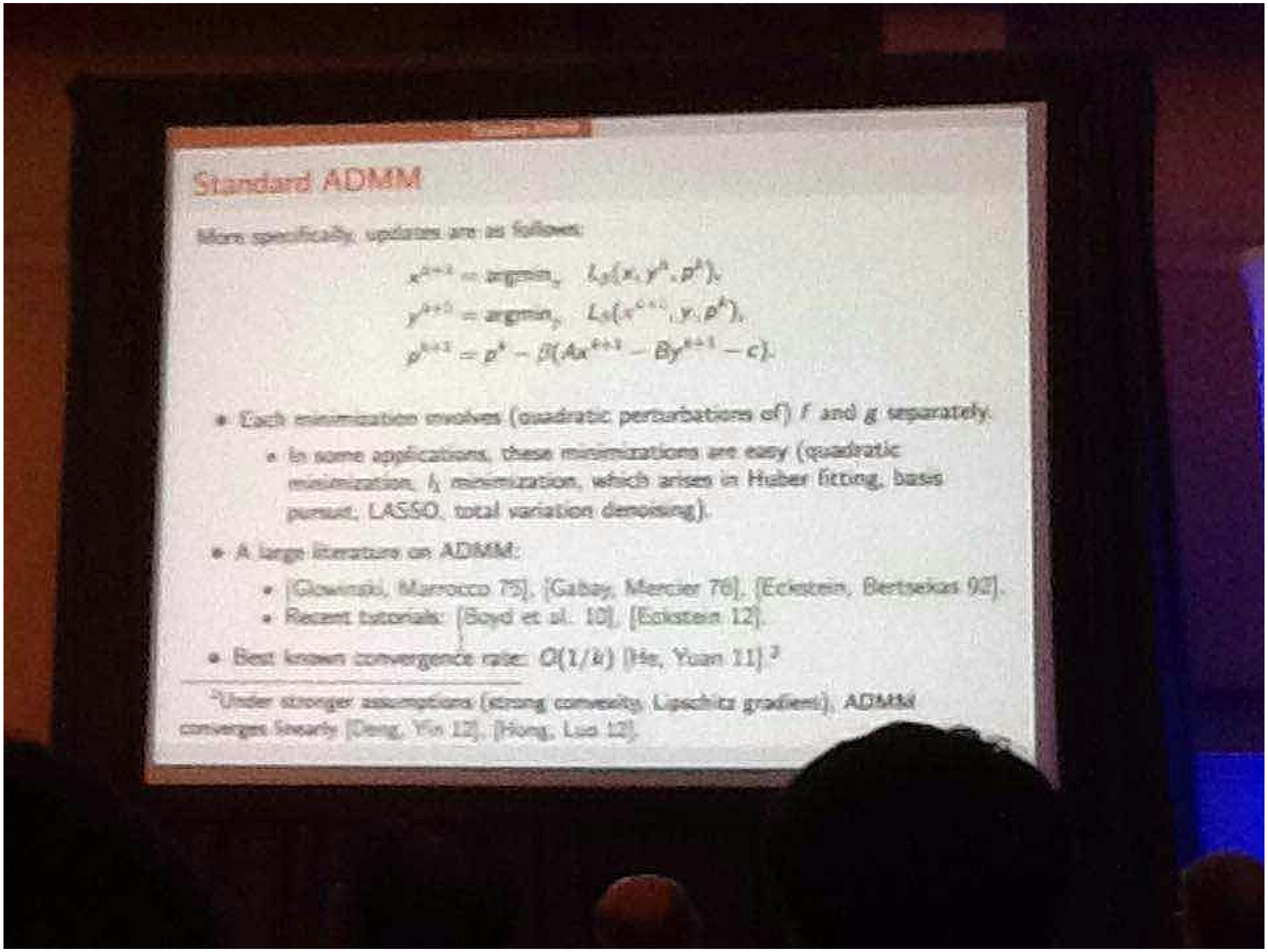
Theory and variations

On the $O(1/t)$ convergence rate of the Douglas-Rachford alternating direction method
B. He and X. Yuan, 2012

Fast alternating direction optimization methods
T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, 2012

Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results
J. Eckstein, 2012

世界数学规划会议上, Plenary Speaker 提到我们的工作:



Numerische Mathematik 2015 文章的结论与影响

除了

$$\left\| \begin{array}{c} B(y^{k+1} - y^*) \\ \lambda^{k+1} - \lambda^* \end{array} \right\|_H^2 \le \left\| \begin{array}{c} B(y^{k} - y^*) \\ \lambda^{k} - \lambda^* \end{array} \right\|_H^2 - \left\| \begin{array}{c} B(y^{k} - y^{k+1}) \\ \lambda^{k} - \lambda^{k+1} \end{array} \right\|_H^2,$$

还有

$$\left\| \begin{array}{c} B(y^{k+1} - y^{k+2}) \\ \lambda^{k+1} - \lambda^{k+2} \end{array} \right\|_H^2 \le \left\| \begin{array}{c} B(y^{k} - y^{k+1}) \\ \lambda^{k} - \lambda^{k+1} \end{array} \right\|_H^2.$$

因此有了点列意义下的收敛速率

ADMM 有 PPA 算法 (除延拓性质以外) 的所有漂亮性质

3.2 邻近点算法和增广 Lagrange 乘子法的延拓性质

凸优化问题

$$\min\{\theta(x) \mid x \in \mathcal{X}\}$$

PPA 算法

$$x^{k+1} = \text{Argmin}\{\theta(x) + \tfrac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}$$

主要性质

- $\|x^{k+1} - x^*\|^2 \le \|x^k - x^*\|^2 - \|x^k - x^{k+1}\|^2,$
- $\|x^{k+1} - x^{k+2}\| \le \|x^k - x^{k+1}\|.$

延拓

$$x^{k+1} := x^k - \gamma(x^k - x^{k+1}), \quad \gamma \in (0, 2).$$

右端的 x^{k+1} 是邻近点算法 (PPA) 的解.

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

增广 Lagrange 函数：$L_\beta(x,\lambda) = \theta(x) - \lambda^T(Ax-b) + \frac{\beta}{2}\|Ax-b\|^2.$

增广 Lagrange 乘子法 (ALM)

$$\begin{cases} x^{k+1} = \text{Argmin}\{L_\beta(x,\lambda^k) \mid x \in \mathcal{X}\} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b) \end{cases}$$

主要性质

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

延拓 $$\lambda^{k+1} := \lambda^k - \gamma(\lambda^k - \lambda^{k+1}), \quad \gamma \in (0,2).$$

右端的 λ^{k+1} 是 增广 Lagrange 乘子法的解.

PPA 和ALM 可以延拓的性质ADMM 能不能有? 适当改造可以有!

3.3 两个算子问题的 ADMM 方法的 (主要) 改进

1. ADMM in sense of PPA 交换顺序并外延 从 (y^k, λ^k) 出发.

$$
\left\{
\begin{array}{l}
\left\{
\begin{array}{rcl}
x^{k+1} & = & \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\
\lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^k - b), \\
y^{k+1} & = & \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+1}) \,|\, y \in \mathcal{Y}\},
\end{array}
\right. \qquad (3.2a) \\
\left\{
\begin{array}{rcl}
y^{k+1} & := & y^k - \gamma(y^k - y^{k+1}), \\
\lambda^{k+1} & := & \lambda^k - \gamma(\lambda^k - \lambda^{k+1}).
\end{array}
\right. \qquad \text{（松弛延拓）} \qquad (3.2b)
\end{array}
\right.
$$

这里 $\gamma \in (0, 2)$. 赋值号 $:=$ 表示 (3.2b) 右端的 (y^{k+1}, λ^{k+1}) 是由算法的前半部分 (3.2a) 产生的. 对多数问题, 这样往往能加快收敛速度.

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

2. Symmetric ADMM 对称的交替方向法

原始变量 x 和 y 本质上是平等的. 所以建议采用对称的交替方向法.

Symmetric Alternating Direction Method of Multipliers is described as

$$\left\{\begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\[2mm] \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), \\[2mm] y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\[2mm] \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b). \end{array}\right. \tag{3.3}$$

wehre $\mu \in (0,1)$ (usually $\mu = 0.9$).

- B.S. He, H. Liu, Z.R. Wang and X.M. Yuan, A strictly contractive Peaceman- Rachford splitting method for convex programming, *SIAM Journal on Optimization* **24**(2014), 1011-1040.

4 多个可分离算子的凸优化问题

我们以 3 个可分离算子的问题为例

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\}. \quad (4.1)$$

这个问题的增广 Lagrange 函数是

$$\mathcal{L}_\beta^3(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b)+\frac{\beta}{2}\|Ax+By+Cz-b\|^2.$$

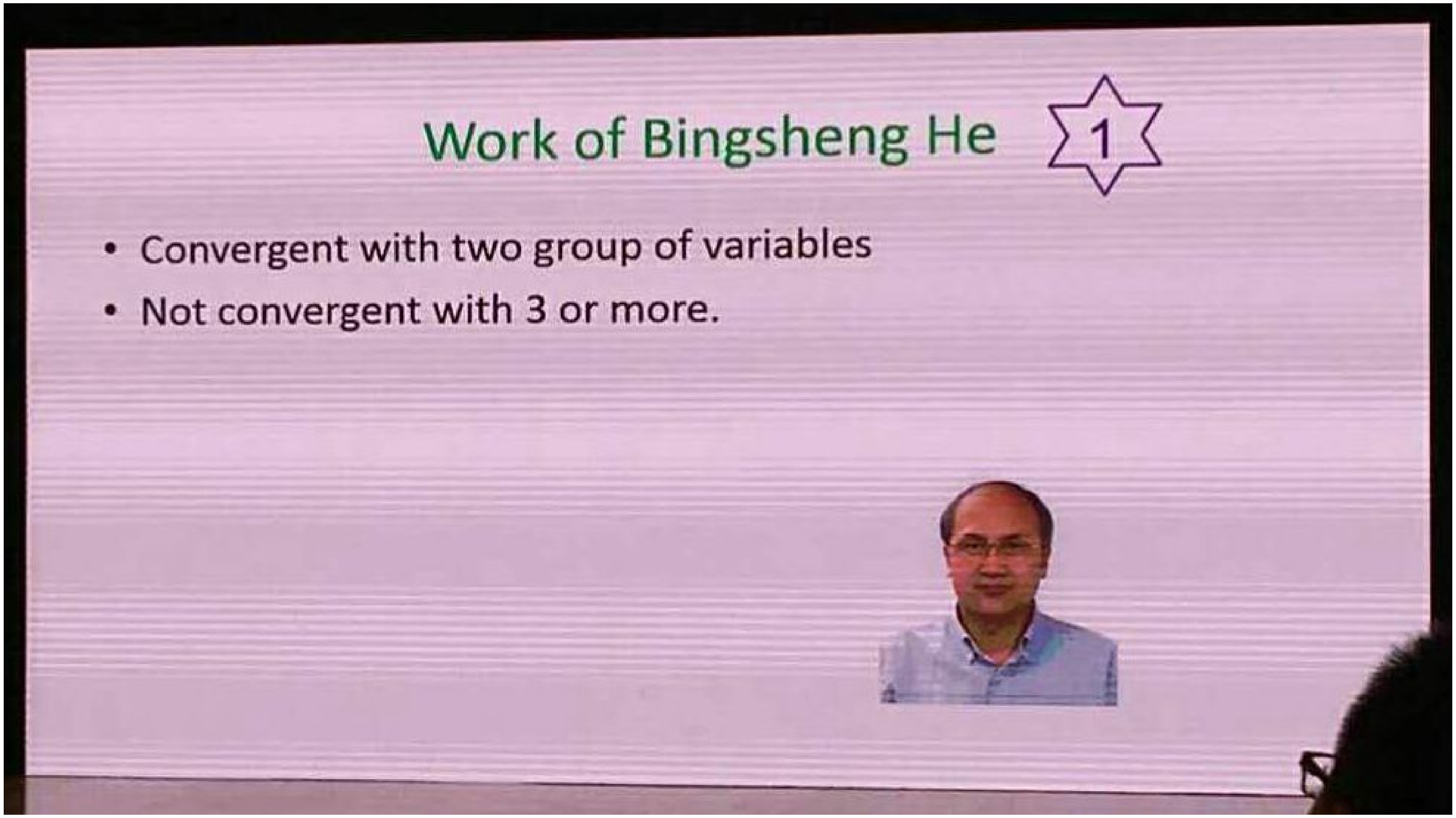
$$\boxed{\left\{\begin{array}{lcl} x^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x,y^k,z^k,\lambda^k)\ \middle|\ x\in\mathcal{X}\right\}, \\ y^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1},y,z^k,\lambda^k)\ \middle|\ y\in\mathcal{Y}\right\}, \\ z^{k+1} & = & \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1},y^{k+1},z,\lambda^k)\ \middle|\ z\in\mathcal{Z}\right\}, \\ \lambda^{k+1} & = & \lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b). \end{array}\right. \quad (4.2)}$$

对 $m\geq 3$, 直接推广的交替方向法不能保证收敛.

- 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.

♣ 2016 年 10 月中国运筹学会第十届年会

感谢堵丁柱教授注意到我们的工作, 还用了一张我最满意的照片.



不过, 准确的说法, 应该是何炳生和他的一些合作者的工作

感谢堵丁柱教授在一些国内论坛同样提到我们在 ADMM 方面的工作

直接推广 ADMM： 我们发表在 2016 Math.Progr. 的三个算子问题

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

的第一个例子中， $\theta_1(x)=\theta_2(y)=\theta_3(z)=0, \mathcal{X}=\mathcal{Y}=\mathcal{Z}=\Re,$

$\mathcal{A}=[A,B,C]\in\Re^{3\times 3}$ 是个非奇异矩阵， $b=0\in\Re^3.$

还有一些据此延伸的例子, 证明了直接推广的 ADMM 并不收敛.
这些例子更多的是在理论方面的意义.

值得继续研究的问题： 三个算子的实际问题中, 线性约束矩阵

$\mathcal{A}=[A,B,C]$ 往往至少有一个是单位矩阵, 即， $\mathcal{A}=[A,B,I].$

直接推广的 ADMM 处理这种更贴近实际的三个算子的问题,
既没有证明收敛，也没有举出反例，至今我们于心不甘！！

举个简单的例子来说：

- 乘子交替方向法 (ADMM) 处理问题

$$\min\{\theta_1(x) + \theta_2(y) | Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\} \quad \text{是收敛的.}$$

- 将等式约束换成不等式约束, 问题就变成

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

- 再化成三个算子的等式约束问题

$$\min\{\theta_1(x) + \theta_2(y) + 0 \,|\, Ax + By + z = b,\ x \in \mathcal{X}, y \in \mathcal{Y}, z \ge 0\}$$

- 直接推广的 ADMM 处理上面这种问题, 不少人做过尝试, 但是至今既没有证明收敛性, 也没有举出反例！

基于上述认知, 我们对三个算子的问题提出了一些修正算法. 注意：我们的方法对问题不加任何条件！ 对 β 不加限制, 只对方法动手术！

4.1 平行处理的-预测校正方法

平行分裂的增广拉格朗日乘子法

$$\left[\begin{array}{c}\text{平行处理}\\ x,\ y,\ z\\ \text{子问题再}\\ \text{更新 }\lambda,\\ \text{当预测点}\end{array}\right]\left\{\begin{array}{lcl}\tilde{x}^k & = & \arg\min\left\{\mathcal{L}^3_\beta(x,y^k,z^k,\lambda^k)\mid x\in\mathcal{X}\right\},\\ \tilde{y}^k & = & \arg\min\left\{\mathcal{L}^3_\beta(x^k,y,z^k,\lambda^k)\mid y\in\mathcal{Y}\right\},\\ \tilde{z}^k & = & \arg\min\left\{\mathcal{L}^3_\beta(x^k,y^k,z,\lambda^k)\mid z\in\mathcal{Z}\right\},\\ \tilde{\lambda}^k & = & \lambda^k-\beta(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b).\end{array}\right.$$

x, y, z 子问题平行了, 太自由. 由这些x, y, z 更新的 λ, 都需要校正！

$$[\text{简单校正}]\qquad w^{k+1}=w^k-\alpha_k(w^k-\tilde{w}^k),\qquad w=\begin{pmatrix}x\\ y\\ z\\ \lambda\end{pmatrix}$$

每步的 α 可以通过

$$\alpha_k = \gamma\alpha_k^*, \qquad \gamma \in (0,2), \qquad \alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|w^k - \tilde{w}^k\|_M^2}$$

给出, 其中

$$M = \begin{pmatrix} \beta A^T A & 0 & 0 & 0 \\ 0 & \beta B^T B & 0 & 0 \\ 0 & 0 & \beta C^T C & 0 \\ 0 & 0 & 0 & \frac{1}{\beta} I \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \\ C^T \\ 0 \end{pmatrix} \begin{pmatrix} A, & B, & C, & 0 \end{pmatrix}$$

是对称半正定矩阵,

$$\begin{aligned} \varphi(w^k, \tilde{w}^k) \;&=\; \|w^k - \tilde{w}^k\|_M^2 \\ &\quad + 2(\lambda^k - \tilde{\lambda}^k)^T [A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)]. \end{aligned} \tag{4.3}$$

校正形式是简单的, 计算步长只是一些简单的运算, 可以证明

$$\alpha_k \geq \frac{2-\sqrt{3}}{2}.$$

预测-校正, 是我们从投影收缩算法开始的算法框架. 预测只是提供收缩方向, 校正时计算步长, 可以确保 $\|w^k - w^*\|_M$ 下降收缩.

文章 2009 年发表在 Computational Optimization and Applications.

- B.S. He, Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities, Computational Optimization and Applications, 42, 195-212, 2009. 这篇文章

- Received: 22 April 2007
- Revised: 29 September 2007
- Published online: 6 November 2007

因此, 也可以看做我和 ADMM 20 年中前后衔接的一个工作.

也可以把 x 当成中间变量, 迭代 $(y^k, z^k, \lambda^k) \to (y^{k+1}, z^k, \lambda^k)$：

[先求 $\tilde{x}^k$, 然后平行处理理 y, z 子问题, 再更新 λ 生成预测点]

$$\left\{\begin{array}{lcl} \tilde{x}^k &=& \arg\min\left\{\mathcal{L}^3_\beta(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\right\}, \\ \tilde{y}^k &=& \arg\min\left\{\mathcal{L}^3_\beta(\tilde{x}^k, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\right\}, \\ \tilde{z}^k &=& \arg\min\left\{\mathcal{L}^3_\beta(\tilde{x}^k, y^k, z, \lambda^k) \mid z \in \mathcal{Z}\right\}, \\ \tilde{\lambda}^k &=& \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b). \end{array}\right.$$

y, z 子问题平行了, 太自由. 包括据此更新的 λ, 都需要校正！

[简单校正] $$v^{k+1} = v^k - \alpha_k(v^k - \tilde{v}^k), \qquad v = \begin{pmatrix} y \\ z \\ \lambda \end{pmatrix}$$

注意这里只要校正 (y, z, λ), 为下一步迭代提供 $(y^{k+1}, z^{k+1}, \lambda^{k+1})$.

每步的 α 可以通过

$$\alpha_k = \gamma \alpha_k^*, \qquad \gamma \in (0,2), \qquad \alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|w^k - \tilde{w}^k\|_M^2}$$

给出, 其中

$$M = \begin{pmatrix} \beta B^T B & 0 & 0 \\ 0 & \beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}$$

是对称半正定矩阵,

$$\begin{aligned} \varphi(w^k, \tilde{w}^k) \quad = \quad & \|w^k - \tilde{w}^k\|_M^2 \\ & + 2(\lambda^k - \tilde{\lambda}^k)^T [B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)]. \end{aligned}$$

校正形式是简单的, 计算步长只是一些简单的运算, 可以证明

$$\alpha_k \geq \frac{2 - \sqrt{2}}{2}.$$

4.2 带高斯回代的 ADMM 方法

以 (4.2) 提供的 (y^{k+1}, z^{k+1}) 为预测, 取 $\alpha \in (0,1)$, 校正公式为

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} := \begin{pmatrix} y^k \\ z^k \end{pmatrix} - \alpha \begin{pmatrix} I & -(B^TB)^{-1}B^TC \\ 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}. \quad (4.4)$$

由于为下一步迭代只要准备 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$, 我们只要做

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \end{pmatrix} := \begin{pmatrix} By^k \\ Cz^k \end{pmatrix} - \alpha \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} B(y^k - y^{k+1}) \\ C(z^k - z^{k+1}) \end{pmatrix}.$$

- 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.

对 y 和 z, 有先后, 不公平, 那就要做找补, 调整

4.3 ADMM + Prox-Parallel Splitting ALM

$$\left[\begin{array}{c}\text{强}\\ \text{制}\\ \text{y, z}\\ \text{平}\\ \text{等}\end{array}\right]\left\{\begin{array}{lcl} x^{k+1} &=& \arg\min\left\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \mid x\in\mathcal{X}\right\},\\ y^{k+1} &=& \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) \mid y\in\mathcal{Y}\right\},\\ z^{k+1} &=& \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) \mid z\in\mathcal{Z}\right\},\\ \lambda^{k+1} &=& \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b).\end{array}\right.$$

y, z 子问题平行, 如果不想做后处理, 就给它们俩预先都加个正则项

$$\left\{\begin{array}{l} x^{k+1} = \arg\min\left\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \mid x\in\mathcal{X}\right\},\\ y^{k+1} = \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\beta\|B(y-y^k)\|^2 \big| y\in\mathcal{Y}\right\},\\ z^{k+1} = \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\beta\|C(z-z^k)\|^2 \big| z\in\mathcal{Z}\right\},\\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b).\end{array}\right.$$

上述做法相当于:

$$
\begin{cases}
x^{k+1} = \text{Argmin}\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b - \frac{1}{\beta}\lambda^k\|^2 \,|\, x \in \mathcal{X}\}, \\
\qquad \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\
y^{k+1} = \text{Argmin}\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, \\
z^{k+1} = \text{Argmin}\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T Cz + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, \\
\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b),
\end{cases}
\tag{4.5}
$$

其中 $\mu > 2$. 例如, 可以取 $\mu = 2.01$.

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

太自由, 又不校正, 就加正则项, 不忘自己昨天的承诺.

This method is accepted by Osher's research group

- E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 21, NO. 7, JULY 2012 3239

A Convex Model for Nonnegative Matrix Factorization and Dimensionality Reduction on Physical Space

Ernie Esser, Michael Möller, Stanley Osher, Guillermo Sapiro, *Senior Member, IEEE*, and Jack Xin

$$\min_{T\geq 0, V_j\in D_j, e\in E} \zeta \sum_i \max_j (T_{i,j}) + \langle R_w \sigma C_w, T\rangle$$

$$\text{such that} \quad YT - X_s = V - X_s \mathrm{diag}(e). \quad (15)$$

Since the convex functional for the extended model (15) is slightly more complicated, it is convenient to use a variant of ADMM that allows the functional to be split into more than two parts. The method proposed by He *et al.* in [34] is appropriate for this application. Again, introduce a new variable Z

Using the ADMM-like method in [34], a saddle point of the augmented Lagrangian can be found by iteratively solving the subproblems with parameters $\delta > 0$ and $\mu > 2$, shown in the

tion refinement step. Due to the different algorithm used to solve the extended model, there is an additional numerical parameter μ, which for this application must be greater than two according to [34]. We set μ equal to 2.01. There are also model parame-

[33] E. Candes, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis,” 2009 [Online]. Available: http://arxiv.org/PS cache/arxiv/pdf/0912/0912.3599v1.pdf
[34] B. He, M. Tao, and X. Yuan, “A splitting method for separate convex programming with linking linear constraints,” Tech. Rep., 2011 [Online]. Available: http://www.optimization-online.org/DB_FILE/2010/06/ 2665.pdf

说明

- 上面提到的方法都与邻近点算法 (PPA) 有关.
- 所有的 (ADMM) 类分裂算法都源于增广 Lagrange 乘子法.
- ADMM 类方法只是对应用中出现一些实际问题有较好的计算效果. 说到底, 只是一阶方法.
- 需要注意的：PPA 和 ALM 具有的短处, ADMM 都有 !

5 分裂收缩算法的统一框架

对变分不等式

$$u^* \in \Omega, \quad (u-u^*)^T F(u^*) \ge 0, \quad \forall u \in \Omega.$$

通过选取适当的 $\beta > 0$ 和投影产生预测点 $\tilde{u}^k$:

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

根据投影的基本性质有

$$(u-\tilde{u}^k)^T \beta F(\tilde{u}^k) \ge (u-\tilde{u}^k)^T d(u^k, \tilde{u}^k), \ \forall u \in \Omega. \tag{5.1}$$

参数 β 选取使得 $(u^k-\tilde{u}^k)^T d(u^k,\tilde{u}^k) \ge \delta\|u^k-\tilde{u}^k\|^2$. $\delta > 0$ 为常数.

我们要求解的是如下的混合单调变分不等式:

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

在结构型凸优化中, $u=(x,y)$, $w=(x,y,\lambda)$, 核心变量是 $v=(y,\lambda)$.

通过求解一些子问题, 产生的预测点 $\tilde{w}^k \in \Omega$ 是与 (5.1) 类似的

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall\, w \in \Omega, \tag{5.2}$$

的解. 其中 $\boxed{Q^T + Q \ \text{正定}}$ 将 (5.1) 抄录于此看得更明白

$$\boxed{(u - \tilde{u}^k)^T \beta F(\tilde{u}^k) \ge (u - \tilde{u}^k)^T d(u^k, \tilde{u}^k), \ \forall u \in \Omega.}$$

因此, (5.2) 中的 $Q(v^k - \tilde{v}^k)$ 相当于 (5.1) 中的 $d(u^k, \tilde{u}^k)$. 对不同的问题, 预测都提供了下降方向. 取正定矩阵 H, 使 $M = H^{-1}Q$, (5.2) 就变成

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

我们用校正公式 $\boxed{v^{k+1} = v^k - M(v^k - \tilde{v}^k),}$ 产生新的迭代点.

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

因为 H 正定, 就可以归结到 PPA 的框架中

对预测和校正公式中的矩阵 Q 和 M, 如果

$$H = QM^{-1} \succ 0, \tag{5.3a}$$

并且

$$G = Q^T + Q - M^T H M \succeq 0, \tag{5.3b}$$

方法就是收敛的, 并且具有 $O(1/t)$ 的收敛速率.

一个简单清晰的框架对设计和评价算法是很有帮助的

前面已经分别提到

普通的大学数学，基本的优化常识 和 必要的社会实践.

中学的数理基础 . 中学生就会验证的一个初等恒等式

$$(a-b)(c-d) = \frac{1}{2}\big[(a-d)^2-(a-c)^2\big]+\frac{1}{2}\big[(c-b)^2-(d-b)^2\big]$$

帮了我的大忙.

✠ 得到这些式子, 要感谢我接受的良好中学教育 ✠

利用上面的恒等式处理：$(v-\tilde{v}^k)^T H(v^k-v^{k+1})$

✠ 有关统一框架下收敛性的简单证明可以从我的主页 maths.nju.edu.cn/~hebma 的《My Talks》中第三个报告的 §4 中找到.

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

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 (适合打印的文本)

My Foundations:

1. 国家自然科学基金   2 教育部博士点基金---江苏省自然科学基金

凸优化和单调变分不等式的收缩算法 (20单元近400页讲义--适合打印的文本)

统一框架与应用---算法研究力求数学之美

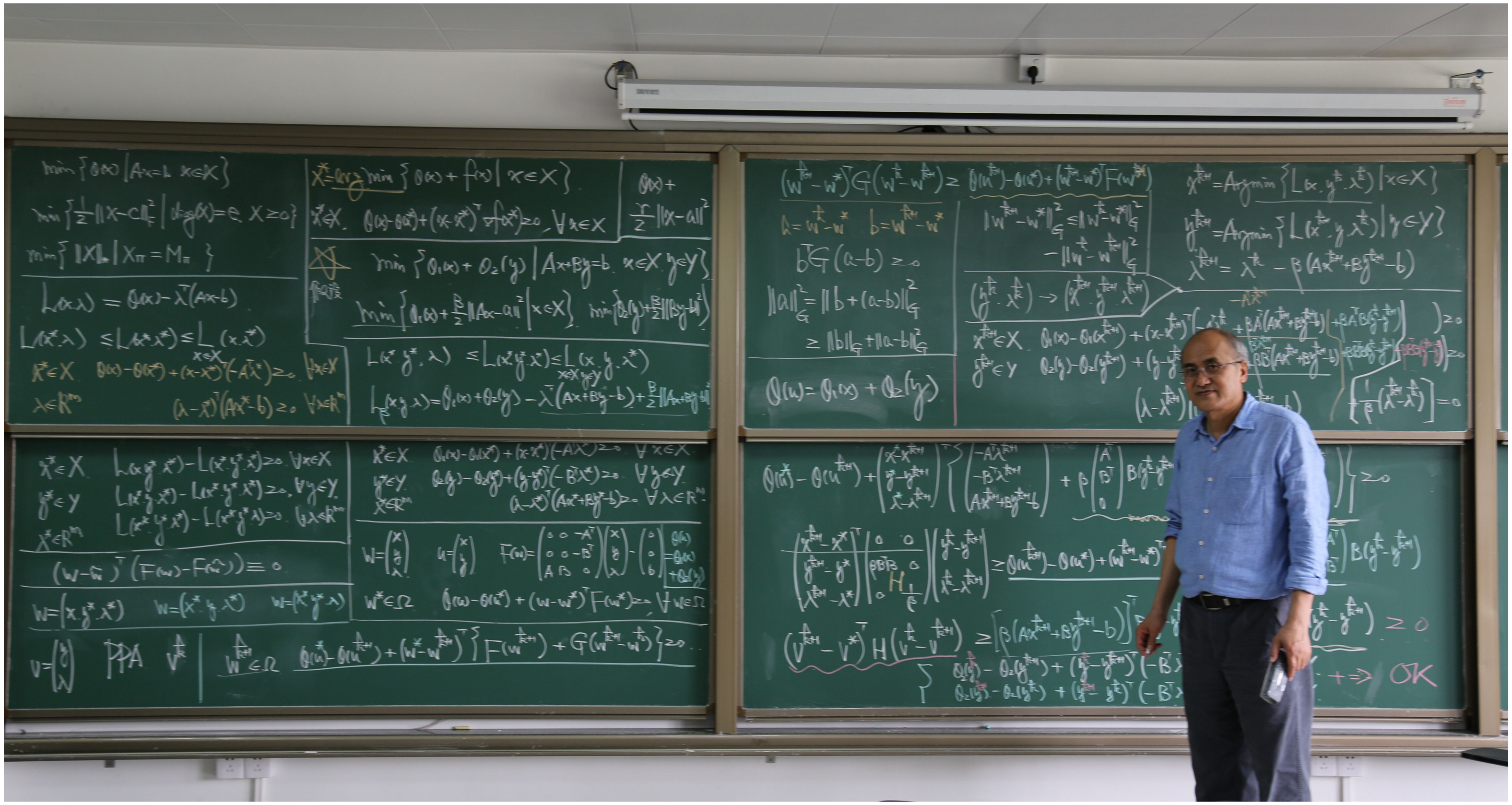
从 2011 年开始，我分别在 大连理工大学，天津大学，四川大学，复旦大学，浙江大学，武汉大学，南开大学，山东大学，中科院计算数学研究所，广西大学，重庆大学等高校宣讲交替方向法.

2012 年5 月, 大连理工大学报告加回代的三个算子的 ADMM

2015 年研究生暑期班, 河南大学, 六块黑板犹如一面墙

抹完一面墙, 不少人照相. 抹墙的人, 心想汗水没有白淌!

2017 年 3 月, 重庆师大, 统一框架下的理论证明



✠ 变分不等式是分析凸优化分裂收缩算法的有力工具.

在变分不等式框架下收敛性证明是非常简单的.

✠ 这幅照片说明, 我们在黑板上用不大的篇幅, 就能把

乘子交替方向法的来龙去脉和收敛性证明全部讲清楚.

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

✠ 追求简单与统一, 是我研究工作欲罢不能的原因 ✠

求解线性约束凸优化，增广拉格朗日乘子法 (ALM) 优于罚函数方法，有点优化基础知识的人都知道。

对两个可分离算子的线性约束凸优化问题，增广拉格朗日乘子法 (ALM) 和罚函数方法, 松弛后分别成了乘子交替方向法 (ADMM) 和交替极小化方法 (AMA)。

人们因此有理由对 ADMM 格外关心。

ADMM 不是我们提出来的。有了 10 年投影收缩算法的研究的基础, 使得我对 ADMM 类方法格外感兴趣。带领学生对ADMM 方法做一些有价值的改进和证明一些重要的理论结果，便顺理成章。

方法上，交换了原始变量 y 和对偶变量 λ 次序，进而得到因需定制的 PPA 意义下的 ADMM (Science in China, Mathematics, 2013)；

平等对待原始变量 x 和 y，两次校正对偶变量 λ，就得到对称型的 ADMM (SIAM Optimization, 2014)。

这些方法，道理上能站住脚，计算表现也不俗。

理论上，我们证明了 ADMM 在遍历意义下 (SIAM Numerical Analysis， 2012) 和点列意义下(Numer. Mathematik, 2015) 的 $O(1/t)$ 的收敛速率. 证明都不复杂.

ADMM 的广泛应用，人们自然想到向三个算子和多个算子的问题推广。

我们在不能证明“直接推广的方法”收敛的时候，提出了一些处理多个算子问题的ADMM 类方法

(SIAM Optimization, 2012; IMA Numerical Analysis, 2015).

被国际著名学者用来解决他们的科学计算问题。

后来我们又给出“直接推广的ADMM 方法处理三个算子问题不保证收敛”的例子(Math. Progr., 2016),说明:

- 我们处理多个算子的 ADMM 类方法中采用“不公平就找补校正”(SIAM Optimization, 2012) 和
- “各自为政处理子问题，就必须加正则项加强自我节制”(IMA Numer. Analysis, 2015)

等策略，手段上是必须的，机制上也是合理的。

出发点正确，往往事半而功倍。这几十年，生活理念的帮助，使得我们一步一步走的大致不错。

References

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

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

[3] A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vison, 40, 120-145, 2011.

[4] C. H. Chen, B. S. He and X. M. Yuan, Matrix completion via alternating direction method, *IMA Journal of Numerical Analysis* **32**(2012), 227-245.

[5] 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*, to appear in *Mathematical Programming, Series A*.

[6] E. Esser, M. Möller, S. Osher, G. Sapiro and J. Xin, A convex model for non-negative matrix factorization and dimensionality reduction on physical space, IEEE Trans. Imag. Process., 21(7), 3239-3252, 2012.

[7] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity problems, Volume I*, Springer Series in Operations Research, Springer-Verlag, 2003.

[8] 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.

[9] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin,

Heidelberg, Tokyo, 1984.

[10] D. Hallac, Ch. Wong, S. Diamond, A. Sharang, R. Sosič, S. Boyd and J. Leskovec, SnapVX: A Network-Based Convex Optimization Solver, Journal of Machine Learning Research 18 (2017) 1-5.

[11] B. S. He, Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities, *Computational Optimization and Applications* **42**(2009), 195–212.

[12] B. S. He, PPA-like contraction methods for convex optimization: a framework using variational inequality approach. J. Oper. Res. Soc. China 3 (2015) 391 – 420.

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

[14] B.S. He, Li-Zhi Liao and Mai-jian Qian, Alternating projection based prediction-correction methods for structured variational inequalities, Journal of Computational Mathematics, 24(6), 693-710, 2006.

[15] 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.

[16] 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.

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

[18] B. S. He, M. H. Xu, and X. M. Yuan, Solving large-scale least squares covariance matrix problems by alternating direction methods, *SIAM Journal on Matrix Analysis and Applications* **32**(2011), 136-152.

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

[20] 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.

[21] 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.

[22] 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.

[23] 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.

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

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

[26] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM J. Optim.* **15** (2004), 229–251.

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

[28] 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.

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

[30] P. Tseng, On accelerated proximal gradient methods for convex-concave optimization, manuscript, 2008.

Thank you very much for your attention ！